

UNIVERSIDAD NACIONAL  
DE ROSARIO  
FACULTAD DE CIENCIAS  
ECONÓMICAS Y ESTADÍSTICA



UNR

TESIS PRESENTADA PARA OPTAR AL GRADO DE DOCTOR  
EN ESTADÍSTICA

**Avances en modelos espacio-estado  
para el análisis de movimiento y  
comportamiento animal**

Por  
**Sofia Ruiz Suarez**

**Director**  
Dr. Juan Manuel Morales  
**Codirectora**  
Dra. Vianey Leos Barajas

**Rosario, Argentina.  
Año 2022**



# Agradecimientos

---

A mis dos directores, por guiarme y enseñarme durante todos estos años. Gracias por el tiempo y esfuerzo que han dedicado a mi formación. Gracias Juan por confiar en mí, por tu paciencia, y por brindarme tus conocimientos. Aprendí muchísimo y estoy segura que seguiré aprendiendo trabajando con vos. Gracias Vianey por animarte a este desafío de dirección a distancia antes de que la pandemia existiera. Gracias por el apoyo que me diste durante todos estos años, por tu inmensa paciencia, por tu motivación y entusiasmo siempre presentes.

A mi mamá, mi papá y mi hermana porque la distancia nunca fue un problema para acompañarme y apoyarme durante estos cinco años. Gracias familia por ser siempre mi gran fortaleza durante todo este proceso.

A la familia Melone. Gracias Sandra, Pablo, Cami, Ampí y Mauro por recibirme en su hogar durante todas mis estadías en Rosario. Gracias por ser mi familia rosarina, por su apoyo y calidez de todos estos años.

A mis compañeros amigos hermanos con los que compartí todos mis años de becaria en Bariloche. Porque sin ellos nada de esto hubiera sido posible. A mi amiga Gime, por alentarme siempre, por escucharme, por su fe en mí, por ser una hermana incondicional. A mi querido Tiño, porque al ser tan parecidos supimos entendernos y acompañarnos. Gracias amigo del alma por estar siempre. A la gran Tere, por todas esas mañanas de charlas en la oficina frente al lago. Por ser tan genuina, y enseñarme que a la vida académica la podíamos combinar con nuestras pasiones por la danza y la montaña.



# Resumen

---

La forma en que se mueven los animales es de gran interés en ecología, ya que afecta a la mayoría de los procesos ecológicos y evolutivos. Analizar estos procesos, implica estudiar sistemas que varían en el espacio y el tiempo a distintas escalas temporales y con distintos niveles de dependencia. Nuevas tecnologías han revolucionado la forma de estudiar y monitorear el movimiento de animales, su comportamiento, y su relación con el medio ambiente, llevando a la necesidad de desarrollar nuevas metodologías estadísticas. Bajo el contexto de modelos espacio-estado (MEE) y utilizando perspectiva Bayesiana, en esta tesis buscamos dar respuestas a este problema.

Presentamos un MEE que permite describir trayectorias formulando el proceso de movimiento a tiempo continuo y la observación a tiempo discreto. Usando Modelos Ocultos de Markov, clasificamos datos temporales de aceleración en distintos comportamientos. Por último, consideramos el error de las observaciones ambientales para describir trayectorias según la selección de recursos disponibles.

Los resultados obtenidos resaltan la importancia de contar con modelos adecuados que permitan describir e interpretar correctamente estos sistemas y diagramar prácticas de manejo apropiadas. Evidenciamos cómo la escala a la que los animales toman las decisiones de movimiento debe tenerse en cuenta a la hora de diseñar protocolos de colecta de datos y, que no siempre es necesario contar con datos de alta frecuencia para tener buenas estimaciones de ciertos procesos de movimiento. Además mostramos cómo el efecto del error en las observaciones ambientales puede sesgar las estimaciones de los análisis llevando a conclusiones erróneas.

**Palabras clave:** *autocorrelación temporal, datos de telemetría, inferencia Bayesiana, modelos espacio estado, movimiento animal.*



# Publicaciones

---

Los estudios desarrollados en los capítulo 2 y 3 de esta tesis se encuentran publicados en revistas indexadas. El trabajo vinculado al modelo de trayectorias individuales con datos de GPS (sección 2 de esta tesis) fue publicado en la revista *Journal of Life and Environmental Sciences-PeerJ* en el año 2020 (Ruiz-Suarez et al., 2020). El trabajo del modelo de clasificación de comportamientos con datos de acelerometría (sección 3 de esta tesis) fue publicado en la revista *Journal of Agricultural, Biological and Environmental Statistics-JABES* en el año 2021 (Ruiz Suarez et al., 2021).



# Índice general

---

<b>1. Introducción</b>	<b>3</b>
<b>2. Modelo de trayectorias individuales con datos GPS</b>	<b>9</b>
2.1. Introducción . . . . .	9
2.2. Modelo . . . . .	11
2.3. Inferencia . . . . .	12
2.3.1. Expresión de la función de verosimilitud . . . . .	12
2.3.2. Inferencia mediante ABC . . . . .	13
2.4. Estudio de simulación . . . . .	15
2.4.1. Evaluación de la capacidad inferencial de las técnicas ABC . . . . .	16
2.4.2. Relación entre las escalas temporales y la capacidad inferencial . . . . .	17
2.4.3. Resultados . . . . .	18
2.5. Análisis datos de oveja . . . . .	22
2.5.1. Resultados . . . . .	24
2.6. Discusión . . . . .	26
<b>3. Modelo de clasificación de comportamientos con datos de acelerometría</b>	<b>29</b>
3.1. Introducción . . . . .	29
3.2. Modelos . . . . .	31
3.2.1. Modelos ocultos de Markov(HMM) . . . . .	31
3.2.2. Modelos ocultos semi-Markovianos(HSMM) . . . . .	33
3.2.3. Estructura Autorregresiva (AR(p)-HMM and AR(p)-HSMM) . . . . .	34
3.3. Clasificación por aprendizaje supervisado . . . . .	35
3.3.1. Inferencia . . . . .	36
3.3.2. Predicción de estados . . . . .	36
3.3.3. Error de clasificación . . . . .	39
3.4. Estudio de simulación . . . . .	40
3.4.1. Resultados . . . . .	41
3.5. Datos de aceleración de ovejas . . . . .	43
3.5.1. Resultados . . . . .	45
3.6. Discusión . . . . .	49
<b>4. Modelo de trayectorias incluyendo error en las covariables ambientales</b>	<b>55</b>
4.1. Introducción . . . . .	55
4.2. Formulación de los modelos iSSA . . . . .	56
4.2.1. Inferencia . . . . .	58
4.3. Considerando error en las observaciones . . . . .	60
4.3.1. Caso continuo . . . . .	60
4.3.2. Caso discreto . . . . .	61
4.3.3. Algoritmo de inferencia . . . . .	62

4.4. Estudio de simulación . . . . .	64
4.5. Resultados . . . . .	66
4.6. Discusión . . . . .	69
<b>5. Discusión</b>	<b>75</b>
<b>6. Bibliografía</b>	<b>81</b>
<b>Anexos</b>	<b>92</b>
<b>A. Anexo I:</b>	<b>95</b>
A.1. Cálculo de la verosimilitud completa de los datos . . . . .	95
A.2. Estadísticos resumen . . . . .	97
A.3. Cobertura empírica . . . . .	100
<b>B. Anexo II:</b>	<b>102</b>
B.1. Inferencia Bayesiana . . . . .	102
B.2. Pruebas . . . . .	103
B.2.1. De la sección de decodificación local . . . . .	104
B.2.2. De la sección de decodificación global . . . . .	107
B.3. Distribuciones previas . . . . .	108
B.4. RECM figura extra . . . . .	110
<b>C. Anexo III:</b>	<b>111</b>
C.1. Aproximación por regresión logística condicional para el modelo iSSA .	111

# Índice de figuras

---

2.1.	Pequeño ejemplo de simulación del modelo de movimiento con tiempos aleatorios entre decisiones de movimiento. Los cambios de dirección se indican con puntos grises y las observaciones con cruces rojas. La trayectoria se inicia en las coordenadas $(0, 0)$ . Antes de las dos primeras cruces rojas (la primera no cuenta, pues $i = 0$ ) hay dos puntos grises, luego el número de cambios previos a las observaciones 1 y 2 son $N_1 = N_2 = 2$ . Antes de la tercera red roja hay cuatro puntos grises, luego $N_3 = 4$ , etc.	13
2.2.	Esquema de muestreo y escala temporal de los cambios de dirección. Con líneas negras el proceso de movimiento y con puntos rojos la observación.(a) caso de sobremuestreo, (b) caso equilibrado, (c) caso de submuestreo.	18
2.3.	Valores para el error de predicción (Ecuación 2.3) y para el índice de dispersión (Ecuación 2.4) para los dos parámetros obtenidos según cada algoritmo de ABC considerado y para cada valor de umbral $\epsilon$	19
2.4.	Resultados del análisis de validación cruzada para el algoritmo de rechazo y las dos versiones corregidas. Se muestra la relación entre los parámetros verdaderos y la mediana de la distribución posterior estimada. Con diferentes colores los resultados según los distintos para el umbral $\epsilon$ : violeta para $\epsilon = 0,1$ , azul para $\epsilon = 0,01$ , verde para $\epsilon = 0,005$ y rojo para $\epsilon = 0,001$ . La línea negra indica la relación ideal $x = y$ .	20
2.5.	Frecuencia relativa $p$ de los parámetros aceptados que fueron inferiores al valor real en los análisis. Por columna los resultados para diferentes valores de $\epsilon$ . Las dos primeras filas corresponden a los resultados del algoritmo de rechazo, las filas tres y cuatro al algoritmo corregido por regresión lineal, y las filas cinco y seis al algoritmo ABC corregido por redes neuronales	21
2.6.	Gráfico de dispersión suavizado (por regresión local) de los errores de predicción del tiempo esperado entre cambios de dirección ( $\lambda$ ) para diferentes relaciones entre la escala temporal de observación y la escala de cambios de dirección ( $R$ ). Los valores altos de $R$ indican que la escala temporal del proceso de observación es mayor que la escala temporal del proceso de decisión del movimiento. Con puntos negros los errores de predicción de $\lambda$ para cada valor de $R$ . La línea azul indica la curva suavizada para esos valores. Los intervalos de 95 % se muestran en gris. La línea vertical de puntos indica $R = 1$ .	22

2.7.	Gráfico de dispersión suavizado (por regresión local) de los errores de predicción del parámetro de dispersión para los ángulos de giro ( $\kappa$ ) para diferentes relaciones entre la escala temporal de observación y la escala de cambios de dirección ( $R$ ). Por filas los diferentes valores de $\kappa$ y por columnas los diferentes algoritmos ABC. Los valores altos de $R$ indican que la escala temporal del proceso de observación es mayor que la escala temporal del proceso de decisión del movimiento. Con puntos negros los errores de predicción de $\kappa$ para cada valor de $R$ . La línea azul indica la curva suavizada para esos valores. Los intervalos de 95% se muestran en gris. La línea vertical de puntos indica $R = 1$ . . . . .	23
2.9.	Con rojo las distribuciones posteriores obtenidos para ambos parámetros $\kappa$ y $\lambda$ mediante MCMC y con gris las obtenidas con cada algoritmo ABC. Por fila los resultados para cada parámetro y por columna los resultados para cada algoritmo ABC. . . . .	25
2.8.	Trayectoria final con una resolución de 1 segundo reconstruida por Dead-Reckoning y corregida con la información del GPS. . . . .	25
3.1.	Diagramas de las estructuras de los modelos (a)HMM: $C_t$ nota el proceso Markoviano latente y $X_t$ el proceso observaciones cuyas distribuciones dependen del estado $C_t$ . (b) Ejemplo HSMM: $C_t$ nota el proceso latente semi-Markoviano y $X_t$ el proceso de observación. $C_t^*$ indica el proceso Markoviano de los tiempos no absorbentes (es decir, el estado a tiempo $t$ es igual al estado a tiempo $t - 1$ ), y para cada uno de ellos $d_t$ da el valor muestreado para los tiempos de permanencia. . . . .	32
3.2.	Valores de los tres índices calculados en el estudio de simulación. Por columna el índice calculado (entropía cruzada, precisión para FB y precisión para Viterbi), por fila el grado de solapamiento entre las distribuciones de observación (alto, medio y bajo). En cada casilla, el eje x se refiere a la configuración de los parámetros de dispersión de los tiempos de permanencia ( $k_1$ y $k_2$ ): para la etiqueta 1 $k_1 = 1$ (un estado es geométrico) y para la etiqueta 2 ni $k_1$ ni $k_2$ es igual a 1 (ningún estado es geométrico). Los grupos de tres valores consecutivos de cada casilla corresponden a las diferentes medias entre $m_1$ y $m_2$ (20,40 y 90), y para cada uno de ellos en orden las tres posibles diferencias entre $m_1$ y $m_2$ (3,15 y 30). Los puntos representan el valor de la mediana y las barras el primer y el tercer cuartil. En verde los valores obtenidos para los HMM y en rojo los valores obtenidos para los HSMM. . . . .	42
3.3.	En el dispositivo colocado sobre la cabeza del animal, el eje x (color rojo) se corresponde con los movimientos laterales de la cabeza (i.e., sway); el eje z (color azul) coincide con los movimientos verticales (i.e., heave); y el eje y (color verde) representa los movimientos hacia atrás y adelante (i.e., surge). . . . .	43
3.4.	Dos ejemplos de las tres variables calculadas a partir de la señal de aceleración registrada de una oveja: $\log(\text{Mean.VedBA})$ , $\text{Mean.Pitch}$ y $\log(\text{Sd.Pitch})$ . Diferentes colores representan las señales correspondientes a cada uno de los cuatro comportamientos: Caminata, Caminata-Rápida, Inactivo y Alimentación. . . . .	47

3.5.	Boxplots de $\log(\text{Mean.VedBA})$ , $\text{Mean.Pitch}$ y $\log(\text{Sd.Pitch})$ para cada uno de los cuatro comportamientos: Inactivo (descanso y vigilancia), Caminada, Alimentación y Caminata-Rápida. . . . .	48
3.6.	Boxplots de los valores de precisión para ambos algoritmos de decodificación y del índice de entropía cruzada obtenidos mediante el análisis de validación cruzada para los cuatro modelos considerados: AR(1)-HMM, HMM, AR(1)-HSMM y HSMM. . . . .	49
3.7.	Distribuciones ajustadas de los tiempos de permanencia para cada comportamiento. El histograma muestra la distribución empírica de los tiempos de permanencia observados, mientras que los cuatro modelos ajustados se muestran en color. Las barras de error indican las estimaciones puntuales del primer y tercer cuartil. . . . .	50
4.1.	Distribución de los valores esperados para el parámetro $\alpha$ cuando la variable ambiental es continua(a) y cuando es categórica binaria(b). Con color rojo las distribuciones para el modelo CE-M1 y con verde para el modelo CE-M2. El boxplot azul indica la distribución de los valores esperados en el modelo base (SE-M1). La línea punteada marca el valor real del parámetro. . . . .	67
4.2.	Resultados obtenidos de RECM e ID para el parámetro $\alpha$ para las simulaciones con variables ambientales continuas (a-c) y para las simulaciones con variables ambientales categóricas binarias (b-d). En el eje x los diferentes grados de error dados por las tablas En el eje x los diferentes grados de error de la tabla 4.2 y 4.1 y en el eje y los valores de los índices calculados. Los puntos indican los valores medios de ambos índices y las barras los intervalos del 95 %. En rojo los valores para el modelo CE-M1 y en verde los valores para el modelo CE-M2. Con líneas grises punteadas se muestran la media y el intervalo del 95 % de los valores del modelo base (SE-M1). . . . .	68
4.3.	Resultados obtenidos de RECM e ID para el parámetro $\beta$ para las simulaciones con variables ambientales continuas (a-c) y para las simulaciones con variables ambientales categóricas binarias (b-d). En el eje x los diferentes grados de error dados por las tablas En el eje x los diferentes grados de error de la tabla 4.2 y 4.1 y en el eje y los valores de los índices calculados. Los puntos indican los valores medios de ambos índices y las barras los intervalos del 95 %. En rojo los valores para el modelo CE-M1 y en verde los valores para el modelo CE-M2. Con líneas grises punteadas se muestran la media y el intervalo del 95 % de los valores del modelo base (SE-M1). . . . .	70
4.4.	Distribución de los valores esperados para el parámetro $\beta$ cuando la variable ambiental es continua(a) y cuando es categórica binaria. Con color rojo las distribuciones para el modelo CE-M1 y con verde para el modelo CE-M2. El boxplot azul indica la distribución de los valores esperados en el modelo base (SE-M1). La línea punteada marca el valor real del parámetro. . . . .	71
A.1.	Gráficos de los estadísticos resumen evaluados para incluir dentro del análisis . . . . .	98

A.2. Gráficos de los estadísticos resumen evaluados para incluir dentro del análisis . . . . .	99
A.3. Gráficos de los estadísticos resumen evaluados para incluir dentro del análisis . . . . .	100
A.4. Cobertura empírica obtenida para $\kappa$ y $\lambda$ para los diferentes valores de umbral. Por fila los resultados para cada parámetros y por columna los resultados para los tres algoritmos de ABC. La línea punteada indica el valor 0,95 . . . . .	101
B.1. Boxplots del error medio cuadrático sobre las observaciones para los cuatro modelos considerados. . . . .	110

# Índice de tablas

---

2.1. Valores para las duraciones de paso y ángulos de giro del ejemplo de la Figura 2.1 . . . . .	12
2.2. Estadísticos resumen seleccionados para ajustar los tres algoritmos de ABC . . . . .	16
4.1. Parámetros de error para la variable ambiental continua . . . . .	65
4.2. Parámetros de error para la variable ambiental categórica binaria . . .	65
B.1. Distribuciones previas utilizadas para realizar el estudio de simulación .	108
B.2. Distribuciones previas utilizadas para realizar el análisis de los datos de aceleración de ovejas . . . . .	109





# 1. Introducción

---

En ecología, se le ha dado mucha importancia a entender y modelar el movimiento de los animales (Mevin B. Hooten et al., 2017), ya que desempeña un papel importante en la aptitud y la evolución de las especies (Nathan, 2008), la estructuración de las poblaciones y las comunidades (Turchin, 1998; Morales et al., 2010; Matthiopoulos et al., 2015), las respuestas al cambio ambiental (Jonsen, 2016) y la propagación de enfermedades (Fèvre et al., 2006). En particular, el estudio del movimiento ovino resulta de gran interés, ya que tanto a nivel mundial como en Patagonia, la actividad ganadera es la principal causa de la degradación de pastizales naturales. La falta de conocimiento sobre los procesos ecológicos que ocurren en los mismos conlleva a la toma de medidas de manejo no adecuadas. De esta manera, estudiar los patrones de movimiento del ganado ovino en paisajes heterogéneos permite comprender cuáles son los procesos que influyen su comportamiento y el uso del espacio, para luego contribuir con el desarrollo de estrategias de manejo sustentables para la producción ganadera y del ambiente.

El interés por el movimiento animal se debe en parte a avances tecnológicos en distintos tipos de medidores remotos (biotelemetría), que están revolucionando la forma en que se trabaja en ecología (Brown et al., 2013; Kays et al., 2015; Williams et al., 2020). En muchos aspectos, las técnicas de recolección de datos se han desarrollado más rápido que los métodos de análisis disponibles para los ecólogos. La cantidad y calidad de los datos de movimiento disponibles hoy en día sobrepasan los análisis tradicionales y muchos estudios sólo realizan estimaciones simples de tamaños de áreas de vida, distribuciones de frecuencia de desplazamientos o estimaciones de selección de hábitat usando por ejemplo técnicas de regresión logística (Millsbaugh y Marzluff, 2001). Sin embargo, los procesos que subyacen a la forma en que se mueven y se comportan los animales son complejos, variables en el tiempo y afectados por el paisaje, el estado interno de los individuos, las interacciones entre ellos y la disponibilidad de recursos.

Analizar estos procesos e intentar comprenderlos, implica estudiar sistemas que varían en el espacio y el tiempo a diferentes escalas temporales y con distinto nivel de dependencia. Los datos de telemetría que permiten hacer inferencia sobre estos sistemas suelen proveer información medida en diversas escalas temporales, en general distintas de las escalas a las que se producen los procesos reales. Además, en general estos dispositivos proveen grandes volúmenes de información, muchas veces con cierto grado de error, agregando cierta complejidad al análisis. Resulta así un desafío incorporar la correlación espacio temporal del sistema dentro del análisis, considerando que los procesos de movimiento, comportamiento y observación ocurren a escalas temporales diferentes y que muchas veces los datos pueden tener cierto grado de error.

En este contexto, los modelos espacio estado (MEE) proporcionan una herramienta conveniente para el análisis de datos de movimiento (Patterson et al., 2008). Los MEE asumen que el desarrollo en el tiempo de un sistema está determinado por una serie de valores no observados ( $n_t$ ) que son asociados con una serie de observaciones ( $x_t$ ). La relación entre ( $n_t$ ) y ( $x_t$ ) es especificada por el modelo (Durbin y Koopman, 2001;

Prado y West, 2016). Luego, los MEE distinguen dos modelos diferentes: un modelo mecanicista hipotético del movimiento no observado, conocido como modelo latente o de proceso, y un modelo de observación. El modelo latente consiste en una o más variables que describen el estado real del sistema, como podría ser la ubicación real del animal o el comportamiento. El modelo de observación proporciona la probabilidad de obtener una determinada observación en función del estado real, posiblemente con algún grado de error. Al incluir explícitamente el modelo de observación, los MEE permiten realizar inferencias estadísticas que tienen en cuenta las incertidumbres de los datos de entrada. Además permiten estimar las probabilidades de los estados, los parámetros del modelo de proceso y los parámetros del modelo de observación de manera simultánea.

El objetivo general de este trabajo apunta al desarrollo de modelos matemáticos que permitan modelar sistemas espacio temporales dentro de un contexto espacio estado para luego probar sus bondades en el desarrollo y ajuste de modelos de movimiento y comportamiento animal. Nos proponemos estudiar cómo son las formulaciones de estos modelos, cuáles son sus limitaciones y virtudes. Bajo el enfoque Bayesiano buscaremos comprender y desarrollar maneras eficientes de hacer inferencia. Trabajaremos con datos de localización GPS y de aceleración de ovejas recolectados en Patagonia Norte Argentina. El actual proyecto de doctorado, se encuentra dentro del marco del PICT 2015-0815 “Conectando comportamiento, ecología de paisajes y dinámica poblacional: hacia una gestión sustentable de producción ovina en sistemas de ganadería extensiva”.

Los objetivos específicos de este trabajo son: (i) Desarrollar y ajustar un modelo espacio estado que permita integrar un proceso latente a tiempo continuo con un proceso observacional a tiempo discreto. Estudiar posibles alternativas de técnicas Bayesianas para hacer inferencia en este contexto. Implementar luego un modelo para describir trayectorias considerando la relación entre la escala temporal a la que se toman los datos y aquella en la que se toman las decisiones de movimiento; (ii) Estudiar, describir y desarrollar métodos espacio estado cuyos procesos latentes sean descritos mediante variables discretas para clasificar datos con autocorrelación temporal. Analizar formas de incluir la dependencia temporal tanto del proceso de observación como del proceso de estado dentro de la estructura del modelo. Implementar luego algoritmos para discriminar comportamientos de animales a partir de datos provenientes de acelerómetros; (iii) Estudiar la posibilidad de extender y formular modelos integrados de selección de pasos con un enfoque espacio estado de manera de incluir covariables medidas con error. Utilizar estos modelos para describir el movimiento de individuos considerando el efecto del paisaje y la dependencia temporal cuando se cuenta con observaciones ambientales con error.

Patterson et al. (2008) define los MEE como modelos de series temporales que por un lado predicen un estado futuro del sistema utilizando información previa del mismo de manera probabilística mediante un modelo latente, y que al mismo tiempo, describen de manera matemática cómo las observaciones del estado del sistema son generadas mediante un modelo de observación. Esencialmente, el proceso latente predice estados futuros dado el estado actual del sistema (condición de Markov) y luego el modelo de observación pondera esas predicciones mediante la verosimilitud de los datos, conectando así el modelo latente con el de observación. Patterson hace énfasis en que los MEE tienen como mayor ventaja el hecho de que integran la corrección del error, el cálculo de métricas y el análisis estadístico dentro del análisis.

Las primeras aplicaciones de los MEE fueron en el campo de la ingeniería, pero luego fueron aplicados en un gran rango de disciplinas incluyendo el campo de la ecología. El uso de estos modelos ha ido en aumento desde la década de los ochenta. Uno de los primeros modelos utilizados, y quizás uno de los más clásicos es el Modelo Lineal Dinámico Normal (NDLM) (West y Harrison, 1997). Este modelo se construye en base a dos modelos lineales: uno para el proceso latente (condicionado al estado anterior), y otro para el proceso de observación (condicionado al estado actual). Este modelo básico tiene como punto a favor que puede ser ajustado utilizando el filtro de Kalman (Harvey, 1990). El filtro de Kalman consiste en un procedimiento analítico recursivo que permite estimar los parámetros del modelo y los estados ocultos desde un enfoque frecuentista. Las primeras aplicaciones en ecología fueron en datos de pesca, entre ellos: Mendelsohn (Mendelsohn y McKee, 1988) quien ajustó un NDLM a datos de estornino (*Scomber japonicus*), una especie de pez perciforme y (Sullivan, 1992) quien también ajustó un NDLM a datos de captura de pesca categorizados por el largo de sus clases. Posteriormente, avances en cuanto al poder computacional combinados con nuevos procedimientos de estimación basados en simulaciones, como las técnicas de Markov chain Monte Carlo (MCMC), permitieron extender el estudio de los MEE a casos no lineales y no gaussianos. Millar y Meyer (Millar y Meyer, 2000) presentaron uno de los primeros modelos de este tipo para datos de biomasa de *Thunnus alalunga* (atún blanco). Utilizaron técnicas de MCMC para ajustar un MEE en donde el proceso latente es considerado univariado no lineal y el proceso de observación usa índices de captura por unidad de esfuerzo.

Los avances en materia de tecnología de rastreo y monitoreo de animales permitieron contar con datos de alta resolución que describen perfiles continuos de movimiento y comportamiento animal y que serían imposibles de obtener por métodos de observación directa. Junto con esta capacidad de generar datos y debido a la estructura de los sistemas que describen movimiento y comportamiento animal es que los MEE resultan herramientas útiles en este contexto. En los últimos años se han propuesto una serie de modelos que permiten estudiar y analizar los procesos de movimiento y comportamiento animal.

Por un lado se han utilizado para modelar trayectorias de animales en una gran variedad de situaciones. Gran parte del tratamiento teórico del movimiento en ecología está constituido sobre “caminatas aleatorias” y sus correspondientes aproximaciones de difusión. En particular (Morales et al., 2004) propone un MEE a tiempo discreto basado en “caminatas aleatorias” en donde las trayectorias son modeladas a partir de largos de paso y ángulos de giro. Morales propone diferentes configuraciones de “caminatas aleatorias” dependiendo del comportamiento específico del animal, lo que le permite modelar movimiento y comportamiento al mismo tiempo. Para el ajuste e inferencia de este modelo se utilizaron técnicas de Markov Chain Monte Carlo. Cada observación debe ser asignada a un tipo de caminata aleatoria en particular dentro de un set pre-definido. El modelo presentado por Morales en 2004 permite incorporar movimiento y comportamiento, a la vez que resulta sencillo de ajustar a datos reales mediante métodos Bayesianos, por lo que fue tomado como base por varios autores para proponer nuevos modelos que describen trayectorias (Jonsen et al., 2005; Johnson et al., 2008a; Fryxell et al., 2008; Patterson et al., 2017). Por ejemplo (Jonsen et al., 2005) hizo uso de los MEE para estudiar el movimiento y uso del espacio de tortugas marinas

(*Dermochelys coriacea*) utilizando diferentes modelos de trayectorias dependiendo del momento del día (diurnas/nocturnas). Estudios como (Breed et al., 2006) distinguieron diferentes modelos de trayectorias dependiendo del sexo específico en datos de foca gris (*Halichoerus grypus*).

Por otro lado, a partir del auge del uso de acelerómetros y magnetómetros se han propuesto modelos del tipo modelos ocultos de Markov (HMM), un tipo específico de MEE, para identificar diferentes estados comportamentales en individuos. Gran parte del foco dentro del análisis de datos de acelerometría ha sido identificar patrones en las observaciones que correspondan a modos de movimiento o comportamientos conocidos. Sin embargo, la mayoría de los trabajos que buscan identificar estados a partir de datos de acelerometría no se han planteado en el marco espacio estado. Muchos estudios han mostrado la capacidad de algoritmos del tipo *Machine Learning*, tales como máquinas de vectores de soporte, árboles de clasificación, bosques aleatorios, entre otros, para clasificar datos de aceleración en animales (Nathan et al., 2012; Carroll et al., 2014; Graf et al., 2015). Por ejemplo Nathan et al., 2012 comparó la efectividad de cinco algoritmos de *Machine Learning* para distinguir entre siete comportamientos diferentes de buitres leonados. Sin embargo, los datos secuenciales de los acelerómetros poseen una dependencia temporal natural entre las observaciones y los comportamientos (los comportamientos de un animal duran un período de tiempo usualmente mayor a la frecuencia de muestreo). Este punto ha sido ignorado en la mayoría de las aplicaciones que buscan hacer clasificaciones. Los estudios en donde la estructura temporal de los datos es considerada han sido abordados desde el enfoque espacio estado mediante el uso de HMM. Por ejemplo, (Phillips et al., 2015) aplicó un HMM dentro de un contexto no supervisado para estudiar el comportamiento de atunes a partir de datos de aceleración vertical y (Leos-Barajas et al., 2017) también utilizó este tipo de herramientas para clasificar datos de aceleración de aves de rapiña.

El contenido de esta tesis se encuentra organizado en 5 capítulos, incluido este primer capítulo introductorio, un apartado de bibliografía y anexos. El capítulo 2 corresponde al desarrollo del objetivo (i) de este trabajo: desarrollar y ajustar un modelo espacio estado que permita integrar un proceso latente a tiempo continuo con un proceso observacional a tiempo discreto. Estudiar posibles alternativas de técnicas Bayesianas para hacer inferencia en este contexto. En el mismo exponemos brevemente los enfoques previos utilizados para modelar trayectorias de animales y presentamos un modelo que considera las variaciones temporales entre el proceso de movimiento y la observación. Para la inferencia utilizamos métodos “Bayesianos de verosimilitud libre”, y evaluamos la capacidad de los mismos para el modelo presentado. Por último aplicamos el modelo a datos reales de ovejas.

En el capítulo 3 desarrollamos el segundo objetivo específico: (ii) estudiar, describir y desarrollar métodos espacio estado cuyos procesos latentes sean descriptos mediante variables discretas para clasificar datos con autocorrelación temporal. Analizar formas de incluir la dependencia temporal tanto del proceso de observación como del proceso de estado dentro de la estructura del modelo. Aquí damos una visión general los modelos ocultos de Markov y algunas extensiones en el contexto de clasificación supervisada para series temporales. Luego estudiamos cómo estos modelos se pueden utilizar para realizar clasificaciones en diferentes escenarios, y finalmente aplicamos los mismos a datos de acelerómetros de ovejas para distinguir entre cuatro estados

comportamentales.

El capítulo 4 se dedica a responder al último objetivo específico del trabajo: (iii) estudiar la posibilidad de extender y formular modelos integrados de selección de pasos con un enfoque espacio estado de manera de incluir covariables medidas con error. Utilizar estos modelos para describir el movimiento de individuos considerando el efecto del paisaje y la dependencia temporal cuando se cuenta con observaciones ambientales con error. En esta sección desarrollamos la formulación de los enfoques comúnmente utilizados para describir movimiento considerando el efecto del paisaje. Luego a partir de un estudio basado en simulaciones estudiamos como influye en las estimaciones la presencia de error en la observación y proponemos un modelo para corregir este sesgo. Por último exponemos un algoritmo para hacer inferencia bajo el enfoque Bayesiano.

Para finalizar en el capítulo 5 discutimos ampliamente los resultados obtenidos, resumiendo las conclusiones y hallazgos más importantes de esta investigación. Planteamos también algunas posibles líneas de trabajos futuros y concluimos el trabajo final.



# 2. Modelo de trayectorias individuales con datos GPS

---

## 2.1. Introducción

El estudio del movimiento animal es desafiante ya que es un proceso modulado por muchos factores que actúan a diferentes escalas espaciales y temporales (Gurarie y Ovaskainen, 2011; Mevin B. Hooten et al., 2017). En general, el proceso de movimiento de los animales se produce en tiempo continuo, pero las localizaciones son observadas en intervalos discretos, las cuales muchas veces están dictaminados por cuestiones relacionadas a logísticas de muestreo tales como la duración de la batería. Al modelar las trayectorias, es común suponer que la escala de tiempo en la que los animales deciden cambiar su movimiento es la misma que la de la observación. Sin embargo, no siempre es así, y es necesario ser consciente de este hecho para evitar sacar conclusiones sobre el proceso de movimiento que dependan de la escala temporal a la que se tomaron las observaciones.

Se han propuesto varios modelos espacio-estado para el análisis de datos de movimiento animal que difieren principalmente en la conceptualización temporal del proceso de movimiento: formulaciones de tiempo discreto y formulaciones de tiempo continuo (McClintock et al., 2014). Por un lado, los modelos de tiempo discreto describen el movimiento como una serie de pasos y giros (o direcciones de movimiento) que se realizan en ocasiones regulares (Morales et al., 2004; Jonsen et al., 2005; McClintock et al., 2012). La ventaja de este enfoque es que permite conceptualizar la dinámica implicada en el proceso de movimiento de una forma sencilla e intuitiva, lo que facilita su aplicación e interpretación. Normalmente, en estos modelos las escalas temporales del proceso de estado y del proceso de observación son las mismas. De esta manera, los tiempos de observación coinciden con los tiempos en los que se supone que el animal toma decisiones de movimiento. Por otro lado, se han propuesto modelos de tiempo continuo (Blackwell, 1999; Johnson et al., 2008b; Harris y Blackwell, 2013) en los que el proceso de movimiento se define para cualquier tiempo. Estos modelos, suelen expresarse mediante ecuaciones diferenciales estocásticas que dan cuenta de la dependencia entre localizaciones sucesivas, lo que significa que no dependen de una escala temporal concreta. En general, siempre que se utilicen límites razonables para los análisis, los modelos de movimiento en tiempo continuo proporcionan los mismos resultados independientemente de la resolución temporal de las observaciones.

El enfoque de tiempo continuo tiene la ventaja de ser más realista y de que evita la dependencia de una escala temporal concreta. Un posible inconveniente está en la interpretación de los parámetros de movimiento instantáneo (por ejemplo, los relacionados con los procesos de Ornstein-Uhlenbeck y otros modelos de difusión). Sin embargo, como se menciona en (McClintock et al., 2014), para que cualquier enfoque de tiempo continuo o discreto sea útil, la resolución temporal de los datos observados debe

ser relevante para los comportamientos de movimiento específicos de interés. Aunque ambos enfoques tienen ventajas y desventajas, los modelos de tiempo discreto son más intuitivos y fáciles de interpretar, pero pueden considerarse menos realistas que los modelos de tiempo continuo (McClintock et al., 2014).

En este capítulo, presentamos un modelo espacio-estado que formula el proceso de movimiento en tiempo continuo y la observación en tiempo discreto (intervalos regulares). Aunque la formulación se basa en largos de pasos y ángulos de giro, la probabilidad de que ocurra un cambio en la trayectoria es positiva para todos los tiempos  $t$ , y por tanto el proceso de movimiento es continuo en el tiempo. En la mayoría de los casos, cuando se intenta hacer inferencias sobre los procesos de movimiento, sólo se dispone de datos de localización observados en momentos regulares, los cuales muchas veces son dictaminados por cuestiones prácticas del muestreo. Por lo tanto, el número y los momentos en los que se produjeron cambios de movimiento entre las sucesivas localizaciones registradas es una información que no está disponible y que se suele ignorar.

Nuestro objetivo aquí es combinar la facilidad de interpretación de los modelos basados en pasos y giros con el realismo de los modelos de tiempo continuo, analizando la relación entre las escalas de las observaciones y las decisiones de movimiento. Utilizamos un paseo aleatorio en el que las decisiones de movimiento (pasos y giros) pueden tomarse en cualquier momento, mientras que el proceso de movimiento se observa a intervalos de tiempo regulares. Se tienen entonces dos escalas de tiempo diferentes: una para el proceso de estado y otra para la observación. La ventaja de este modelo es que nos permite diferenciar entre los tiempos en los que los animales toman decisiones de movimiento y los tiempos en los que se realizan las observaciones.

Parton y Blackwell (2017) presentan un modelo de movimiento en tiempo continuo que tiene algunas similitudes con el modelo que presentamos aquí. El modelo de Parton y Blackwell se formula mediante ecuaciones diferenciales estocásticas en términos de dirección y velocidades. Las localizaciones de los animales se observan en tiempos discretos con cierto error. Ambos modelos son similares ya que suponen que los cambios en la dirección de la trayectoria pueden ser en cualquier momento, que la trayectoria está correlacionada y que las observaciones son a tiempos discretos. Sin embargo las formulaciones son diferentes, ya que en el modelo de Parton y Blackwell la velocidad es estocástica y hay error de observación.

Debido a la complicada relación entre las localizaciones y los parámetros cuando los cambios de dirección y velocidad no se observan, tanto para el trabajo de Parton y Blackwell como para el modelo aquí propuesto, el cálculo de la función de verosimilitud resulta ser intratable. Sin embargo, simular trayectorias de movimiento y su observación es sencillo, lo cual sugiere el uso de técnicas de computación Bayesiana aproximada para la inferencia (Beaumont, 2010; Csilléry et al., 2010). Si bien estas técnicas ya se han empleado para ajustar modelos que implican procesos de movimiento, como el índice de movimiento global (Van der Vaart et al., 2015) o el movimiento entre parches en modelos individuales de dinámica de metapoblación (Sirén et al., 2018), hasta donde sabemos, no se han utilizado para hacer inferencia sobre trayectorias de movimiento.

En lo que sigue de este capítulo describimos, formalizamos y exponemos las posibles complicaciones de un modelo de movimiento espacio-estado con dos escalas

temporales diferentes (sección 2.2). Desarrollamos brevemente las técnicas de ABC y explicamos cómo utilizarlas para hacer inferencia en el modelo propuesto (sección 2.3). Utilizando simulaciones estocásticas, evaluamos la capacidad de las mismas para recuperar los valores de los parámetros que conducen el proceso de movimiento y, considerando el propósito ecológico subyacente, analizamos la calidad de las mismas con respecto a la relación entre las dos escalas temporales involucradas (sección 2.4). Por último, aplicamos el modelo a una trayectoria de ovejas de alta resolución para evaluar el rendimiento de la inferencia por ABC con datos reales (sección 2.5).

## 2.2. Modelo

Muchos estudios sobre movimiento animal asumen caminatas aleatorias correlacionadas en tiempo discreto como bloques para construir modelos más complejos (Turchin, 1998; Morales et al., 2004). En cada paso de tiempo, el individuo elige un ángulo de giro (diferencia entre la dirección de movimiento anterior y la nueva) y una velocidad de paso. La distribución del ángulo de giro se concentra en cero, dando lugar a direcciones de movimiento correlacionadas en el corto plazo (persistencia en la dirección de movimiento). En el caso del modelo aquí utilizado, permitiremos que los cambios de dirección se produzcan en cualquier instante de tiempo. Así, el individuo se mueve en una cierta dirección por un cierto período de tiempo, luego gira y empieza a moverse en una nueva dirección durante otro período de tiempo. Consideramos luego que las observaciones de la trayectoria se producen en intervalos regulares en el tiempo ( $\Delta t$ ), y para simplificar el modelo, asumiremos observaciones sin error.

Asumiendo velocidad constante, sea  $M_i$  la variable que describe la posición del proceso de movimiento a tiempo  $i$  representada en coordenadas cartesianas, es decir,  $M_i = (m_{i,1}, m_{i,2})$  donde  $i$  representa el índice del tiempo sobre los pasos para  $i = 0, \dots, N_{steps}$ . Dado  $m_{0,1} = 0$  y  $m_{0,2} = 0$ , para  $i = 1, \dots, N_{steps}$  se tiene,

$$\begin{aligned} m_{i,1} &= m_{i-1,1} + \cos(\phi_{i-1})t_{i-1} \\ m_{i,2} &= m_{i-1,2} + \sin(\phi_{i-1})t_{i-1} \\ \phi_i &= \sum_{k=1}^i \omega_k \end{aligned} \tag{2.1}$$

En donde  $t_i$  es la duración del paso  $i$ ,  $\omega_i$  el ángulo de giro entre los pasos  $i$  e  $i+1$ , y  $\phi_i$  la dirección del paso  $i$ . Para los tiempos  $t_i$  asumimos distribuciones exponenciales independientes con parámetro  $\lambda$ , y para los ángulos de giro  $\omega_i$  distribuciones von Mises con media cero y concentración dada por el parámetro  $\kappa$  (Wu et al., 2000; Codling et al., 2008). Si bien el modelo puede ampliarse para permitir que  $\kappa$  y  $\lambda$  dependan del paisaje, del entorno o del comportamiento de los animales, para este trabajo sólo consideramos el caso inicial.

