



ESCUELA DE DOCTORADO
INTERNACIONAL DE LA USC

Sara
Martínez Santalla

Tesis doctoral

ANÁLISIS DE LA DIVERSIDAD
BETA PARA ENTENDER
PATRONES
MACROECOLÓGICOS

Santiago de Compostela, 2024

TESIS DOCTORAL

**ANÁLISIS DE LA
DIVERSIDAD BETA
PARA ENTENDER PATRONES
MACROECOLÓGICOS**

Autora

Sara Martínez Santalla

Directores: Andrés Baselga y Carola Gómez-Rodríguez

Tutor: Andrés Baselga

**PROGRAMA DE DOCTORADO EN BIODIVERSIDAD Y CONSERVACIÓN DEL MEDIO NATURAL
SANTIAGO DE COMPOSTELA**



AGRADECIMIENTOS

Esta tesis que ahora lees es una gotita en los océanos de la ciencia y, a la vez, mares enteros en los míos. Me ha cambiado fundamentalmente, como supongo que debía hacer. Sea como fuere, en la ciencia y en la vida todos mis pasos han ido a hombros de gigantes, y a ellos les ofrezco el fruto de ese cambio:

En primer lugar, agradecerle al Ministerio de Ciencia e Innovación la financiación (BES-2017-081643) y la oportunidad, la cual debo agradecer también y enormemente a mis directores, Andrés y Carola. De manera personal hacia vosotros, valoro cada lección, cada momento de paciencia, cada palabra de ánimo, cada río de ayuda. Todo lo que se ve, y lo que no. Gracias.

Gracias también al profesor José Castillejo por todo. Y a la profesora Rosa Crujeiras, que me llevó de la mano por los pasos más difíciles de la estadística.

En nuestro laboratorio, agradezco la suerte que tuve con Andrea y Adrián, y la herencia de sus conocimientos. A Erika, a Alba, a Aitor, a Jose. Por ser ejemplos, amigos y compañeros en aguas extrañas. A Victoria, que es pura luz y llama, que me ha inspirado y hecho crecer a diario, gracias. Por favor, lleva con orgullo todos tus caminos andados. A Ramiro. No sé cómo habría hecho todo esto sin ti. Gracias por darme alas, fe, compañía y propósito, incluso en días oscuros, sin pizca de juicio. Gracias por disfrutar conmigo de discutir durante horas frente a una pizarra sobre simulaciones. Gracias por vivir conmigo la ciencia con la pasión que merece. Espero seguir leyendo tus trabajos extraordinarios muchos años más, compañero. A Mari, gracias por ser una puerta de vuelta a la realidad cada tarde. Me pasaría la vida escuchándote.

A mi Casa en la mayoría de esta etapa, a Vittoria y a Raquel, gracias por hacer un hogar resistente a penas y pandemias y lleno de celebración y alegría por las cosas mundanas. A mí también me gustan estas “borrachas”. Gracias por hacerme crecer, gracias por ser mi familia. A Raquel, en particular, gracias por ser la hermana mayor en todos mis dilemas. Gracias por ser el referente científico que eres, por compartir conmigo tu curiosidad insaciable.

A Ale, ser de luz, genio gentil, gracias por todo lo que he aprendido de ti. Como compañera de clase, como científica, como amiga. Si algún día soy la mitad de buena que tú explicando, será por haberte visto a ti.

A Su y a Jandro, gracias por ser hermanos sin necesidad de sangre por el medio, por el apoyo, los paseos, la confianza inquebrantable. Gracias por traer a Gaia, y por iluminar de nuevo todo lo que se apaga. Por empujarme a ser mejor, y siempre con bondad.

A Azu, trasniño, también hermana en la vida y compañera de las más salvajes aventuras. Gracias por compartir tu vida y tu hogar conmigo.

A Yai y Leandro, gracias por el apoyo de tantos años, por ser un rincón seguro siempre, por ser un sitio siempre fascinante. A Yai, gracias por toda tu sensatez, y gracias por mostrarme cómo puede convivir con la más pura fantasía. Alegras cada habitación en la que entras. A Leandro, gracias por no dejar que me rinda nunca si la lucha es noble, gracias por darme siempre refugio en tus mundos extraordinarios.

A María, gracias amiga por la telepatía, la honestidad amable, la ausencia de máscaras y las tardes sentadas sobre la hierba. Eres extraordinaria, y yo muy afortunada.

A I, gracias por todo. Y, particularmente, gracias por explicarme hace tantos años que cuando no se puede hay que dar un pasito de cada vez hasta que se pueda. Lo recuerdo todos los días.

A nuestro maravilloso Sarastre, gracias por trascender los años y ser un punto de encuentro de tantas vidas distintas.

A mis padres, Gerardo y Mónica, a mi abuela Pilar y a Lía. Gracias por quererme con uñas y dientes incluso contra mi propio juicio, por darme ejemplo y unas raíces inquebrantables. También me habéis dado un futuro libre, hecho de mis propias elecciones, y no habría sido capaz sin vuestro apoyo.

A Sari, mi biutifulest, mi hermana del alma. La fuerte, la dulce, la lista, la extraordinaria, el goblin. Y a Kini, mi pareja que de hecho es pareja, el del corazón enorme y la mente brillante. Sois mi referencia, mi humanidad, mi anclaje frente a huracanes. El faro más luminoso de la costa. Gracias por acompañarme en esta aventura, en esta vida, con todo lo que ha implicado.