A continuación definiremos el proceso de observación y su conexión con el proceso de movimiento latente. Para ello, llamemos  $T_i$  al tiempo en el que se produjo el cambio de dirección  $i$ :

Tabla 2.1

Valores para las duraciones de paso y ángulos de giro del ejemplo de la Figura 2.1

<b>Duración de paso (<math>t'_i s</math>)</b>	0.2	0.2	0.7	0.4	0.4	0.8
<b>Ángulo de giro (<math>\omega'_i s</math>)</b>	0.32	5.65	5.81	0.02	0.11	5.81

$$T_0 = 0$$

$$T_i = \sum_{k=0}^{i-1} t_k \text{ for } j = 1, \dots, N_{steps}$$

Sea  $O_j = (o_{j,1}, o_{j,2})$  la posición de la observación  $j$  en coordenadas x-y, con  $j = 0, \dots, N_{obs}$ . Para notar el índice del tiempo sobre las observaciones utilizamos un segundo índice  $j$ .

Para completar la formulación es necesario determinar el número de cambios de dirección que se produjeron antes de una determinada observación. Definimos entonces  $N_j$  como el número de pasos (o cambios de dirección) que dio el animal desde el inicio de la trayectoria hasta el tiempo  $j\Delta t$ . Tenemos así que

$$O_0 = M_0$$

Y para  $j = 1, \dots, N_{obs}$

$$\begin{aligned} o_{j,1} &= m_{N_j,1} + \cos(\phi_{N_j}) (j\Delta t - T_{N_j}) \\ o_{j,2} &= m_{N_j,2} + \sin(\phi_{N_j}) (j\Delta t - T_{N_j}) \end{aligned} \quad (2.2)$$

Obsérvese que  $N_j$  es el índice que corresponde al tiempo máximo  $T_i$  menor o igual a  $j\Delta t$ , es decir,  $N_j = \max \{m \text{ tq } T_m \leq j\Delta t\}$ . Por lo tanto, la ubicación  $j$  es la última ubicación del proceso latente dado por  $N_j$  más la diferencia entre  $j\Delta t$  y el momento en que se produjo el paso  $N_j$  en la dirección  $\phi_{N_j}$ . Para comprender mejor esta relación, veamos un ejemplo de unos pocos pasos. La Figura 2.1 muestra una trayectoria suponiendo  $\Delta t = 0,5$  y duraciones de paso y ángulos de giro dados por la Tabla 2.1. En ese caso  $N_1 = 2$ , ya que  $T_1 = t_0 = 0,2 \leq 1\Delta t$ ,  $T_2 = t_0 + t_1 = 0,4 \leq 1\Delta t$  pero  $T_3 = t_0 + t_1 + t_2 = 1,1 \not\leq 1\Delta t$ . Con el mismo razonamiento  $N_2 = 2$ ,  $N_3 = 4$ ,  $N_4 = 5$ ,  $N_5 = 5$ , etc.

## 2.3. Inferencia

### 2.3.1. Expresión de la función de verosimilitud

Para poder obtener una expresión de la función de verosimilitud, formulamos primero la *verosimilitud completa de los datos*, es decir, asumimos que conocemos todos

### Ejemplo de pocos pasos

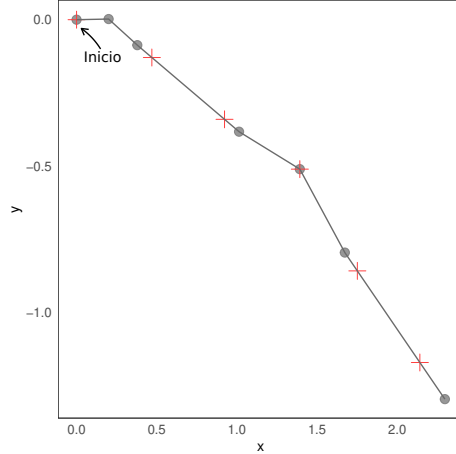


Figura 2.1: Pequeño ejemplo de simulación del modelo de movimiento con tiempos aleatorios entre decisiones de movimiento. Los cambios de dirección se indican con puntos grises y las observaciones con cruces rojas. La trayectoria se inicia en las coordenadas  $(0, 0)$ . Antes de las dos primeras cruces rojas (la primera no cuenta, pues  $i = 0$ ) hay dos puntos grises, luego el número de cambios previos a las observaciones 1 y 2 son  $N_1 = N_2 = 2$ . Antes de la tercera red roja hay cuatro puntos grises, luego  $N_3 = 4$ , etc.

los valores de  $t_i$  y  $\omega_i$ . En ese caso, como  $O|M$  es determinístico, se tiene

$$\begin{aligned}
 L(\kappa, \lambda; M, O) &= P(O, M|\kappa, \lambda) \\
 &= P(O|M, \kappa, \lambda)P(M|\kappa, \lambda) \\
 &= P(M|\kappa, \lambda) \\
 &= P(t_1, \dots, t_{N_{steps}}|\lambda)P(\omega_1, \dots, \omega_{N_{steps}}|\kappa)
 \end{aligned}$$

Supongamos ahora que no conocemos los valores de  $t_i$  y  $\omega_i$ , pero si conocemos el número de pasos que da el animal entre observaciones consecutivas,  $N_j \forall j$ . En ese caso, para completar la formulación sería necesario obtener la distribución de las variables  $M_i$  (Eq. 2.1) (o al menos una proporción de ellas), las cuales no poseen forma cerrada. El desarrollo de cuales serían los pasos necesarios para llegar a tal formulación están desarrollados en el [Apendice A.1](#).

Finalmente, para formular la probabilidad marginal,  $L(\kappa, \lambda)$ , es necesario además integrar sobre todos los posibles valores de  $N_j$ , lo que implica determinar las probabilidades  $P(N_j = r)$  para  $r \in \mathcal{N}$  (las cuales podrían ser infinito). Se tiene así que obtener la expresión de la función de verosimilitud, resulta ser una tarea compleja. Los métodos de “verosimilitud libre” evitan evaluar la función de verosimilitud, siendo útiles en casos como este.

### 2.3.2. Inferencia mediante ABC

Las metodologías de cómputo Bayesiano aproximado (ABC) son una familia de técnicas basadas en simulaciones que permiten obtener muestras de las distribuciones

posteriores de modelos con cuyas funciones de verosimilitud son intratables. En los últimos años, las técnicas de ABC se han popularizado en varios campos de estudio, tales como genética molecular (Marjoram y Tavaré, 2006), epidemiología (Tanaka et al., 2006; McKinley et al., 2009; Lopes y Beaumont, 2010), biología evolutiva (Bertorelle et al., 2010; Csilléry et al., 2010; Baudet et al., 2015), y ecología (Beaumont, 2010; Sirén et al., 2018). Este enfoque resulta útil también en los casos en que el esfuerzo de muestreo computacional para calcular la función de verosimilitud es grande en comparación con el de la simulación del modelo de interés. Veamos brevemente cómo funcionan.

Sea  $\boldsymbol{\theta}$  el vector de parámetros de interés, y sea  $\mathbf{y}$  el vector de observaciones. Se tiene que la distribución posterior  $p(\boldsymbol{\theta} | \mathbf{y})$  es proporcional al producto de la distribución previa  $p(\boldsymbol{\theta})$  y la función de verosimilitud  $p(\mathbf{y} | \boldsymbol{\theta})$

$$p(\boldsymbol{\theta} | \mathbf{y}) \propto p(\boldsymbol{\theta})p(\mathbf{y} | \boldsymbol{\theta})$$

La idea básica de los métodos ABC es obtener simulaciones de la distribución conjunta,  $p(\mathbf{y}, \boldsymbol{\theta})$  seleccionando aquellos valores de los parámetros que generan datos simulados “parecidos” a los datos observados ( $\mathbf{y}$ ). De esta manera, los métodos ABC intentan sustituir la función de verosimilitud por una medida de similitud entre los datos simulados y los reales.

El *algoritmo de rechazo simple* fue el primer método que se propuso para llevar a cabo ABC, siendo también el más sencillo de implementar (Tavaré et al., 1997; Pritchard et al., 1999). Considerando alguna medida resumen  $S(\cdot)$  para los datos, una medida de distancia  $\rho$ , un valor umbral  $\delta$ , y un valor natural  $K$  suficientemente grande, el método ABC de rechazo simple puede ser descrito de la siguiente manera:

1. Calcular el vector de estadísticos resumen de los datos observados,  $S(\mathbf{y})$ .
2. Simular parámetros  $\boldsymbol{\theta}_*$  a partir de  $p(\boldsymbol{\theta})$ , y datos  $\mathbf{y}_*$  a partir  $p(\cdot | \boldsymbol{\theta}_*)$ .
3. Calcular el vector de estadísticos resumen de los datos simulados,  $S(\mathbf{y}_*)$ .
4. Aceptar  $\boldsymbol{\theta}_*$  como muestra de la distribución posterior, si  $\rho(S(\mathbf{y}_*), S(\mathbf{y})) < \delta$ .
5. Repetir 2-4  $K$  veces.

El algoritmo de rechazo anterior produce muestras de  $p(\boldsymbol{\theta} | \rho(S(\mathbf{y}), S(\mathbf{y}_*)) < \delta$ , siendo esta una aproximación de  $p(\boldsymbol{\theta} | \mathbf{y})$ . Sucede que, cuando los estadísticos resumen son suficientes o casi suficientes para  $\rho$ , la distribución posterior aproximada converge a la verdadera distribución posterior cuando  $\delta \rightarrow 0$  (Marjoram et al., 2003). En lugar de seleccionar un valor para  $\delta$ , se suele establecer un umbral  $\epsilon$  como nivel de tolerancia para definir la proporción de simulaciones aceptadas. Para una revisión completa de los métodos y técnicas ABC, se puede consultar (Csilléry et al., 2010; Beaumont, 2010; Scott, A Sisson y Beaumont, 2018).

Además del método de rechazo simple, para este trabajo, consideramos dos métodos de corrección basados en regresión: uno por regresión lineal simple y el otro por corrección mediante redes neuronales. Los mismos implementan un paso adicional para corregir la coincidencia imperfecta entre los estadísticos resumen aceptados y los observados. Uno de ellos lo hace vía regresión lineal local (Beaumont et al., 2002), y

el otro utiliza redes neuronales (Blum y François, 2010). Para realizar la corrección, ambos métodos utilizan la ecuación de regresión dada por

$$\boldsymbol{\theta}_i = r(S(\mathbf{y}_i)) + \xi_i$$

donde  $r$  es la función de regresión y las  $\xi_i$  son variables aleatorias centradas en cero con igual varianza. Para la corrección lineal se asume que  $r$  es una función lineal, mientras que para la corrección mediante redes neuronales  $r$  no es necesariamente lineal. Se asigna un peso  $K[\rho(S(\mathbf{y}_i), S(\mathbf{y}))]$  (con  $K$  un cierto núcleo) a cada simulación, de modo que las más cercanas a los estadísticos resumen observados tienen mayor peso. Los valores de  $r$  y  $\xi$  pueden estimarse ajustando una regresión lineal en el primer caso y una regresión mediante redes neuronales prealimentadas (*feed-forward neural network regression*) en el segundo caso. Una vez ajustada la regresión, se obtiene una muestra ponderada de la distribución posterior considerando  $\theta_i^{corr}$  de la siguiente manera

$$\theta_i^{corr} = \hat{r}(S(\mathbf{y})) + \hat{\xi}_i$$

En donde  $\hat{r}(\cdot)$  es la media condicional estimada y  $\hat{\xi}_i$  son los residuos empíricos de la regresión.

Para ajustar el modelo de trayectorias propuesto, evaluamos 20 estadísticos resumen diferentes, y seleccionamos aquellos cuatro que mejor caracterizaban a las trayectorias según sus parámetros de movimiento. Buscando medidas resumen que capten diversas características de movimiento, graficamos los resúmenes propuestos vs los parámetros. Luego seleccionamos aquellos estadísticos resumen que cambiaban monótonamente según los valores de los parámetros. Los gráficos de todas las medidas resumen evaluadas se incluyen en el Apéndice A.2. Finalmente, los cuatro resúmenes seleccionados fueron: la inversa de la longitud media de los pasos observados (donde un paso observado es la distancia entre posiciones de tiempos observados consecutivos); un estimador puntual para  $\kappa$ ; la desviación estándar de los ángulos de giro observados y, por último, la desviación estándar de las longitudes de paso observadas (Tabla 2.2).

Para ajustar el modelo de trayectorias propuesto en este trabajo, utilizamos los dos métodos de corrección basados en regresión, y el método ABC de rechazo simple. Para realizar los análisis utilizamos el paquete de R “abc” (Csilléry et al. (2012), <http://cran.r-project.org/web/packages/abc/index.html>), el cual utiliza como medida de distancia para comparar los estadísticos resumen entre los datos simulados y los observados, la distancia euclidiana estandarizada.

## 2.4. Estudio de simulación

De manera de evaluar la capacidad de ajuste de las técnicas ABC para el modelo de trayectorias planteado, realizamos un análisis de simulación en dos pasos. Primero analizamos el rendimiento de los tres métodos ABC de manera general, y luego estudiamos que tan bien aproximan estos métodos la probabilidad posterior según la relación entre las escalas temporales de las trayectorias simuladas y sus observaciones. Para ambos experimentos utilizamos un conjunto de un millón de trayectorias simuladas, con parámetros  $\kappa$  (parámetro de dispersión para los ángulos de giro) y  $\lambda$  (parámetro para los tiempos entre cambios de dirección) muestreados según las previas  $p(\kappa) = U[0, 100]$

Tabla 2.2

*Estadísticos resumen seleccionados para ajustar los tres algoritmos de ABC*

Estadístico resumen	Fórmula
(1) Inversa de la longitud media de los pasos observados	$1 / \sum_{j=0}^{N_{obs}} \sqrt{(o_{j+1,1} - o_{j,1})^2 + (o_{j+1,2} - o_{j,2})^2}$ ,
(2) Estimador puntual de $\kappa$	$A^{-1} \left( \frac{\sum_{j=1}^{N_{obs}} \cos(\omega_{obs,j})}{N_{obs}} \right)$ Donde $\omega_{obs,j} = \arctan\left(\frac{o_{j+1,1}-o_{j,1}}{o_{j+1,2}-o_{j,2}}\right)$ y $A(x) = \frac{I_1(x)}{I_0(x)}$ (Kurt Hornik y Bettina Grün, 2014)
(3) Desviación estándar de los ángulos de giro observados	$\sqrt{\frac{\sum_{j=0}^{N_{obs}} \omega_{obs,j} - \omega_{obs}^-}{N_{obs} - 1}}$
(4) Desviación estándar de las longitudes de paso observadas	$\sqrt{\frac{\sum_{j=0}^{N_{obs}} t_{obs,j} - t_{obs}^-}{N_{obs} - 1}}$

y  $p(\lambda) = U[0, 50]$ . La elección de previas uniformes se debe a que el objetivo de nuestro estudio de simulación es analizar el rendimiento de los métodos ABC para ajustar el modelo bajo diversos escenarios. Las previas uniformes nos permiten considerar diferentes valores para los parámetros con igual probabilidad. El número de pasos simulados fue tal que todas las trayectorias tuvieron al menos 1500 observaciones y consideramos como tiempo fijo entre duraciones consecutivas  $\Delta t = 0, 5$ . El código R de este análisis está disponible en <https://github.com/sofiar/ABC-steps-and-turns->.

### 2.4.1. Evaluación de la capacidad inferencial de las técnicas ABC

Evaluamos el rendimiento de las tres versiones ABC: rechazo simple, rechazo corregido mediante regresión lineal y rechazo corregido mediante una red neuronal. Para diferentes valores de umbral ( $\epsilon$ ) y para cada versión del algoritmo realizamos un análisis de validación cruzada. Es decir, seleccionamos una trayectoria del conjunto de un millón de referencia y la utilizamos como si fuera la trayectoria real. La selección de la trayectoria real fue de manera aleatoria pero, para evitar el sesgo producido por el uso de previas uniformes, incluimos la condición de que los parámetros de la misma no estuvieran cerca del límite superior de la distribución previa ( $\lambda \leq 25$  y  $\kappa \leq 70$ ). Luego, estimamos los parámetros utilizando diferentes valores de umbral ( $\epsilon$ ) con los tres algoritmos y utilizando todas las simulaciones excepto la elegida. Este proceso se

replicó  $N_{rep} = 100$  veces. Para cada método y valor de  $\epsilon$ , registramos las muestras posteriores obtenidas tanto para  $\lambda$  como para  $\kappa$  y calculamos el error de predicción como

$$\sqrt{\frac{\sum_i (\tilde{\theta}_i - \theta_i)^2}{N_{rep}}} \quad \theta = (\lambda, \kappa); i = 1, \dots, N_{rep} \quad (2.3)$$

donde  $\theta_i$  es el valor verdadero del parámetro de la  $i$ -ésima corrida y  $\tilde{\theta}_i$  es la mediana de la posterior del parámetro. También calculamos una medida de dispersión de los errores en relación con la magnitud de los parámetros para cada método y cada valor de tolerancia. Llamamos índice de dispersión ( $ID$ ) a esta cantidad y la calculamos como

$$ID = \left[ \sum_i \frac{|\tilde{\theta}_i - \theta_i|}{\theta_i} \right] / N_{rep} \quad \theta = (\lambda, \kappa); i = 1, \dots, N_{rep} \quad (2.4)$$

Además, para evaluar si la dispersión de las distribuciones posteriores no era excesivamente grande o pequeña, calculamos la cobertura empírica del intervalo de credibilidad para  $\alpha = 95$  de los dos parámetros involucrados y para los diferentes umbrales ( $\epsilon$ ). La cobertura empírica es la proporción de simulaciones en las que el valor real del parámetro cae dentro del intervalo de credibilidad establecido. Si los niveles de confianza nominales fueran precisos, esta proporción debería estar cerca de 0,95. Si esto es cierto para todos los valores de  $\alpha$ , se dice que el análisis satisface la *propiedad de cobertura*. Una forma de comprobar esta propiedad es realizar una Prueba de Cobertura (*Coverage Test*), la cual sirve también para elegir el valor del umbral  $\epsilon$ . La Prueba de Cobertura fue introducida por primera vez por [Prangle et al. \(2014\)](#) y la idea básica es aplicar ABC en muchos conjuntos de datos simulados, y para cada uno de ellos calcular  $p$ , la proporción de valores de la distribución posterior que resultan más pequeños que el parámetro verdadero. Lo ideal es que estos valores se distribuyan como una  $U(0, 1)$ . Para una descripción completa de esta prueba, se puede consultar ([Prangle et al., 2014](#)). Para realizar esta prueba utilizamos el paquete de R “abctools” ([Nunes y Prangle \(2015\)](#)); <https://cran.r-project.org/web/packages/abctools/index.html>).

## 2.4.2. Relación entre las escalas temporales y la capacidad inferencial

Continuamos el análisis evaluando que tan bien estos métodos aproximan las probabilidades posteriores en función de la relación entre la escala temporal de observación ( $\Delta t$ ) y la escala temporal de los cambios de dirección ( $1/\lambda$ ) (Figura 2.2). Definimos

$$R = \frac{\lambda}{1/\Delta t}$$

Si  $R = 1$  entonces  $\lambda = 1/\Delta t$ , lo que significa que el tiempo entre observaciones consecutivas es igual al tiempo medio entre cambios de dirección. Por el contrario, si

## Ejemplos de simulación

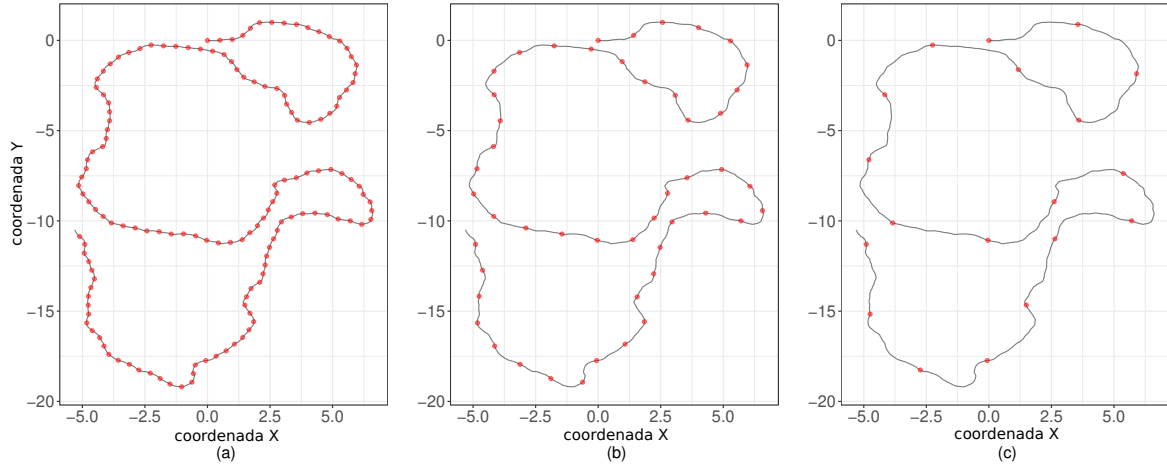


Figura 2.2: Esquema de muestreo y escala temporal de los cambios de dirección. Con líneas negras el proceso de movimiento y con puntos rojos la observación. (a) caso de sobremuestreo, (b) caso equilibrado, (c) caso de submuestreo.

$R < 1$  entonces la escala de tiempo entre observaciones consecutivas es menor que la escala de tiempo en la que los animales deciden cambiar de dirección (caso de sobremuestreo), y lo contrario ocurre si  $R > 1$  (caso de submuestreo) (Figura 2.2). Consideramos diferentes valores de  $R$  (entre 0,06 y 5) y para cada uno de ellos simulamos 50 trayectorias con valores de  $\kappa \in \{10, 20, 30, \dots, 70\}$ . Luego, utilizando el millón de trayectorias originales, calculamos las distribuciones posteriores mediante los tres métodos de ABC considerando estas nuevas trayectorias como las verdaderas. Por último, calculamos el error de predicción de  $\kappa$  y  $\lambda$  para cada combinación de  $R$  y  $\kappa$ .

### 2.4.3. Resultados

La figura 2.3 muestra los valores de los errores de predicción y el índice de dispersión ( $ID$ ) para cada método ABC y para cada valor  $\epsilon$  obtenido según el análisis de validación cruzada. En todos los casos, los errores de predicción disminuyen cuando el valor del umbral ( $\epsilon$ ) también disminuye. Sin embargo, los algoritmos corregidos mediante regresión lineal y redes neuronales obtienen menor error de predicción con niveles altos de  $\epsilon$ . Algo similar ocurre con el índice  $MD$ : valores de umbral más bajos implican valores más pequeños de este índice (Figura 2.3). Este índice nos permite tener una idea de la amplitud de las distribuciones posteriores. Es evidente que para el caso del algoritmo de rechazo las distribuciones posteriores son bastante amplias, especialmente cuando  $\epsilon = 0, 1$ . Sin embargo, para los algoritmos corregidos la diferencia entre los parámetros estimados y los verdaderos es de hasta aproximadamente 1,3 unidades para  $\kappa$  y 0,3 para  $\lambda$  en el mejor caso ( $\epsilon = 0,001$ ). La figura 2.4 muestra que la estimación de  $\lambda$  mejora cuando toma valores más bajos, especialmente para el algoritmo corregido mediante regresión lineal. Discutiremos este punto en la sección 2.6. En base a estos resultados, el algoritmo corregido vía regresión lineal parece ser el de mejor rendimiento.

### Resultados para el error de predicción e índice de dispersión

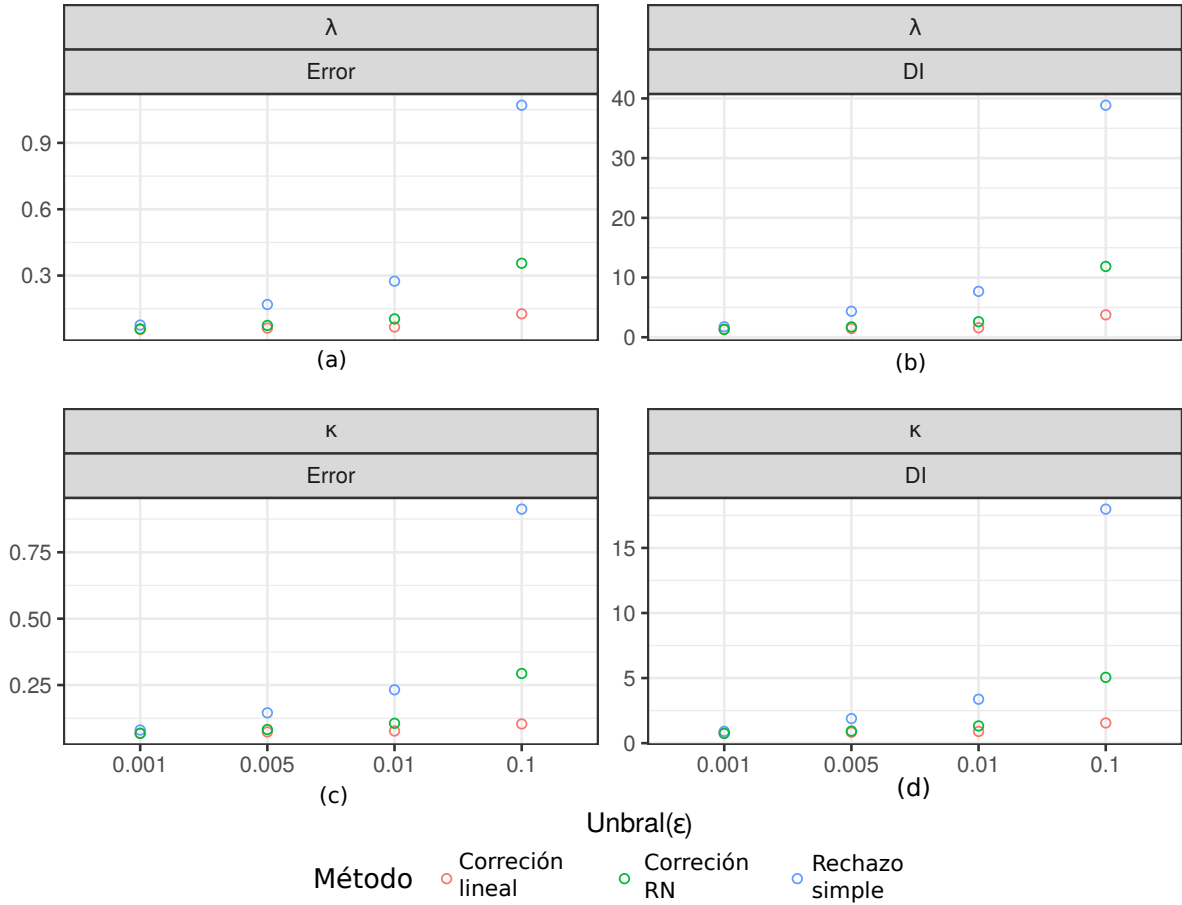


Figura 2.3: Valores para el error de predicción (Ecuación 2.3) y para el índice de dispersión (Ecuación 2.4) para los dos parámetros obtenidos según cada algoritmo de ABC considerado y para cada valor de umbral  $\epsilon$

Con respecto a la cobertura empírica, obtuvimos que casi siempre estos índices fueron superiores al 95%, excepto para los casos en donde  $\epsilon = 0,1$  y el algoritmo de ABC propuesto es rechazo simple o el corregido mediante red neuronal. En esos dos casos, las coberturas empíricas estuvieron un poco por debajo de 0,95, indicando que las dispersiones de las posteriores obtenidas podrían ser demasiado grandes. Los gráficos de este análisis se presentan en el Apéndice A.3.

Para comprobar la propiedad de cobertura realizamos la prueba de cobertura para ambos parámetros (Figura 2.5). En la mayoría de los casos, las distribuciones obtenidas no muestran una clara aproximación a una  $U(0,1)$ . Sin embargo, hay una evidente diferencia entre los histogramas obtenidos para el modelo ABC de rechazo simple y los histogramas obtenidos con los otros dos algoritmos. Las formas del ABC de rechazo simple son las que más se alejan de la uniformidad: para ambos parámetros las distribuciones de los valores  $p$  son sesgadas a la izquierda, lo que indica que el algoritmo tiende a sobreestimar los parámetros. Para los otros dos algoritmos el sesgo a la izquierda es mucho más moderado, e incluso en el caso de los valores más bajos de  $\epsilon$  para el algoritmo lineal los histogramas son más uniformes, lo que indica que se podría estar alcanzando su *cobertura*.

## Resultados del análisis de validación cruzada

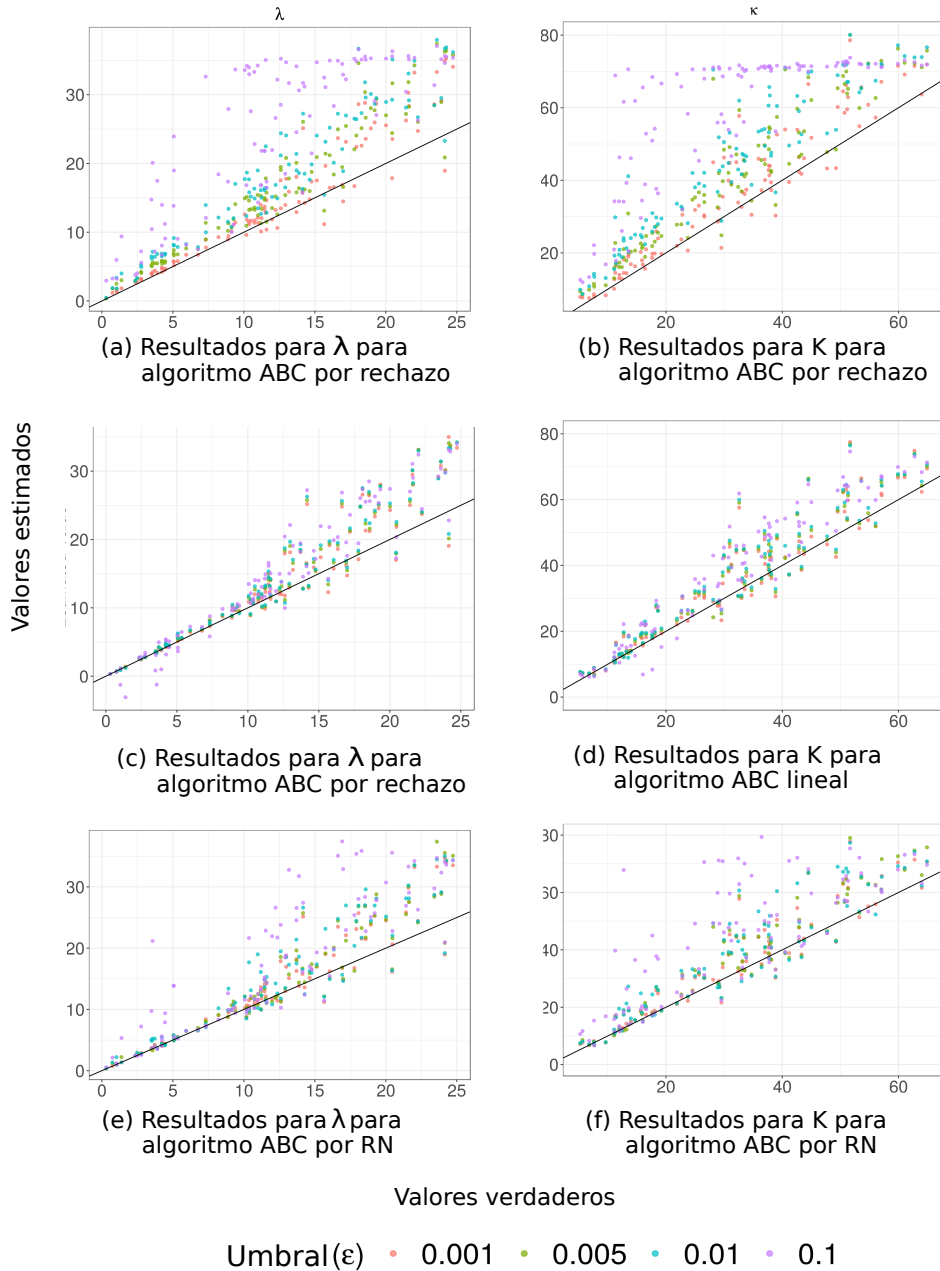


Figura 2.4: Resultados del análisis de validación cruzada para el algoritmo de rechazo y las dos versiones corregidas. Se muestra la relación entre los parámetros verdaderos y la mediana de la distribución posterior estimada. Con diferentes colores los resultados según los distintos para el umbral  $\epsilon$ : violeta para  $\epsilon = 0,1$ , azul para  $\epsilon = 0,01$ , verde para  $\epsilon = 0,005$  y rojo para  $\epsilon = 0,001$ . La línea negra indica la relación ideal  $x = y$ .

## Análisis de cobertura

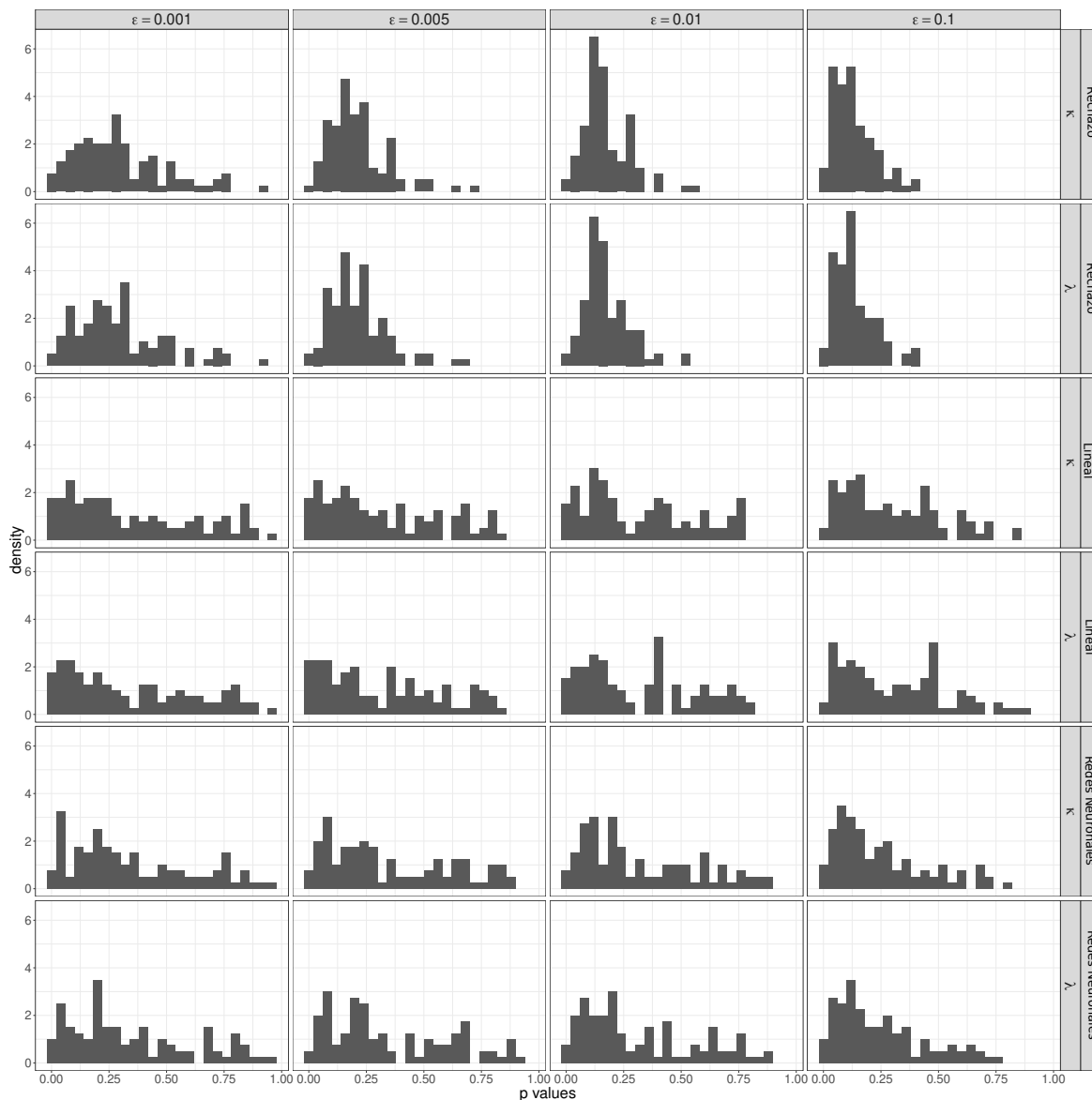


Figura 2.5: Frecuencia relativa  $p$  de los parámetros aceptados que fueron inferiores al valor real en los análisis. Por columna los resultados para diferentes valores de  $\epsilon$ . Las dos primeras filas corresponden a los resultados del algoritmo de rechazo, las filas tres y cuatro al algoritmo corregido por regresión lineal, y las filas cinco y seis al algoritmo ABC corregido por redes neuronales

## Errores de predicción para $\lambda$

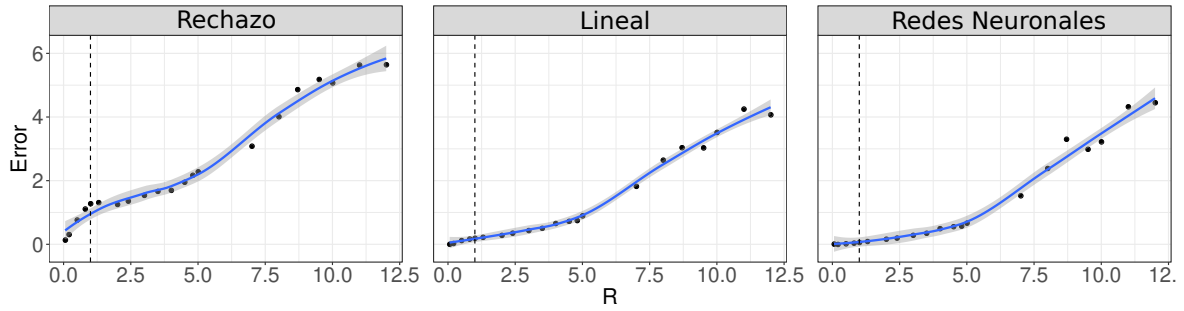


Figura 2.6: Gráfico de dispersión suavizado (por regresión local) de los errores de predicción del tiempo esperado entre cambios de dirección ( $\lambda$ ) para diferentes relaciones entre la escala temporal de observación y la escala de cambios de dirección ( $R$ ). Los valores altos de  $R$  indican que la escala temporal del proceso de observación es mayor que la escala temporal del proceso de decisión del movimiento. Con puntos negros los errores de predicción de  $\lambda$  para cada valor de  $R$ . La línea azul indica la curva suavizada para esos valores. Los intervalos de 95 % se muestran en gris. La línea vertical de puntos indica  $R = 1$ .

Para evaluar la importancia de la relación entre la escala temporal del proceso de observación y aquella en la que se producen los cambios en el proceso latente, evaluamos el ajuste de los dos parámetros en relación con la tasa  $R$ . Obtuvimos que los errores de predicción para  $\lambda$  aumentaron a medida que el valor de  $R$  aumenta (Figura 2.6). Para el caso de  $\kappa$ , esta misma relación se puede observar cuando el valor real de este parámetro toma valores grandes ( $\kappa > 60$ ) (Figura 2.7). Nuevamente los algoritmos corregidos presentan errores menores para ambos parámetros.

Según los resultados anteriores, es evidente que existe una relación entre el coeficiente  $R$  y la capacidad de estos métodos para estimar los parámetros. Para valores de  $R$  aproximadamente inferiores a 5 los errores son pequeños y es posible obtener buenas estimaciones. Para ello es necesario que la escala temporal del proceso de observación sea aproximadamente inferior a 5 veces la escala temporal en la que los animales deciden cambiar de dirección. Para valores más grandes de  $\Delta t$  resulta más difícil hacer inferencias con esta técnica.

## 2.5. Análisis datos de oveja

Evaluamos el rendimiento de este modelo utilizando datos de una trayectoria real reconstruida en alta resolución con información de dispositivos DailyDiary y GPS. Utilizando una trayectoria de alta resolución es posible deducir los momentos en los que el animal cambia de dirección para luego utilizar esta información para estimar la media de la longitud del paso y la dispersión de los ángulos de giro. De esta manera se pueden obtener ciertos valores de parámetros “verdaderos” para luego ver si se pueden recuperar con las técnicas ABC.

Para este análisis utilizamos datos de rastreo proveniente de una oveja tomados

### Errores de predicción para $\kappa$

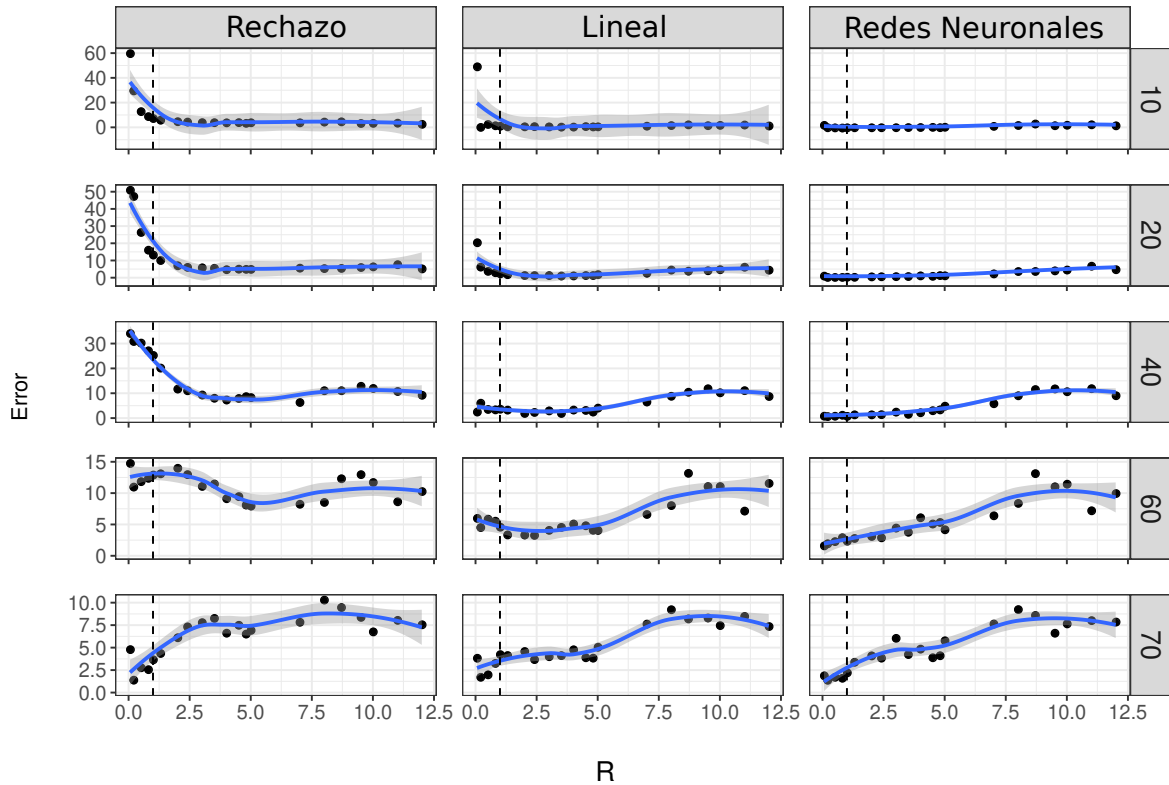


Figura 2.7: Gráfico de dispersión suavizado (por regresión local) de los errores de predicción del parámetro de dispersión para los ángulos de giro ( $\kappa$ ) para diferentes relaciones entre la escala temporal de observación y la escala de cambios de dirección ( $R$ ). Por filas los diferentes valores de  $\kappa$  y por columnas los diferentes algoritmos ABC. Los valores altos de  $R$  indican que la escala temporal del proceso de observación es mayor que la escala temporal del proceso de decisión del movimiento. Con puntos negros los errores de predicción de  $\kappa$  para cada valor de  $R$ . La línea azul indica la curva suavizada para esos valores. Los intervalos de 95% se muestran en gris. La línea vertical de puntos indica  $R = 1$ .

en Bariloche, Argentina durante los meses de febrero y marzo de 2019. La oveja estaba equipada con un collar que contenía un GPS (CatLog-B, Perthold Engineering, [www.perthold.de](http://www.perthold.de); USA), que fue programado para registrar datos de localización cada cinco minutos, y un dispositivo DailyDiary (DD, [Wilson et al. \(2008a\)](#)), que fue programado para registrar 40 datos de aceleración por segundo (frecuencia de 40Hz) y 13 datos de magnetómetro por segundo (frecuencia de 13Hz). Los DD son dispositivos electrónicos que miden la aceleración y el magnetismo en relación al cuerpo del animal en tres dimensiones. Estos datos permiten reconstruir con alta resolución la trayectoria de un animal mediante el uso de las técnicas de Dead-Reckoning ([Wilson y Wilson, 1988](#); [Wilson et al., 2007](#)). El objetivo aquí es utilizar esta trayectoria detallada para obtener los momentos reales de cambio de dirección, para luego inferir sobre los parámetros de las distribuciones de largo de paso y ángulo de giro. Posteriormente, podemos comparar las estimaciones anteriores con las obtenidas mediante el enfoque ABC.

De los datos originales seleccionamos al azar un segmento de 6 horas. Utilizando la información del DD, primero estimamos el camino recorrido (pseudotrack) por la oveja utilizando la técnica de Dead-Reckoning. ([Wilson y Wilson, 1988](#); [Wilson et al., 2007](#)). Para ello utilizamos el paquete de R “TrackReconstruction” (<https://cran.r-project.org/web/packages/TrackReconstruction/index.html>). Después, corregimos el posible sesgo de estas estimaciones utilizando los datos de GPS ([Liu, Y. et al., 2015](#)) mediante el paquete de R “BayesianAnimalTracker” (<https://cran.r-project.org/web/packages/BayesianAnimalTracker/index.html>). Así, obtuvimos una trayectoria muestreada con una resolución de 1 segundo. Para satisfacer las hipótesis del modelo, seleccionamos una parte de esa trayectoria que parecía proceder del mismo comportamiento, es decir, seleccionamos un trozo de la trayectoria que visualmente parecía tener la misma distribución de ángulos de giro y longitudes de paso.

Para estimar los parámetros de la trayectoria fue necesario determinar los puntos en los que había un cambio de dirección del movimiento. Para ello, aplicamos el algoritmo propuesto por Potts. et al. ([Potts et al., 2018](#)), el cual detecta los cambios de giro de la trayectoria utilizando los datos de las direcciones de los animales para posteriormente estimar los largos de paso y los ángulos de giro. Así de esta manera es posible obtener, no solo los valores de los  $N_j$ , sino también muestras para la longitud de paso y los ángulos de giro. Con esa información fue realmente sencillo inferir los valores de los parámetros mediante técnicas de MCMC y obtener muestras de la distribución posterior conjunta utilizando el software Stan ([Carpenter et al., 2017b](#)).

Luego calculamos los estadísticos resumen de la trayectoria observada considerando  $\Delta t = 50$  segundos (una observación cada 50 de la trayectoria reconstruida) y aplicamos los tres algoritmos ABC. Por último, comparamos ambas estimaciones.

### 2.5.1. Resultados

La trayectoria seleccionada fue del día 27 de Febrero de 2019 desde las 19 : 01 : 21hs hasta las 20 : 02 : 00hs, un total de 1,01 horas (Figura 2.8). Estimamos la distribución posterior de cada parámetro a partir de una muestra de 3000 observaciones MCMC independientes. Como los parámetros obtenidos mediante esta estimación se

## Resultados para el ejemplo con datos reales

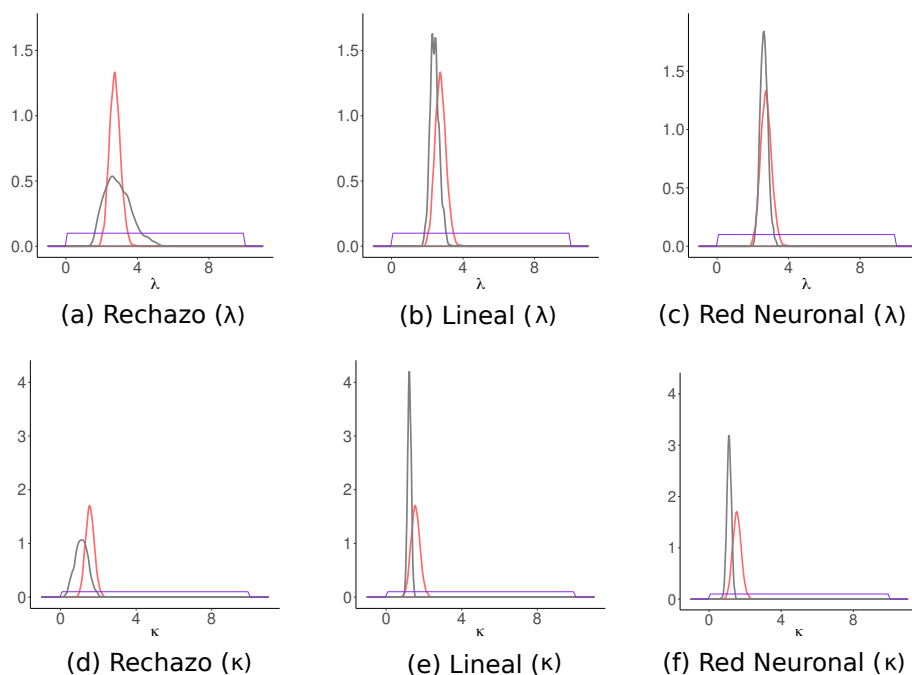


Figura 2.9: Con rojo las distribuciones posteriores obtenidos para ambos parámetros  $\kappa$  y  $\lambda$  mediante MCMC y con gris las obtenidas con cada algoritmo ABC. Por fila los resultados para cada parámetro y por columna los resultados para cada algoritmo ABC.

encontraban en los límites inferiores de las distribuciones previas consideradas para el esquema de simulación, simulamos un nuevo conjunto de trayectorias con previas de  $\kappa \sim U[0, 10]$  y  $\lambda \sim U[0, 10]$  y calculamos las inferencias ABC con esas simulaciones.

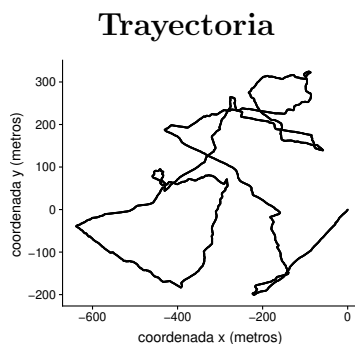


Figura 2.8: Trayectoria final con una resolución de 1 segundo reconstruida por Dead-Reckoning y corregida con la información del GPS.

Las distribuciones posteriores obtenidas mediante MCMC y mediante los algoritmos ABC dieron resultados similares (Figura 2.9). De nuevo, el algoritmo ABC de rechazo simple produjo la estimación menos exacta, es decir, la posterior es la más alejada de la obtenida por MCMC. Aunque esta trayectoria es sólo un ejemplo simple, muestra que es posible aplicar este modelo a trayectorias reales de animales.

## 2.6. Discusión

Los estudios de datos de movimiento animal pueden ser analizados o bien a partir de formulaciones continuas o bien a partir de formulaciones discretas. Los modelos en tiempo continuo son más realistas, pero a menudo más difíciles de interpretar que las versiones discretas (McClintock et al., 2014). Un compromiso entre estos enfoques es modelar movimiento como pasos y giros, pero que la duración de los pasos (o los momentos en que se realizan los giros) se produzca en tiempo continuo. Aquí consideramos que el proceso de movimiento subyacente, que evoluciona en tiempo continuo se observa a intervalos de tiempo regulares, como sería habitual en un animal terrestre provisto de un collar GPS. La función de verosimilitud resulta compleja de calcular, pero es posible generar rápidamente simulaciones a partir del proceso de movimiento y de la observación. Por ello, propusimos utilizar métodos ABC para hacer inferencia bajo este modelo. Aunque estas técnicas mostraron ciertas limitaciones, fue posible obtener estimaciones precisas de los parámetros cuando la escala temporal de las observaciones no era demasiado grande en comparación con la escala de los cambios de dirección.

Nuestro estudio de simulación mostró que el ABC de rechazo simple no funciona bien para el modelo espacio estado propuesto, pero la versión corregida de este algoritmo mejora realmente las estimaciones (figuras 2.3 y 2.4). En general, el mejor rendimiento se obtuvo con algoritmo ABC por corrección lineal. Sin embargo, la aplicabilidad de estos métodos depende en gran medida de la relación entre la escala temporal del proceso de observación y el tiempo medio entre los cambios de dirección del movimiento. Encontramos que cuando esta relación es inferior a 5 unidades, es posible recuperar los parámetros que gobierna las trayectorias (figuras 2.6 y 2.7). Es decir, resulta necesario que los tiempos en los que se toman las observaciones sean como mucho 5 veces menores que la media de los tiempos reales entre cambios de dirección para poder generar buenas estimaciones de este modelo.

El modelo de movimiento presentado aquí es bastante simple, ya que suponemos velocidad de movimiento constante. No obstante, el modelo supone una mejora con respecto a las versiones de tiempo discreto en las que la escala temporal del movimiento tiene que coincidir con la escala de las observaciones. Futuros desarrollos de estos métodos deberían considerar características adicionales que son comunes en los estudios de movimiento, como la inclusión de más de un comportamiento de movimiento y el efecto de las características del hábitat tanto en los parámetros de movimiento como en los cambios entre comportamientos (Morales et al., 2004; Mevin B. Hooten et al., 2017). Aunque esto signifique estimar varios parámetros, tales modelos implicarán una mayor estructura en las trayectorias que podría utilizarse como parte de los estadísticos resumen utilizados para caracterizar los datos. Por lo tanto, se podría reducir la combinación de valores de los parámetros capaces de reproducir las características presentes en los datos, permitiendo así buenas inferencias con ABC.

En general, los procesos que subyacen el movimiento de un individuo y los procesos que afectan la forma en que registramos su trayectoria suelen operar a escalas de tiempo diferentes, lo que dificulta el análisis y la comprensión de los primeros en función de los segundos. El modelo espacio-estado utilizado aquí nos permitió conectar estas dos escalas de forma intuitiva e interpretable. Nuestros resultados subrayan la

idea de que la escala temporal a la que toman las decisiones de movimiento los animales debe tenerse en cuenta a la hora de diseñar los protocolos de colecta de datos, y que no siempre es necesario contar con datos de alta frecuencia para tener buenas estimaciones de ciertos procesos de movimiento.



# 3. Modelo de clasificación de comportamientos con datos de acelerometría

---

## 3.1. Introducción

El objetivo en el problema de clasificación es asignar datos a diferentes grupos o descubrir conjuntos de patrones que reflejen dinámicas de interés importantes. En el caso de los sistemas que presentan dependencia temporal, el objetivo es asignar correctamente diferentes segmentos de los datos a un conjunto finito de grupos, teniendo en cuenta que es probable que observaciones cercanas entre sí (en el tiempo) correspondan al mismo grupo. Por ejemplo, en lingüística computacional el objetivo es identificar palabras y frases en el lenguaje hablado ([Juang y Rabiner, 1991](#); [Deng y Li, 2013](#)); en meteorología, el cambio del tiempo puede ser monitoreado analizando medidas secuenciales de radares meteorológicos ([Rico-Ramirez y Cluckie, 2008](#); [Ruiz-Suarez et al., 2019](#)); en neurofisiología, se pueden distinguir diferentes actividades cerebrales evaluando variables fisiológicas como la frecuencia cardíaca o los electrocardiogramas a través del tiempo ([Cheng y Chan, 1998](#); [Inan et al., 2006](#)); y en ecología, es posible identificar diversos comportamientos de los animales biológicamente relevantes a partir de datos de aceleración ([Nathan et al., 2012](#); [Leos-Barajas et al., 2017](#)).

Para entender cómo se mueven los animales en respuesta a las condiciones internas y a los entornos externos, es esencial poder distinguir entre un conjunto de comportamientos biológicamente relevantes. En la actualidad, existen datos de aceleración triaxial (ACC), los cuales habitualmente se registran mediante dispositivos de biotelemetría que permiten investigar el rendimiento, el gasto energético y el comportamiento de los animales de vida silvestre ([Williams et al., 2020](#)). Estos dispositivos miden el cambio de velocidad a lo largo del tiempo en tres direcciones que pueden describirse en relación al cuerpo del individuo. Con esta información es posible identificar diferentes patrones de actividad para luego distinguir entre diferentes comportamientos ([Wilson et al., 2008b](#); [Williams et al., 2015](#)).

Existen varias técnicas utilizadas para resolver problemas de clasificación ([Trevor Hastie et al., 2001](#)): árboles de clasificación, regresión logística, análisis discriminante, redes neuronales, árboles de regresión potenciados, bosques aleatorios, métodos de aprendizaje profundo, vecinos más cercanos, máquinas de vectores de soporte, etc. Muchas de estas técnicas han sido propuestas para clasificar el comportamiento de animales a partir de datos de acelerómetros. Por ejemplo, [Nathan et al. \(2012\)](#) identificó modos de comportamiento de buitres leonados utilizando una selección de métodos no lineales y de árboles de decisión; ([Carroll et al., 2014](#)) entrenó una máquina de vectores de soporte para clasificar el comportamiento de pingüinos según “manipulación de presas” o “natación”; ([Williams et al., 2015](#)) examinó la capacidad de los algorit-

mos de  $k$ -vecinos más cercanos para distinguir entre los comportamientos de vuelo de los cóndores andinos y los buitres leonados; (Chakravarty et al., 2019) desarrolló un modelo híbrido que combina características biomecánicas y máquinas de vectores de soporte para identificar entre cuatro posibles comportamientos de suricatas de Kalahari; y (Studd et al., 2019) utilizó un algoritmo de bosque aleatorio y un árbol de decisión creado manualmente para asociar los comportamientos observados de ardillas rojas norteamericanas según aceleración y temperatura. Todos estos métodos suponen que las observaciones son independientes, pero los datos de las series temporales presentan una correlación secuencial. Esta característica de los datos puede aprovecharse para mejorar la precisión de la predicción de los clasificadores (Geurts, 2001; Dietterich, 2002).

La clase de modelos ocultos de Markov (HMM) (Zucchini et al., 2017; Frühwirth-Schnatter et al., 2019) proporciona un marco intuitivo para la clasificación de sistemas con dependencia temporal que experimentan cambios en los patrones a lo largo del tiempo conectados con cambios subyacentes en un proceso latente de interés. Los HMM asumen que la(s) observación(es) en cada punto del tiempo son el resultado del “estado” (o clase) desconocido (oculto) del sistema. Se supone que el estado del sistema cambia a lo largo del tiempo según la propiedad de Markov, es decir, que la distribución de probabilidad condicional de los estados futuros (condicionada a los estados pasados y presentes) sólo depende del estado presente. Así, un HMM se define como un doble proceso estocástico compuesto por un proceso de observación  $X_t$  y un proceso de estado (latente)  $C_t$ , donde el proceso de estado se suele considerar una cadena de Markov de primer orden y la distribución de  $X_t$  depende sólo del estado actual  $C_t$  y no del estado u observaciones anteriores. Como tal, los HMM proporcionan una forma clara de hacer clasificaciones en procesos que evolucionan en el tiempo, ya que tienen en cuenta la dependencia secuencial presente en los datos y la estructura temporal de los estados consecutivos.

Los HMM pueden utilizarse para la predicción de estados (enfoque supervisado) o para hacer inferencias sobre los procesos comportamentales (enfoque no supervisado). En el contexto del aprendizaje supervisado, la información sobre las clases de las observaciones está disponible por adelantado (hay un conjunto de datos preetiquetados), y el número de estados posibles es conocido por el usuario. En cambio, en el contexto del aprendizaje no supervisado no hay datos etiquetados y el número de estados no está predefinido. Los HMM se han implementado con éxito para clasificar datos de acelerómetros: (Li et al., 2010) utilizó un HMM para reconocer actividades físicas humanas utilizando características resumen de dos segundos a partir de datos de aceleración triaxial; (Wang et al., 2011) presentó un HMM para reconocer seis actividades diarias humanas a partir de señales de sensores recogidas de un único acelerómetro triaxial de cintura; (Leos-Barajas et al., 2017) proporcionó los detalles necesarios para implementar y evaluar un HMM tanto en el contexto de aprendizaje supervisado como no supervisado y esbozó dos aplicaciones a sistemas marinos y aéreos (tiburón y águila) utilizando aprendizaje no supervisado.

Los HMMs son modelos formulados mayoritariamente en tiempo discreto, y como se supone que el proceso de estado es una cadena de Markov de primer orden, el número de tiempos consecutivos que el sistema pasa en un estado determinado (tiempo de permanencia), sigue una distribución geométrica (Zucchini et al., 2017; Langrock

y Zucchini, 2011). La popularidad de los HMM se debe, en parte, a la facilidad con la que se pueden ampliar para dar lugar a otras formas de dependencia y estructuras. Por ejemplo, la distribución de duración de los estados puede generalizarse de manera que el proceso estocástico subyacente sea una cadena semi-Markoviana. Estos modelos se denominan modelos ocultos semi-Markovianos (HSMM) (Yu, 2010), y al ser más flexibles permiten un mayor realismo, mejorando las clasificaciones cuando las distribuciones de los tiempos de permanencia están lejos de ser geométricas. Los HSMMs se han aplicado con éxito en muchas áreas, sobre todo para reconocimiento del habla (Chen et al., 2006; Hung-Yan Gu et al., 1991; Hieronymus et al., 1992; Oura et al., 2006), pero también para clasificar actividades diarias de humanos (Duong et al., 2005; Chung y Liu, 2008), identificar escritura a mano (Kashi et al., 1997; Benouareth et al., 2007), y para el reconocimiento de genes en ADN (Kulp et al., 1996). Los HSMM también se han utilizado para la identificación de idiomas (Marcheret y Savic, 1997), la predicción de estructuras de proteínas (Aydin et al., 2006), el reconocimiento de eventos en vídeos (Hongeng y Nevatia, 2003), la modelización de series temporales financieras (Bulla y Bulla, 2006), la clasificación de música (XiaoBing Liu et al., 2008), y la teledetección (Pieczynski, 2007), entre otros.

Los HMMs y HSMMs clásicos asumen independencia entre las observaciones condicionadas por el estado, pero a veces los datos se toman con alta resolución temporal, lo que hace que esta suposición no sea realista y pueda afectar al rendimiento de los clasificadores. En estos casos, se puede considerar estructura autorregresiva para modelar las observaciones de cada estado, dando lugar así a los HMMs y HSMMs autorregresivos (AR(p)-HMM y AR(p)-HSMM) (Xu y Liu, 2020), también conocidos comúnmente como modelos de conmutación de Markov. Aunque estos modelos se han estudiado ampliamente, hasta donde sabemos, estas extensiones de los HMM no se han utilizado para clasificar comportamientos de animales según datos de aceleración.

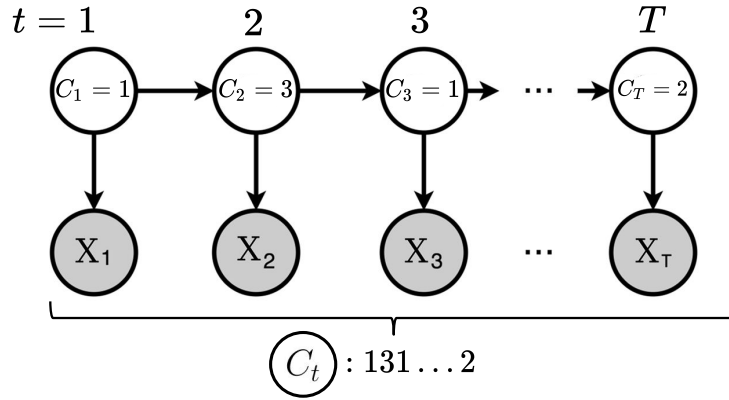
En este capítulo, damos una visión general de los modelos HMM, HSMM, AR(p)-HMM y AR(p)-HSMM; explicamos la estructura de estos modelos en el contexto de la clasificación supervisada, cómo pueden relajarse y sus diferencias en el modelado, la inferencia, la estimación y la predicción (sección 3.3). Presentamos sus formulaciones y derivaciones como una recopilación de la literatura publicada sobre este tema extendiéndolas a HSMMs con o sin estructura autorregresiva (sección 3.2). A continuación, estudiamos cómo estos modelos realizan clasificaciones en diferentes escenarios para caracterizar cuándo sería importante extender un HMM a un HSMM (sección 3.4). Finalmente, los utilizamos para clasificar datos de acelerómetros de ovejas merinas domésticas y distinguir entre cuatro estados de comportamiento diferentes (sección 3.5).

## 3.2. Modelos

### 3.2.1. Modelos ocultos de Markov (HMM)

Los HMM se componen de dos capas: un proceso latente  $\{C_t\}_{t=1}^T$ , comúnmente denominado proceso de estado, que satisface la propiedad de Markov; y un proceso ob-

**(a) HMM**



**(b) HSMM**

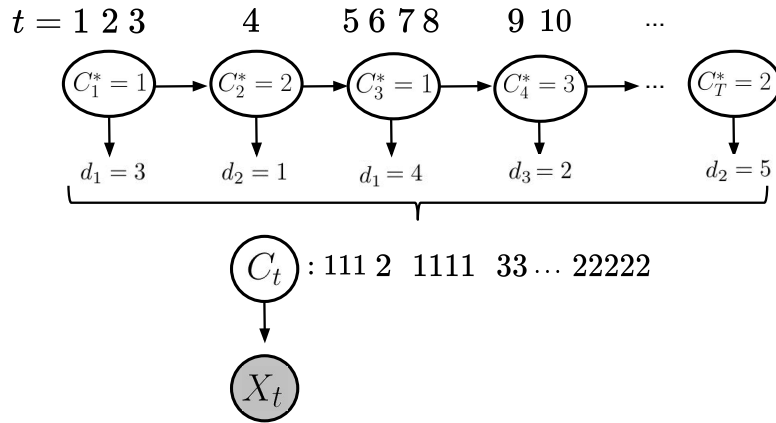


Figura 3.1: Diagramas de las estructuras de los modelos (a)HMM:  $C_t$  nota el proceso Markoviano latente y  $X_t$  el proceso observaciones cuyas distribuciones dependen del estado  $C_t$ . (b) Ejemplo HSMM:  $C_t$  nota el proceso latente semi-Markoviano y  $X_t$  el proceso de observación.  $C_t^*$  indica el proceso Markoviano de los tiempos no absorbentes (es decir, el estado a tiempo  $t$  es igual al estado a tiempo  $t - 1$ ), y para cada uno de ellos  $d_t$  da el valor muestreado para los tiempos de permanencia.

servable dependiente del estado  $\{X_t\}_{t=1}^T$ , donde  $P(X_t|X_1, X_2, \dots, X_{t-1}, C_t) = P(X_t|C_t)$  (Figura 3.1 (a)). En cada punto del tiempo  $t$ , se supone que  $C_t$  toma uno de los  $J$  valores posibles,  $C_t \in \{1, 2, \dots, J\}$ , donde  $J$  denota el número de estados.

La dinámica del proceso de estado  $C$  está gobernada por una matriz de probabilidades de transición de dimensión  $J \times J$  (m.p.t),  $\mathbf{\Gamma}$ , con entradas  $\gamma_{ij} = \Pr(C_{t+1} = j|C_t = i)$ , for  $i, j \in \{1, \dots, J\}$ . Las probabilidades de los estados iniciales se notan como  $\delta_i = \Pr(C_1 = i)$ ,  $i, j = 1, \dots, J$ .

Por definición, el supuesto Markoviano de primer orden para el proceso de estado implica que el número de tiempos consecutivos que el proceso pasa en un estado dado antes de cambiar, es decir el tiempo de permanencia, sigue una distribución geométrica con parámetro  $\gamma_{ii}$ . En consecuencia, el tiempo de permanencia más probable para cada estado en un HMM es uno, y la probabilidad de permanecer en un estado de-

terminado decae geoméricamente. Para completar la definición de un HMM, notamos las distribuciones dependientes del estado del proceso de observación (univariante o multivariante) como  $f_j(x) = \Pr(X_t = x|C_t = j)$ , para  $j = 1, \dots, J$ . Las distribuciones dependientes del estado pueden ser discretas o continuas y es común asumir la misma distribución paramétrica en todos los estados. Por lo tanto, un HMM se define por los dos procesos estocásticos  $\{X_t, C_t\}_{t=1}^T$ , las distribuciones dependientes del estado  $f_1(x), \dots, f_J(x)$ , la matriz de probabilidad de transición  $\Gamma$ , y el vector de probabilidad inicial  $\delta = (\delta_1, \dots, \delta_J)$ .

### 3.2.2. Modelos ocultos semi-Markovianos(HSMM)

Una de las limitaciones de un HMM básico es la suposición de que el tiempo de permanencia sigue una distribución geométrica. Se ha trabajado mucho para ampliar el HMM básico y modelar explícitamente el tiempo de permanencia mediante otras distribuciones de valores discretos, lo que ha dado lugar a la clase de modelos ocultos semi-Markovianos (HSMM) (Yu, 2010). Al igual que un HMM, un HSMM es un proceso doblemente estocástico compuesto por dos capas: un proceso de estado  $\{C_t\}_{t=1}^T$  y un proceso observable dependiente del estado  $\{X_t\}_{t=1}^T$ . Para un HSMM, sin embargo, el proceso de estado se supone es una cadena semi-Markoviana donde el tiempo de permanencia en un estado dado puede seguir cualquier distribución discreta válida (por ejemplo, Poisson o Binomial Negativa). En lo que resta de esta sección daremos las definiciones y notaciones necesarias para describir un HSMM.

De manera de formalizar la estructura del modelo, primero dividimos los tiempos  $t$  en dos categorías disjuntas dependiendo del estado  $C_t$ :

1. **Tiempo no absorbente(NAT)**: El estado a tiempo  $t$  es diferente que el estado a tiempo  $t - 1$  ( $C_t \neq C_{t-1}$ )
2. **Tiempo absorbente(AT)**: El estado a tiempo  $t$  es igual que el estado a tiempo  $t - 1$ . ( $C_t = C_{t-1}$ ).

La sucesión de estados de tiempos no absorbentes es un proceso de Markov de primer orden, mientras que la secuencia de valores de  $C_t$  es una cadena semi-Markoviana de estado finito. Para simplificar, en este trabajo asumimos que los cambios de estado ocurren en el tiempo  $t = 1$  y  $t = T$ , de tal manera que  $C_1 \neq C_0$  y  $C_T \neq C_{T+1}$ . Un HSMM bajo este supuesto se denomina HSMM censurado a la derecha (right censored HSMM)(Guédon, 2003).

Para cada estado  $i$  de los tiempos no absorbentes, el tiempo de permanencia es la realización de su distribución correspondiente de tiempos de permanencia  $d_i$ . Supongamos que  $t^*$  es un tiempo no absorbente del estado  $i$  entonces la probabilidad de que el estado  $i$  dure  $u$  pasos es,  $d_i(u) = \Pr(C_{t^*+u+1} \neq i \wedge C_{t^*+v} = i, v = 1, \dots, u | C_{t^*} = i)$ ,  $u \geq 1$ . Para un HSMM, la matriz de probabilidad de transiciones contiene las probabilidades de transición condicional  $(\Gamma)_{ij} = \gamma_{ij} = \Pr(C_{t+1} = j | C_t = i, C_{t+1} \neq i)$   $i \neq j$ ,  $j = 1, \dots, J$  con  $\gamma_{ii} = 0 \quad \forall i = 1, \dots, J$ . Las probabilidades de las autotransiciones están determinadas por las distribuciones  $d_i$ .

En la Figura 3.1(b) se representa un ejemplo de un HSMM de 3 estados ( $C_t \in$

$\{1, 2, 3\}$ ). La capa markoviana  $C_t^*$  gobierna el cambio de estados y da lugar a una serie de valores  $d$  todos iguales a  $C_t^*$ . Estos valores  $d$  son realizaciones de las distribuciones de tiempo de permanencia  $d_i$  ( $i = 1, 2, 3$ ). En este ejemplo  $C_1^* = 1$  y  $d_1 = 3$ , lo que indica que el proceso semi-Markoviano comienza con tres unos, es decir,  $C_1 = 1$ ,  $C_2 = 1$  y  $C_3 = 1$ . Luego  $C_2^* = 2$  y  $d_2 = 1$ , lo que indica que la secuencia sigue con un solo dos,  $C_4 = 2$ . A continuación  $C_3^* = 1$  y  $d_1 = 4$ , por lo que se añaden cuatro unos a la secuencia,  $C_5 = 1$ ,  $C_6 = 1$ ,  $C_7 = 1$ ,  $C_8 = 1$  y así sucesivamente. El proceso de observación es similar a un HMM en el sentido de que las observaciones  $X_t$  son variables aleatorias que dependen de los valores de  $C_t$ . Por lo tanto, un HSMM puede definirse por los dos procesos estocásticos  $\{X_t, C_t\}_{t=1}^T$ , con  $X_t$  dependiente de la cadena Semi-Markoviana  $C_t$ , las distribuciones dependientes del estado  $f_1(x), \dots, f_J(x)$ , las distribuciones de tiempo de permanencia  $d_1(u) \dots, d_J(u)$ , la m.p.t.  $\Gamma$ , y la distribución de estado inicial  $\delta = (\delta_1, \dots, \delta_J)$ .

Para el resto de este trabajo, utilizamos la siguiente notación. La secuencia observada de longitud  $l$ ,  $(X_t, X_{t+1}, \dots, X_{t+l})$  se nota como  $\mathbf{X}_{t:t+l}$ , y la secuencia de estados  $(C_t, C_{t+1}, \dots, C_{t+l})$  como  $\mathbf{C}_{t:t+l}$ . Usamos  $\mathbf{C}_{t_1:t_2} = i$  para denotar que el vector de variables de estado  $\{C_{t_1} = i, C_{t_1+1} = i, \dots, C_{t_2} = i\}$  toma el valor  $i$  con cambios de estado que ocurren en  $t_1$  y  $t_2 + 1$ ,  $C_{t_1-1} \neq i$  y  $C_{t_2+1} \neq i$ . Similarmente  $\mathbf{C}_{[t_1;t_2]} = i$  denota la secuencia de variables de estado  $\{C_{t_1} = i, \dots, C_{t_2} = i\}$  que toma el valor  $i$  con un cambio de estado que ocurre en el tiempo  $t_1$ , es decir,  $C_{t_1-1} \neq i$ , pero permite que  $C_{t_2+1} \in \{1, \dots, J\}$ . Del mismo modo  $\mathbf{C}_{t_1;t_2+1} = i$  denota un cambio de estado en el tiempo  $t_2 + 1$ , pero permite  $C_{t_1-1} \in \{1, \dots, J\}$ . Por  $\mathbf{C}_{[t]=i}$  queremos decir que en el tiempo  $t$  el sistema cambió de algún otro estado a  $i$ , y por  $\mathbf{C}_{[t]=i}$  que en el tiempo  $t$  el estado terminará y transitará a algún otro estado. Por último, sea  $\theta = (\theta_\delta, \theta_P, \theta_\Gamma, \theta_d)$  el vector de parámetros del HSMM:  $\theta_\delta$  las probabilidades iniciales,  $\theta_P$  los parámetros de las distribuciones de dependencia de estado  $f_i(x)$ ,  $\theta_\Gamma$  las probabilidades de transición condicional de la m.p.t. y  $\theta_d$  los parámetros de las distribuciones de tiempo de permanencia  $d_i(u)$ .

### 3.2.3. Estructura Autorregresiva (AR(p)-HMM and AR(p)-HSMM)

En los HMMs y HSMMs básicos, asumimos la independencia condicional de  $\mathbf{X}$  dado  $\mathbf{C}$ . Sin embargo, para las series de tiempo tomadas en escalas temporales finas, la estructura de autocorrelación puede no ser capturada adecuadamente bajo este supuesto. Una extensión común es asumir que la distribución dependiente del estado de  $\{X_t\}_{t=2}^T$  depende del estado actual  $C_t$  así como de un subconjunto de observaciones anteriores. Este nuevo supuesto conduce a una importante clase de modelos denominados modelos de conmutación de Markov o modelos ocultos autorregresivos de (semi)Markov ([Hamilton, 1994](#)).

En este enfoque, dado el estado actual, se asume que  $\{X_t\}$  sigue un proceso autorregresivo gaussiano de orden  $p$  (AR(p)) dependiente del estado:

$$x_t = \mu_i + \sum_{k=1}^p \omega_{ki} x_{t-k} + \epsilon_{ti} \quad \text{with } \epsilon_{ti} \sim N(0, \Sigma_i)$$

donde  $\mu_i$ ,  $\omega_{ki}$  y  $\Sigma_i$  son la media condicional, los parámetros del proceso autorregresivo y la matriz de covarianza condicionada a  $C_t = i$ . Notamos a las distribuciones estado dependientes de un AR(p)-HSMM como

$$f_i(x_t) = f(x_t | C_t = i, \mathbf{x}_{t-p:t-1})$$

y la distribución conjunta del vector  $\mathbf{x}_{t:t'}$  con  $t < t'$  como,

$$f_i(\mathbf{x}_{t:t'}) = \prod_{k=t}^{t'} f(x_k | C_k = i, \mathbf{x}_{k-p:k-1}) = \prod_{k=t}^{t'} f_i(x_k).$$

El uso y las propiedades estadísticas de los HMMs con estructura autorregresiva han sido estudiadas y formalizadas (Yang, 2000; Francq y Zakoian, 2001), así como también generalizadas para los HSMMs (Xu y Liu, 2020). Los AR(p)-HSMM también se han utilizado para realizar clasificación de manera no supervisada; (Bryan y Levinson, 2015) demuestra su aplicación en reconocimiento de señales de voz y (Duong et al., 2005) propuso el uso de un modelo oculto semi-Markoviano para reconocer actividades humanas de la vida diaria. Sin embargo, hasta donde sabemos, no se ha estudiado ni aplicado el uso de AR(p)-HSMM para realizar clasificación supervisada.

### 3.3. Clasificación por aprendizaje supervisado

En muchas aplicaciones, el objetivo principal del análisis es predecir de manera precisa los estados latentes del sistema. Las observaciones en sí mismas no son de verdadero interés, ya que su función es simplemente proporcionar información sobre los estados (Zucchini et al., 2017). Para los procesos que evolucionan en el tiempo (o en secuencia), los HMM (y sus extensiones) pueden aplicarse para clasificar las observaciones en diferentes estados latentes, teniendo en cuenta que es probable que las observaciones *cercanas entre sí* en el tiempo hayan surgido del mismo estado. En esta sección discutimos la implementación de los HMMs, HSMMs, y sus respectivas versiones autorregresivas bajo el enfoque de aprendizaje supervisado.

Los métodos de aprendizaje supervisado parten de la base de que existe un conjunto de datos para el que se conocen las clases verdaderas (datos etiquetados). Este conjunto de datos etiquetados se utiliza para ajustar el modelo y evaluar su capacidad de predicción. A continuación, el conjunto de datos etiquetados se distribuye entre el conjunto de entrenamiento para ajustar el modelo y el conjunto de prueba para medir la precisión del modelo final. Por último, una vez evaluado el modelo, puede utilizarse para clasificar datos no etiquetados. Así, cuando el objetivo es realizar la clasificación utilizando HMMs o HSMMs con un enfoque supervisado, los pasos son: dividir los datos etiquetados, entrenar el modelo (sección 3.3.1), predecir los estados ocultos (sección 3.3.2), estimar el error de clasificación (sección 3.3.3), y finalmente clasificar los datos no etiquetados.

### 3.3.1. Inferencia

La inferencia para  $\theta$  se realiza mediante la verosimilitud completa de los datos, ya que tanto las observaciones como los valores de los estados se conocen en los datos de entrenamiento. La verosimilitud completa de los datos se expresa como la distribución conjunta de las observaciones,  $X_t$ , y los estados,  $C_t$ ,

$$\begin{aligned} f(\mathbf{c}_{1:T}, \mathbf{x}_{1:T} | \theta) &= \Pr(\mathbf{C}_{1:T} = \mathbf{c}_{1:T}, \mathbf{X}_{1:T} = \mathbf{x}_{1:T} | \theta) \\ &= \delta_{C_1} d_{C_1}(u_1) \prod_{t=1}^T f_{C_t}(x_t) \prod_{r \text{ is NAT}} \gamma_{c_{r-1}, c_r} d_{C_r}(u_r) \end{aligned}$$

Notemos que para el caso autorregresivo  $f_{C_t}(x_t)$  depende de las  $p$  observaciones previas, es decir,  $f_{C_t}(x_t) = f_{C_t}(x_t | \mathbf{x}_{t-p:t-1})$ . La verosimilitud completa de los datos en un HMM, tiene necesariamente  $d_{C_r}(\cdot) \sim \text{Geom}(\lambda_r)$ .

La verosimilitud completa de los datos puede expresarse como un producto entre los términos de la m.p.t., la distribución inicial, las distribuciones dependientes del estado y las distribuciones del tiempo de permanencia, permitiendo así escribir la misma en escala logarítmica como la siguiente suma,

$$\begin{aligned} \log(f(\mathbf{c}_{1:T}, \mathbf{x}_{1:T}, \theta)) &= \log(\delta_{C_1}) + \log(d_{C_1}(u_1)) + \sum_{t=1}^T \log(f_{C_t}(x_t)) \\ &\quad + \sum_{t \in \text{NAT}} \log(\gamma_{c_{t-1}, c_t}) + \log(d_{C_t}(u_t)) \end{aligned}$$

Para  $K$  series de tiempo independientes, la verosimilitud completa de los datos es el producto de las verosimilitudes completas individuales. Como realizamos la inferencia en un marco Bayesiano, primero especificamos las distribuciones previas para todos los parámetros  $f(\theta)$  y derivamos luego la distribución posterior conjunta  $f(\theta | \mathbf{x}_{1:T}, \mathbf{c}_{1:T}) \propto f(\mathbf{c}_{1:T}, \mathbf{x}_{1:T}, \theta)$ . En particular, la distribución posterior conjunta logarítmica puede expresarse como una suma sobre los componentes del logaritmo de la verosimilitud completa y las distribuciones previas. Los detalles completos de las especificaciones para las distribuciones previas y la derivación de la distribución posterior conjunta se proporciona en el Apéndice B.1.

### 3.3.2. Predicción de estados

El objetivo de la clasificación, en este contexto, es determinar cuales son los estados más probables que generaron la secuencia de observaciones. Este proceso se conoce como *decodificación de estados* y puede realizarse mediante una de las dos siguientes formas generales (i) *decodificación local de estados* – identificar el estado más probable

en cada momento o (ii) *decodificación global de estados* – identificar la secuencia de estados con mayor probabilidad de haber generado la secuencia completa de observaciones. A continuación formalizamos y discutimos los algoritmos más comunes y sus adaptaciones a los HSMMs y AR-HSMMs para la decodificación de estados.

## Decodificación local de estados

La decodificación local de estados es el proceso de determinar para cada momento  $t$ , cuál es la etiqueta del estado  $C_t$  con mayor probabilidad de haber generado  $x_t$ . Para ello, se calculan las probabilidades marginales  $\Pr(C_t = i | \mathbf{X}_{1:T})$  mediante el algoritmo de Forward-Backward (FB), las cuales pueden expresarse como función de las probabilidades *forward*  $\{\alpha_t\}_{t=1}^T$  y las probabilidades *backward*  $\{\beta_t\}_{t=1}^T$ . Para el HMM básico, estas cantidades vienen dadas como

$$\alpha_t(j) = \Pr(\mathbf{X}_{1:t}, C_t = j) \quad \beta_t(j) = \Pr(\mathbf{X}_{t+1:T} | C_t = j)$$

donde  $\beta_T(j) = 1$ , para todo  $j \in \{1, \dots, J\}$  entonces

$$\alpha_t(j) = f_j(x_t) \sum_i \gamma_{i,j} \alpha_{t-1}(i) \quad \beta_t(j) = \sum_i \beta_{t+1}(i) \gamma_{i,j} f_i(x_{t+1})$$

Dado  $\alpha_t(j)$  y  $\beta_t(j)$ , podemos expresar la probabilidad marginal como,

$$\Pr(C_t = j | \mathbf{X}_{1:T}) = \frac{\alpha_t(j) \beta_t(j)}{\Pr(\mathbf{X}_{1:T})}.$$

Para más detalles se puede consultar (Bishop, 2012) y (Zucchini et al., 2017). La derivación de las probabilidades marginales de estado para el HSMM son similares a las del HMM. Sin embargo, es necesario considerar ciertas adaptaciones para el caso del AR(p)-HSMM. Primero expresamos las probabilidades *forward* y *backward* de la siguiente manera:

$$\alpha_t(j) = \Pr(\mathbf{X}_{1:t}, C_{t_j} = j) \quad \beta_t(j) = \Pr(\mathbf{X}_{t+1:T} | C_{t_j} = j, \mathbf{X}_{t-p:t})$$

Es posible también formular las ecuaciones de recurrencia para expresar  $\alpha_t(\cdot)$  en términos de  $\alpha_{t-d}(\cdot)$  y  $\beta_t(\cdot)$  en términos de  $\beta_{t+d}(\cdot)$ . Para  $t = 2, \dots, T$  y  $j = 1, \dots, J$  se puede mostrar que,

$$\begin{aligned} \alpha_t(j) &= \sum_{d \in \mathcal{D}} \sum_{i \neq j} \alpha_{t-d}(i) \gamma_{ij} d_j(d) f_j(\mathbf{x}_{t-d+1:t}) \\ \beta_t(j) &= \sum_{d \in \mathcal{D}} \sum_{i \neq j} \beta_{t+d}(i) \gamma_{ji} d_i(d) f_i(\mathbf{x}_{t+1:t+d}) \end{aligned}$$

Los detalles se encuentran desarrollados en el Apéndice B.2.1. Expresamos las probabilidades marginales como  $\xi_t(j) = \Pr(C_t = j, \mathbf{X}_{1:T})$  y sea  $\beta_t^*(j) = \Pr(\mathbf{X}_{t+1:T} | C_{[t+1} = j)$  (o  $\beta_t^*(j) = \Pr(\mathbf{X}_{t+1:T} | C_{[t+1} = j, \mathbf{X}_{t-p:t})$  para el caso de AR(p)-HSMM). Se sigue de la definición de  $\beta_t^*(j)$  que,

$$\beta_t(j) = \sum_{i \neq j} \gamma_{ji} \beta_t^*(j)$$

$$\beta_t^*(j) = \sum_{d \in \mathcal{D}} d_j(d) \beta_{t+d}(j) f_j(\mathbf{x}_{t+1:t+d})$$

Podemos luego calcular la probabilidad marginal como (detalles completos en el Apéndice B.2.1). Para  $t = 2, \dots, T$  y  $j = 1, \dots, J$

$$\xi_t(j) = \xi_{t+1}(j) + \alpha_t(j) \sum_{i \neq j} \gamma_{ji} \beta_t^*(i) - \beta_t^*(j) \sum_{i \neq j} \alpha_t(i) \gamma_{ij}$$

Luego para asignar los estados locales, usamos  $\hat{C}_t = \arg \max_{j \in 1, \dots, J} \{\xi_t(j)\}$ .

## Decodificación global de estados

La decodificación global de estados es una forma de identificar la secuencia de estados más probable que podría haber dado lugar a los datos temporales observados. Para esta tarea, determinamos la secuencia  $c_1, c_2, \dots, c_T$  que maximiza la probabilidad condicional  $\Pr(\mathbf{C}_{1:T} | \mathbf{X}_{1:T})$  o equivalentemente la probabilidad conjunta  $\Pr(\mathbf{C}_{1:T}, \mathbf{X}_{1:T})$  haciendo uso del algoritmo de Viterbi. Para un HMM básico definimos primero

$$\psi_{1i} = \Pr(C_1 = i, X_1) = \delta_i f_i(x_1)$$

y para  $t = 2, \dots, T$ ,

$$\psi_{ti} = \max_{c_{1:t-1}} \Pr(\mathbf{C}_{1:t-1}, C_t = i, \mathbf{X}_{1:t})$$

Dada la siguiente recursión,

$$\psi_{tj} = \left( \max_i (\psi_{t-1,i} \gamma_{ij}) \right) f_j(x_t) \quad (3.1)$$

la secuencia de estados más probables se obtiene como

$$\hat{i}_T = \operatorname{argmax}_{i=1, \dots, m} \psi_{Ti}$$

y para  $t = T-1, T-2, \dots, 1$

$$\hat{i}_t = \operatorname{argmax}_{i=1, \dots, m} (\psi_{ti} \gamma_{i, \hat{i}_{t+1}})$$

La decodificación global de estados para un HSMM es similar a la de un HMM básico, pero también requiere que se determinen los tiempos de permanencia más probables además de la secuencia de estados más probable. Definimos

$$\psi_t(j, d) = \max_{c_{1:t-d}} \Pr(\mathbf{C}_{1:t-d}, \mathbf{C}_{[t-d+1:t]} = j, \mathbf{X}_{1:T})$$

De forma similar al HMM básico, es posible obtener la siguiente recursión (los detalles completos se proporcionan en el Apéndice B.2.2). Para  $t = 2, \dots, T$ ,  $d = 1, \dots, D$  y  $j = 1, \dots, J$

$$\psi_t(j, d) = \max_{\substack{i \neq j \\ d' \leq t}} \{\psi_{t-d}(i, d') \gamma_{ij} d_j(d) f_j(\mathbf{x}_{t-d+1:t})\}$$

Luego, comenzando en tiempo  $T$ , la secuencia de estados y tiempos de duración más probables es determinada como

$$(\hat{i}_T, \hat{d}_T) = \operatorname{argmax}_{\substack{j=1, \dots, J \\ d=1, \dots, D}} \psi_T(j, d)$$

y para  $t = T - 1, \dots, 1$

$$(\hat{i}_t, \hat{d}_t) = \operatorname{argmax}_{\substack{j \neq \hat{i}_{t+1} \\ d=1, \dots, D}} \left( \psi_{t-d}(j, d') \gamma_{j, \hat{i}_{t+1}} d_{\hat{i}_{t+1}}(d) f_{\hat{i}_{t+1}}(\mathbf{x}_{t-d+1:t}) \right)$$

Los resultados obtenidos mediante la decodificación de estados local y global suelen ser muy similares pero no idénticas (Zucchini et al., 2017). Una ventaja de utilizar decodificación local a partir del algoritmo FB sobre utilizar decodificación global mediante el algoritmo de Viterbi, es que el primero proporciona una estimación de la probabilidad de cada estado posible para cada uno de los tiempos. Estas probabilidades proporcionan información sobre la incertidumbre de las predicciones que puede ser útil a la hora de comparar modelos.

### 3.3.3. Error de clasificación

Dado un modelo ajustado, podemos evaluar su capacidad para clasificar los estados correctamente mediante la decodificación de estados local y global. Una de las medidas más comunes para estimar la precisión de la predicción es el índice de precisión de clasificación: el número de estados correctamente predichos dividido por el número total de observaciones en la serie temporal. Podemos calcular esta medida a partir de los resultados de la decodificación local y global. Sin embargo, aunque esta forma de estimar la precisión de la predicción da una medida general del rendimiento del modelo, no tiene en cuenta la incertidumbre sobre la clasificación. Otra forma de evaluar la capacidad de predicción de nuestros modelos es utilizar el índice entropía cruzada, haciendo uso de las probabilidades de predicción de los estados obtenidas a través de la decodificación local de los estados se puede calcular como,

$$EC = - \sum_{c=1}^J y_{x,c} \log(p_{x,c})$$

donde  $y_{x,c}$  es un indicador binario que indicia si la etiqueta de estado  $c$  es la clasificación correcta para la observación  $x$  ( $y_{x,c} = 1$  cuando la clasificación es correcta),

y  $p_{x,c}$  es la probabilidad predicha de que la observación  $x$  sea generada por el estado  $c$ . El índice de entropía cruzada aumenta a medida que la probabilidad predicha diverge de la etiqueta real, de modo que un modelo perfecto tendría un valor EC de 0.

Para medir la precisión (o el error) de la predicción fuera de la muestra, adaptamos la técnica de validación cruzada de dejar una observación fuera a nuestra aplicación, dejando una serie temporal afuera *validación-cruzada-dejar-una-serie-fuera* (Geisser, 1975; Refaeilzadeh et al., 2009). Al hacerlo, asumimos que cada serie temporal es independiente de la otra y cumplimos con el requisito de que los conjuntos de entrenamiento y de prueba sean independientes.

### 3.4. Estudio de simulación

En algunas aplicaciones, el supuesto de que los tiempos de permanencia siguen distribuciones geométricas (HMMs) está lejos de ser realista. En estos casos, los HSMMs pueden considerarse más apropiados ya que permiten que cada estado tenga un tiempo de duración variable (Yu, 2010). Sin embargo, cuando el objetivo principal es hacer clasificación, puede no ser necesario extender el modelo para obtener predicciones de estado correctas. En este estudio pretendemos caracterizar en qué escenarios las predicciones de estado de un HMM difieren fuertemente de las de un HSMM, con el fin de entender cuándo la violación del supuesto de tiempos de permanencia geométricos puede afectar la precisión de las clasificaciones.

Simulamos series temporales univariadas de longitud 3000 de un HSMM de 2 estados con diferentes tiempos de permanencia y distribuciones para las observaciones. En todos los casos asumimos distribuciones gaussianas para las observaciones y distribuciones binomiales negativas para los tiempos de permanencia. Propusimos tres escenarios que difieren en el solapamiento entre las distribuciones de observación: el primero supone que  $f_1(x_t) \sim N(0, 1)$  y  $f_2(x_t) \sim N(0, 3, 1)$  (solapamiento alto), el segundo supone que  $f_1(x_t) \sim N(0, 1)$  y  $f_2(x_t) \sim N(1, 1)$  (solapamiento medio) y el último supone que  $f_1(x_t) \sim N(0, 1)$  y  $f_2(x_t) \sim N(3, 1)$  (solapamiento bajo). El grado de solapamiento da una idea de la similitud de las distribuciones de observación. Un solapamiento alto implica que las distribuciones son casi iguales, es decir, el área entre ellas incluye al menos 80 % de ellas. Un solapamiento medio indica que las distribuciones son similares pero difieren en al menos un 40 %, y finalmente un solapamiento bajo indica que las distribuciones son diferentes, compartiendo como mucho un 15 % del total. Para cada escenario fijamos diferentes valores para los parámetros de las distribuciones Binomiales negativas:  $d_1 \sim BN(m_1, k_1)$  y  $d_2 \sim BN(m_2, k_2)$ , donde  $m_1$  y  $m_2$  son las medias y  $k_1$  y  $k_2$  los parámetros de dispersión. Cuando  $k = 1$  la binomial negativa es una distribución geométrica con probabilidad  $1/(m + 1)$ , y a medida que  $k$  toma valores mayores a 1 la dispersión de la distribución aumenta, alejándose de la distribución geométrica. De manera de caracterizar diferentes escenarios, primero fijamos un promedio y una diferencia entre  $m_1$  y  $m_2$  ( $(m_1 + m_2)/2$  y  $m_1 - m_2$ ). Consideramos como promedio entre  $m_1$  y  $m_2$ : 90, 40 y 20, y como diferencia entre ellos: 3, 15 y 30. Luego variamos  $k_1$  y  $k_2$  de manera que o bien  $k_1 = 1$  (una de las distribuciones es geométrica:  $k_1 = 1$  y  $k_2 = 10$  o  $k_2 = 30$ ) o bien ambos valores  $k_1$  y  $k_2$  son mayores que uno (ninguna de las distribuciones es geométrica:  $k_1 = 30$  y  $k_2 = 50$  o  $k_1 = 80$  y

$k_2 = 100$ ). Los valores de  $k_i$  y  $m_i$  ( $i = 1, 2$ ) fueron seleccionados buscando considerar la mayor cantidad de escenarios posible. Cuando  $k_i = 1$  el valor más probable es 1 y a medida que  $m_i$  aumenta la distribución posee colas más pesadas. Cuando  $k_i = 10$  o  $k_i = 30$  el valor más probable es más cercano a  $m_i$ , obteniendo así distribuciones más concentradas para  $k_i = 30$  y con mayor dispersión para  $k_i = 10$ . La diferencia entre  $m_1$  y  $m_2$  indica que tan similares son las medias de los tiempos de permanencia de los dos estados.

Para cada combinación de parámetros, simulamos diez series temporales de longitud 3000. De manera de evaluar y comparar la precisión de la clasificación bajo ambos modelos (HMM y HSMM) utilizamos el enfoque de validación cruzada *dejar-una-serie-fuera* de la siguiente manera: ajustamos ambos modelos con todas las series temporales simuladas menos una, luego muestreando de la distribución posterior conjunta obtenida, predecimos 30 veces los estados ocultos de las series temporales simuladas que dejamos fuera utilizando los algoritmos Viterbi y FB. Finalmente, calculamos el índice de precisión de ambas predicciones y el índice de entropía cruzada a partir de la salida de decodificación local. Para ajustar el modelo consideramos previas no informativas:  $N(0, 5)$  y  $NT(0, 5)$  para la media y la desviación estándar de las distribuciones de observación y  $NT(20, 50)$  para la media y el parámetro de dispersión de las distribuciones de tiempo de permanencia. Para realizar los análisis utilizamos R (R Core Team, 2019), para el ajuste de los modelos usamos Stan (Carpenter et al., 2017a) y para implementar el algoritmo de Viterbi para los modelos HSMM usamos la función “hsmm.viterb” del paquete de R “hsmm” (<https://cran.r-project.org/web/packages/hsmm/index.html>). Los códigos se encuentran disponibles en <https://github.com/sofiar/HMM-HSMM-classification>.

### 3.4.1. Resultados

La figura 3.2 muestra los valores de precisión estimados y el índice de entropía cruzada para cada escenario obtenido a partir del análisis de simulación. Es evidente que el rendimiento de los modelos es menor cuando el solapamiento entre las observaciones es mayor: los valores de entropía cruzada aumentan y los índices de precisión disminuyen. Cuando la media entre  $m_1$  y  $m_2$  (las medias de las distribuciones del tiempo de permanencia) es grande, las diferencias entre los tres índices de clasificación son bajas. Sin embargo, cuando esta media disminuye el HSMM supera al HMM, ya que se obtienen valores más bajos de entropía cruzada y valores más altos de precisión. En general, para ambos modelos, los tres índices mejoran a medida que aumenta la media entre  $m_1$  y  $m_2$ . En cuanto a los valores de los parámetros de dispersión ( $k_1$  y  $k_2$ ) de los tiempos de permanencia, es evidente que cuando una de las distribuciones es geométrica (caso 1 en la etiqueta del eje), las diferencias entre los modelos son menores. Este último hecho tiene sentido considerando la hipótesis de los modelos: para un HMM los tiempos de permanencia siguen una distribución geométrica y para un HSMM siguen una distribución Binomial Negativa.

Los valores de precisión del algoritmo FB y del algoritmo de Viterbi resultaron ser prácticamente iguales en todos los casos, excepto cuando hay un alto solapamiento entre las distribuciones de las observaciones y la media entre  $m_1$  y  $m_2$  es baja. En ese

caso los valores de precisión para el HSMM presentan una mayor dispersión: rango intercuartil entre 0,05 y 0,55 para el HSMM, y 0,04 y 0,004 para el HMM. Para la mayoría de los escenarios en términos de capacidad de clasificación, el HSMM superó al HMM, obteniendo mayores diferencias cuando el solapamiento de las distribuciones de observación aumenta, la media entre los parámetros medios de las distribuciones de tiempos de permanencia disminuye y cuando los parámetros de dispersión son mayores que uno. Sólo cuando el solapamiento entre las distribuciones de observación es bajo y las medias entre  $m_1$  y  $m_2$  no son demasiado pequeñas, ambos modelos se comportan prácticamente igual.

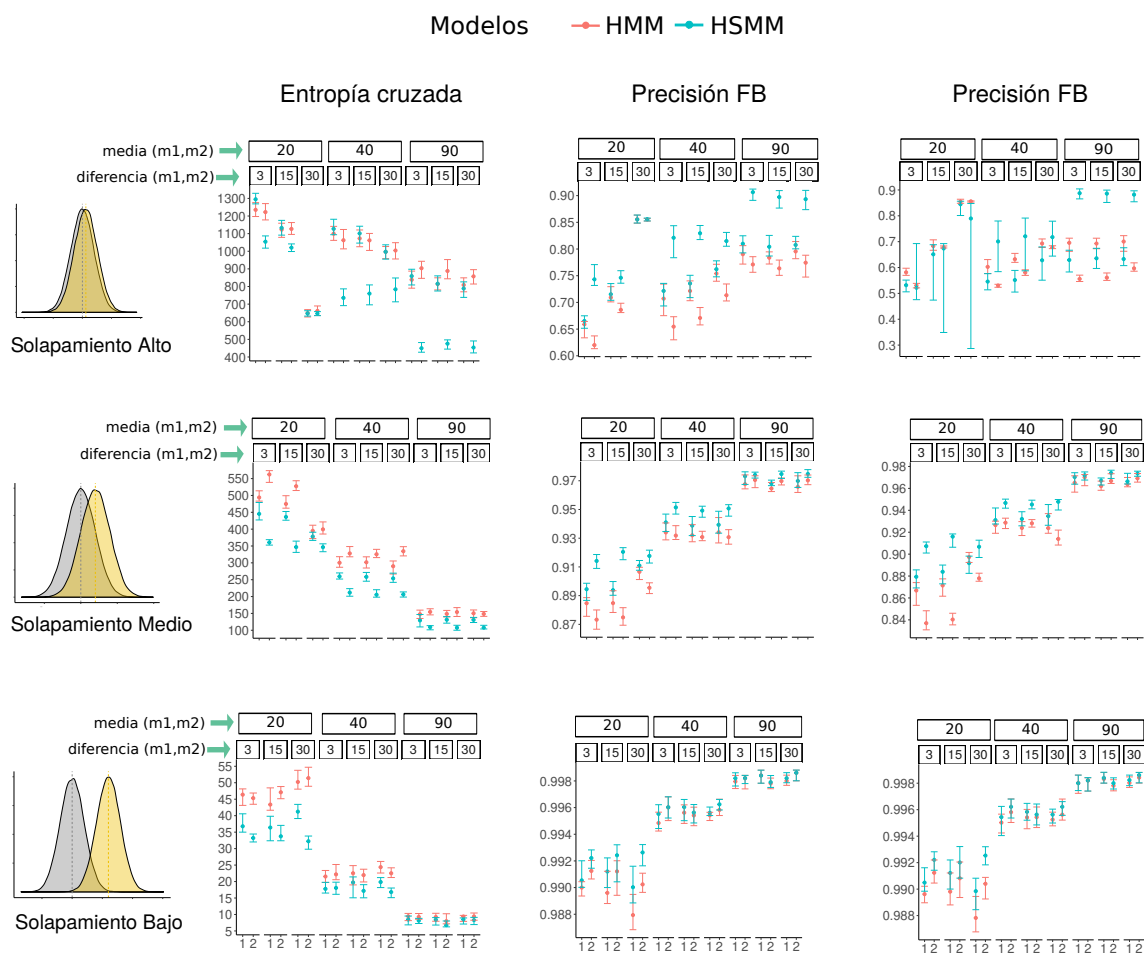


Figura 3.2: Valores de los tres índices calculados en el estudio de simulación. Por columna el índice calculado (entropía cruzada, precisión para FB y precisión para Viterbi), por fila el grado de solapamiento entre las distribuciones de observación (alto, medio y bajo). En cada casilla, el eje x se refiere a la configuración de los parámetros de dispersión de los tiempos de permanencia ( $k_1$  y  $k_2$ ): para la etiqueta 1  $k_1 = 1$  (un estado es geométrico) y para la etiqueta 2 ni  $k_1$  ni  $k_2$  es igual a 1 (ningún estado es geométrico). Los grupos de tres valores consecutivos de cada casilla corresponden a las diferentes medias entre  $m_1$  y  $m_2$  (20,40 y 90), y para cada uno de ellos en orden las tres posibles diferencias entre  $m_1$  y  $m_2$  (3,15 y 30). Los puntos representan el valor de la mediana y las barras el primer y el tercer cuartil. En verde los valores obtenidos para los HMM y en rojo los valores obtenidos para los HSMM.

### 3.5. Datos de aceleración de ovejas

Nuestro objetivo es clasificar los datos del acelerómetro de ovejas merinas domésticas para distinguir entre cuatro estados comportamentales diferentes. Como los datos de los acelerómetros de los animales se recogen típicamente con una resolución temporal fina, lo que lleva a una alta dependencia temporal del proceso de observación, propusimos el uso de HMMs y HSMMs para clasificar dichos estados. El trabajo de campo se realizó en el “Campo Anexo Pilcaniyeu” de INTA (Instituto Nacional de Tecnología Agropecuaria) Bariloche, Patagonia Argentina. Se recolectaron datos de 25 ovejas diferentes durante los meses de septiembre y diciembre de 2019. Los animales fueron equipados con collares que contenían un dispositivo DailyDiary (Wilson et al., 2008b) desarrollado en la Universidad de Swansea, y fueron programados para registrar 40 datos de aceleración por segundo (frecuencia de 40 Hz) y 13 datos de magnetómetro por segundo (frecuencia de 13 Hz). Las unidades se fijaron en la parte superior del cuello del animal y registraron la aceleración en tres ejes: anterior-posterior (*surge*), lateral (*sway*) y dorso-ventral (*heave*) (figura 3.3).

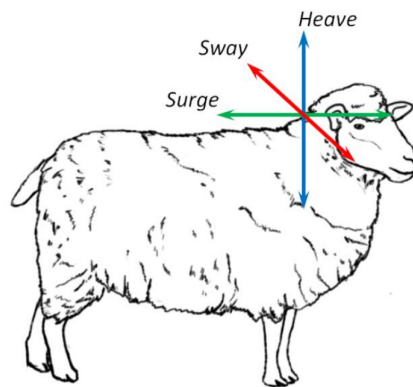


Figura 3.3: En el dispositivo colocado sobre la cabeza del animal, el eje x (color rojo) se corresponde con los movimientos laterales de la cabeza (i.e., sway); el eje z (color azul) coincide con los movimientos verticales (i.e., heave); y el eje y (color verde) representa los movimientos hacia atrás y adelante (i.e., surge).

Para obtener el conjunto de datos clasificados, se realizaron dos sesiones de grabación durante los días 24 de Septiembre de 2019 y 17 de Diciembre de 2019. Estas grabaciones sirvieron como base para nuestro esquema de reconocimiento de comportamiento. Los datos de aceleración obtenidos se suavizaron considerando una media móvil de ventanas de diez observaciones (un cuarto de segundo). A partir de los datos observados distinguimos cuatro comportamientos (i) Caminata (marcha a velocidad normal con la cabeza levantada), (ii) Caminata-Rápida (cuando el animal corre o se mueve rápido), (iii) Inactivo (cuando el animal está quieto: descansando o vigilante) y, (iv) Alimentación (cuando el animal come o busca comida, lo que puede implicar algo de marcha, pero es lenta y con la cabeza baja). La clasificación manual se realizó con una resolución temporal de 1 segundo.

Para definir las variables del proceso de observación, buscamos identificar características de los datos que diferencien cada comportamiento, es decir, determinamos

características derivadas de la aceleración cuyas distribuciones poseían bajo solapamiento entre los estados. Para ello, analizamos los ejes de aceleración y dos valores derivados, la aceleración dinámica vectorial del cuerpo (VeDBA, (Qasem et al., 2012)) y el ángulo de inclinación de la cabeza (Pitch (Wilson et al., 2008b)). El valor VeDBA es la raíz cuadrada de la suma de los cuadrados de los tres ejes de aceleración; esta cantidad puede considerarse como una aproximación al gasto energético del animal (Qasem et al., 2012). El Pitch puede derivarse como una aproximación de la media de los sensores dorso ventrales y resulta en una aproximación del ángulo de inclinación de la cabeza. Se utiliza principalmente para determinar si la cabeza del animal está orientada hacia arriba, en posición neutral o hacia abajo.

En el caso de las ovejas, el ángulo de la cabeza proporciona información importante. La cabeza del animal mirando hacia abajo (valores negativos de Pitch) indica un comportamiento de búsqueda o forrajeo, la cabeza en posición neutra (valores de Pitch cercanos a cero) indica que la oveja está probablemente en estado Inactivo y, por último, una orientación variable de la cabeza medida a uno o dos segundos (alta varianza de los valores de Pitch) implica que la oveja podría estar corriendo o caminando. El índice VeDBA también proporciona información útil: los valores bajos de VeDBA indican menos actividad (comportamiento Inactivo) y los valores altos indican más actividad (comportamientos de caminar y caminar rápido). Tras analizar varios estadísticos de resumen derivados de estas cantidades y buscando diferenciar lo mejor posible los comportamientos preestablecidos, seleccionamos tres estadísticos resumen sobre una ventana de un segundo:  $\log(\text{Mean.VeDBA})$ , el logaritmo del VeDBA medio durante un segundo;  $\text{Mean.Pitch}$ , la media del ángulo Pitch durante un segundo; y el  $\log(\text{Sd.Pitch})$ , el logaritmo de la desviación estándar del ángulo de Pitch durante un segundo.

Para distinguir entre los cuatro estados comportamentales y considerando una escala temporal de 1s, analizamos las clasificaciones derivadas de cuatro modelos (i) HMM, (ii)AR(1)-HMM, (iii) HSMM y (iv)AR(1)-HSMM. Asumimos los siguientes supuestos:

- $\forall t = 1, \dots, T, C_t \in \{1, 2, 3, 4\}$ , con 1 = Caminata, 2 = Caminata-Rápida, 3 = Inactivo y 4 = Alimentación
- $\boldsymbol{\delta} = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$
- $\forall t = 1, \dots, T, \mathbf{X}_t = (X_t^1, X_t^2, X_t^3)$  con,  $X_t^1 \log(\text{Mean.VeDBA})$ ,  $X_t^2 \text{Mean.Pitch}$ , y  $X_t^3 \log(\text{Sd.Pitch})$

Para los modelos sin estructura autorregresiva (i-iii)

- Para  $i \in \{1, 2, 3, 4\}$   $P(\mathbf{X}_t | C_t = i) \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ , con  $\boldsymbol{\Sigma}_i$  matriz diagonal.

Para los modelos con estructura autorregresiva (ii-iv)

- Para  $i \in \{1, 2, 3, 4\}$   $P(\mathbf{X}_t | C_t = i) \sim N(\boldsymbol{\alpha}_i + \beta_i x_{t-1}, \boldsymbol{\Sigma}_i)$ , con  $\boldsymbol{\Sigma}_i$  matriz diagonal.

y para ambos HSMMs (iii-iv)

- Para  $i \in \{1, 2, 3, 4\}$   $d_i \sim NB(m_i, k_i)$ , donde  $m_i$  es la media y  $k_i$  el parámetro de dispersión.

Para evaluar el rendimiento de la clasificación de los cuatro modelos descriptos anteriormente, realizamos un enfoque de validación cruzada similar al detallado en la sección anterior. Para simplificar el análisis, aquí consideramos las series temporales de cada vídeo como independientes. Sin embargo, como varios vídeos corresponden al mismo animal, sería posible también añadir un efecto aleatorio de manera de considerar una posible correlación entre las series temporales del mismo individuo. En este caso, primero ajustamos los cuatro modelos con todas las series temporales menos una, y luego predecimos las series temporales restantes con los dos algoritmos de decodificación 100 veces (muestreando 100 veces de las distribuciones posteriores conjuntas). Por último, calculamos los dos índices de precisión y el valor de la entropía cruzada.

Al hacer inferencia bajo el enfoque Bayesiano, es posible incorporar el conocimiento previo sobre la media, la moda y/o la varianza en la distribución de las observaciones y los tiempos de permanencia para cada comportamiento mediante las distribuciones previas. Por ejemplo, en el caso de las ovejas se puede suponer que la distribución del tiempo de permanencia de la clase de descanso tendrá valores más grandes en comparación con las distribuciones de otros estados y que, por ejemplo, la distribución del la media del ángulo Pitch para el caso de la Alimentación debe poseer casi toda la masa de probabilidad en valores negativos. Por ello, para ajustar los modelos hemos considerado previas normales con desviación estándar alta y valores de la media que difieren según cada comportamiento. La lista de las previas utilizadas se proporciona en el Apéndice B.3.

Por último, ajustamos los cuatro modelos con todas las series temporales disponibles de manera de analizar la capacidad de captar la estructura del sistema de los mismos. Para medir la capacidad de predicción, calculamos la raíz de error cuadrático medio (RECM) sobre las observaciones: para cada serie temporal y modelo, utilizando una muestra de 100 valores de la distribución posterior obtenida, calculamos 100 predicciones (utilizando el algoritmo FB) de los estados latentes y el proceso de observación. A continuación, calculamos el RECM entre las predicciones y las observaciones. Para evaluar la bondad del ajuste, finalmente comparamos las estimaciones de los tiempos de permanencia asumiendo distribuciones geométricas (HMM y AR(1)-HMM) y como distribuciones Binomiales Negativas (HSMM y AR(1)-HSMM) con las distribuciones empíricas obtenidas de los tiempos de permanencia observados.

### 3.5.1. Resultados

Tras descartar los vídeos que no habían capturado ovejas o aquellos para los que no se disponía de datos de aceleración, se obtuvieron un total de 18 vídeos de ocho animales diferentes de la sesión de Septiembre y un total de 49 vídeos de 17 animales diferentes de la sesión de Diciembre. De las 67 series temporales con una duración total de dos horas y 53 minutos, un 62,27% correspondió al comportamiento Inactivo, un 33,51% al estado de Alimentación, un 3,75% a la clase de Caminata, y el 0,47% restante al estado de Caminata-Rápida. Los datos utilizados están disponibles en <https://doi.org/10.6084/m9.figshare.14473455>.

La figura 3.4 muestra dos ejemplos de patrones en la señal obtenidos para los cuatro comportamientos. Cuando la oveja está inactiva, los valores de  $\log(\text{Sd.Pitch})$

y  $\log(\text{Mean.VedBA})$  son bajos, lo que significa que la orientación de la cabeza es casi constante y el gasto de energía es muy bajo. Cuando la oveja está comiendo o buscando alimento, los valores de  $\text{VeDBA}$  son mayores y el ángulo de cabeceo es menor. Los patrones del comportamiento de la Caminata son similares al anterior, sin embargo, los valores del ángulo de cabeceo pueden ser más altos y menos constantes. Los valores más altos de la desviación estándar de  $\log(\text{Mean.Pitch})$  y de  $\log(\text{Mean.VedBA})$  se obtienen cuando la oveja está corriendo. También puede observarse que los cuatro comportamientos tienen tiempos de duración diferentes: el estado Inactivo es el que más dura, seguido del estado de Alimentación, luego la Caminata y, por último, el estado de Caminata-Rápida.

La figura 3.5 muestra los boxplots obtenidos considerando todos los datos etiquetados disponibles para cada comportamiento para las tres variables consideradas. Se puede observar que para las tres variables las distribuciones son bastante diferentes entre los cuatro estados. Aunque los valores de  $\log(\text{Mean.VedBA})$  para los estados de Caminata y Alimentación son similares, para el comportamiento Inactivo son claramente inferiores y para el comportamiento de Caminata-Rápida son superiores. En cuanto al valor de  $\text{Medio.Pitch}$ , los comportamientos de inactividad y corrida muestran distribuciones similares, pero para el estado de Caminata los valores de  $\text{Medio.Pitch}$  son menores y para el comportamiento de Alimentación son incluso más pequeños (con dispersión). Por último, a pesar de que los valores de  $\log(\text{Sd.Pitch})$  presentan mayor dispersión, están bien diferenciados entre las cuatro clases.

La figura 3.6 muestra los valores de los tres índices de clasificación calculados para cada modelo. En todos los casos, los modelos se comportan de forma similar y predicen correctamente. Sin embargo, los modelos con estructura autorregresiva (AR(1)-HMM y AR(1)-HSMM) muestran una menor dispersión en todos los índices, lo que sugiere una menor incertidumbre sobre las clasificaciones. Al comparar las predicciones del algoritmo FB y del algoritmo Viterbi no se aprecian diferencias significativas en los valores del índice de precisión, lo que sugiere que la decodificación global y la local tienen un rendimiento similar en cuanto a las clasificaciones.

Los valores de RECM estimados (que figuran en el Apéndice B.4) revelan que no existen diferencias importantes en la capacidad de predicción entre los cuatro modelos. El RECM estimado se comporta de forma similar para las tres variables, excepto en el caso de  $\log(\text{Mean.VedBA})$  para el que los modelos no autorregresivos (HMM y HSMM) muestran una menor incertidumbre. Sin embargo, al analizar la bondad de ajuste del proceso latente, se observa que las estimaciones de los tiempos de permanencia de los HMM difieren de las estimaciones de los HSMM (Figura 3.7). La distribución empírica de los tiempos de permanencia para los comportamientos de Caminata y Caminata-Rápida evidencian que el tiempo de permanencia más probable es diferente de uno, lo que indica que modelar estos tiempos con una distribución geométrica no sería apropiado. Aunque en el caso de los comportamientos Inactivo y Alimentación este patrón es menos claro, las estimaciones de los HSMMs proporcionan mejores aproximaciones al proceso observado.

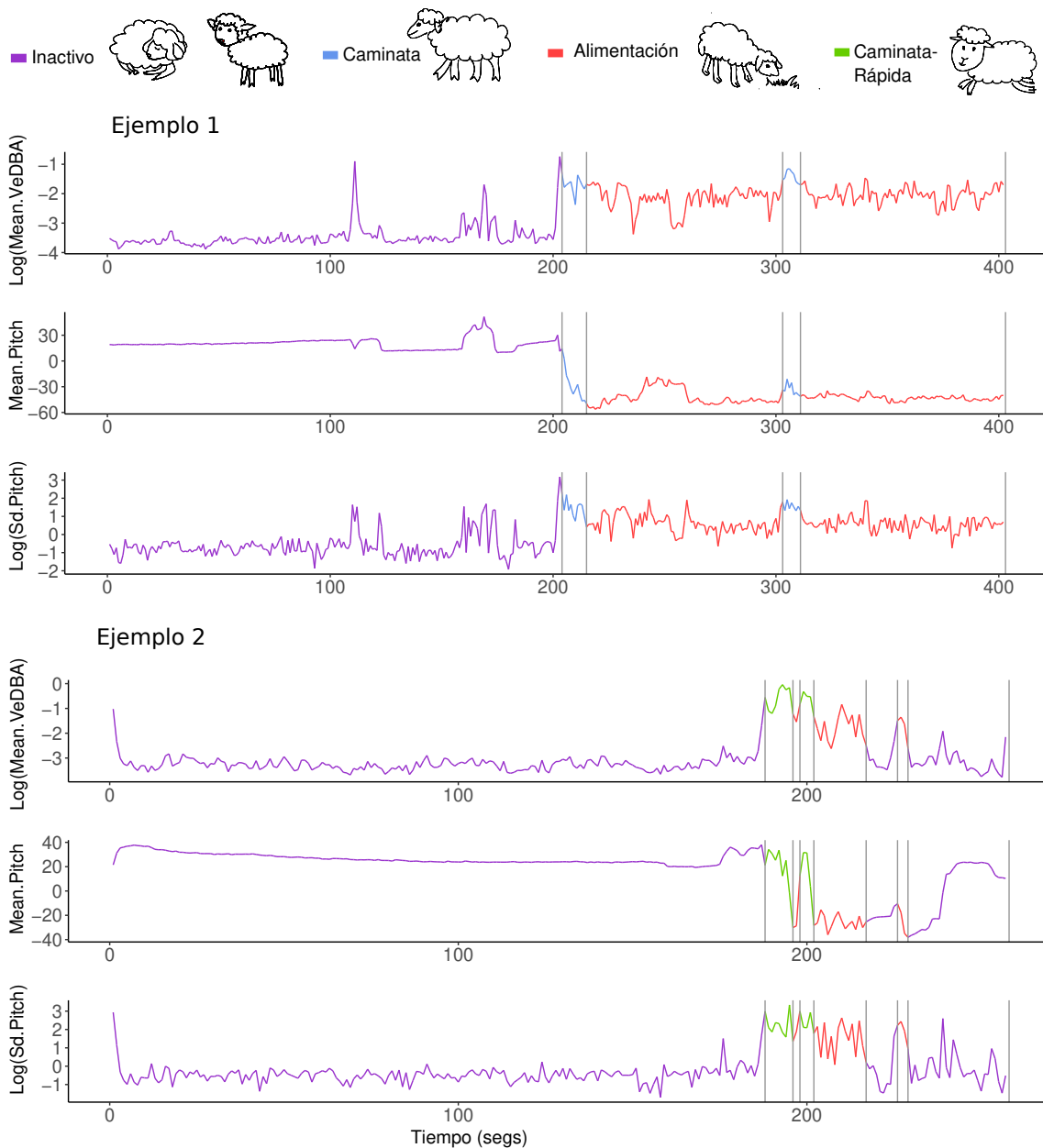


Figura 3.4: Dos ejemplos de las tres variables calculadas a partir de la señal de aceleración registrada de una oveja:  $\log(\text{Mean.VeDBA})$ , Mean.Pitch y  $\log(\text{Sd.Pitch})$ . Diferentes colores representan las señales correspondientes a cada uno de los cuatro comportamientos: Caminata, Caminata-Rápida, Inactivo y Alimentación.

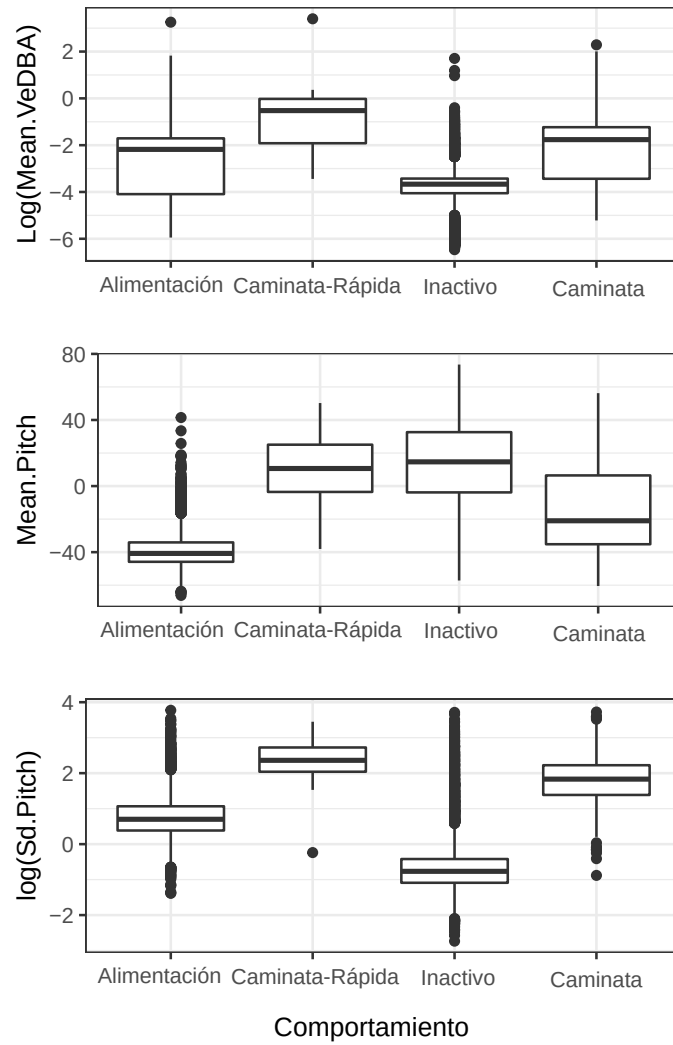


Figura 3.5: Boxplots de  $\log(\text{Mean.VedBA})$ ,  $\text{Mean.Pitch}$  y  $\log(\text{Sd.Pitch})$  para cada uno de los cuatro comportamientos: Inactivo (descanso y vigilancia), Caminada, Alimentación y Caminata-Rápida.

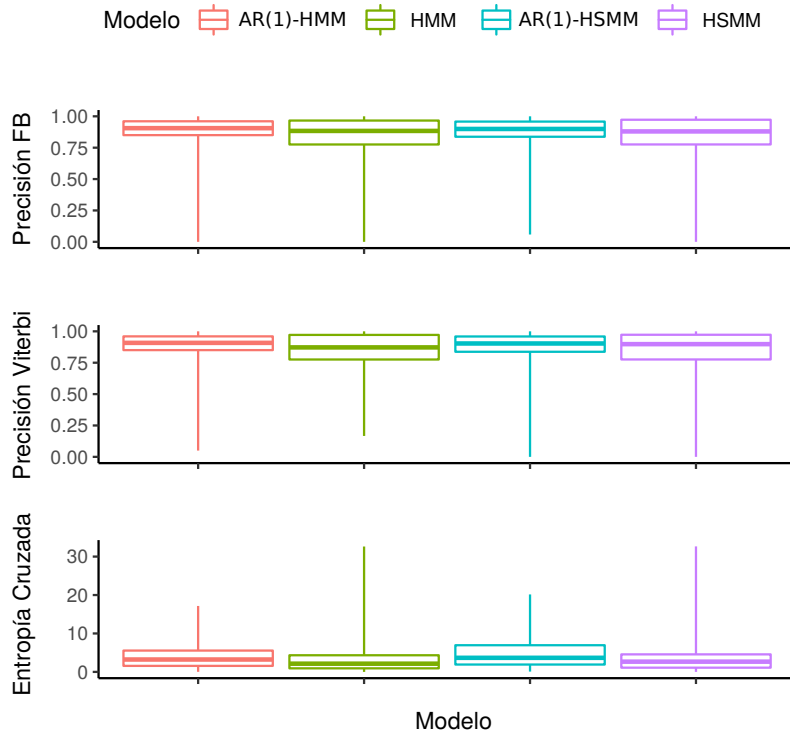


Figura 3.6: Boxplots de los valores de precisión para ambos algoritmos de decodificación y del índice de entropía cruzada obtenidos mediante el análisis de validación cruzada para los cuatro modelos considerados: AR(1)-HMM, HMM, AR(1)-HSMM y HSMM.

### 3.6. Discusión

En este trabajo presentamos una visión general de cómo utilizar los modelos HMMs y HSMMs con o sin estructura autorregresiva en el proceso de observación para realizar una clasificación supervisada de datos temporales. Describimos cómo extender un HMM a un HSMM, detallamos las estructuras del modelo y presentamos los algoritmos para predecir los estados latentes. A continuación, estudiamos los casos en los que las clasificaciones de ambos modelos aplicados a los datos de las series temporales difieren realmente.

Cuando el objetivo final es realizar una clasificación de estados, a partir del estudio de simulación podemos concluir que cuando las distribuciones dependientes del estado tienen un alto solapamiento y el tiempo de permanencia más probable es claramente diferente de uno, diferenciándose de la distribución geométrica, los HSMM superan a los HMM ya que el índice de precisión es mayor y la incertidumbre asociada es menor. Sin embargo, cuando el solapamiento entre las distribuciones de observación es pequeño o cuando la diferencia entre las medias de las distribuciones de tiempo de permanencia es de al menos 25-35 unidades de tiempo (independientemente de su forma), la diferencia entre los modelos es menor.

Aplicamos dos modelos HMM y dos modelos HSMM para clasificar datos del acelerómetro de ovejas obteniendo predicciones precisas para todos ellos: en todos los casos obtuvimos valores de índice de precisión superiores a 0,87 de media. No obstante,

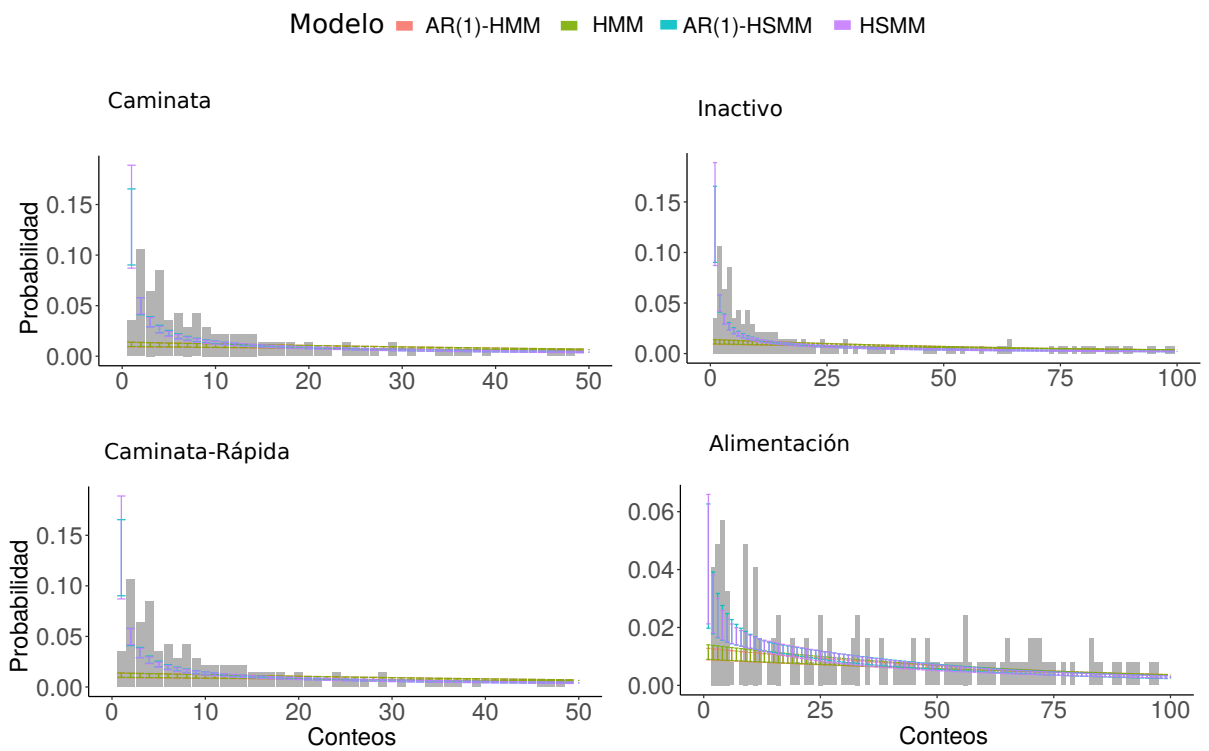


Figura 3.7: Distribuciones ajustadas de los tiempos de permanencia para cada comportamiento. El histograma muestra la distribución empírica de los tiempos de permanencia observados, mientras que los cuatro modelos ajustados se muestran en color. Las barras de error indican las estimaciones puntuales del primer y tercer cuartil.

no pudimos observar diferencias importantes entre el rendimiento de los cuatro modelos, ya que las medidas de error fueron similares. Según los resultados del estudio de simulación, esto podría explicarse principalmente por el escaso solapamiento entre las distribuciones de las observaciones, aunque las estimaciones de las distribuciones del tiempo de permanencia fueron más consistentes con el enfoque semi-Markoviano. Sin embargo, debido a la forma en que se recogieron los datos, no fue posible obtener suficiente información para distinguir entre vigilancia y descanso, por lo que hubo que considerar ambos como un solo comportamiento (Inactivo). Los datos de aceleración de estos dos comportamientos son prácticamente idénticos, sin embargo, los tiempos que las ovejas pasan en estos dos comportamientos son muy diferentes: largos tiempos de descanso y tiempos de vigilancia mucho más cortos. Suponemos que si el conjunto de datos pudiera ser más completo, ampliando las sesiones de grabación para observar suficientes muestras de reposo, se obtendría un rendimiento diferente entre los cuatro modelos.

A la hora de decidir qué modelo utilizar, la principal consideración es el objetivo final del análisis. Cuando el objetivo es realizar una clasificación, la atención se centra en la capacidad de distinguir entre diferentes categorías para luego predecir los estados no observados. Sin embargo, si el objetivo es también proporcionar un mecanismo para la generación de datos, la robustez de las hipótesis del modelo adquiere relevancia. En esos casos, hay que tener en cuenta que las especificaciones sean correctas para obtener estimaciones precisas y lograr los objetivos del análisis.

Cuando el objetivo es seleccionar un modelo con buen rendimiento de clasificación, varios grados de especificación errónea pueden no dar lugar a predicciones que difieran mucho de aquellas obtenidas con un modelo correctamente especificado. Por ejemplo, de nuestras simulaciones se puede concluir que si la media de las distribuciones del tiempo de permanencia difiere en más de 25-35 unidades de tiempo, los modelos simples incorrectamente especificados pueden funcionar igual de bien que los correctamente especificados. Sin embargo, si hay un gran solapamiento entre las distribuciones de estado, los modelos correctamente especificados mejorarán el rendimiento del análisis.

Cuando el objetivo es clasificar observaciones provenientes de datos temporales, los HMM y los HSMM resultan apropiados, ya que incluyen explícitamente la correlación secuencial dentro del análisis. Debido a su estructura, estos modelos permiten distinguir entre la dependencia secuencial presente en los datos y la estructura temporal de los estados consecutivos. La diferencia entre un HMM y un HSMM es que este último permite modelar explícitamente la duración del proceso de comportamiento. Cuando hay un solapamiento significativo entre las distribuciones de las observaciones y las distribuciones de los tiempos de permanencia no tienen forma geométrica, considerar la ampliación del modelo para mejorar las especificaciones del mismo debería beneficiar la precisión de las clasificaciones. Sin embargo, cuando las distribuciones de observación son diferentes entre los estados y aún más, si las distribuciones de tiempo de permanencia también difieren, incluso si el modelo falla en algunos supuestos, las clasificaciones de un HMM pueden ser tan precisas como las clasificaciones de un HSMM con mejor especificación. Para estos casos, recomendamos el uso de los HMM aunque no estén correctamente especificados. Como los HSMM son modelos más complejos con un mayor número de parámetros que los HMM, tienen el coste de un proceso de inferencia más caro computacionalmente.

En la práctica, cuando se trabaja con datos de aceleración de animales bajo un enfoque supervisado, puede pasar que los estados pre-asignados no se observen de manera perfecta. En ciertos casos, los animales pueden ser observados en un entorno de laboratorio (Wilson et al., 2008b), lo que reduce la posibilidad de etiquetas ruidosas. Sin embargo, en otros casos, cuando el animal estudiado no puede ser observado en un entorno controlado, no es fácil recoger un conjunto etiquetado robusto. La preclasificación ruidosa puede afectar al rendimiento de los modelos de clasificación, conduciendo a una menor precisión de la predicción (Frenay y Verleysen, 2014). Muchos trabajos se han dedicado al estudio del ruido de las etiquetas y al desarrollo de técnicas para tratarlo (Sáez et al., 2014; Garcia et al., 2015; Wang et al., 2021), sin embargo hasta ahora no hay estudios que consideren este punto para realizar la clasificación de los datos de aceleración. La inclusión del error de medida en el proceso de estado podría ser un punto interesante para futuras investigaciones en el ámbito de la clasificación supervisada con modelos HMM y HSMM.





# 4. Modelo de trayectorias incluyendo error en las covariables ambientales

---

## 4.1. Introducción

Comprender el proceso que subyace a la forma en que se mueven los animales, implica comprender un sistema variable en el tiempo y en el espacio. Como resultado de este proceso de movimiento se obtiene una selección de recursos y un uso del espacio disponible (Mevin B. Hooten et al., 2017). De esta manera, para poder analizar los procesos involucrados en el movimiento animal, resulta fundamental poder hacer inferencia sobre la relación entre las decisiones de movimiento de los individuos dado el tipo de entorno en el que se encuentran. Durante las últimas décadas, junto con el avance de tecnologías de rastreo animal y el uso de dispositivos de telemetría, se han comenzado a utilizar para este propósito los *análisis de funciones de selección de pasos o recursos* (Fieberg et al., 2021). Los mismos permiten vincular a los animales con su entorno a través de funciones de selección de hábitat o de selección de pasos. En este contexto, un “paso” refiere a la distancia euclídea entre dos ubicaciones muestreadas de manera consecutiva.

Los animales normalmente habitan paisajes espacialmente heterogéneos de manera que emplean distintas estrategias de movimiento y uso del hábitat para satisfacer sus necesidades. De esta manera, el movimiento observado está condicionado por la variabilidad de recursos, condiciones y riesgos presentes en el ambiente. Es posible por un lado analizar cómo los animales seleccionan los recursos, considerando un modelo que sea solo espacial y estimando la distribución de probabilidad de uso asociada. También, por otro lado, se puede analizar cómo los animales modifican su ubicación a lo largo del tiempo, utilizando las ubicaciones observadas y estimando parámetros que permitan describir estas trayectorias. Estos dos análisis por separado, pueden dar una idea general del comportamiento del sistema, sin embargo al no integrar el proceso de movimiento con el de selección de recursos, se impide una comprensión más precisa y completa. Existen luego los análisis integrados de selección por pasos (iSSA) (Avgar et al., 2016), que incorporan características de movimiento al análisis de selección de recursos. Estos modelos posibilitan una mayor comprensión del sistema de movimiento ya que permiten modelizar tanto los procesos de selección de paso como los de selección de hábitat, contemplando así el proceso temporal de cambio de ubicación y el proceso espacial de selección de recursos. La idea básica de estos modelos es comparar las características ambientales de las ubicaciones visitadas por el animal con las características ambientales de un conjunto de ubicaciones que se supone están disponibles, utilizando una regresión logística condicional (Boyce y McDonald, 1999; Fortin et al., 2005; Thurfjell et al., 2014). La interpretación de los parámetros de regresión resulta sencilla, proporcionando así una manera simple y eficaz de conectar el proceso de movimiento

con el uso del paisaje.

La formulación de estos modelos presupone que las observaciones no presentan error. Es decir, se asume que las ubicaciones registradas y la información sobre la disponibilidad y uso de recursos en cada paso de la trayectoria, es completamente conocida y exacta. Sin embargo, en la práctica suele existir cierto grado de error en los datos, que puede llevar a sesgos en las estimaciones y falsas conclusiones sobre el comportamiento del sistema (Visscher, 2006; Johnson y Gillingham, 2008; Hefley et al., 2014; Brost et al., 2015). Interpretaciones erróneas sobre cómo los animales hacen uso y seleccionan el paisaje pueden derivar en diseños de manejo equivocados generando efectos negativos sobre los sistemas naturales.

A partir de un estudio basado en simulaciones, en este trabajo nos proponemos estudiar cómo influye en las estimaciones de los modelos iSSA, la presencia de error en la observación de las variables ambientales. Posteriormente proponemos una forma de contemplar el mismo dentro de la formulación del modelo, tanto para cuando las covariables que describen el paisaje son continuas como para cuando son discretas. Exponemos métodos de inferencia para ambos casos y estudiamos las limitaciones de los mismos en cuanto al grado de error que resisten. Lo que sigue de este capítulo se organiza de la siguiente manera: en la sección 4.2 desarrollamos la formulación de los modelos de iSSA y explicamos los pasos para hacer inferencia; en 4.3 extendemos el modelo anterior de manera de contemplar el error asociado a la observación de las covariables ambientales y explicamos cómo hacer inferencia distinguiendo los casos entre covariables continuas y discretas; en 4.4 exponemos el esquema de simulación utilizado para el análisis; en 4.5 mostramos los resultados obtenidos; y por último en 4.6 concluimos y discutimos los mismos.

## 4.2. Formulación de los modelos iSSA

Los datos de localización GPS permiten obtener datos de las ubicaciones de los animales cada un cierto período de tiempo. Además es posible obtener información ambiental mediante el uso de dispositivos remotos tales como radares o satélites. Es posible luego, utilizar esta información para describir la trayectoria de un individuo, mediante el uso de los modelos iSSA. Estos modelan las transiciones que conectan ubicaciones consecutivas en el espacio mediante una función núcleo espacio temporal de movimiento,  $\phi$ , multiplicado por una función núcleo espacial de selección de hábitat,  $\omega$  (Avgar et al., 2016; Forester et al., 2009). Estos modelos pueden ser formulados siguiendo la estructura de un modelo de procesos de puntos (Johnson et al., 2013). Sea  $\mathbf{s} \in D$  una coordenada espacial, y sea  $t > 0$  un instante de tiempo. Si  $\boldsymbol{\theta}_\omega$  es el vector de parámetros del núcleo de hábitat  $\omega$  y  $\boldsymbol{\theta}_\phi$  lo es para el núcleo de movimiento  $\phi$ , siendo  $H_t$  la historia de la trayectoria hasta tiempo  $t$ , la función de intensidad  $\lambda$  asociada al modelo de proceso de puntos resulta:

$$\lambda(\mathbf{s}, t|H_t) = \omega(\mathbf{s}; \boldsymbol{\theta}_\omega)\phi(\mathbf{s}, t|\boldsymbol{\theta}_\phi, H_t)$$

Al núcleo  $\phi$  se lo denomina función de disponibilidad, y describe como se movería

el animal en un paisaje sin ninguna característica. El núcleo  $\omega$  se llama función de selección de recursos, y resulta proporcional a la densidad de las ubicaciones en el caso en que todas estuvieran disponibles, es decir en el caso en que  $\phi$  fuera uniforme sobre  $D$ .

Supongamos ahora que se observan las ubicaciones,  $\mathbf{s}_1, \dots, \mathbf{s}_n$  en los tiempos  $t_1, \dots, t_n$  (asumiremos  $\Delta_t = t_i - t_{i-1}$  constante). Es posible entonces formular la función de verosimilitud condicional a los tiempos observados como

$$\prod_{i=1}^n \frac{\omega(\mathbf{s}_i; \theta_\omega) \phi(\mathbf{s}_i, t_i | \theta_\phi, H_{t_i})}{\int_D \omega(\mathbf{u}; \theta_\omega) \phi(\mathbf{u}, t_i | \theta_\phi, H_{t_i}) d\mathbf{u}} \quad (4.1)$$

Para lo que sigue de este trabajo consideraremos como función de selección de recursos  $\omega$  a la función exponencial

$$\omega(\mathbf{s}; \theta_\omega) = \exp(\boldsymbol{\alpha}^T \mathbf{X}_s) \quad (4.2)$$

en donde  $\mathbf{X}_s$  es el vector de covariables ambientales medidas en la ubicación  $\mathbf{s}$ . Y como función de disponibilidad, consideráramos la siguiente expresión motivada a partir de un movimiento Browniano (Johnson et al., 2013)

$$\phi(\mathbf{s}, t | \theta_\phi, H_t) = \sqrt{\frac{\beta}{2\pi\Delta_t}} \exp\left(-\frac{\beta d_s^2}{2\Delta_t}\right) \quad (4.3)$$

donde  $d_s$  es la distancia entre ubicaciones consecutivas, y  $\Delta_t$  el tiempo transcurrido entre ambas. Obtenemos así la siguiente función de intensidad

$$\lambda(\mathbf{s}, t | H_t) = \exp(\boldsymbol{\alpha}^T \mathbf{X}_s) \sqrt{\frac{\beta}{2\pi\Delta_t}} \exp\left(-\frac{\beta d_s^2}{2\Delta_t}\right) \quad (4.4)$$

La ecuación 4.1 resulta relativamente sencilla excepto por la expresión de la integral que debe ser evaluada para cada tiempo observado. El cálculo de esta integral en general resulta complejo, pues suele suceder que no existe una forma funcional cerrada sobre el espacio  $D$ , siendo necesarios métodos de integración numérica. Los métodos de integración numérica resultan muy caros computacionalmente y más en este contexto en donde la integral debe ser calculada para cada tiempo observado y para cada combinación de parámetros que se quiera evaluar. Debido a estas complicaciones, para hacer inferencia sobre este modelo se utiliza una aproximación por regresión logística condicional (Fortin et al., 2005; Forester et al., 2009). Para llegar a esta aproximación son necesarios una serie de pasos y derivaciones que, considerando los objetivos de este análisis y la extensión de los mismos, evitaremos desarrollar en esta sección. Sin embargo, la derivación completa de esta aproximación se encuentra disponible en el Apéndice C.1. La idea básica consiste en muestrear para cada tiempo observado  $t_i$  un conjunto de ubicaciones  $m_i = \{\mathbf{l}_{i1}, \dots, \mathbf{l}_{iK}\}$  aleatoriamente. Luego, se considera cada paso  $i$  como un grupo o estrato diferente, las ubicaciones  $m_i$  como el conjunto “conjunto control” de cada grupo y la ubicación observada  $\mathbf{s}_i$  como el caso verdadero. Bajo estos supuestos

se ajusta una regresión logística condicional. En lo que sigue detallamos la formulación aproximada resultante para el caso del modelo con función de intensidad dada por 4.4.

Supongamos que para cada tiempo observados  $t_i$ , muestreamos de manera aleatoria  $K$  ubicaciones control  $\mathbf{l}_{i1}, \dots, \mathbf{l}_{iK}$  utilizando alguna función de disponibilidad  $\phi^*(\mathbf{s}_{i-1}, \mathbf{l}, \theta_{\phi^*})$ . En principio  $\phi^*$  podría tener cualquier forma funcional, aunque en la práctica es útil utilizar una distribución de la familia exponencial (Forester et al., 2009). Sea  $i = 1, \dots, n$  y  $j = 0, \dots, K$ , se definen luego los conjuntos  $m_i = \{\mathbf{l}_{i0}, \mathbf{l}_{i1}, \dots, \mathbf{l}_{iK}\}$ , en donde  $\mathbf{l}_{i0} = \mathbf{s}_i$  y el resto de los elementos son las nuevas ubicaciones muestreadas para el paso  $i$ . Sea entonces  $Y_{ij} \in \{0, 1\}$  tal que

$$Y_{ij} = \begin{cases} 1 & \text{si } j = 0 \\ 0 & \text{si } j \neq 0 \end{cases} \quad (4.5)$$

Siguiendo la aproximación propuesta en (Forester et al., 2009) y desarrollada en C.1, se tiene que el modelo con función de intensidad  $\lambda$  dada por 4.4, puede aproximarse por un modelo de regresión logística condicional, con estratos dados por los pasos de la trayectoria y casos control dados por las  $K$  ubicaciones muestreadas en cada paso. Se considera entonces la siguiente función condicional de verosimilitud

$$P(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}) = \prod_{i=1}^n \frac{\exp\{\boldsymbol{\alpha}^T \mathbf{X}_{\mathbf{l}_{i0}} + \log(\phi(\mathbf{s}_{i-1}, \mathbf{l}_{i0})/\phi^*(\mathbf{s}_{i-1}, \mathbf{l}_{i0}))\}}{\sum_{\mathbf{l} \in m_i} \exp\{\boldsymbol{\alpha}^T \mathbf{X}_{\mathbf{l}} + \log(\phi(\mathbf{s}_{i-1}, \mathbf{l})/\phi^*(\mathbf{s}_{i-1}, \mathbf{l}))\}} \quad (4.6)$$

Si asumimos para  $\phi^*$  la misma distribución paramétrica que para  $\phi$  pero con parámetro  $\eta$ ,

$$\phi^*(\mathbf{s}, t) = \sqrt{\frac{\eta}{2\pi\Delta_t}} \exp\left(\frac{-\eta d_s^2}{2\Delta_t}\right) \quad (4.7)$$

reemplazando en 4.6, resulta que

$$P(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}) = \prod_{i=1}^n \frac{\exp(\boldsymbol{\alpha}^T \mathbf{X}_{i0} - \nu r_{i0})}{\sum_{j=0}^K \exp(\boldsymbol{\alpha}^T \mathbf{X}_{ij} - \nu r_{ij})} \quad (4.8)$$

en donde  $\nu = \eta - \beta$ ,  $\mathbf{X}_{ij}$  es el vector de covariables ambientales en la ubicación  $\mathbf{l}_{ij}$  y  $r_{ij} = d(\mathbf{l}_{ij})^2 \Delta_t^{-1}/2$ , siendo  $d(\mathbf{l}_{ij})$  la distancia euclídea entre la observación  $\mathbf{s}_{i-1}$  y la ubicación  $\mathbf{l}_{ij}$

### 4.2.1. Inferencia

En general los parámetros de movimiento y selección de hábitat suelen estimarse por el método de máxima verosimilitud mediante un proceso de varios pasos (Fieberg et al., 2021; Avgar et al., 2016). Para este trabajo utilizaremos la perspectiva Bayesiana adaptando el proceso anterior a este enfoque.

Consideramos el modelo cuya función de intensidad está dada por 4.4 y una trayectoria observada con ubicaciones  $\mathbf{s}_1, \dots, \mathbf{s}_n$  y tiempos  $t_1, \dots, t_n$ . Al definir la función de selección de paso  $\phi$  mediante la expresión 4.3, se tiene que, si el paisaje es constante, las distancias entre observaciones consecutivas (largos de paso) siguen una distribución Rayleigh con parámetro de escala  $(\Delta_t/\beta)$ . Es decir, siendo  $d_i = d(\mathbf{s}_i, \mathbf{s}_{i-1})$  la distancia euclídea entre las ubicación  $\mathbf{s}_i$  y  $\mathbf{s}_{i-1}$ , se tiene que

$$D \sim \text{Rayleigh}(\Delta_t/\beta) \quad (4.9)$$

cuya función de intensidad es

$$f(d, \Delta_t/\beta) = \frac{d}{(\frac{\Delta_t}{\beta})^2} \exp(-d^2/2(\frac{\Delta_t}{\beta})^2)$$

La distribución Rayleigh equivale a una distribución Weibull con parámetro de forma igual a 2.

Luego, para el modelo dado por 4.4, los pasos para hacer inferencia son

1. Se propone como distribución posterior  $p(\eta|\cdot)$  la distribución obtenida a partir de los largos de paso observados  $d_i$  (expresión 4.9). Siendo  $p(\eta)$  una distribución previa para  $\eta$ , calculamos

$$\begin{aligned} p(\eta|\cdot) &\propto p(d|\eta)p(\eta) \\ p(d|\eta) &\sim \text{Rayleigh}(\Delta_t/\eta) \end{aligned} \quad (4.10)$$

2. Para cada paso  $i = 1, \dots, n$ , a partir de 4.7 se muestrean  $K$  ubicaciones control  $\mathbf{l}_{i1}, \dots, \mathbf{l}_{iK}$  considerando el paso previo observado  $\mathbf{s}_{i-1}$  y la posterior  $p(\eta|\cdot)$  del punto anterior
3. Utilizando las muestras  $\mathbf{l}_{i1}, \dots, \mathbf{l}_{iK}$  del punto anterior, se ajusta la regresión logística condicional 4.8 y se obtienen posteriores  $p(\boldsymbol{\alpha}|\cdot)$  y  $p(\nu|\cdot)$ . Sean  $p(\boldsymbol{\alpha})$  y  $p(\nu)$  las distribuciones previas para estos parámetros y  $p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\alpha}, \nu)$  dada por la ecuación 4.8, entonces

$$\begin{aligned} p(\boldsymbol{\alpha}|\cdot) &\propto p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\alpha}, \nu)p(\boldsymbol{\alpha}) \\ p(\nu|\cdot) &\propto p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\alpha}, \nu)p(\nu) \end{aligned}$$

4. Se calcula la posterior  $p(\beta|\cdot)$  utilizando  $p(\nu|\cdot)$  y  $p(\eta|\cdot)$ . Siendo que  $\nu = \eta - \beta$ , se puede considerar

$$p(\beta|\cdot) = p(\eta|\cdot) - p(\nu|\cdot) \quad (4.11)$$

### 4.3. Considerando error en las observaciones

Tanto los modelos de selección de pasos o recursos, como el modelo iSSA expuesto en las secciones anteriores, consideran que las variables ambientales se observan sin error. Es decir, se asume que el valor del vector  $\mathbf{X}_s$  es completamente preciso para cada ubicación  $s \in D$ . En la práctica, para obtener información sobre variables ambientales y recursos disponibles, se suelen utilizar o bien datos satelitales o de radar, o datos obtenidos por métodos de observación directa. En cualquier caso, la información obtenida generalmente posee algún grado de error, ya sea por desperfectos en los dispositivos de medición o debido a fallas en el registro de los datos. Por ejemplo, para recolectar información sobre variables discretas tales como “tipo de habitat” se suelen hacer clasificaciones a partir de imágenes satelitales, y si bien muchas veces se hacen chequeos a campo posteriores, siempre resulta haber algún nivel de error en las clasificaciones. Variables continuas como la elevación del terreno generalmente se derivan a partir de imágenes satelitales combinadas con modelos de predicción, llevando así también a valores con cierto error.

Como la presencia de error en las observaciones puede sesgar las estimaciones del modelo, buscaremos proponer una extensión del modelo iSSA que contemple la presencia de error. Bajo la perspectiva Bayesiana, trataremos los valores verdaderos de las variables ambientales como datos faltantes (Clayton, 1992; Richardson y Gilks, 1993), siendo entonces necesario definir la derivación de los mismos según las observaciones. La estructura del modelo, dependerá de la naturaleza de las variables ambientales. Distinguiremos así el caso en donde  $\mathbf{X}_s$  es un vector de valores continuos (4.3.1) y el caso en donde  $\mathbf{X}_s$  corresponde a un vector de variables categóricas (4.3.2). Haremos dos simplificaciones de manera de reducir la complejidad del análisis: por un lado asumiremos que el vector de covariables  $\mathbf{X}_s$  es de dimensión 1 y por el otro consideraremos sólo el caso de variables discretas binarias.

Ya sea si consideramos covariables continuas o discretas, partiremos del modelo con función de intensidad dada por 4.4 y cuya aproximación resulta 4.8. Para ambas formulaciones utilizaremos las siguientes definiciones: sean  $\mathbf{s}_1, \dots, \mathbf{s}_n$  y  $t_1, \dots, t_n$  las ubicaciones y tiempos observados de un cierta trayectoria. Sean las ubicaciones  $\mathbf{l}_{ij}$  tales que  $\mathbf{l}_{i0} = \mathbf{s}_i$ , y el resto las ubicaciones muestreadas como disponibles para cada paso  $i$ . Sea entonces  $m = n(K + 1)$  la cantidad ubicaciones totales (entre casos y controles). Sean además  $Y_{ij} \in \{0, 1\}$  definidas como en 4.5, y sea  $X_{\mathbf{l}}^*$  la variable ambiental observada (con error) en la ubicación  $\mathbf{l}$ . Por último sea  $X_{\mathbf{l}}$  el valor de la variable ambiental real (sin error) de la ubicación en  $\mathbf{l}$ .

#### 4.3.1. Caso continuo

Supongamos entonces que los valores verdaderos de la variable predictora  $X_{\mathbf{l}}$  no se conocen, pero tenemos disponibles valores  $X_{\mathbf{l}}^*$  que difieren de los valores verdaderos con una cierta medida. Vamos a suponer que la medida de este error se describe a partir de distribuciones normales independientes con cierta varianza  $\sigma^2$  conocida. Entonces

$$\begin{aligned}
\mathbf{X}^* &= \mathbf{X} + \mathcal{E} \\
\mathcal{E} &\sim \mathcal{N}(0, \Sigma) \\
\Sigma &= \sigma^2 \mathcal{I}_m
\end{aligned} \tag{4.12}$$

en donde  $\mathcal{I}_m$  es la matriz identidad de  $m \times m$ .

### 4.3.2. Caso discreto

Consideraremos nuevamente el vector de valores verdaderos  $\mathbf{X}$  como faltante y nos proponemos modelar la relación entre éste y los valores observados  $\mathbf{X}^*$ . Supongamos entonces que el vector de variables ambientales es categórico binario, es decir  $X_l \in \{0, 1\}$  y que se distribuye como Bernoulli de parámetro  $\rho$ ,

$$X_l \sim \text{Bernoulli}(\rho) \quad \forall l \in D$$

Llamaremos  $\theta_i = P(X_l = i)$ , la cual vale  $\rho$  si  $i = 1$  y  $1 - \rho$  cuando  $i = 0$ . Definimos las probabilidades de error  $p_{ij}$  de observar  $i$  cuando la variable verdadera toma valor  $j$  como

$$p_{ij} = P(X_l^* = i | X_l = j) \quad \forall l \in D \quad \text{e} \quad i, j \in \{0, 1\}$$

Sean también las probabilidades de reclasificación  $q_{ij}$  que dan la probabilidad de que la variable verdadera sea  $i$  dado que se observa  $j$ ,

$$q_{ij} = P(X_l = i | X_l^* = j) \quad \forall l \in D \quad \text{e} \quad i, j \in \{0, 1\}$$

Las probabilidades de reclasificación  $q_{ij}$  pueden ser formuladas en términos de las probabilidades de error  $p_{ij}$  y las probabilidades  $\theta_i$ :

$$q_{ij} = \frac{P(X_l^* = j | X_l = i)P(X_l = i)}{P(X_l^* = j)} = \frac{p_{ij}\theta_i}{P(X_l^* = j)}$$

La probabilidad  $P(X_l^* = j)$  puede ser calculada condicionando a los posibles valores de  $X_l$

$$\begin{aligned}
P(X_l^* = j) &= P(X_l^* = j | X_l = 0)P(X_l = 0) + P(X_l^* = j | X_l = 1)P(X_l = 1) \\
&= p_{j0}\theta_0 + p_{j1}\theta_1 \\
&= \sum_{i=0}^1 p_{ij}\theta_i
\end{aligned}$$

Obtenemos así la siguiente expresión para las probabilidades  $q_{ij}$

$$q_{ij} = \frac{p_{ij}\theta_i}{\sum_{w=0}^1 p_{wj}\theta_w} \quad (4.13)$$

Luego, suponiendo conocidos los valores de las probabilidades  $p_{ij}$  y la probabilidad de éxito  $\rho$  podemos establecer la conexión entre el valor observado  $X_i^*$  y su correspondiente valor real  $X_i$  a partir de las probabilidades de reclasificación  $q_{ij}$ .

### 4.3.3. Algoritmo de inferencia

Para hacer inferencia bajo el enfoque Bayesiano, es necesario formular las distribuciones conjuntas posteriores de los parámetros del modelo como producto de la función de verosimilitud y de las previas. En este caso, la verosimilitud del modelo resulta

$$\begin{aligned} L(\alpha, \nu, \mathbf{X}) &= p(\mathbf{Y}|\mathbf{X}, \alpha, \nu)p(\mathbf{X}|\mathbf{X}^*) \\ &= \prod_{i=1}^n p(\mathbf{Y}_i|\mathbf{X}_i, \alpha, \nu)p(\mathbf{X}_i|\mathbf{X}_i^*) \\ &= \prod_{i=1}^n \frac{\exp(\alpha X_{i0} - \nu r_{i0})}{\sum_{j=0}^K \exp(\alpha X_{ij} - \nu r_{ij})} P(\mathbf{X}_i|\mathbf{X}_i^*) \end{aligned} \quad (4.14)$$

En donde  $\mathbf{Y}_i = (Y_{i0}, Y_{i1}, \dots, Y_{iK})$  y  $\mathbf{X}_i = (X_{i0}, X_{i1}, \dots, X_{iK})$ . La expresión para  $P(\mathbf{X}_i|\mathbf{X}_i^*)$  dependerá de si la variable es continua o si es discreta. En el caso continuo, siguiendo 4.12 tenemos que

$$\begin{aligned} P(\mathbf{X}_i|\mathbf{X}_i^*) &= \prod_{j=0}^K p(X_{ij}|X_{ij}^*) \\ &= \frac{1}{\sigma(2\pi)^{(K+1)/2}} \prod_{j=0}^K \exp\left(-\frac{(X_{ij} - X_{ij}^*)^2}{2\sigma^2}\right) \end{aligned}$$

y para el caso discreto siguiendo 4.13

$$\begin{aligned} P(\mathbf{X}_i|\mathbf{X}_i^*) &= \prod_{j=0}^K p(X_{ij}|X_{ij}^*) \\ &= \prod_{j=0}^K \frac{p_{ij}\theta_i}{\sum_{w=0}^1 p_{wj}\theta_w} \end{aligned}$$

Para poder hacer inferencia utilizando técnicas de MCMC, se puede o bien partir de la distribución posterior conjunta de  $\mathbf{X}$  y los parámetros,  $P(\mathbf{X}, \alpha, \nu | \mathbf{Y}, \mathbf{X}^*)$  o bien trabajar con la distribución marginal de los parámetros dado las observaciones  $P(\alpha, \nu | \mathbf{Y}, \mathbf{X}^*)$ . En el primer caso, una manera de hacer inferencia sería utilizando el método *Gibbs sampling* para lo que se necesita muestrear iterando entre  $P(\mathbf{X} | \mathbf{Y}, \mathbf{X}^*, \alpha, \nu)$  y  $P(\alpha, \nu | \mathbf{Y}, \mathbf{X}^*, \mathbf{X})$ . En el segundo caso, se necesita marginalizar sobre los posibles valores de  $\mathbf{X}$ , es decir en el caso discreto,  $\sum_{\mathbf{X}} P(\alpha, \nu, \mathbf{X} | \mathbf{Y}, \mathbf{X}^*)$ , y en el caso continuo  $\int P(\alpha, \nu, \mathbf{X} | \mathbf{Y}, \mathbf{X}^*) d\mathbf{X}$ . La suma en el denominador de la expresión de verosimilitud 4.14, denota una dependencia entre las observaciones de cada uno de los  $n$  grupos. Esta dependencia hace que tanto muestrear valores para  $\mathbf{X}$ , como marginalizar sobre los mismos resulte un desafío.

Recordemos que cada uno de los vectores  $\mathbf{X}_i$  posee dimensión  $K$ , y que  $K$  debe ser lo suficientemente grande como para que la aproximación por regresión logística condicional funcione. Supongamos que se quiere hacer inferencia a partir de la distribución posterior conjunta  $P(\mathbf{X}, \alpha, \nu | \mathbf{Y}, \mathbf{X}^*)$  bajo el esquema de *Gibbs sampling*. Se debe entonces en cada paso de la cadena Markoviana muestrear valores para todos los vectores  $\mathbf{X}_i$ . En el caso discreto, obtener una forma cerrada para la expresión  $P(\mathbf{X} | \mathbf{Y}, \mathbf{X}^*, \alpha, \nu)$  resulta complejo ya que implica calcular la distribución del vector  $\mathbf{X}_i$  de dimensión  $K$ , para lo cual se debe condicionar sobre todos los posibles valores de cada elemento del vector. Si la variable  $X$  es binaria, esto implicaría calcular  $2^K$  posibles combinaciones, complicando la inferencia por este camino cuando la magnitud de  $K$  es grande. Otra opción sería evitar calcular la expresión cerrada para  $P(\mathbf{X} | \mathbf{Y}, \mathbf{X}^*, \alpha, \nu)$  utilizando *Metropolis-Hasting* dentro del esquema de *Gibbs sampling* muestreando valores de  $\mathbf{X}_i$  a partir de una distribución de propuesta en cada paso del algoritmo Markoviano. Nuevamente la dimensión del vector  $\mathbf{X}_i$  supone un problema, pues este método implica utilizar una distribución de propuesta para un vector de parámetros demasiado grande. Supongamos ahora que se busca hacer inferencia a partir de la distribución marginal de los parámetros dado las observaciones  $P(\alpha, \nu | \mathbf{Y}, \mathbf{X}^*)$ . Para ello es necesario condicionar sobre todos los posibles valores del vector  $\mathbf{X}_i$ . Nuevamente cuando la variable es discreta esta marginalización implica considerar  $2^K$  posibles combinaciones, lo que hace inviable este camino cuando  $K$  es grande.

Como posible solución, en este trabajo reparametrizaremos la expresión 4.8 de la regresión logística condicional mediante una regresión logística simple (Faraway, 2016), evitando así la dependencia entre el vector  $\mathbf{X}$  como:

$$P(\mathbf{Y} | \mathbf{X}, \boldsymbol{\theta}) = \prod_{i=1}^n \prod_{j=0}^K \frac{\exp(\gamma_i + \alpha X_{ij} - \nu r_{ij})}{1 + (\gamma_i + \alpha X_{ij} - \nu r_{ij})} \quad (4.15)$$

en donde  $\gamma_i$  modela un efecto para cada uno de los  $n$  grupos. De esta manera, es posible obtener la distribución marginal de los parámetros dado las observaciones, condicionando de manera independiente sobre los valores de  $\mathbf{X}$ ,  $\sum_{\mathbf{X}} P(\alpha, \nu, \mathbf{X} | \mathbf{Y}, \mathbf{X}^*) = \sum_{ij} P(\alpha, \nu, \mathbf{X}_{ij} | \mathbf{Y}) P(\mathbf{X}_{ij} | \mathbf{X}_{ij}^*)$ .

Si bien esta nueva parametrización tiene más parámetros, hace posible la inferencia mediante técnicas de *Hamiltonian Monte Carlo*, las cuales se caracterizan por su eficiencia en problemas de alta dimensión y con parámetros continuos (Gelman et al., 2013). Además como esta técnica se encuentra disponible en Stan (Carpenter

et al., 2017a) resulta más sencilla su implementación. Es posible entonces reescribir la verosimilitud como

$$L(\alpha, \nu, \mathbf{X}) = \prod_{i=1}^n \prod_{j=0}^K \frac{\exp(\gamma_i + \alpha X_{i0} - \nu r_{i0})}{1 + \exp(\gamma_i + \alpha X_{ij} - \nu r_{ij})} P(X_{ij}|X_{ij}^*) \quad (4.16)$$

De esta manera, marginalizamos el valor del vector aleatorio  $\mathbf{X}$  de manera independiente sobre cada observación, condicionando sólo sobre dos valores ( $X_{ij} = 0$  ó  $X_{ij} = 1$ ). Si bien esta reparametrización es sólo necesaria para el caso discreto, de manera de ser consistente la utilizaremos también para el caso continuo. Así, las distribuciones posteriores resultan el producto entre la verosimilitud (4.16) y las distribuciones previas  $p(\alpha)$  y  $p(\nu)$

$$\begin{aligned} p(\alpha|\cdot) &\propto L(\alpha, \nu)p(\alpha) \\ p(\nu|\cdot) &\propto L(\alpha, \nu)p(\nu) \end{aligned} \quad (4.17)$$

En donde  $p(X_{ij}|X_{ij}^*)$  está como en 4.13 en el caso discreto y como en 4.12 en el caso continuo. Considerando esta nueva parametrización, el esquema para hacer inferencia en el modelo iSSA contemplando la presencia de error en las variables ambientales resulta similar al de la sección 4.2.1, excepto por el paso 3:

1. Se calcula la distribución posterior preliminar  $p(\eta|\cdot)$  a partir de los largos de paso observados  $d_i$  y la ecuación 4.9. Siendo  $p(\eta)$  la distribución previa para  $\eta$ , se calcula entonces la posterior como en 4.10.
2. Para cada paso  $i = 1, \dots, n$ , a partir de 4.7 se muestrean  $K$  ubicaciones control  $\mathbf{l}_{i1}, \dots, \mathbf{l}_{iK}$  considerando el paso previo observado  $\mathbf{s}_{i-1}$  y la posterior  $p(\eta|\cdot)$  del punto anterior.
3. Utilizando las muestras  $\mathbf{l}_{i1}, \dots, \mathbf{l}_{iK}$  del punto anterior se obtienen posteriores  $p(\alpha|\cdot)$  y  $p(\nu|\cdot)$  según 4.17.
4. Se calcula la posterior  $p(\beta|\cdot)$  a partir de  $p(\nu|\cdot)$  y  $p(\eta|\cdot)$  como en 4.11

## 4.4. Estudio de simulación

Este análisis de simulación tiene dos metas. Por un lado buscamos medir el sesgo en las estimaciones del efecto del paisaje en las trayectorias que produce la presencia de error de las variables ambientales cuando no se lo considera dentro de la estructura del modelo. Por otro lado, queremos evaluar la capacidad de inferencia del modelo que contempla el error, para luego comparar su desempeño con la versión inicial. Para ello, simulamos trayectorias en donde las variables ambientales son medidas con diferentes grados de error y ajustamos el modelo que no contempla el error y el modelo que si lo contempla. Luego consideramos el caso en donde las variables han sido medidas de manera exacta y ajustamos el modelo inicial. Comparamos así tres escenarios diferentes:

Variable Continua	
Grado	$\sigma^2$
g1	0.05
g2	0.065
g3	0.12
g4	0.25
g5	0.35
g6	0.5
g7	0.65

Tabla 4.1

*Parámetros de error para la variable ambiental continua*

Variable Categórica					
Grado	$p_{00}$	$p_{11}$	$q_{00}$	$q_{11}$	$q_{00} \times q_{11}$
g1	0.99	0.99	0.96	0.99	0.95
g2	0.98	0.98	0.92	0.99	0.92
g3	0.9	0.98	0.91	0.97	0.89
g4	0.95	0.97	0.88	0.98	0.87
g5	0.96	0.96	0.85	0.98	0.84
g6	0.8	0.97	0.86	0.95	0.82
g7	0.95	0.95	0.82	0.98	0.81

Tabla 4.2

*Parámetros de error para la variable ambiental categórica binaria*

(i) Observaciones exactas y modelo sin error (SE-M1); (ii) Observaciones con error y modelo sin contemplar el error (CE-M1); y (iii) Observaciones con error y modelo contemplando el error (CE-M2).

Con este objetivo entonces, simulamos trayectorias según el modelo de proceso de puntos dado por 4.4 considerando valores  $\alpha = 2$  y  $\beta = 1$ , y donde el vector de covariables ambientales es de dimensión 1. Simulamos el caso en donde la variable ambiental es continua y el caso donde es categórica binaria. En ambos casos simulamos un total de 25 trayectorias, todas de largo  $N = 250$  pasos, con  $\Delta_t = 1$ , y con inicio en el centro del paisaje. Para muestrear los casos control consideramos  $K = 80$  para cada uno de los 250 pasos, obteniendo así 20 mil puntos para ajustar cada regresión logística condicional. Los grados de error considerados dependen de las probabilidades de observar de manera incorrecta la covariable ambiental. En el caso discreto, este valor está dado por las probabilidades  $q_{00}$  y  $q_{11}$ , y en el caso continuo por el valor del desvío del error  $\sigma$ .

Para el caso discreto, asumimos que  $X_t \sim \text{Bernoulli}(\rho = 0,8)$  y consideramos como grados de error en las observaciones los dados por la tabla 4.2. Para el caso continuo, asumimos  $X_t \sim N(0, 1)$  con grados de error dados por la tabla 4.1. Para cada una de las trayectorias simuladas, ajustamos el modelo SE-M1, en donde la observación ambiental es considerada exacta. Luego para cada escenario de error considerado ajustamos los otros dos modelos: CE-M1 y CE-M2.

Una vez ajustados los modelos sobre las 25 trayectorias para cada escenario de simulación, calculamos la distribución de los valores esperados resultante para ambos parámetros ( $\alpha$  y  $\beta$ ). Además calculamos la raíz del error cuadrático medio (RECM) y el índice de dispersión (ID) de la siguiente manera:

$$\text{RECM} = \sqrt{\frac{\sum_i (\tilde{\theta}_i - \theta_i)^2}{N}} \quad \text{ID} = \frac{1}{N} \sum_i \frac{|\tilde{\theta}_i - \theta_i|}{\theta_i}$$

en donde  $\theta = \beta$  ó  $\theta = \alpha$ ,  $\tilde{\theta}$  es el valor verdadero del parámetro (1 para  $\beta$  y 2 para  $\alpha$ ), y  $N$  el número total de muestras de la distribución posterior.

Como distribuciones previas para  $\alpha$ ,  $\beta$  y  $\nu$  consideramos  $N(0, 1)$ , y para los valores  $\gamma$  utilizamos una previa no informativa,  $N(0, 5)$ . En todos los casos corrimos 2 cadenas Markovianas con 4000 iteraciones cada una eliminando las primeras 1000 como el período de *burn in*.

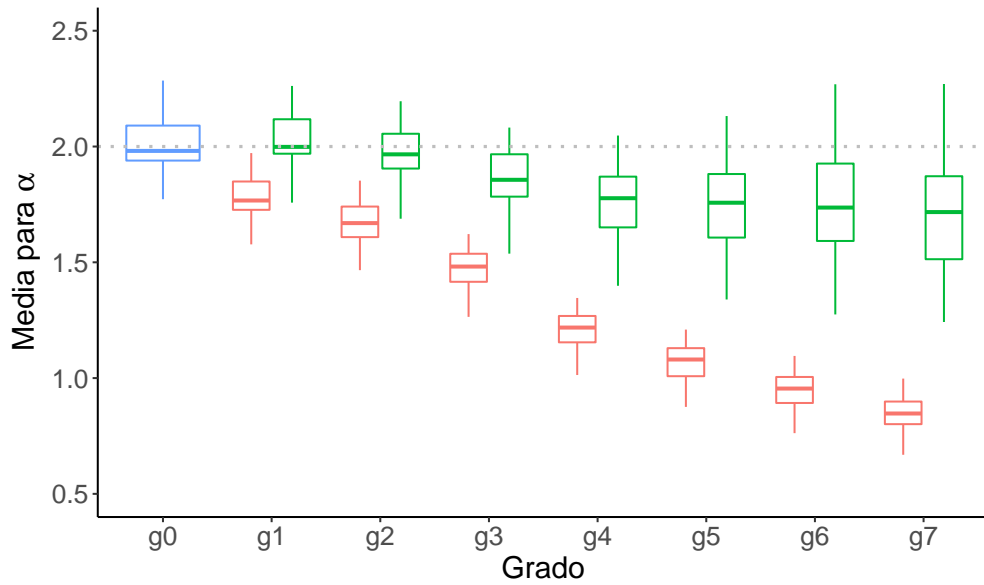
Todos los análisis se realizaron en R y para correr los ajustes Bayesianos utilizamos Stan (Carpenter et al., 2017b). Los códigos de las simulaciones y los ajustes se encuentran disponibles en el repositorio digital [https://github.com/sofiar/SSF\\_ErrorCov](https://github.com/sofiar/SSF_ErrorCov).

## 4.5. Resultados

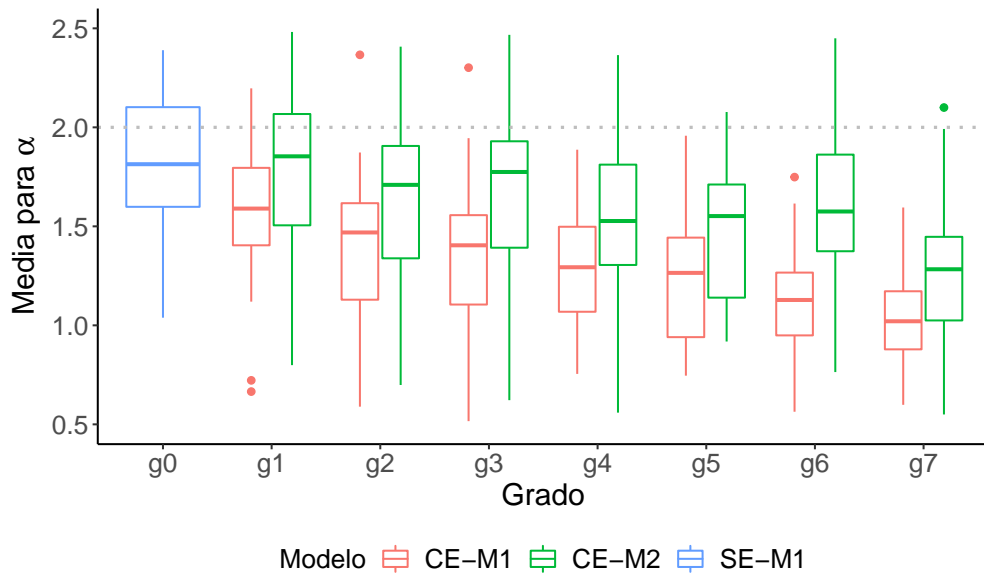
Los valores estimados para el coeficiente de selección de paisaje ( $\alpha$ ) en el modelo SE-M1 fueron precisos, aunque se lograron mejores resultados para el caso en donde la covariable ambiental es continua: se obtuvieron valores medios de 0.17 de RECM y 0.26 de ID para el caso continuo y de 0.54 de RECM y 0.47 de ID para el caso discreto.

En la figura 4.1 se muestran las distribuciones de los valores esperados para el coeficiente de selección de paisaje  $\alpha$  para los diferentes grados de error según si la variable ambiental es continua o categórica. Se puede observar que la presencia de error en la variable ambiental lleva a una subestimación de  $\alpha$ , la cual incrementa a medida que el nivel del grado aumenta. Las diferencias entre los modelos CE-M1 y CE-M2 se hacen más evidentes con el aumento del error, obteniéndose distribuciones más cercanas al valor real con el modelo CE-M2, sobretodo en el caso de variables continuas.

En las figuras 4.2 se muestran los valores de RECM e ID obtenidos para el parámetro  $\alpha$  en cada uno de los escenarios de simulación. Los resultados obtenidos son coherentes con los analizados con respecto a la distribución de los valores esperados. En todos los casos los valores de RECM para el modelo CE-M2 son más bajos que para el modelo CE-M1, aumentando la diferencia con el aumento del grado de error y sobretodo en el caso continuo (figura 4.2a). El comportamiento para el índice de dispersión (ID) resulta similar: índices más bajos a menor presencia de error y valores más pequeños para el modelo CE-M2. Nuevamente se observan mayores diferencias entre los modelos CE-M1 y CE-M2 en el caso continuo.

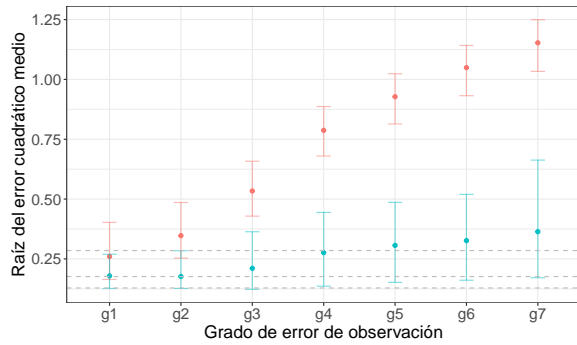


(a) Resultados en el caso de que la variable ambiental es continua. En el eje x los diferentes grados de error de la tabla 4.1

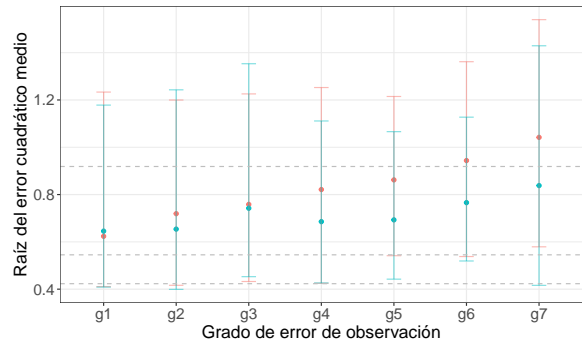


(b) Resultados en el caso de que la variable ambiental es categórica binaria. En el eje x los diferentes grados de error de la tabla 4.2.

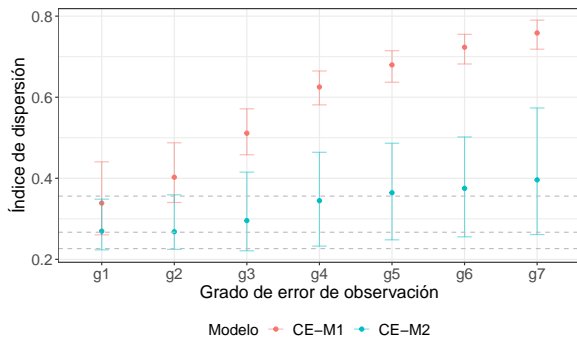
Figura 4.1: Distribución de los valores esperados para el parámetro  $\alpha$  cuando la variable ambiental es continua(a) y cuando es categórica binaria(b). Con color rojo las distribuciones para el modelo CE-M1 y con verde para el modelo CE-M2. El boxplot azul indica la distribución de los valores esperados en el modelo base (SE-M1). La línea punteada marca el valor real del parámetro.



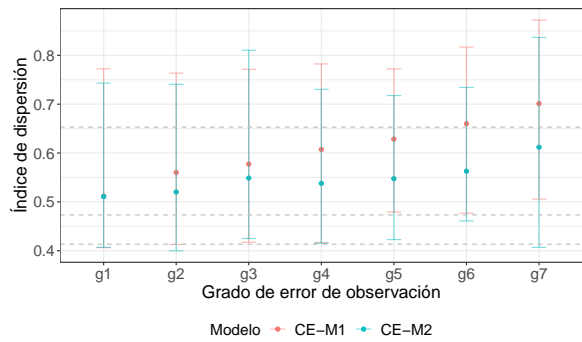
(a) RECM para el caso continuo



(b) RECM para el caso categórico binario



(c) ID para el caso continuo



(d) ID para el caso categórico binario

Figura 4.2: Resultados obtenidos de RECM e ID para el parámetro  $\alpha$  para las simulaciones con variables ambientales continuas (a-c) y para las simulaciones con variables ambientales categóricas binarias (b-d). En el eje x los diferentes grados de error dados por las tablas En el eje x los diferentes grados de error de la tabla 4.2 y 4.1 y en el eje y los valores de los índices calculados. Los puntos indican los valores medios de ambos índices y las barras los intervalos del 95%. En rojo los valores para el modelo CE-M1 y en verde los valores para el modelo CE-M2. Con líneas grises punteadas se muestran la media y el intervalo del 95% de los valores del modelo base (SE-M1).

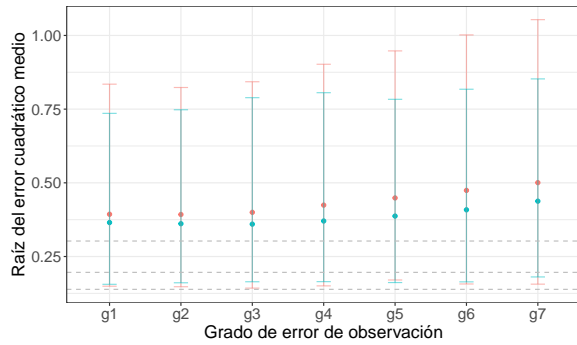
Se tiene entonces que el modelo CE-M2 presenta estimaciones para  $\alpha$  menos sesgadas y con menor dispersión que el modelo CE-M1, obteniéndose mayores diferencias cuando las covariables ambientales son continuas. Sin embargo, si bien en todos los casos las distribuciones posteriores para  $\alpha$  del modelo CE-M2 contienen al valor verdadero, a medida que aumenta el grado del error las mismas se sesgan a izquierda indicando una subestimación en el parámetro de selección de ambiente.

Si bien lo interesante de este análisis es ver como se comportan las estimaciones del parámetro que mide el efecto del paisaje en las trayectorias ( $\alpha$ ), también evaluamos la capacidad de los modelos de recuperar el valor del parámetro  $\beta$  que mide la dispersión del núcleo de movimiento. Las distribuciones de los valores esperados para  $\beta$  se presentan en la figura 4.4 y los valores de RECM e ID en la figura 4.3. Las estimaciones para este parámetro son bastante estables en todos los escenarios de simulación presentados. Las distribuciones para el valor esperado no varían demasiado dependiendo del modelo, el grado de error o la naturaleza de la variable ambiental (figura 4.4), aunque se observa un poco de sesgo cuando el grado del error es grande y la variable es continua. En lo que respecta a los índices RECM e ID cuando la variable es continua el modelo base SE-M1 supera ampliamente a los modelos CE-M1 y CE-M2 (4.3a y 4.3c) indicando que la presencia de error en la variable ambiental puede llevar a mayor sesgo y dispersiones en las estimaciones de  $\beta$ . En el caso categórico no se observa este comportamiento: las estimaciones son similares en todos los grados y modelos (figuras 4.3b y 4.3d). De todas maneras, el comportamiento para los valores esperados es estable también para el caso continuo, excepto para el grado más alto de error, indicando entonces que en principio la presencia de error en las variables ambientales no modifica la estimación media del parámetro de movimiento.

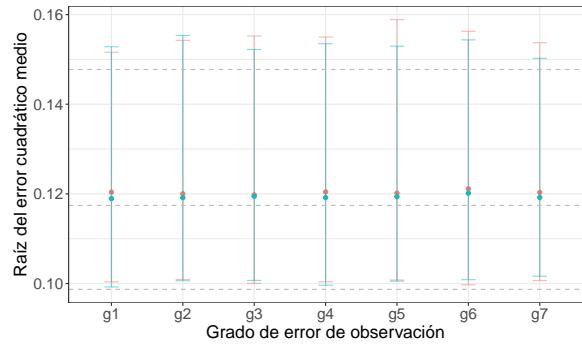
## 4.6. Discusión

La dificultad para hacer inferencia en los modelos de selección de paso subyace en la relación entre el proceso de movimiento y la interacción con el ambiente, que lleva a una expresión de verosimilitud compleja de calcular (ecuación 4.1). Esto hace que sea necesario utilizar aproximaciones para el cálculo de la verosimilitud, siendo la aproximación por regresión logística condicional la más utilizada. Bajo este contexto, incluir errores en las variables ambientales lleva a un nuevo desafío a la hora de hacer inferencia, ya que implica incluir una variable latente de gran dimensión dentro de la formulación. En este trabajo propusimos sustituir el modelo aproximado de regresión logística condicional por un modelo de regresión logística simple de manera de evitar la dependencia entre las observaciones, lo que nos permitió marginalizar la función de verosimilitud de manera más sencilla y utilizar técnicas de *Hamiltonian Monte Carlo* para la inferencia. Desarrollamos la formulación del modelo y los pasos para hacer inferencia tanto cuando el ambiente es medido por variables continuas como cuando lo es por variables categóricas binarias. Mediante un estudio de simulaciones analizamos cómo influye la presencia de error en las observaciones ambientales cuando este error no se considera en la formulación y comparamos con el modelo que si lo contempla.

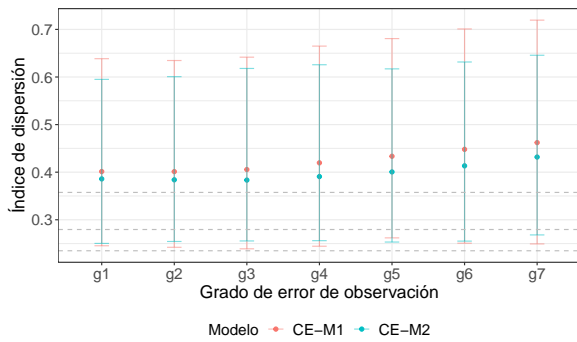
Nuestro estudio de simulación evidencia cómo la presencia de error en las observaciones ambientales lleva a estimaciones incorrectas, subestimado el efecto del ambiente



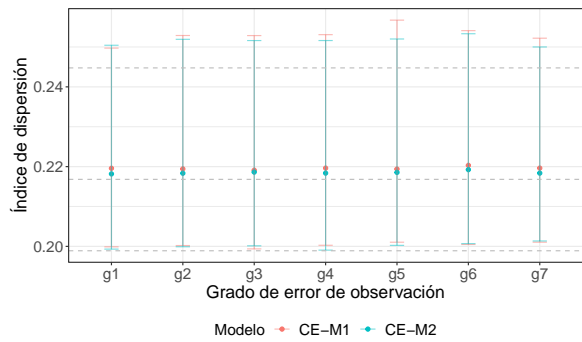
(a) RECM para el caso continuo



(b) RECM para el caso categórico binario

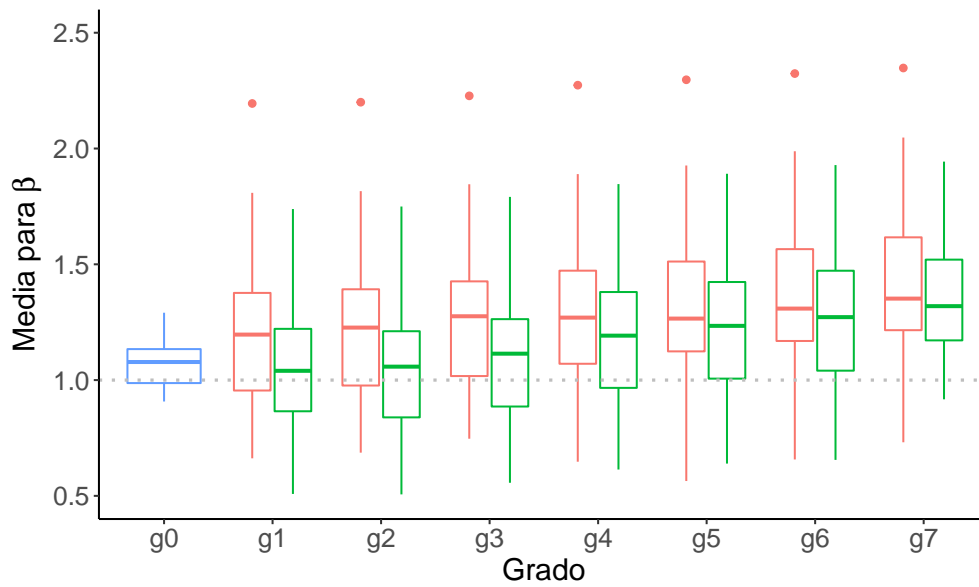


(c) ID para el caso continuo

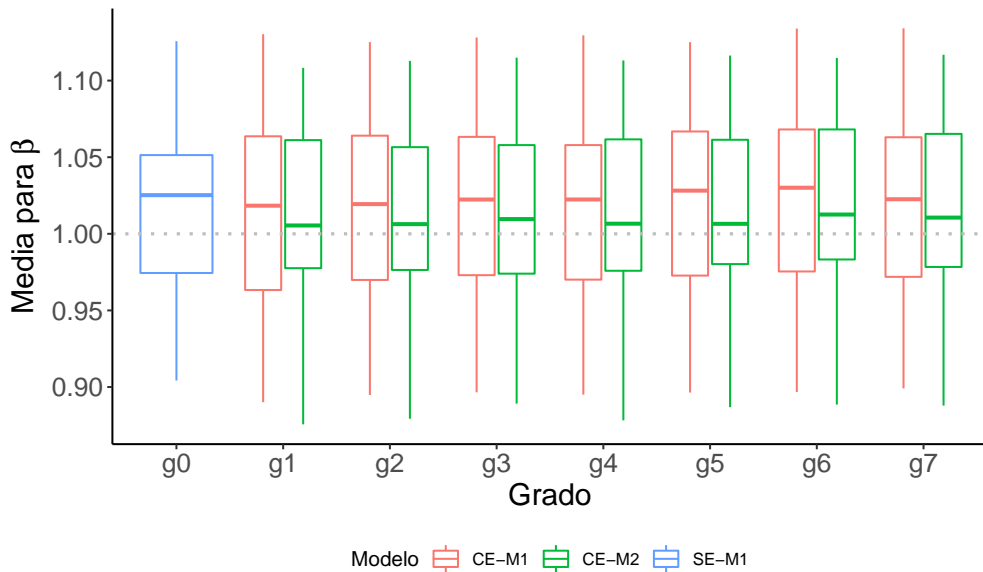


(d) ID para el caso categórico binario

Figura 4.3: Resultados obtenidos de RECM e ID para el parámetro  $\beta$  para las simulaciones con variables ambientales continuas (a-c) y para las simulaciones con variables ambientales categóricas binarias (b-d). En el eje x los diferentes grados de error dados por las tablas En el eje x los diferentes grados de error de la tabla 4.2 y 4.1 y en el eje y los valores de los índices calculados. Los puntos indican los valores medios de ambos índices y las barras los intervalos del 95%. En rojo los valores para el modelo CE-M1 y en verde los valores para el modelo CE-M2. Con líneas grises punteadas se muestran la media y el intervalo del 95% de los valores del modelo base (SE-M1).



(a) Resultados en el caso de que la variable ambiental es continua. En el eje x los diferentes grados de error de la tabla 4.1



(b) Resultados en el caso de que la variable ambiental es categórica binaria. En el eje x los diferentes grados de error de la tabla 4.2

Figura 4.4: Distribución de los valores esperados para el parámetro  $\beta$  cuando la variable ambiental es continua(a) y cuando es categórica binaria. Con color rojo las distribuciones para el modelo CE-M1 y con verde para el modelo CE-M2. El boxplot azul indica la distribución de los valores esperados en el modelo base (SE-M1). La línea punteada marca el valor real del parámetro.

y la selección del paisaje a medida que el grado del error aumenta. Pudimos observar que cuando el grado del error es grande y la variable ambiental es de naturaleza continua, se producen sobreestimaciones del parámetro de selección de paso  $\beta$ . De esta manera, evidenciamos como la presencia de error en las variables ambientales puede afectar no solo las conclusiones sobre el efecto del ambiente en el movimiento, sino que además en ciertas ocasiones puede trasladarse también a las conclusiones sobre la disponibilidad del ambiente. El modelo aquí propuesto para contemplar dicho error (CE-M2) funcionó siempre mejor que el modelo que no lo contempla (CE-M1). Cuando el grado de error no es demasiado grande (15% en relación a la magnitud de la variable ambiental), nuestro modelo logra corregir completamente el sesgo. Sin embargo, cuando el grado del error es grande y a medida que este aumenta, CE-M2 también subestima el parámetro de selección de paisaje (aunque siempre en menor medida que el modelo inicial CE-M1).

El modelo CE-M2 presentado en este capítulo, constituye un primer paso en la inclusión de error en las observaciones ambientales. Si bien hemos demostrado que este modelo permite mejorar las estimaciones, también hemos visto que posee ciertas limitaciones. Creemos que son necesarios análisis más exhaustivos para estudiarlo en mayor profundidad y mejorar su capacidad inferencial. En principio conjeturamos que incluir muestras más grandes podría llevar a estimaciones más precisas, reduciendo el sesgo cuando el grado de error es grande. Sin embargo, aumentar el número de muestras conlleva un coste computacional que debe ser tenido en cuenta a la hora de la implementación. Por otro lado, el modelo CE-M2 aquí presentado asume que el error presente en las variables ambientales no posee ninguna dependencia espacial. Sin embargo, es razonable pensar que observaciones cercanas en el espacio posean grados de error similares. Consideramos entonces que sería interesante analizar si el sesgo presente de las estimaciones varía cuando el error en las variables ambientales poseen cierto grado de dependencia espacial, y estudiar la forma de incluir esa dependencia dentro del análisis.

En este trabajo utilizamos el enfoque Bayesiano para hacer inferencia, lo que nos permitió incluir la presencia de error en la observación del ambiente de manera relativamente simple. En este caso asumimos que la magnitud del error es conocido por el usuario, supuesto muchas veces razonable, sobretodo cuando se trabaja con datos satelitales o de radar. Sin embargo, si bien en ocasiones la medida del error puede no saberse de manera exacta, es posible obtener una aproximación. En ese caso, se pueden incluir como parámetros del modelo las magnitudes del error, otorgándoles previas informativas basadas en las estimaciones empíricas de los errores.

Al trabajar con datos de telemetría, suele suceder que los datos de localización observados también tienen cierto grado de error. Muchas veces desperfectos en los dispositivos GPS pueden llevar a fallos en las localizaciones de los individuos, o incluso pueden no guardar información para ciertos tiempos, generando bases con datos faltantes. La pregunta de cómo incluir estos errores dentro del modelo resulta útil e interesante a la hora de estudiar estos sistemas. Existen diversos trabajos que buscan dar una respuesta a este problema ([Ganskopp y Johnson, 2007](#); [Jerde y Visscher, 2005](#); [Brost et al., 2015](#)). Sin embargo hasta el momento no existen estudios que busquen incluir los errores presentes en los datos de localización y los errores de las variables ambientales de manera simultánea.

El capítulo de esta tesis busca analizar cómo la presencia de error en las observaciones ambientales puede llevar a conclusiones erróneas sobre el comportamiento de un sistema de movimiento que varía en el tiempo y el espacio. En general, la mayoría de los estudios que apuntan a describir trayectorias de animales ignoran la presencia de error en las observaciones ambientales, desestimando las consecuencias que pueden tener en los resultados. La forma en que se mueven los animales está intrínsecamente relacionada con la forma en que habitan el espacio y seleccionan los recursos disponibles. Por esto, muchas veces las estimaciones de selección de recursos se utilizan para informar estrategias de conservación y manejo de animales. Por lo tanto, conclusiones o interpretaciones erróneas sobre cómo los individuos seleccionan el ambiente pueden llevar a prácticas de manejo incorrectas. Considerando los resultados obtenidos en este trabajo, creemos es necesario prestar mayor atención al efecto del error en los datos ambientales para evitar sacar conclusiones erróneas que lleven a prácticas de manejo equivocadas.



## 5. Discusión

---

Entender cuáles son los procesos que subyacen a la forma en que se mueven los animales permite en parte mejorar las estrategias de producción animal y la conservación de recursos naturales (Clemmons y Buchholz, 1997; Gosling y Sutherland, 2000). Comprender cómo los animales silvestres habitan el paisaje, usan el ambiente y emplean diferentes estrategias de movimiento posee importancia a nivel socioambiental y económico, ya que proporciona herramientas para poder manejar de manera eficiente y sustentable poblaciones sujetas a cambios en el ambiente y el clima (Hall y Paruelo, 2006; Andersson et al., 2011). Los patrones que gobiernan la forma en que se mueven los animales pueden depender de factores tales como la relación con el medio ambiente, el estado interno de los mismos, las interacciones con otros individuos o la disponibilidad de recursos. Así, el estudio de estos sistemas resulta un desafío, ya que implica integrar diferentes fenómenos que ocurren a distintas escalas espaciales y temporales (Gurarie y Ovaskainen, 2011; McClintock et al., 2014).

En la actualidad, nuevas tecnologías permiten el monitoreo y rastreo de animales, así como también posibilitan la obtención de datos relacionados al paisaje y a la fisiología de los individuos (Kays et al., 2015). Estos avances tecnológicos han generado grandes expectativas en cuanto a la posibilidad de lograr importantes progresos en el entendimiento de cómo los animales reaccionan al medio ambiente y cómo el comportamiento de los individuos se traduce en distintas dinámicas poblacionales (Morales et al., 2010). Sin embargo, para alcanzar estos objetivos es necesario sortear las dificultades propias de relacionar e integrar las distintas escalas de los procesos de movimiento, comportamiento y disponibilidad de recursos, con las escalas de las fuentes de información. Por ejemplo, los GPS permiten obtener datos de localización a frecuencias de minutos o segundos, pero en general se programan para coleccionar datos con mucha menor frecuencia (horas) por limitaciones de las baterías. Los dispositivos de tipo *Daily Diary* registran datos de aceleración y geomagnetismo en tres ejes con resoluciones temporales de alrededor de 40 datos por segundo, pero los cambios de movimiento o comportamiento de los animales pueden ocurrir en cualquier instante de tiempo, siendo en general los primeros más frecuentes que los segundos. Cómo integrar y utilizar la información de estos dispositivos en modelos de movimiento y uso del espacio sigue siendo un gran desafío. La complejidad principal está en incorporar la estructura de estos datos dentro del análisis, teniendo en cuenta la correlación espacio-temporal, la presencia de error de medición y la enorme cantidad de datos coleccionados.

Ocurre que las escalas temporales a las que tomamos los datos en general no suelen ser las mismas a las que suceden los procesos de movimiento. Por eso es que los modelos espacio estado (MEE) con los que trabajamos en esta tesis, al distinguir el modelo estadístico para el método de observación del modelo para la dinámica de movimiento, permiten diferenciar y conectar estas dos escalas evitando sacar conclusiones erradas debido a supuestos incorrectos sobre los tiempos a los que ocurren los procesos. Los MEE han demostrado ser útiles, no solamente porque permiten modelar las diferentes escalas temporales, sino porque además permiten incluir la presencia de error en las

observaciones de manera sencilla.

En esta tesis nos abocamos a estudiar cómo hacer uso de los modelos espacio estado de manera de integrar la información proveniente de los datos de telemetría para analizar sistemas de movimiento y comportamiento animal. En la primera parte presentamos un MEE que permitió describir trayectorias formulando el proceso de movimiento a tiempo continuo y la observación a tiempo discreto, combinando la facilidad de interpretación de los modelos basados en pasos y giros con el realismo de los modelos a tiempo continuo. Debido precisamente a la combinación de escalas temporales, la función de verosimilitud resultó difícil de calcular. Sin embargo, a la hora de hacer inferencia los métodos de cómputo Bayesiano aproximado mostraron un buen desempeño. Encontramos que la capacidad de recuperar correctamente los parámetros del modelo está intrínsecamente vinculada con la relación entre la escala temporal del proceso de observación y el tiempo medio entre los cambios de dirección de movimiento. Estos resultados por un lado resaltan la importancia de que la escala temporal a la que los animales toman las decisiones de movimiento debe tenerse en cuenta a la hora de diseñar protocolos de colecta de datos. Además, evidencian el hecho de que no siempre es necesario contar con datos de alta frecuencia para tener buenas estimaciones de ciertos procesos de movimiento. Si bien los métodos ABC resultaron útiles en este contexto, consideramos que es posible estudiar si la implementación de otras técnicas de verosimilitud libre permiten mejorar aun más la capacidad inferencial. Por ejemplo, podría ser interesante estudiar diferentes configuraciones de estadísticos resumen que permitan aplicar técnicas “Bayesianas de verosimilitud sintética” (Price et al., 2018) y estudiar si la precisión y la eficiencia computacional mejoran con respecto a las técnicas ABC. Además futuros desarrollos deberían considerar características adicionales que son comunes a los estudios de movimiento, como la inclusión de más de un comportamiento en el movimiento y el efecto de las características del paisaje (Morales et al., 2004; Mevin B. Hooten et al., 2017). La ampliación de los modelos, llevaría a contar con trayectorias con mayor estructura, lo que permitiría incrementar los posibles estadísticos resumen para caracterizar los datos.

En la segunda parte de la tesis presentamos cómo utilizar modelos ocultos de Markov (HMM), y modelos ocultos semi-Markovianos (HSMM) con o sin estructura autorregresiva para clasificar de manera supervisada datos con dependencia temporal. Desarrollamos y explicamos sus diferencias en cuanto a su estructura, la manera de hacer inferencia y predicción. Estudiamos la capacidad de estos modelos de hacer clasificaciones bajo diferentes escenarios de manera de comprender bajo que condiciones extender un HMM a un HSMM mejora las predicciones. Por último utilizamos estas técnicas para clasificar datos de aceleración de ovejas con alta correlación temporal en distintos estados comportamentales. Una de las bondades de estos modelos es que nos permitieron distinguir la dependencia temporal presente en los datos de aceleración de la del proceso oculto de cambios de comportamiento, diferenciando así dos escalas temporales diferentes. A la hora de hacer clasificación supervisada en datos con dependencia temporal, la atención se centra en la capacidad de distinguir entre diferentes categorías predefinidas para luego predecir los estados no observados. Según el estudio de simulación realizado y los resultados obtenidos, concluimos que cuando las distribuciones de las observaciones de las diferentes categorías poseen solapamiento significativo y las distribuciones de los tiempos de permanencia difieren de la distribución

geométrica, considerar extender el modelo básico HMM a un modelo HSMM con mejores especificaciones, mejora las predicciones. Sin embargo, también observamos que varios grados de especificación errónea pueden no dar lugar a predicciones que difieran mucho de aquellas obtenidas con un modelo más complejo pero mejor especificado. A la hora de clasificar diferentes estados comportamentales a partir de datos de aceleración, concluimos que tanto los modelos HMM como sus extensiones HSMM resultan adecuados. No obstante, en el estudio puntual de la base de datos de aceleración de ovejas, si bien obtuvimos clasificaciones precisas para todos los modelos estudiados, no observamos diferencias importantes en el rendimiento de estos modelos. Si bien esto puede explicarse debido al bajo solapamiento entre las distribuciones de las observaciones, creemos que ampliar la base de datos existente podría aumentar la diferencia entre la capacidad predictiva de estos modelos. En el estudio realizado, debido a falta de muestras, unificamos dos comportamientos diferentes en uno solo: actividad de descanso y vigilancia como inactivo. Si bien estos dos comportamientos poseen datos de aceleración similares, los tiempos en que las ovejas pasan en estos comportamientos son diferentes: largos periodos de descanso y periodos muy breves de vigilancia. Por lo tanto, la diferencia entre estos dos estados está dada por los tiempos de permanencia. Conjeturamos que en este contexto, los modelos HSMM al poder caracterizar estos tiempos de manera más precisa, mejorarían las clasificaciones en relación con las predicciones de los modelos HMM.

Por último, estudiamos cómo describir trayectorias considerando la selección de recursos disponibles e incluyendo la presencia de error en las variables ambientales. Para este objetivo extendimos los modelos integrados de selección de pasos considerando los valores verdaderos de la variable ambiental como faltantes. Utilizamos así un enfoque espacio estado que nos permitió diferenciar el modelo de observación de covariables con error del modelo del sistema real. Los resultados obtenidos evidencian cómo la presencia de error en las observaciones ambientales lleva a subestimar el efecto del ambiente en el proceso de movimiento. La forma en que se mueven los animales está intrínsecamente relacionada con la forma en que habitan el espacio y seleccionan los recursos ambientales. Contar con modelos adecuados para describir e interpretar correctamente la interacción de estos sistemas resulta crucial a la hora de diagramar prácticas de manejo apropiadas. Los resultados obtenidos evidencian la importancia de incorporar el efecto del error en los datos ambientales para evitar sacar conclusiones erróneas que lleven a prácticas de manejo equivocadas.

Para hacer inferencia en todos los modelos utilizados a lo largo de esta tesis hicimos uso del enfoque Bayesiano. Esta perspectiva viene siendo cada vez más utilizada en el análisis de modelos ecológicos ([Schuwirth et al., 2019](#); [Van de Schoot et al., 2021](#)) debido a que posee una serie de ventajas. En primer lugar el enfoque Bayesiano permite ajustar modelos complejos que muchas veces pueden no ser abordables por los métodos clásicos. Por ejemplo, en nuestro caso pudimos ajustar el modelo propuesto en el primer capítulo cuya verosimilitud era intratable mediante técnicas Bayesianas de verosimilitud libre (sección 2). En segundo lugar, permite incluir medidas de incertidumbre y la presencia de datos perdidos de manera directa, como el caso del modelo aquí presentado para describir trayectorias con error en las variables ambientales (sección 4). Además, la perspectiva Bayesiana proporciona una forma de incluir información conocida acerca del sistema de manera natural mediante el uso de las distribuciones previas.

Por ejemplo, nosotros pudimos incorporar conocimientos previos sobre los tiempos de duración de los diferentes comportamientos de las ovejas a la hora de ajustar los modelos HMM para hacer clasificación en la sección 3. Además, se tiene que las limitaciones en cuanto a las implementaciones en el uso de la estadística Bayesiana están siendo subsanadas gracias al aumento de las capacidades de computación en paralelo y a la mejora de programas informáticos cada vez más asequibles (Little, 2006; Wagenmakers et al., 2008). Así, ha habido un reciente desarrollo de programas informáticos de código abierto tales como Stan (programa que usamos en esta tesis), que hacen uso de algoritmos de muestreo eficientes (p. ej. Hamiltonian Monte Carlo o Gibbs sampling) y permiten utilizar las técnicas Bayesinas de manera sencilla. Asimismo, con la aparición y desarrollo de nuevos paquetes, han proliferado de forma notable los recursos para implementar distintas aplicaciones Bayesianas en R (el caso del paquete “abctools” utilizado en la sección 2) facilitando así su uso para los usuarios. Consideramos que, en esta tesis, la perspectiva Bayesiana resultó beneficiosa para la inferencia de modelos de movimiento y comportamiento animal bajo el enfoque espacio estado.

Si bien creemos que en este trabajo hemos hecho considerables progresos en cuanto al modelado y análisis estadísticos de datos de movimiento, quedan abiertas preguntas interesantes para trabajos futuros, algunas que constituyen ampliaciones de lo realizado y otras que han surgido a partir de los resultados obtenidos. En primer lugar queda abierto el interrogante de cómo conectar diferentes modelos de movimiento según comportamientos asociados, combinando datos de localización GPS con datos de acelerometría considerando diferentes escalas temporales. Por un lado hemos visto que los datos de acelerómetros proveen una forma de clasificar distintos comportamientos, y por el otro que los datos de localización GPS permiten describir trayectorias de movimiento. Como la forma en que se mueven los animales depende en parte de su comportamiento, muchos estudios han integrado distintos modelos de movimiento según diversos comportamientos (Morales et al., 2004; Jonsen et al., 2005; McClintock et al., 2012; Wang, 2019). Sin embargo, son todavía pocos los análisis que incluyen datos con diferentes escalas temporales provenientes de diferentes fuentes de información. En segundo lugar, hemos estudiado la capacidad de los modelos ocultos de Markov y sus extensiones para clasificar datos temporales bajo el enfoque supervisado. En la práctica, cuando se trabaja con bases de datos pre-clasificadas, puede suceder que los estados pre-asignados no sean correctos para todas las observaciones, es decir, que posean cierto grado de error. Existen trabajos en donde se ha estudiado el efecto de las etiquetas ruidosas a la hora de hacer clasificación supervisada (Sáez et al., 2014; Garcia et al., 2015; Wang et al., 2021). Sin embargo, queda aun pendiente estudiar como incluir este ruido en clasificadores por modelos ocultos de Markov para datos temporales. Por último, hemos visto que no incluir el error presente en las observaciones ambientales puede muchas veces sesgar las estimaciones de los análisis. Al trabajar con datos de telemetría, suele suceder que los datos de localización observados también posean cierto grado de error. Si bien la pregunta de cómo incluir estos errores dentro del modelo ya ha sido estudiada (Ganskopp y Johnson, 2007; Jerde y Visscher, 2005; Brost et al., 2015), todavía queda pendiente analizar la forma de incluir los errores presentes en los datos de localización y los errores de las variables ambientales de manera simultánea.

El desarrollo de herramientas estadísticas para la modelización de datos de movimiento animal recién está empezando a ponerse al día con la explosión del volumen de

datos correspondientes y los retos de modelización asociados. Cómo coleccionar los datos telemétricos y cómo integrar esta información en los análisis considerando su autocorrelación espacio temporal, las diferencias entre las escalas a las que se toman los datos y suceden los procesos y la complejidad de los sistemas de movimiento animal, ha generado una enorme demanda de conocimientos estadísticos en esta área. El principal reto reside en el equilibrio adecuado entre los enfoques de modelización excesivamente complejos e inaccesibles y los accesibles pero excesivamente simplistas. Consideramos que el progreso hacia este fin requerirá una estrecha colaboración y compromiso entre las disciplinas de estadística y ecología.



## 6. Bibliografía

---

- Andersson, E., Brogaard, S., y Olsson, L. (2011). The political ecology of land degradation. Annual Review of Environment and Resources, 36(1):295–319.
- Avgar, T., Potts, J. R., Lewis, M. A., y Boyce, M. S. (2016). Integrated step selection analysis: bridging the gap between resource selection and animal movement. Methods in Ecology and Evolution, 7(5):619–630.
- Aydin, Z., Altunbasak, Y., y Borodovsky, M. (2006). Protein secondary structure prediction for a single-sequence using hidden semi-Markov models. BMC Bioinformatics, 7(1):178.
- Baudet, C., Donati, B., Sinimeri, B., Crescenzi, P., Gautier, C., Matias, C., y Sagot, M.-F. (2015). Cophylogeny reconstruction via an approximate Bayesian computation. Syst. Biol., 64:416–431.
- Beaumont, M. A. (2010). Approximate Bayesian Computation in Evolution and Ecology. Annual Review of Ecology, Evolution, and Systematics, 41:379–406.
- Beaumont, M. A., Zhang, W., y Balding, D. J. (2002). Approximate Bayesian Computation in Population Genetics. Genetics, 162:2025–2035.
- Benouareth, A., Ennaji, A., y Sellami, M. (2007). Arabic Handwritten Word Recognition Using HMMs with Explicit State Duration. EURASIP Journal on Advances in Signal Processing, 2008(1):247354.
- Bertorelle, G., Benazzo, A., y Mona, S. (2010). ABC as a flexible framework to estimate demography over space and time: some cons, many pros. Molecular Ecology, 19:2609–2625.
- Bishop, C. M. (2012). Model-based machine learning. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 371(1984):20120222–20120222.
- Blackwell, P. (1999). Random diffusion models for animal movement. Ecological Modelling, 100:87–102.
- Blum, M. G. B. y François, O. (2010). Non-linear regression models for Approximate Bayesian Computation. Stat Comput, 20:63–73.
- Boyce, M. S. y McDonald, L. L. (1999). Relating populations to habitats using resource selection functions. Trends in Ecology & Evolution, 14(7):268–272.
- Breed, G. A., Bowen, W. D., Mcmillan, J. I., y Leonard, M. L. (2006). Sexual segregation of seasonal foraging habitats in a non-migratory marine mammal. Proceedings of the Royal Society B: Biological Sciences, 273:2319 – 2326.

- Brost, B. M., Hooten, M. B., Hanks, E. M., y Small, R. J. (2015). Animal movement constraints improve resource selection inference in the presence of telemetry error. Ecology, 96(10):2590–2597.
- Brown, D. D., Kays, R., Wikelski, M., Wilson, R., y Klimley, A. P. (2013). Observing the unwatchable through acceleration logging of animal behavior. Animal Biotelemetry, 1:1–20.
- Bryan, J. D. y Levinson, S. E. (2015). Autoregressive Hidden Markov Model and the Speech Signal. Procedia Computer Science, 61:328–333.
- Bulla, J. y Bulla, I. (2006). Stylized facts of financial time series and hidden semi-markov models. Computational Statistics & Data Analysis, 51(4):2192–2209.
- Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., y Riddell, A. (2017a). Stan: A probabilistic programming language. Journal of Statistical Software, 76:1.
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., y Riddell, A. (2017b). Stan: A Probabilistic Programming Language. Journal of Statistical Software, 76:1–32.
- Carroll, G., Slip, D., Jonsen, I., y Harcourt, R. (2014). Supervised accelerometry analysis can identify prey capture by penguins at sea. Journal of Experimental Biology, 217(24):4295–4302.
- Chakravarty, P., Cozzi, G., Ozugul, A., y Aminian, K. (2019). A novel biomechanical approach for animal behaviour recognition using accelerometers. Methods in Ecology and Evolution, 10(6):802–814.
- Chen, K., Hasegawa-Johnson, M., Cohen, A., Borys, S., Sung-Suk Kim, Cole, J., y Jeung-Yoon Choi (2006). Prosody dependent speech recognition on radio news corpus of american english. IEEE Transactions on Audio, Speech, and Language Processing, 14(1):232–245.
- Cheng, W. T. y Chan, K. L. (1998). Classification of electrocardiogram using hidden markov models. In Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol.20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No.98CH36286), pages 143–146 vol.1.
- Chung, P.-C. y Liu, C.-D. (2008). A daily behavior enabled hidden markov model for human behavior understanding. Pattern Recognition, 41(5):1572–1580.
- Clayton, D. G. (1992). Models for the analysis of cohort and case-control studies with inaccurately measured exposures. n Statistical Models for Longitudinal Studies of Exposure and Health, edited by James H. Dwyer, Manning Feinleib, Peter Lippert, and Hans Hoffmeister. New York: Oxford University Press., pages 301–31.
- Clemmons, J. y Buchholz, R. (1997). Behavioral approaches to conservation in the wild. Cambridge University Press.

- Codling, E. A., Plank, M. J., y Benhamou, S. (2008). Random walk models in biology. J R Soc Interface, 5:813–834.
- Csilléry, K., Blum, M. G. B., Gaggiotti, O. E., y François, O. (2010). Approximate Bayesian Computation (ABC) in practice. Trends in Ecology & Evolution, 25:410–418.
- Csilléry, K., François, O., y Blum, M. G. B. (2012). abc: an R package for approximate Bayesian computation (ABC). Methods in Ecology and Evolution, 3:475–479.
- Deng, L. y Li, X. (2013). Machine learning paradigms for speech recognition: An overview. IEEE Transactions on Audio, Speech, and Language Processing, 21(5):1060–1089.
- Dietterich, T. G. (2002). Machine Learning for Sequential Data: A Review. In Caelli, T., Amin, A., Duin, R. P. W., de Ridder, D., y Kamel, M., editors, Structural, Syntactic, and Statistical Pattern Recognition, Lecture Notes in Computer Science, pages 15–30. Springer Berlin Heidelberg.
- Duong, T. V., Bui, H. H., Phung, D. Q., y Venkatesh, S. (2005). Activity recognition and abnormality detection with the switching hidden semi-markov model. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), volume 1, pages 838–845 vol. 1.
- Durbin, J. y Koopman, S. (2001). Time Series Analysis by State Space Methods. Oxford Statistical Science Series.
- Faraway, J. J. (2016). Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models. Chapman and Hall/CRC.
- Fieberg, J., Signer, J., Smith, B., y Avgar, T. (2021). A ‘how to’ guide for interpreting parameters in habitat-selection analyses. Journal of Animal Ecology, 90(5):1027–1043.
- Forester, J. D., Im, H. K., y Rathouz, P. J. (2009). Accounting for animal movement in estimation of resource selection functions: sampling and data analysis. Ecology, 90(12):3554–3565.
- Fortin, D., Beyer, H. L., Boyce, M. S., Smith, D. W., Duchesne, T., y Mao, J. S. (2005). Wolves influence elk movements: Behavior shapes a trophic cascade in yellowstone national park. Ecology, 86(5):1320–1330.
- Francq, C. y Zakoian, J.-M. (2001). Stationarity of multivariate Markov-switching ARMA models. Journal of Econometrics, 102(2):339–364.
- Frenay, B. y Verleysen, M. (2014). Classification in the presence of label noise: A survey. IEEE Transactions on Neural Networks and Learning Systems, 25(5):845–869.
- Fryxell, J. M., Hazell, M., Börger, L., Dalziel, B. D., Haydon, D. T., Morales, J. M., McIntosh, T., y Rosatte, R. C. (2008). Multiple movement modes by large herbivores at multiple spatiotemporal scales. Proceedings of the National Academy of Sciences, 105(49):19114–19119.

- Frühwirth-Schnatter, S., Celeux, G., y Robert, C. P., editors (2019). Handbook of mixture analysis. CRC Press, Taylor & Francis Group, Boca Raton.
- Fèvre, E. M., Bronsvoort, B. M. d. C., Hamilton, K. A., y Cleaveland, S. (2006). Animal movements and the spread of infectious diseases. Trends Microbiol., 14:125–131.
- Ganskopp, D. C. y Johnson, D. D. (2007). Gps error in studies addressing animal movements and activities. Rangeland Ecology & Management, 60(4):350–358.
- Garcia, L. P., de Carvalho, A. C., y Lorena, A. C. (2015). Effect of label noise in the complexity of classification problems. Neurocomputing, 160:108–119.
- Geisser, S. (1975). The predictive sample reuse method with applications. Journal of the American Statistical Association, 70(350):320–328.
- Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., y Rubin, D. (2013). Bayesian Data Analysis, Third Edition. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis.
- Geurts, P. (2001). Pattern extraction for time series classification. In De Raedt, L. y Siebes, A., editors, Principles of Data Mining and Knowledge Discovery, pages 115–127, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Gosling, L. M. y Sutherland, W. J. (2000). Behaviour and conservation. Cambridge University Press.
- Graf, P. M., Wilson, R. P., Qasem, L., Hackländer, K., y Rosell, F. (2015). The use of acceleration to code for animal behaviours; a case study in free-ranging eurasian beavers *Castor fiber*. PLOS ONE, 10(8):1–17.
- Guédon, Y. (2003). Estimating hidden semi-Markov chains from discrete sequences. Journal of Computational and Graphical Statistics, 12(3):604–639.
- Gurarie, E. y Ovaskainen, O. (2011). Characteristic Spatial and Temporal Scales Unify Models of Animal Movement. The American Naturalist, 178:113–123.
- Hall, S. y Paruelo, J. (2006). Environmental controls on lambing rate in patagonia (argentina): A regional approach. Journal of Arid Environments, 64(4):713–735.
- Hamilton, J. (1994). Time Series Analysis. Princeton University Press, Princeton, NJ.
- Harris, K. J. y Blackwell, P. G. (2013). Flexible continuous-time modelling for heterogeneous animal movement. Ecological Modelling, 255:29–37.
- Harvey, A. C. (1990). Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press.
- Hefley, T. J., Baasch, D. M., Tyre, A. J., y Blankenship, E. E. (2014). Correction of location errors for presence-only species distribution models. Methods in Ecology and Evolution, 5(3):207–214.

- Hieronimus, J. L., McKelvie, D., y McInnes, F. R. (1992). Use of acoustic sentence level and lexical stress in hsmm speech recognition. In Proceedings of the 1992 IEEE International Conference on Acoustics, Speech and Signal Processing - Volume 1, ICASSP'92, page 225–227, USA. IEEE Computer Society.
- Hongeng, S. y Nevatia, R. (2003). Large-scale event detection using semi-hidden markov models. In Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2, ICCV '03, page 1455, USA. IEEE Computer Society.
- Hung-Yan Gu, Chiu-Yu Tseng, y Lin-Shan Lee (1991). Isolated-utterance speech recognition using hidden markov models with bounded state durations. IEEE Transactions on Signal Processing, 39(8):1743–1752.
- Inan, O. T., Giovangrandi, L., y Kovacs, G. T. A. (2006). Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features. IEEE Transactions on Biomedical Engineering, 53(12):2507–2515.
- Jerde, C. L. y Visscher, D. R. (2005). Gps measurement error influences on movement model parameterization. Ecological Applications, 15(3):806–810.
- Johnson, C. J. y Gillingham, M. P. (2008). Sensitivity of species-distribution models to error, bias, and model design: An application to resource selection functions for woodland caribou. Ecological Modelling, 213(2):143–155.
- Johnson, D. S., Hooten, M. B., y Kuhn, C. E. (2013). Estimating animal resource selection from telemetry data using point process models. Journal of Animal Ecology, 82(6):1155–1164.
- Johnson, D. S., London, J. M., Lea, M.-A., y Durban, J. W. (2008a). Continuous time correlated random walk model for animal telemetry data. Ecology, 89:1208–1215.
- Johnson, D. S., London, J. M., Lea, M.-A., y Durban, J. W. (2008b). Continuous-time correlated random walk model for animal telemetry data. Ecology, 89:1208–1215.
- Jonsen, I. (2016). Joint estimation over multiple individuals improves behavioural state inference from animal movement data. Scientific Reports, 6(1):20625.
- Jonsen, I. D., Flemming, J. M., y Myers, R. A. (2005). Robust State-Space Modeling of Animal Movement Data. Ecology, 86:2874–2880.
- Juang, B. H. y Rabiner, L. R. (1991). Hidden markov models for speech recognition. Technometrics, 33(3):251–272.
- Kashi, R. S., Hu, J., Nelson, W. L., y Turin, W. (1997). On-line handwritten signature verification using hidden markov model features. In Proceedings of the Fourth International Conference on Document Analysis and Recognition, volume 1, pages 253–257 vol.1.
- Kays, R., Crofoot, M., Jetz, W., y Wikelski, M. (2015). Terrestrial animal tracking as an eye on life and planet. Science, 348:aaa2478–1.

- Kulp, D., Haussler, D., Reese, M. G., y Eeckman, F. H. (1996). A generalized hidden Markov model for the recognition of human genes in DNA. Proceedings. International Conference on Intelligent Systems for Molecular Biology, 4:134–142.
- Kurt Hornik y Bettina Grün (2014). On maximum likelihood estimation of the concentration parameter of von Mises–Fisher distributions. Comput Stat, 29:945–957.
- Langrock, R. y Zucchini, W. (2011). Hidden markov models with arbitrary state dwell-time distributions. Computational Statistics & Data Analysis, 55(1):715–724.
- Leos-Barajas, V., Photopoulou, T., Langrock, R., Patterson, T., Watanabe, Y., Murgatroyd, M., y Papastamatiou, Y. (2017). Analysis of animal accelerometer data using hidden Markov models. Methods in Ecology and Evolution.
- Li, A., Ji, L., Wang, S., y Wu, J. (2010). Physical activity classification using a single triaxial accelerometer based on HMM. In IET International Conference on Wireless Sensor Network 2010 (IET-WSN 2010), pages 155–160.
- Little, R. J. (2006). Calibrated bayes. The American Statistician, 60(3):213–223.
- Liu, Y., Battaile, B.C., Trite, A W., y Zidek, J.V. (2015). Bias correction and uncertainty characterization of Dead-Reckoned paths of marine mammals. Animal Biotelemetry, 3:51.
- Lopes, J. S. y Beaumont, M. A. (2010). ABC: a useful Bayesian tool for the analysis of population data. Infect. Genet. Evol., 10:826–833.
- Marcheret, E. y Savic, M. (1997). Random walk theory applied to language identification. In 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing, volume 2, pages 1119–1122 vol.2.
- Marjoram, P., Molitor, J., Plagnol, V., y Tavaré, S. (2003). Markov chain Monte Carlo without likelihoods. PNAS, 100:15324–15328.
- Marjoram, P. y Tavaré, S. (2006). Modern computational approaches for analysing molecular genetic variation data. Nat. Rev. Genet., 7:759–770.
- Matthiopoulos, J., Fieberg, J., Aarts, G., Beyer, H. L., Morales, J. M., y Haydon, D. T. (2015). Establishing the link between habitat selection and animal population dynamics. Ecological Monographs, 85:413–436.
- McClintock, B. T., Johnson, D. S., Hooten, M. B., Ver Hoef, J. M., y Morales, J. M. (2014). When to be discrete: the importance of time formulation in understanding animal movement. Movement Ecology, 2:21.
- McClintock, B. T., King, R., Thomas, L., Matthiopoulos, J., McConnell, B. J., y Morales, J. M. (2012). A general discrete-time modeling framework for animal movement using multistate random walks. Ecological Monographs, 82:335–349.
- McKinley, T., Cook, A. R., y Deardon, R. (2009). Inference in Epidemic Models without Likelihoods. The International Journal of Biostatistics, 5:24.

- Mendelssohn, I. A. y McKee, K. L. (1988). *Spartina alterniflora* die-back in louisiana: Time-course investigation of soil waterlogging effects. Journal of Ecology, 76(2):509–521.
- Mevin B. Hooten, Devin S. Johnson, Brett T. McClintock, y Juan M. Morales (2017). Animal Movement: Statistical Models for Telemetry Data. CRC Press.
- Millar, R. B. y Meyer, R. (2000). Non-linear state space modelling of fisheries biomass dynamics by using metropolis-hastings within-gibbs sampling. Journal of the Royal Statistical Society. Series C (Applied Statistics), 49(3):327–342.
- Millsbaugh, J. J. y Marzluff, J. M. (2001). Preface. In Millsbaugh, J. J. y Marzluff, J. M., editors, Radio Tracking and Animal Populations, pages xv–xvii. Academic Press, San Diego.
- Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E., y Fryxell, J. M. (2004). Extracting More Out of Relocation Data: Building Movement Models as Mixtures of Random Walks. Ecology, 85:2436–2445.
- Morales, J. M., Moorcroft, P. R., Matthiopoulos, J., Frair, J. L., Kie, J. G., Powell, R. A., Merrill, E. H., y Haydon, D. T. (2010). Building the bridge between animal movement and population dynamics. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 365:2289–2301.
- Nathan, R. (2008). PNAS-2008-Nathan-19050-1. Proceedings of the National Academy of Sciences, 105(49):19050–19051.
- Nathan, R., Spiegel, O., Fortmann-Roe, S., Harel, R., Wikelski, M., y Getz, W. M. (2012). Using tri-axial acceleration data to identify behavioral modes of free-ranging animals: general concepts and tools illustrated for griffon vultures. Journal of Experimental Biology, 215(6):986–996.
- Nunes, M. A. y Prangle, D. (2015). abctools: An R Package for Tuning Approximate Bayesian Computation Analyses. The R Journal, 7:17.
- Oura, K., Heiga Zen, Nankaku, Y., Akinobu Lee, y Tokuda, K. (2006). Hidden semi-markov model based speech recognition system using weighted finite-state transducer. In 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, volume 1, pages I–I.
- Parton, A. y Blackwell, P. G. (2017). Bayesian Inference for Multistate ‘Step and Turn’ Animal Movement in Continuous Time. JABES, 22:373–392.
- Patterson, T. A., Parton, A., Langrock, R., Blackwell, P. G., Thomas, L., y King, R. (2017). Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges. AStA Advances in Statistical Analysis, 101:399–438.
- Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O., y Matthiopoulos, J. (2008). State-space models of individual animal movement. Trends Ecol. Evol. (Amst.), 23:87–94.

- Phillips, J. S., Patterson, T. A., Leroy, B., Pilling, G. M., y Nicol, S. J. (2015). Objective classification of latent behavioral states in bio-logging data using multivariate-normal hidden markov models. Ecological Applications, 25(5):1244–1258.
- Pieczynski, W. (2007). Multisensor triplet markov chains and theory of evidence. Int. J. Approx. Reasoning, 45(1):1–16.
- Potts, J. R., Börger, L., Scantlebury, D. M., Bennett, N. C., Alagaili, A., y Wilson, R. P. (2018). Finding turning-points in ultra-high-resolution animal movement data. Methods in Ecology and Evolution, 9:2091–2101.
- Prado, R. y West, M. (2016). Time Series Modeling, Computation and Inference. CRC Press.
- Prangle, D., Blum, M. G. B., Popovic, G., y Sisson, S. A. (2014). Diagnostic tools for approximate Bayesian computation using the coverage property. Australian & New Zealand Journal of Statistics, 56:309–329.
- Price, L. F., Drovandi, C. C., Lee, A., y Nott, D. J. (2018). Bayesian synthetic likelihood. Journal of Computational and Graphical Statistics, 27(1):1–11.
- Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A., y Feldman, M. W. (1999). Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol. Biol. Evol., 16:1791–1798.
- Qasem, L., Cardew, A., Wilson, A., Griffiths, I., Halsey, L., Shepard, E., Gleiss, A., y Wilson, R. (2012). Tri-axial dynamic acceleration as a proxy for animal energy expenditure; should we be summing values or calculating the vector? PloS one, 7:e31187.
- R Core Team (2019). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Refaeilzadeh, P., Tang, L., y Liu, H. (2009). Cross-Validation, pages 532–538. Springer US, Boston, MA.
- Richardson, S. y Gilks, W. R. (1993). A bayesian approach to measurement error problems in epidemiology using conditional independence models. American journal of epidemiology, 138(6):430–442.
- Rico-Ramirez, M. A. y Cluckie, I. D. (2008). Classification of ground clutter and anomalous propagation using dual-polarization weather radar. IEEE Transactions on Geoscience and Remote Sensing, 46(7):1892–1904.
- Ruiz-Suarez, S., Leos-Barajas, V., Alvarez-Castr, I., y JM., M. (2020). Using approximate bayesian inference for a “steps and turns” continuous-time random walk observed at regular time intervals. PeerJ, 8:e8452.
- Ruiz Suarez, S., Leos-Barajas, V., y Morales, J. (2021). Hidden markov and semi-markov models when and why are these models useful for classifying states in time series data? Journal of Agricultural, Biological and Environmental Statistics, pages 1537–2693.

- Ruiz-Suarez, S., Sued, M., Vidal, L., Salio, P., Rodriguez, D., Nesbitt, S., y Garcia Skabar, Y. (2019). Técnicas de clasificación supervisada para la discriminación entre ecos meteorológicos y no meteorológicos usando información de un radar de banda C. Meteorológica, 44:45–65.
- Schuwirth, N., Borgwardt, F., Domisch, S., Friedrichs, M., Kattwinkel, M., Kneis, D., Kuemmerlen, M., Langhans, S. D., Martínez-López, J., y Vermeiren, P. (2019). How to make ecological models useful for environmental management. Ecological Modelling, 411:108784.
- Scott, A. Sisson, Y. F. y Beaumont, M. A. (2018). Handbook of Approximate Bayesian Computation.
- Sirén, J., Lens, L., Cousseau, L., y Ovaskainen, O. (2018). Assessing the dynamics of natural populations by fitting individual-based models with approximate Bayesian computation. Methods in Ecology and Evolution, 9:1286–1295.
- Studd, E. K., Landry-Cuerrier, M., Menzies, A. K., Boutin, S., McAdam, A. G., Lane, J. E., y Humphries, M. M. (2019). Behavioral classification of low-frequency acceleration and temperature data from a free-ranging small mammal. Ecology and Evolution, 9(1):619–630. Publisher: John Wiley & Sons, Ltd.
- Sullivan, P. J. (1992). A kalman filter approach to catch-at-length analysis. Biometrics, 48(1):237–257.
- Sáez, J. A., Galar, M., Luengo, J., y Herrera, F. (2014). Analyzing the presence of noise in multi-class problems: alleviating its influence with the one-vs-one decomposition. Knowl Inf Syst, 38:179–206.
- Tanaka, M. M., Francis, A. R., Luciani, F., y Sisson, S. A. (2006). Using Approximate Bayesian Computation to Estimate Tuberculosis Transmission Parameters From Genotype Data. Genetics, 173:1511–1520.
- Tavaré, S., Balding, D. J., Griffiths, R. C., y Donnelly, P. (1997). Inferring Coalescence Times From DNA Sequence Data. Genetics, 145:505–518.
- Thurfjell, H., Ciuti, S., y Boyce, M. S. (2014). Applications of step-selection functions in ecology and conservation. Ecology, 2(4):2051–3933.
- Trevor Hastie, Robert Tibshirani, y Jerome Friedman (2001). The Elements of Statistical Learning. Springer, New York, NY, USA.
- Turchin, P. (1998). Quantitative Analysis of Movement: Measuring and Modeling Population Redistribution in Animals and Plants. Sinauer Associates, Sunderland, Massachusetts, USA.
- Van de Schoot, R., Depaoli, S., King, R., Kramer, Biancaand Märtens, K., Tadesse, M. G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J., y Yau, C. (2021). Bayesian statistics and modelling. Nature Reviews Methods Primers, 1:1.

- Van der Vaart, E., Beaumont, M. A., Johnston, A. S. A., y Sibly, R. M. (2015). Calibration and evaluation of individual-based models using Approximate Bayesian Computation. Ecological Modelling, 312:182–190.
- Visscher, D. R. (2006). Gps measurement error and resource selection functions in a fragmented landscape. Ecography, 29(3):458–464.
- Wagenmakers, E.-J., Lee, M., Lodewyckx, T., y Iverson, G. J. (2008). Bayesian Versus Frequentist Inference, pages 181–207. Springer New York.
- Wang, G. (2019). Machine learning for inferring animal behavior from location and movement data. Ecological Informatics, 49:69–76.
- Wang, J., Chen, R., Sun, X., She, M. F. H., y Wu, Y. (2011). Recognizing Human Daily Activities From Accelerometer Signal. Procedia Engineering, 15:1780–1786.
- Wang, J., Liu, Y., y Levy, C. (2021). Fair classification with group-dependent label noise. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, pages 526–536.
- West, M. y Harrison, J. (1997). Bayesian Forecasting and Dynamic Models (2nd Ed.). Springer-Verlag, Berlin, Heidelberg.
- Williams, H. J., Shepard, E. L. C., Duriez, O., y Lambertucci, S. A. (2015). Can accelerometry be used to distinguish between flight types in soaring birds? Animal Biotelemetry, 3(1).
- Williams, H. J., Taylor, L. A., Benhamou, S., Bijleveld, A. I., Clay, T. A., de Grissac, S., Demšar, U., English, H. M., Franconi, N., Gómez-Laich, A., Griffiths, R. C., Kay, W. P., Morales, J. M., Potts, J. R., Rogerson, K. F., Rutz, C., Spelt, A., Trevail, A. M., Wilson, R. P., y Börger, L. (2020). Optimizing the use of biologists for movement ecology research. Journal of Animal Ecology, 89(1):186–206.
- Wilson, J., Liebsch, N., Davies, I. M., Quintana, F., Weimerskirch, H., Storch, S., Lucke, K., Siebert, U., Zankl, S., Müller, G., Zimmer, I., Scolaro, A., Campagna, C., Plötz, J., Bornemann, H., Teilmann, J., y McMahon, C. R. (2007). All at sea with animal tracks; methodological and analytical solutions for the resolution of movement. ScienceDirect, 54:193–210.
- Wilson, R., Shepard, E., y Liebsch, N. (2008a). Prying into the intimate details of animal lives: use of a daily diary on animals. Endangered Species Research, 4:123–137.
- Wilson, R., Shepard, E., y Liebsch, N. (2008b). Prying into the intimate details of animal lives: Use of a daily diary on animals. Endangered Species Research, 4:123–137.
- Wilson, R. y Wilson, M. P. (1988). Dead reckoning a new technique for determining penguin movements at sea. Meeresforschung, 32(2):155–158.

- Wu, H.-i., Li, B.-L., Springer, T., y Neill, W. (2000). Modelling animal movement as a persistent random walk in two dimensions: Expected magnitude of net displacement. Ecological Modelling, 132:115–124.
- XiaoBing Liu, DeShun Yang, y XiaoOu Chen (2008). New approach to classification of chinese folk music based on extension of hmm. In 2008 International Conference on Audio, Language and Image Processing, pages 1172–1179.
- Xu, Z. y Liu, Y. (2020). A Regularized Vector Autoregressive Hidden Semi-Markov Model, with Application to Multivariate Financial Data. arXiv:1804.10308 [stat]. arXiv: 1804.10308.
- Yang, M. (2000). Some properties of vector autoregressive processes with Markov-switching coefficients. Econometric Theory, 16:23–43.
- Yu, S.-Z. (2010). Hidden semi-Markov models. Artificial Intelligence, 174(2):215–243.
- Zucchini, W., MacDonald, I. L., y Langrock, R. (2017). Hidden Markov Models for Time Series: An Introduction Using R, Second Edition. CRC Press.



# Anexos



# A. Anexo I:

---

## A.1. Cálculo de la verosimilitud completa de los datos

Consideremos la variable  $M_i = (\mu_{i,1}, \mu_{i,2})$  para describir la posición (en coordenadas cartesianas) del proceso latente dado por el paso  $i$ , y consideremos la variable  $O_j = (o_{j,1}, o_{j,2})$  para describir la posición  $j$ . Recordemos que habíamos definido  $N_j$  como la cantidad de pasos (o cambios de dirección) que el animal de desde tiempo 1 hasta tiempo  $j(\Delta t)$ .

Tenemos que  $\mu_{0,1} = 0$  y  $\mu_{0,2} = 0$

Para  $i = 1, \dots, N_{steps}$

$$\mu_{i,1} = \mu_{i-1,1} + \cos(\phi_{i-1})t_{i-1}$$

$$\mu_{i,2} = \mu_{i-1,2} + \sin(\phi_{i-1})t_{i-1}$$

Luego es posible parametrizar el proceso de observación como

$o_{0,1} = 0$  y  $o_{0,2} = 0$ , y para para  $j = 1, \dots, N_{obs}$

$$o_{j,1} = \mu_{N_j,1} + \cos(\phi_{N_j}) \left( j\Delta t - \sum_{k < N_j-1} t_k \right)$$

$$o_{j,2} = \mu_{N_j,2} + \sin(\phi_{N_j}) \left( j\Delta t - \sum_{k < N_j-1} t_k \right)$$

Entonces  $o_j$  es una función de todas las posiciones  $M_i$  para  $i = 0$  hasta  $i = N_j$ . Entonces  $O_j = h(M_{0:N_j})$ , donde  $M_{0:D} = (M_0, M_1, M_2, \dots, M_D)$ . Supongamos que sabemos el numero de cambios de dirección que toma el animal entre observaciones consecutivas, entonces sabemos el valor de  $N_j \forall j$ . Luego la función de verosimilitud del modelo espacio estado con diferentes escalas temporales para describir la trayectoria de un individuo esta dado por

$$\begin{aligned} L(\kappa, \lambda, M, O) &= P(O_0 = o_0, O_1 = o_1, \dots, O_{N_{obs}} = o_{N_{obs}}) \\ &= P\left(h(M_{0:N_1}) = o_1, h(M_{0:N_2}) = o_2, \dots, h(M_{0:N_{N_{obs}}}) = o_{N_{obs}}\right) \end{aligned}$$

Como  $O_j = h(M_{0:N_j})$ , de manera de obtener una formulación para  $L$  es necesario obtener las distribuciones de  $M_i$  (1) y de  $O_j = h(M_{0:N_j})$  (2).

## Step 1: Formulation of (1)

Buscamos entonces una formulación para  $M_i = (\mu_{i,1}, \mu_{i,2})$ . Vamos a considerar solo la variable correspondiente a la coordenada  $x$  ( $\mu_{i,1}$ ), ya que la derivación para la coordenada  $y$  es análoga.

Tenemos que

$$\mu_{i,1} = \mu_{i-1,1} + \cos(\phi_{i-1})t_{i-1}$$

con  $\phi_i \sim \text{vonMises}(\phi_{i-1}, \kappa)$  y  $t_i \sim \text{Exp}(\lambda)$ . Para obtener las distribuciones de  $\mu_i | \mu_{i-1}$  es necesario obtener las distribuciones para  $Z = \cos(\phi)t$ . Utilizando el teorema de cambio de variable, es posible calcular esta distribución. Para hacer esto, consideremos primero  $V = g(\phi) = \cos(\phi)$ . Queremos obtener una expresión para  $f_V$ . Dividiendo el dominio de  $g$  y aplicando el método de transformación de variables, obtenemos:

$$f_V = (f_\phi(-\text{acos}(v)) + f_\phi(\text{acos}(v))) \frac{1}{\sqrt{1-v^2}} I_{-1 \leq v \leq 1}(v)$$

Ahora podemos calcular  $f_Z$  como  $f_Z = Vt$ . Nuevamente, haciendo uso del método de transformación de variables, y utilizando el hecho de que los tiempos y los ángulos son independientes, es posible obtener la siguiente expresión

$$f_Z(z_1) = \int f_V\left(\frac{z_1}{z_2}\right) f_t(z_2) I_{\{-z_2 \leq z_1 \leq z_2\}}(z_1) I_{\{z_2 > 0\}}(z_2) \cdot dz_2$$

teniendo  $f_Z$ ,  $p(\mu_i | \mu_{i-1})$  resulta inmediato.

## Step 2: Formulation of (2)

Ahora buscamos una formulación para  $O_j = h(M_{0:N_j})$ . Tenemos

$$O_j = (o_{j,1}, o_{j,2}) = (h_1(\mu_{0:N_j,1}), h_2(\mu_{0:N_j,2}))$$

Nuevamente, vamos a considerar solamente la variable correspondiente a la coordenada  $x$  ( $o_{j,1}$ ), la segunda es análoga.

Tenemos que

$$\begin{aligned} o_{j,1} &= h(\mu_{0:N_j,1}) \\ &= \mu_{N_j,1} + \cos(\phi_{N_j}) \left( j\Delta t - \sum_{k < N_j-1} t_k \right) \\ &= \mu_{N_j,1} + V_{N_j}(c_j - W_{N_j-1}) \end{aligned}$$

Ya conocemos la distribución de  $V_{N_j}$ . La distribución de  $W_{N_j-1}$  es simplemente la suma de  $N_j - 2$   $\text{Exp}(\lambda)$ , y una distribución  $\Gamma(N_j - 2, \lambda)$ . Si consideramos  $\tilde{W}_{N_j-1} =$

$c_j - W_{N_j-1}$  (que difiere de  $W_{N_j-1}$  sólo por una constante), tenemos que  $f_{\tilde{W}_{N_j-1}}(v) = f_{W_{N_j-1}}(c_j - v)$

Entonces, podemos reescribir  $o_{j,1}$  como

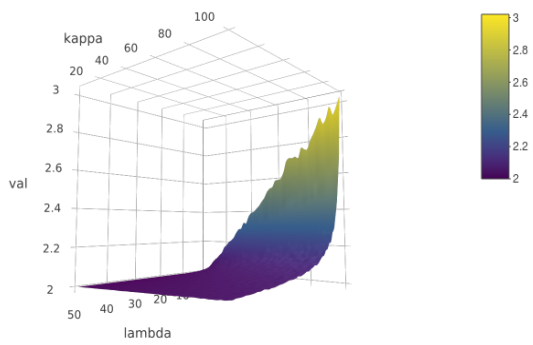
$$o_{j,1} = \mu_{N_j,1} + (V_{N_j})(\tilde{W}_{N_j-1}) = \mu_{N_j,1} + S_{N_j-1}$$

Para obtener las distribuciones de  $S_{N_j-1}$ , nuevamente hacemos uso del método de transformación de variables y la independencia ente tiempos y ángulos.

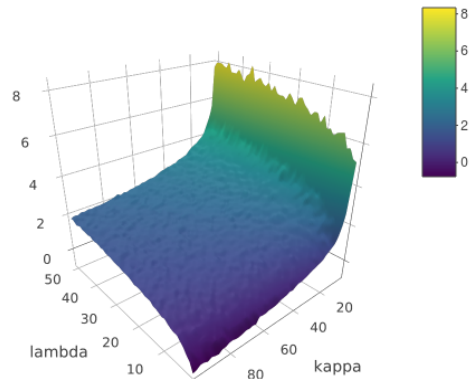
$$f_{S_{N_j-1}}(s_1) = \int f_{V_{N_j}}\left(\frac{s_1}{s_2}\right) f_{\tilde{W}_{N_j-1}}(s_2) I_{\{-s_2 \leq s_1 \leq w_2\}}(s_1) I_{\{s_2 \leq -c_j\}}(s_2) \cdot ds_2$$

## A.2. Estadísticos resumen

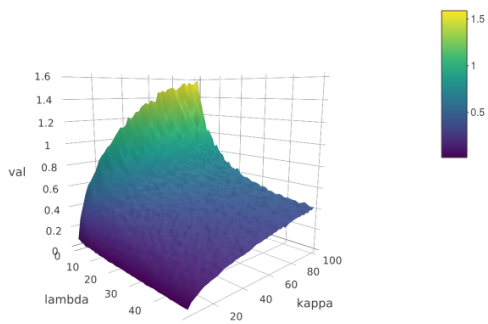
. A continuación exponemos los gráficos de los estadísticos resumen vs los parámetros del modelo (figuras [A.1](#), [A.2](#) y [A.3](#)). Seleccionamos cuatro de ellos que logran describir las trayectorias de forma integral y las caracterizan según los valores de los parámetros. Los estadísticos seleccionados son los de la figura [A.1](#)(a-d).



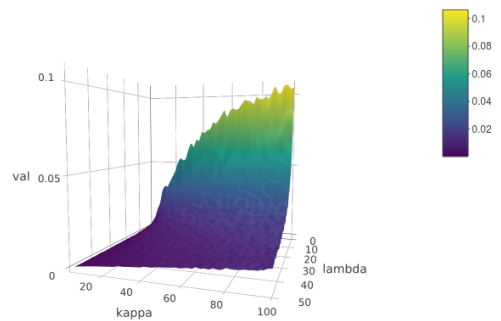
(a) Estimador puntual para  $\lambda$



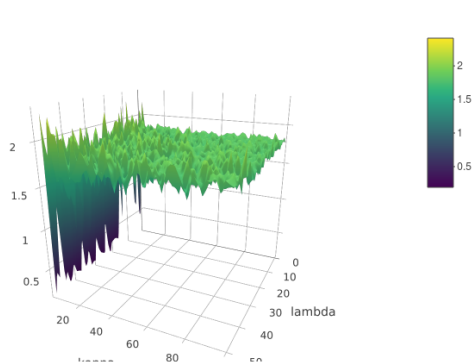
(b) Estimador puntual para K



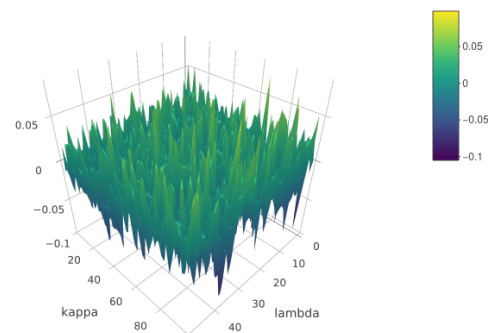
(c) Desvío estándar del ángulo de giro



(d) Desvío estándar del largo de paso

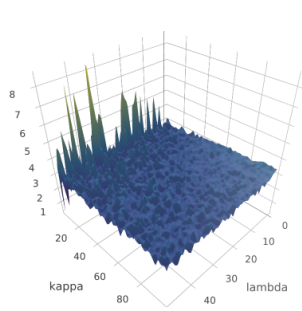


(e) Desvío estándar del las direcciones

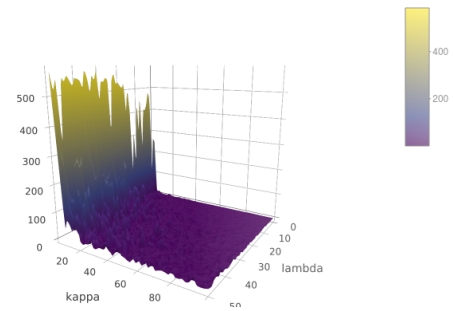


(f) Tercer período de la función de autocorrelación de los ángulos de giro

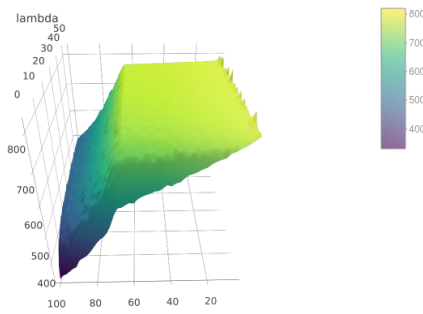
Figura A.1: Gráficos de los estadísticos resumen evaluados para incluir dentro del análisis



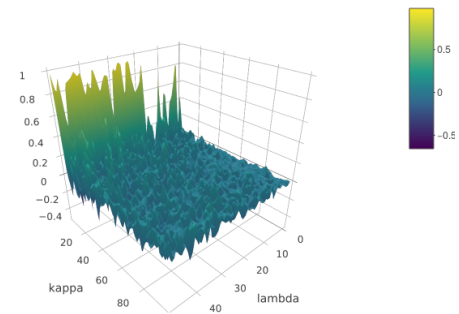
(a) Índice de tortuosidad



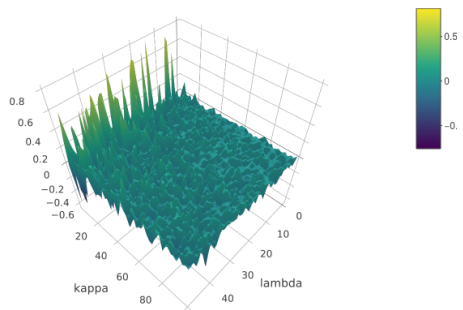
(b) Desplazamiento cuadrático medio



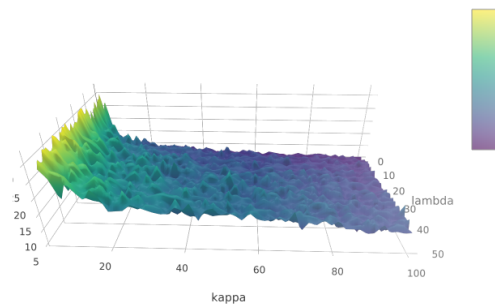
(c) Largo de la trayectoria



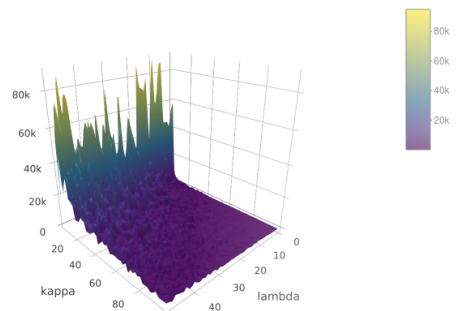
(d) Media del coseno del ángulo de giro



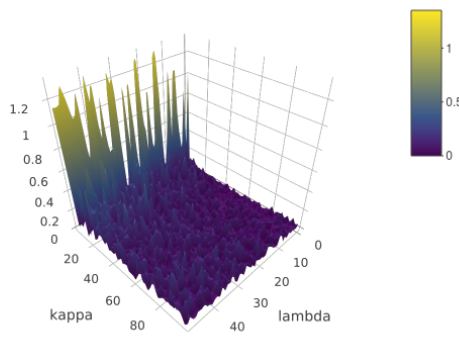
(e) Media del seno del ángulo de giro



(f)  $\sqrt{|y_{.,1max} - y_{.,1min}| + |y_{.,2max} - y_{.,2min}|}$

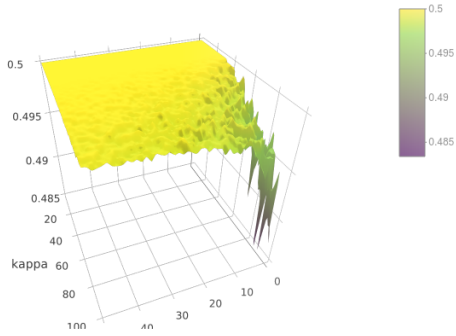


(g) Área determinada por la trayectoria

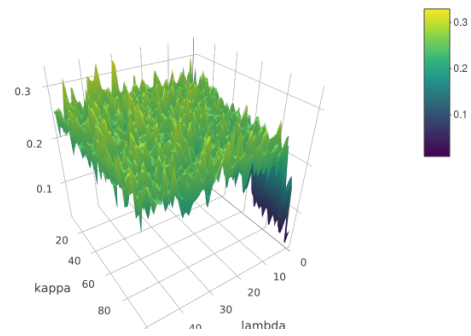


(h) Índice de toruosidad

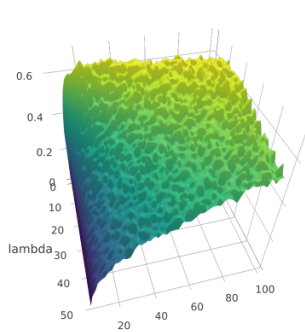
Figura A.2: Gráficos de los estadísticos resumen evaluados para incluir dentro del análisis



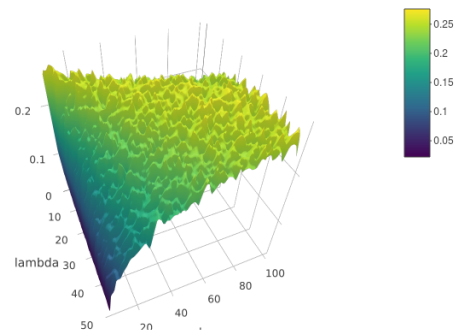
(a) Largo de paso máximo



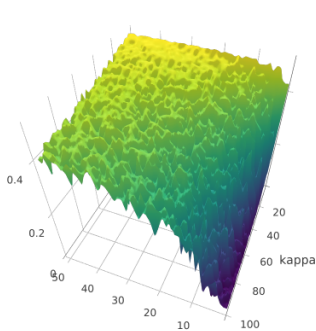
(b) Tercer período de la función de autocorrelación de los ángulos de giro



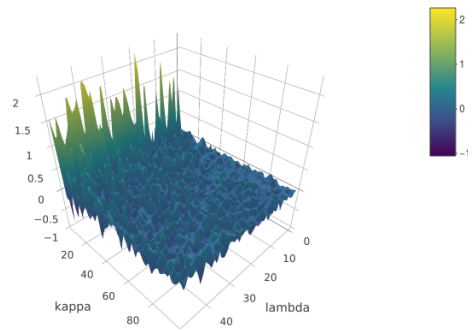
(c) Media del área encerrada cada 5 pasos



(d) Desvío estandar del área encerrada cada 5 pasos



(e) Largo de paso mínimo



(f) Dirección media

Figura A.3: Gráficos de los estadísticos resumen evaluados para incluir dentro del análisis

### A.3. Cobertura empírica

Presentamos los resultados de la cobertura empírica de los intervalos del 95 % para los dos parámetros. Este valor es la proporción de simulaciones en las que el valor verdadero del parámetro cae dentro del intervalo de credibilidad del 95 %. Si

las distribuciones posteriores se correctamente estimadas, esta proporción debería ser cercana a 0,95. Calculamos este índice para ambos parámetros ( $\kappa$  y  $\lambda$ ) y para los tres algoritmos ABC: Rechazo simple, corregido mediante regresión lineal y corregido mediante red neuronal. Lo hicimos para valores de umbral ( $\epsilon$ ) de: 0,001, 0,005, 0,01 y 0,1.

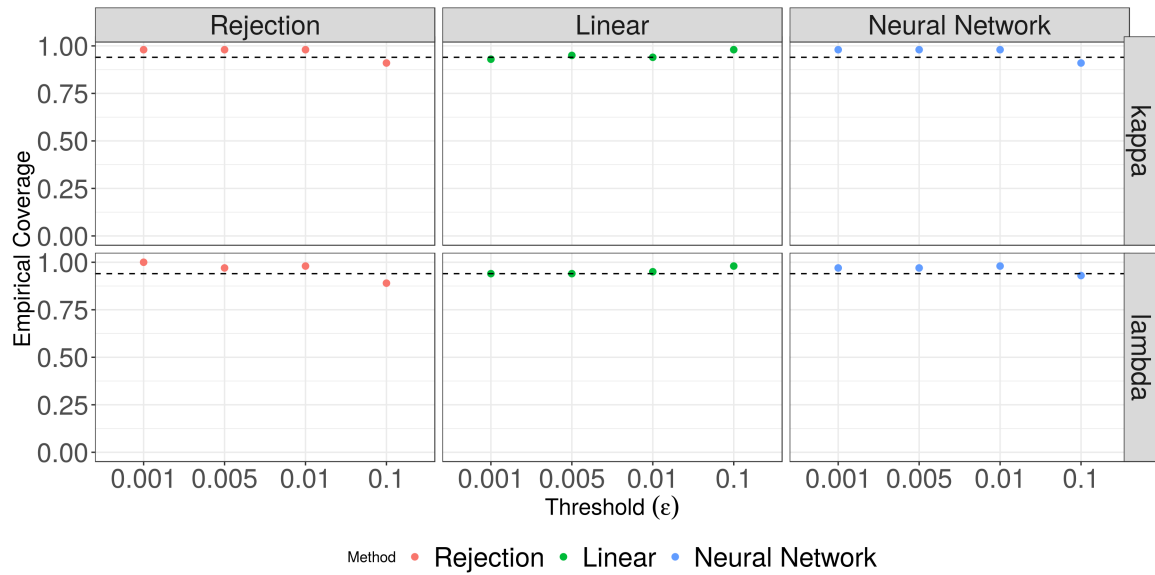


Figura A.4: Cobertura empírica obtenida para  $\kappa$  y  $\lambda$  para los diferentes valores de umbral. Por fila los resultados para cada parámetros y por columna los resultados para los tres algoritmos de ABC. La línea punteada indica el valor 0,95

# B. Anexo II:

---

## B.1. Inferencia Bayesiana

Para realizar la clasificación supervisada, tanto las observaciones como los valores de los estados se conocen en los datos de entrenamiento. Para ajustar los valores del modelo según el enfoque Bayesiano, se calculan las distribuciones posteriores de los parámetros. En la sección 3.1 del manuscrito principal, presentamos la distribución posterior conjunta de los parámetros como el producto de la verosimilitud completa de los datos y las distribuciones previas

$$f(\boldsymbol{\theta}|\mathbf{x}_{1:T}, \mathbf{c}_{1:T}) \propto f(\mathbf{c}_{1:T}, \mathbf{x}_{1:T}, \boldsymbol{\theta})f(\boldsymbol{\theta})$$

Distingamos los diferentes elementos del vector de parámetros. Sea  $\boldsymbol{\theta} = (\boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{\theta}_{obs}, \boldsymbol{\theta}_d)$ , donde  $\boldsymbol{\delta}$  es el parámetros de las distribuciones iniciales,  $\boldsymbol{\gamma}$  el vector de parámetros de la matriz de probabilidad de transición,  $\boldsymbol{\theta}_{obs}$  el vector de parámetros de las distribuciones de observación y  $\boldsymbol{\theta}_d$  el vector de parámetros de los tiempos de permanencia. Si consideramos previas independientes

$$f(\boldsymbol{\theta}) = f(\boldsymbol{\delta})f(\boldsymbol{\gamma})f(\boldsymbol{\theta}_{obs})f(\boldsymbol{\theta}_d)$$

la distribución posterior puede dividirse en componentes separadas

$$f(\boldsymbol{\theta}|\mathbf{x}_{1:T}, \mathbf{c}_{1:T}) = f(\boldsymbol{\delta}|\mathbf{x}_{1:T}, \mathbf{c}_{1:T})f(\boldsymbol{\gamma}|\mathbf{x}_{1:T}, \mathbf{c}_{1:T})f(\boldsymbol{\theta}_{obs}|\mathbf{x}_{1:T}, \mathbf{c}_{1:T})f(\boldsymbol{\theta}_d|\mathbf{x}_{1:T}, \mathbf{c}_{1:T})$$

A continuación detallamos la formulación de cada componente

### Especificaciones para las transiciones y las distribuciones iniciales

Si consideramos  $f(\boldsymbol{\gamma})$  sigue una distribución Dirichlet, es decir  $f(\boldsymbol{\gamma}) \sim \mathcal{D}(\kappa_{j1}, \dots, \kappa_{jJ})$  para  $j = 1 \dots J$ , como los cambios de estado tienen distribución multinomial, la distribución posterior para  $\boldsymbol{\gamma}$  resulta también Dirichlet

$$f(\boldsymbol{\gamma}|\mathbf{x}_{1:T}, \mathbf{c}_{1:T}) \propto \mathcal{D}(\kappa_{j1} + \nu_{j1}, \dots, \kappa_{jJ} + \nu_{jJ})$$

con  $\nu_{ji}$  = número de ocurrencias para del estado  $j$  al  $i$ . De manera similar para  $f(\boldsymbol{\delta}|\mathbf{x}_{1:T}, \mathbf{c}_{1:T})$ , considerando  $f(\boldsymbol{\delta}) \sim \mathcal{D}(\omega_1, \dots, \omega_J)$ , se tiene

$$f(\boldsymbol{\delta}|\mathbf{x}_{1:T}, \mathbf{c}_{1:T}) \propto \mathcal{D}(\omega_1 + \tau_1, \dots, \omega_J + \tau_J)$$

con  $\tau_j$  = número de veces una serie observada comienza en estado  $j$

## Especificaciones para las distribuciones de las observaciones

Dados  $J$  estados, es necesario hacer inferencia sobre los parámetros de  $J$  distribuciones estado dependientes,  $f_1(x) \dots f_J(x)$ . Cada  $j = 1 \dots J$  grupo de parámetros,  $\boldsymbol{\theta}_{obs_j}$ , es estimado utilizando solo las observaciones correspondientes al estado  $j$ . Luego

$$f_j(\boldsymbol{\theta}_{obs} | \mathbf{x}_{1:T}, \mathbf{c}_{1:T}) = f_j(\boldsymbol{\theta}_{obs_j} | \mathbf{x}_{[j]})$$

donde  $\mathbf{x}_{[j]} = \{x_t / C_t = j\}$ .

En este trabajo asumimos que las observaciones estado dependientes se distribuyen normalmente

$$f_j(x) \sim N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$$

Consideramos previas normales independientes para las  $J$  distribuciones de las observaciones.  $f(\boldsymbol{\theta}_{obs_j}) = f(\boldsymbol{\mu}_j) f(\boldsymbol{\Sigma}_j)$

$$f(\boldsymbol{\mu}_j) \sim N(\boldsymbol{\mu}_{0j}, \boldsymbol{\Sigma}_{0j}) \text{ y } f(\boldsymbol{\Sigma}_j) \sim NT(\boldsymbol{\mu}_{0j}, \boldsymbol{\Sigma}_{0j}, 0, \infty)$$

donde  $NT$  indica una distribución normal truncada. En este caso, las distribuciones posteriores no posee forma cerrada. Pueden ser formuladas como

$$f_j(\boldsymbol{\theta}_{obs_j} | \mathbf{x}_{1:T}, \mathbf{c}_{1:T}) \propto f(\boldsymbol{\theta}_{obs_j}) \prod_{t/C_t=j} f_j(x_t)$$

## Especificaciones para las distribuciones de los tiempos de permanencia

Una vez mas, es necesario hacer inferencia sobre los parámetros de las  $J$  distribuciones estado dependientes  $d_1(x) \dots d_J(x)$ . Cada  $j = 1 \dots J$  grupo de parámetros,  $\boldsymbol{\theta}_{d_j}$ , es estimado usando solo los tiempos de permanencia de las observaciones correspondiente al estado  $j$ .

Como consideramos distribuciones binomiales negativas para los  $J$  tiempos de permanencia, las distribuciones posteriores no poseen forma cerrada. Pueden ser formuladas como

$$f_j(\boldsymbol{\theta}_{d_j} | \mathbf{x}_{1:T}, \mathbf{c}_{1:T}) \propto f(\boldsymbol{\theta}_{d_j}) \prod_{\substack{r \text{ is NAT} \\ \text{and } C_r=j}} d_{C_r}(u_r)$$

## B.2. Pruebas

En la siguiente sección se prueban algunos de los resultados utilizados en el manuscrito principal. Con [azul](#) se marcan las consideraciones para el caso autorregresivo.

### B.2.1. De la sección de decodificación local

**Lemma 1.** Para  $t = 2, \dots, T$  y  $j = 1, \dots, J$

$$\alpha_t(j) = \sum_{d \in \mathcal{D}} \sum_{i \neq j} \alpha_{t-d}(i) \gamma_{ij} d_j(d) f_j(\mathbf{x}_{t-d+1:t})$$

*Demostración.*

$$\begin{aligned} \alpha_t(j) &= \Pr(C_t = j, \mathbf{X}_{1:t}) \\ &= \sum_{d \in \mathcal{D}} \Pr(\mathbf{C}_{[t-d+1:t]} = j, \mathbf{X}_{1:t}) \\ &= \sum_{d \in \mathcal{D}} \Pr(\mathbf{X}_{t-d+1:t} | \mathbf{X}_{1:t-d}, \mathbf{C}_{[t-d+1:t]} = j) \Pr(\mathbf{X}_{1:t-d}, \mathbf{C}_{[t-d+1:t]} = j) \\ &= \sum_{d \in \mathcal{D}} f_j(\mathbf{x}_{t-d+1:t}) \Pr(\mathbf{X}_{1:t-d}, \mathbf{C}_{[t-d+1:t]} = j) \end{aligned}$$

y tenemos

$$\begin{aligned} \Pr(\mathbf{X}_{1:t-d}, \mathbf{C}_{[t-d+1:t]} = j) &= \Pr(\mathbf{X}_{1:t-d} | \mathbf{C}_{[t-d+1:t]} = j) d_j(d) \\ &= \sum_{i \neq j} \Pr(\mathbf{X}_{1:t-d}, C_{t-d} = i | \mathbf{C}_{[t-d+1:t]} = j) d_j(d) \\ &= \sum_{i \neq j} \Pr(\mathbf{X}_{1:t-d}, C_{t-d} = i) \gamma_{ij} d_j(d) \\ &= \sum_{i \neq j} \alpha_{t-d}(i) \gamma_{ij} d_j(d) \end{aligned}$$

□

**Lemma 2.** Para  $t = 2, \dots, T$  y  $j = 1, \dots, J$

$$\beta_t(j) = \sum_{d \in \mathcal{D}} \sum_{i \neq j} \beta_{t+d}(i) \gamma_{ji} d_i(d) f_i(\mathbf{x}_{t+1:t+d})$$

*Demostración.*

$$\begin{aligned}
\beta_t(j) &= \Pr(\mathbf{X}_{t+1:T} | C_t = j, \mathbf{X}_{t-p:t}) \\
&= \sum_{d \in \mathcal{D}} \sum_{i \neq j} \Pr(\mathbf{X}_{t+1:T}, \mathbf{C}_{[t+1:t+d]} = i | C_t = j, \mathbf{X}_{t-p:t}) \\
&= \sum_{d \in \mathcal{D}} \sum_{i \neq j} \Pr(\mathbf{X}_{t+1:T} | \mathbf{C}_{[t+1:t+d]} = i, C_t = j, \mathbf{X}_{t-p:t}) \Pr(\mathbf{C}_{[t+1:t+d]} = i | C_t = j, \mathbf{X}_{t-p:t}) \\
&= \sum_{d \in \mathcal{D}} \sum_{i \neq j} \Pr(\mathbf{X}_{t+1:t+d}, \mathbf{X}_{t+d+1:T} | \mathbf{C}_{[t+1:t+d]} = i, C_t = j, \mathbf{X}_{t-p:t}) \gamma_{ji} d_i(d) \\
&= \sum_{d \in \mathcal{D}} \sum_{i \neq j} \Pr(\mathbf{X}_{t+d+1:T} | \mathbf{X}_{t+1:t+d}, \mathbf{C}_{[t+1:t+d]} = i, C_t = j, \mathbf{X}_{t-p:t}) \Pr(\mathbf{X}_{t+1:t+d} | \mathbf{C}_{[t+1:t+d]} = i, C_t = j, \mathbf{X}_{t-p:t}) \\
&= \sum_{d \in \mathcal{D}} \sum_{i \neq q} \Pr(\mathbf{X}_{t+d+1:T} | \mathbf{X}_{t+d+1-p:t+d}, C_{t+d} = i) \gamma_{ji} d_i(d) f_i(\mathbf{x}_{t+1:t+d}) \\
&= \sum_{d \in \mathcal{D}} \sum_{i \neq j} \beta_{t+d}(i) \gamma_{ji} d_i(d) f_i(\mathbf{x}_{t+1:t+d})
\end{aligned}$$

□

**Lemma 3.** Dado  $\beta_q^*(j) = \Pr(\mathbf{X}_{t+1:T} | C_{t+1} = j, \mathbf{X}_{t-p:t})$ , para  $t = 2, \dots, T$  y  $j = 1, \dots, J$

$$\beta_t^*(j) = \sum_{d \in \mathcal{D}} d_j(d) \beta_{t+d}(j) f_j(\mathbf{x}_{t+1:t+d})$$

*Demostración.*

$$\begin{aligned}
\beta_t^*(j) &= \Pr(\mathbf{X}_{t+1:T} | C_{t+1} = j, \mathbf{X}_{t-p:t}) \\
&= \sum_{d \in \mathcal{D}} \Pr(\mathbf{X}_{t+1:T}, \mathbf{C}_{[t+2:t+d]} = j | C_{t+1} = j, \mathbf{X}_{t-p:t}) \\
&= \sum_{d \in \mathcal{D}} \Pr(\mathbf{X}_{t+1:T} | \mathbf{C}_{[t+1:t+d]} = j, \mathbf{X}_{t-p:t}) \Pr(\mathbf{C}_{[t+2:t+d]} = j | C_{t+1} = j, \mathbf{X}_{t-p:t}) \\
&= \sum_{d \in \mathcal{D}} \Pr(\mathbf{X}_{t+1:t+d}, \mathbf{X}_{t+d+1:T} | \mathbf{C}_{[t+1:t+d]} = j, \mathbf{X}_{t-p:t}) d_j(d) \\
&= \sum_{d \in \mathcal{D}} \Pr(\mathbf{X}_{t+d+1:T} | \mathbf{C}_{[t+1:t+d]} = j, \mathbf{X}_{t-p:t+p}) \Pr(\mathbf{X}_{t+1:t+d} | \mathbf{C}_{[t+1:t+d]} = j, \mathbf{X}_{t-p:t}) d_j(d) \\
&= \sum_{d \in \mathcal{D}} \Pr(\mathbf{X}_{t+d+1:T} | C_{t+d} = j, \mathbf{X}_{t+1+d-p:t+d}) d_j(d) f_j(\mathbf{x}_{t+1:t+d}) \\
&= \sum_{d \in \mathcal{D}} \beta_{t+d}(j) d_j(d) f_j(\mathbf{x}_{t+1:t+d})
\end{aligned}$$

□

**Lemma 4.** Dadas las definiciones para  $\xi_t(j)$ ,  $\alpha_t(j)$  y  $\beta_t^*(j)$ , para  $t = 2, \dots, T$  y  $j = 1, \dots, J$

$$\xi_t(j) = \xi_{t+1}(j) + \alpha_t(j) \sum_{i \neq j} \gamma_{ji} \beta_t^*(i) - \beta_t^*(j) \sum_{i \neq j} \alpha_t(i) \gamma_{ij}$$

*Demostración.* Necesitamos las siguientes ecuaciones para la prueba

- $\Pr(C_t = j, \mathbf{X}_{1:T}) = \Pr(C_{t+1} = j, \mathbf{X}_{1:T}) + \Pr(C_{[t]} = j, \mathbf{X}_{1:T}) - \Pr(C_{[t+1]} = j, \mathbf{X}_{1:T})$
- $\Pr(C_{[t]} = i, \mathbf{C}_{[t+1:t+d]} = j, \mathbf{X}_{1:T}) = \alpha_t(i)\gamma_{ij}d_j(d)f_j(\mathbf{x}_{t+1:t+d})\beta_{t+d}(j)$  (Lemma 5)

Luego

$$\begin{aligned}
\xi_t(j) &= \Pr(C_t = j, \mathbf{X}_{1:T}) \\
&= \Pr(C_{t+1} = j, \mathbf{X}_{1:T}) + \Pr(C_{[t]} = j, \mathbf{X}_{1:T}) - \Pr(C_{[t+1]} = j, \mathbf{X}_{1:T}) \\
&= \xi_{t+1}(j) + \Pr(C_{[t]} = j, \mathbf{X}_{1:T}) - \Pr(C_{[t+1]} = j, \mathbf{X}_{1:T}) \\
&= \xi_{t+1}(j) + \sum_{i \neq j} \Pr(C_{[t]} = j, C_{[t+1]} = i, \mathbf{X}_{1:T}) - \sum_{i \neq j} \Pr(C_{[t]} = i, C_{[t+1]} = j, \mathbf{X}_{1:T}) \\
&= \xi_{t+1}(j) + \sum_{i \neq j} \sum_{d \in D} \Pr(C_{[t]} = j, \mathbf{C}_{[t+1:t+d]} = i, \mathbf{X}_{1:T}) - \sum_{i \neq j} \sum_{d \in D} \Pr(C_{[t]} = i, \mathbf{C}_{[t+1:t+d]} = j, \mathbf{X}_{1:T}) \\
&= \xi_{t+1}(j) + \sum_{i \neq j} \sum_{d \in D} \alpha_t(j)\gamma_{ji}d_j(d)f_j(\mathbf{x}_{t+1:t+d})\beta_{t+d}(i) - \sum_{i \neq j} \sum_{d \in D} \alpha_t(i)\gamma_{ij}d_i(d)f_i(\mathbf{x}_{t+1:t+d})\beta_{t+d}(j) \\
&= \xi_{t+1}(j) + \alpha_t(j) \sum_{i \neq j} \gamma_{ji} \sum_{d \in D} d_j(d)f_j(\mathbf{x}_{t+1:t+d})\beta_{t+d}(i) - \sum_{i \neq j} \alpha_t(i)\gamma_{ij} \sum_{d \in D} d_i(d)f_i(\mathbf{x}_{t+1:t+d})\beta_{t+d}(j)
\end{aligned}$$

y finalmente, usando que  $\beta_t(j) = \sum_{i \neq j} \gamma_{ij}\beta_t^*(j)$ , tenemos

$$\xi_t(j) = \xi_{t+1}(j) + \alpha_t(j) \sum_{i \neq j} \gamma_{ji}\beta_t^*(i) - \beta_t^*(j) \sum_{i \neq j} \alpha_t(i)\gamma_{ij}$$

□

El siguiente lema fue necesario para probar la proposición anterior 4

**Lemma 5.**

$$\Pr(C_{[t]} = i, \mathbf{C}_{[t+1:t+d]} = j, \mathbf{X}_{1:T}) = \alpha_t(i)\gamma_{ij}d_j(d)f_j(\mathbf{x}_{t+1:t+d})\beta_{t+d}(j)$$

*Demostración.*

$$\begin{aligned}
\Pr(C_t = i, \mathbf{C}_{[t+1:t+d]} = j, \mathbf{X}_{1:T}) &= \Pr(C_t = i, \mathbf{C}_{[t+1:t+d]} = j, \mathbf{X}_{1:t}, \mathbf{X}_{t+1:T}) \\
&= \Pr(\mathbf{C}_{[t+1:t+d]} = j, \mathbf{X}_{t+1:T} | C_t = i, \mathbf{X}_{1:t}) \Pr(C_t = i, \mathbf{X}_{1:t}) \\
&= \Pr(\mathbf{C}_{[t+1:t+d]} = j, \mathbf{X}_{t+1:T} | C_t = i, \mathbf{X}_{1:t}) \alpha_t(i) \\
&= \Pr(\mathbf{X}_{t+1:T} | C_t = i, \mathbf{X}_{1:t}, \mathbf{C}_{[t+1:t+d]} = j) \\
&\Pr(\mathbf{C}_{[t+1:t+d]} = j | C_t = i, \mathbf{X}_{1:t}) \alpha_t(i) \\
&= \Pr(\mathbf{X}_{t+1:t+d}, \mathbf{X}_{t+d+1:T} | C_t = i, \mathbf{X}_{1:t}, \mathbf{C}_{[t+1:t+d]} = j) \gamma_{ij} d_j(d) \alpha_t(i) \\
&= \Pr(\mathbf{X}_{t+d+1:T} | C_t = i, \mathbf{X}_{1:t}, \mathbf{C}_{[t+1:t+d]} = j, \mathbf{X}_{t+1:t+d}) \\
&\Pr(\mathbf{X}_{t+1:t+d} | C_t = i, \mathbf{X}_{1:t}, \mathbf{C}_{[t+1:t+d]} = j) \gamma_{ij} d_j(d) \alpha_t(i) \\
&= \Pr(\mathbf{X}_{t+d+1:T} | C_{t+d} = j, \mathbf{X}_{t-p:t-1}) \\
&\Pr(\mathbf{X}_{t+1:t+d} | \mathbf{X}_{t+1-p:t+d-1}) \gamma_{ij} d_j(d) \alpha_t(i) \\
&= \beta_{t+d}(j) f_j(\mathbf{x}_{t+1:t+d}) \gamma_{ij} d_j(d) \alpha_t(i)
\end{aligned}$$

□

## B.2.2. De la sección de decodificación global

**Lemma 6.** *Sea*

$$\psi_t(j, d) = \max_{c_{1:t-d}} \Pr(\mathbf{C}_{1:t-d}, \mathbf{C}_{[t-d+1:t]} = j, \mathbf{X}_{1:T})$$

para  $t = 2, \dots, T$ ,  $d = 1, \dots, D$  y  $j = 1, \dots, J$ , luego

$$\psi_t(j, d) = \max_{\substack{i \neq j \\ d' \leq t}} \{\psi_{t-d}(i, d') \gamma_{ij} d_j(d) f_j(\mathbf{x}_{t-d+1:t})\}$$

*Demostración.* Por definición tenemos

$$\psi_{t-d}(i, d') = \max_{c_{1:t-d-d'}} \Pr(\mathbf{C}_{1:t-d-d'}, \mathbf{C}_{[t-d-d'+1:t-d]} = i, \mathbf{X}_{1:t-d})$$

De la misma manera,

$$\begin{aligned}
\psi_t(j, d) &= \max_{c_{1:t-d}} \Pr(\mathbf{C}_{1:t-d}, \mathbf{C}_{[t-d+1:t]} = j, \mathbf{X}_{1:t}) \\
&= \max_{\substack{c_{1:t-d-d'} \\ c_{t-d'-d+1:t-d} \\ d' \leq d}} \Pr(\mathbf{C}_{1:t-d'-d}, \mathbf{C}_{t-d'-d+1:t-d}, \mathbf{C}_{[t-d+1:t]} = j, \mathbf{X}_{1:t-d}, \mathbf{X}_{t-d+1:t}) \\
&= \max_{\substack{c_{1:t-d-d'} \\ i \neq j \\ d' \leq d}} \Pr(\mathbf{C}_{1:t-d'-d}, \mathbf{C}_{[t-d'-d+1:t-d]} = i, \mathbf{C}_{[t-d:t]} = j, \mathbf{X}_{1:t-d}, \mathbf{X}_{t-d+1:t}) \\
&= \max_{\substack{c_{1:t-d-d'} \\ i \neq j \\ d' \leq d}} \Pr(\mathbf{C}_{1:t-d'-d}, \mathbf{C}_{[t-d'-d+1:t-d]} = i, \mathbf{X}_{1:t-d}) \gamma_{ij} d_j(d) \Pr(\mathbf{X}_{t-d+1:t} | \mathbf{X}_{t-d+1-p:t+d}) \\
&= \max_{\substack{c_{1:t-d-d'} \\ i \neq j \\ d' \leq d}} \Pr(\mathbf{C}_{1:t-d'-d}, \mathbf{C}_{[t-d'-d+1:t-d]} = i, \mathbf{X}_{1:t-d}) \gamma_{ij} d_j(d) f_j(\mathbf{x}_{t-d+1:t}) \\
&= \max_{\substack{i \neq j \\ d' \leq d}} \psi_{t-d}(i, d') \gamma_{i,j} d_j(d) f_j(\mathbf{x}_{t-d+1:t})
\end{aligned}$$

□

### B.3. Distribuciones previas

En esta sección mostramos las distribuciones previas utilizadas en el estudio de simulación (tabla B.1) y para el análisis de datos de acelerómetros de ovejas (table B.2)

Parámetro	Distribución previa	Interpretación
$\mu_i$	Normal(0,5)	Media de la distribución Normal que describe la observación dado el estado $i$ .
$\sigma_i$	Normal trunc(0,5)	Desviación estándar para la distribución Normal que describe la observación dado el estado $i$ .
$m_i$	Normal trunc(20,50)	Media de las distribuciones Binomiales negativas que describe el tiempo de permanencia de cada estado.
$k_i$	Normal trunc (20,50)	Parámetro de dispersión para las distribuciones Binomiales negativas que describe el tiempo de permanencia de cada estado.

Tabla B.1

*Distribuciones previas utilizadas para realizar el estudio de simulación*

Parámetro	Distribución previa	Interpretación
$\alpha_1$	MultiNormal( $(0, 20, 0)^T, diag(3, 5, 10)$ )	Vector medias de la distribución Normal que describe la observación del comportamiento de caminata
$\alpha_2$	MultiNormal( $(0, 20, 0)^T, diag(3, 5, 10)$ )	Vector medias de la distribución Normal que describe la observación del comportamiento de caminata rápida
$\alpha_3$	MultiNormal( $(0, 20, 0)^T, diag(3, 5, 10)$ )	Vector medias de la distribución Normal que describe la observación del comportamiento inactivo
$\alpha_4$	MultiNormal( $(0, -20, 0)^T, diag(3, 5, 10)$ )	Vector medias de la distribución Normal que describe la observación del comportamiento de alimentación
$\sigma_i$	MultiNormalT( $(0, 0, 0)^T, diag(3, 10, 3)$ )	Vector de desviación estándar de la distribución Normal que describe la observación para cada comportamiento $i = 1, 2, 3, 4$
$m_1$	Normal trunc(0,5)	Media de las distribuciones Binomiales negativas que describen el tiempo de permanencia del comportamiento de caminata
$m_2$	Normal trunc(0,5)	Media de las distribuciones Binomiales negativas que describen el tiempo de permanencia del comportamiento de caminata rápida
$m_3$	Normal trunc(80,30)	Media de las distribuciones Binomiales negativas que describen el tiempo de permanencia del comportamiento inactivo
$m_4$	Normal trunc(25,20)	Media de las distribuciones Binomiales negativas que describen el tiempo de permanencia del comportamiento de alimentación
$k_i$	Normal trunc(0,5)	Parámetro de dispersión para la distribuciones Binomiales negativas que describen los tiempos de permanencia de cada comportamiento $i = 1, 2, 3, 4$
$\beta_i$	Normal(0,5)	Pendiente del término autorregresivo para cada comportamiento $i = 1, 2, 3, 4$

Tabla B.2

*Distribuciones previas utilizadas para realizar el análisis de los datos de aceleración de ovejas*

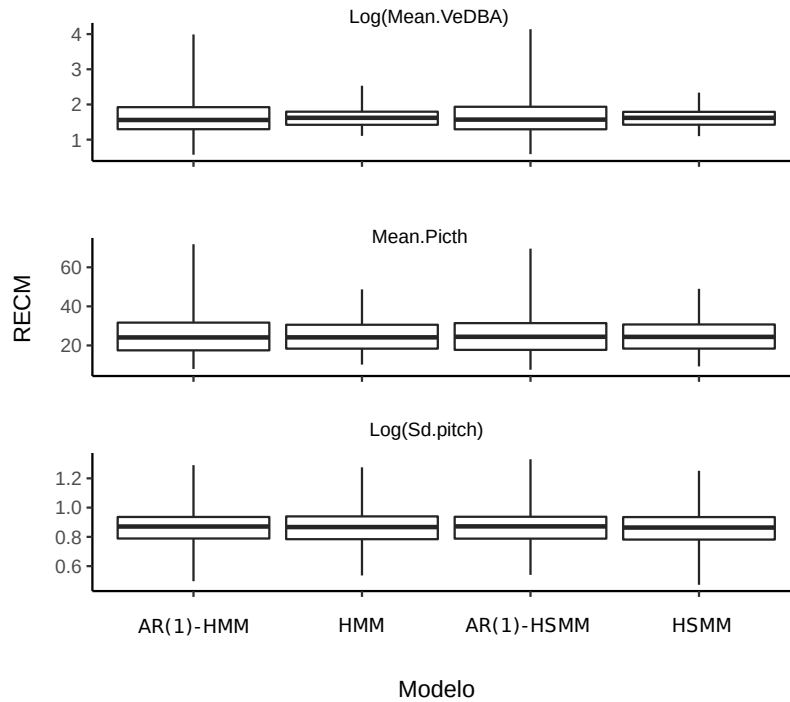


Figura B.1: Boxplots del error medio cuadrático sobre las observaciones para los cuatro modelos considerados.

## B.4. RECM figura extra

Para medir la capacidad de predicción de los modelos propuestos para analizar los datos de acelerómetros, se calculó la raíz del error cuadrático medio (RECM) de los cuatro modelos: AR(1)-HMM, HMM, AR(1)-HSMM y HSMM. Para cada serie temporal observada y modelo, utilizando una muestra de 100 de la posterior ajustada, calculamos 100 predicciones (utilizando el algoritmo FB) de los estados ocultos y el proceso de observación. La figura B.1 muestra los valores obtenidos

## C. Anexo III:

---

### C.1. Aproximación por regresión logística condicional para el modelo iSSA

Dado la función condicional del log-verosimilitud de la ecuación 4.1, vamos a derivar su aproximación por la verosimilitud de una regresión logística condicional. Llamamos  $f$  a la función de densidad de que el animal se mueva desde la ubicación  $\mathbf{s}^*$  a la ubicación  $\mathbf{s}$  condicional al paisaje como

$$f(\mathbf{s}|\mathbf{s}^*, \theta_\omega) = \frac{\phi(\mathbf{s}, \mathbf{s}^*)\omega(\mathbf{s}, \theta_\omega)}{\int_D \phi(\mathbf{u}, \mathbf{s}^*)\omega(\mathbf{u}, \theta_\omega) d\mathbf{u}}$$

Sean  $\mathbf{s}_1, \dots, \mathbf{s}_n$  las ubicaciones observadas en los tiempos  $t_1, \dots, t_n$ . Supongamos que para cada ubicación observada  $\mathbf{s}_i$ , muestreamos de manera aleatoria  $K$  casos control utilizando alguna función de selección de paso  $\phi^*$  independiente del paisaje (en principio  $\phi^*$  podría ser cualquiera). Para cada  $i = 1, \dots, n$ , se tiene el conjunto  $m_i = \{l_{i0}, l_{i1}, \dots, l_{iK}\}$ , en donde  $l_{i0} = \mathbf{s}_i$  y el resto de los elementos son las nuevas ubicaciones muestreadas para el paso  $i$ . Para cada conjunto  $m_i$ , la densidad condicional al paisaje resulta ser

$$g(m_i|\mathbf{s}_i, \mathbf{s}_{i-1}) = \frac{1}{\phi^*(\mathbf{s}_{i-1}, \mathbf{s}_i)} (K-1)! \prod_{l \in m_i} \phi^*(\mathbf{s}_{i-1}, l)$$

Luego, considerando  $\omega(\mathbf{s}, \theta_\omega) = \exp(\boldsymbol{\alpha}^T \mathbf{X}_\mathbf{s})$  la probabilidad condicional del observar  $\mathbf{s}_i$  dado  $\mathbf{s}_{i-1}$ , el conjunto  $m_i$  y el paisaje es

$$\begin{aligned} P(\text{caso} = \mathbf{s}_i | m_i, \mathbf{s}_{i-1}) &= \frac{f(\mathbf{s}_i | \mathbf{s}_{i-1}) g(m_i | \mathbf{s}_i, \mathbf{s}_{i-1})}{\sum_{l \in m_i} f(l | \mathbf{s}_{i-1}) g(m_i | l, \mathbf{s}_{i-1})} \\ &= \frac{f(\mathbf{s}_i | \mathbf{s}_{i-1}) / \phi^*(\mathbf{s}_i, \mathbf{s}_{i-1})}{\sum_{l \in m_i} f(l | \mathbf{s}_{i-1}) / \phi^*(\mathbf{s}_{i-1}, l)} \\ &= \frac{\phi(\mathbf{s}_{i-1}, \mathbf{s}_i) \exp(\boldsymbol{\alpha}^T \mathbf{X}_{\mathbf{s}_i}) / \phi^*(\mathbf{s}_{i-1}, \mathbf{s}_i)}{\sum_{l \in m_i} \phi(\mathbf{s}_{i-1}, l) \exp(\boldsymbol{\alpha}^T \mathbf{X}_l) / \phi^*(\mathbf{s}_{i-1}, l)} \end{aligned}$$

En lo anterior, la segunda igualdad se debe a que el producto en la expresión de  $g()$  es el mismo en el numerador y en todos los términos del denominador. La tercera igualdad se debe que el denominador de la expresión de  $f()$  es el mismo para el numerador y todos los términos del denominador. Simplificando lo anterior se tiene:

$$P(\text{caso} = \mathbf{s}_i | m_i, \mathbf{s}_{i-1}) = \frac{\exp\{\boldsymbol{\alpha}^T \mathbf{X}_{\mathbf{s}_i} + \log(\phi(\mathbf{s}_{i-1}, l_{i0}) / \phi^*(\mathbf{s}_{i-1}, l_{i0}))\}}{\sum_{l \in m_i} \exp\{\boldsymbol{\alpha}^T \mathbf{X}_l + \log(\phi(\mathbf{s}_{i-1}, l) / \phi^*(\mathbf{s}_{i-1}, l))\}}$$

Asumiendo que  $\phi^*$  es conocido, para una trayectoria de  $n$  pasos, los parámetros del modelo inicial pueden ser estimados a partir de la función de verosimilitud condicional

$$L(\theta) = \prod_{i=1}^n \frac{\exp\{\boldsymbol{\alpha}^T \mathbf{X}_{\mathbf{s}_i} + \log(\phi(\mathbf{s}_{i-1}, l_{i0})/\phi^*(\mathbf{s}_{i-1}, l_{i0}))\}}{\sum_{l \in m_i} \exp\{\boldsymbol{\alpha}^T \mathbf{X}_l + \log(\phi(\mathbf{s}_{i-1}, l)/\phi^*(\mathbf{s}_{i-1}, l))\}} \quad (\text{C.1})$$

Llegamos así a una expresión equivalente a la función de verosimilitud de un modelo logístico condicional, en donde los casos control son  $l_{i1}, \dots, l_{iK}$  y el caso verdadero es  $l_{i0}$ . Luego dados los tiempos y ubicaciones observadas  $t_1, \dots, t_n$ , y  $s_1, \dots, s_n$ , la función de verosimilitud [4.1](#), puede aproximarse por [C.1](#)