Por último, gracias Trocolita. Eres un gato, así que no entiendes de qué va esto, pero para tener sólo un año y pesar menos de cinco quilos me has traído más alegrías de las que puedo contar.

A todos vosotros, gracias de nuevo y que vuestra curiosidad nunca se sacie.

ÍNDICE

1. RESUMEN.....	1
2. INTRODUCCIÓN.....	3
3. OBJETIVOS.....	15
4. METODOLOGÍA.....	17
4.1. Datos de distribución de coleópteros en Europa.....	17
4.1.1. Cálculo de la similitud entre faunas y distancias entre territorios europeos.....	19
4.2. Estudio no paramétrico de las diferencias norte-sur en los patrones de <i>distance-decay</i> de coleópteros europeos.....	21
4.3. Estructura del error en modelos paramétricos (GLM) de <i>distance-decay</i>	23
4.4. Métodos <i>bootstrap</i> : exploración y aproximación inicial.....	25
4.4.1. Modelo teórico inicial.....	25
4.4.2. Métodos <i>bootstrap</i> por bloques propuestos por autores previos.....	27
4.4.3. Estudio de simulación.....	29
4.4.4. Contraste de significación del modelo basado en la <i>deviance</i>	30
4.4.5 Selección de modelos mediante AIC.....	31
4.5. Implementación de un método para contraste de hipótesis en modelos no lineales de <i>distance-decay</i> basado en permutaciones y remuestreo por bloques de localidad.....	32
4.5.1. Una nueva familia de simulaciones con dependencia por pares.....	32
4.5.2. Generación de matrices de similitud/distancia y ajuste de modelos <i>distance-decay</i>	34
4.5.3. El contraste de la pseudo-R ² : R_{KLD}^2	35
4.5.4. Evaluación del funcionamiento del contraste de significación.....	37
4.5.5. Ejemplo de aplicación a datos reales.....	39
5. RESULTADOS.....	41
5.1. Estudio no paramétrico de las diferencias norte-sur en los patrones de <i>distance-decay</i> de coleópteros europeos.....	41
5.2. Estructura del error en modelos paramétricos (GLM) de <i>distance-decay</i>	45
5.3. Métodos <i>bootstrap</i> : exploración y aproximación inicial.....	57
5.3.1. Estimación de los parámetros del modelo y su desviación típica mediante métodos de <i>bootstrap</i> propuestos por autores previos.....	57
5.3.2. Contraste de significación del modelo basado en la <i>deviance</i>	64
5.3.3. Selección de modelos mediante AIC.....	65

5.4. Implementación de un método para contraste de hipótesis en modelos no lineales de <i>distance-decay</i> basado en permutaciones y remuestreo por bloques de localidad.....	67
5.4.1. Simulaciones para la estimación de los errores de tipo I y II... ..	67
5.4.2. Aplicación a datos empíricos: coleópteros europeos.....	71
5.4.3. Implementación del contraste estadístico en la librería “betapart” de R.....	84
6. DISCUSIÓN.....	87
7. CONCLUSIONES.....	93
8. BIBLIOGRAFÍA.....	95
9. APÉNDICES.....	105
9.1. Código de R para comparación no paramétrica de los patrones de <i>distance-decay</i> de coleópteros en el norte y el sur de Europa.....	105
9.2. Código de R para el estudio de la estructura del error.....	111
9.3. Código de R para el estudio de aplicación de los métodos de <i>bootstrap</i> disponibles en la literatura.....	124
9.4. Código de R para la implementación de permutaciones y remuestreo por bloques de localidad.....	136
9.5. Aplicación de los métodos <i>distance-decay</i> desarrollados en estas tesis a datos reales...	149
9.6. Relación de publicaciones y autorización de la revista para inclusión en esta tesis.....	179

1 RESUMEN

El descenso de la similitud biótica con la distancia espacial, también conocido por su término en inglés, *distance-decay*, es uno de los patrones universales en Ecología y Biogeografía. Este patrón surge porque, en numerosos taxones y territorios, se ha observado que cuanto más alejadas están dos comunidades biológicas, menor es el número de especies que tienen en común y, por tanto, menos similares son dichas comunidades. El descenso en la similitud que existe entre comunidades se mide con índices de similitud biótica que se calculan para cada par de comunidades, resultando así en una semimatriz (matriz de datos triangular), en la que cada comunidad biológica inicial está involucrada en el cálculo de múltiples valores de similitud, uno por cada comunidad biológica con la que es comparada. Este hecho, que se da tanto en la matriz de similitud biótica como en la matriz de distancia espacial entre las comunidades, genera pseudorreplicación en los datos y con ello la violación de una de las principales asunciones de los modelos estadísticos clásicos: la independencia de las observaciones. Existen métodos diseñados para ajustar modelos estadísticos para datos de estas características cuando la relación entre variables es lineal (por ejemplo, el test de Mantel). Sin embargo, la forma de la relación de los modelos *distance-decay* no se ajusta a un modelo lineal, predominando los patrones de tipo exponencial negativo o *power-law*. Por ello, la ausencia de un procedimiento estandarizado para ajustar los modelos *distance-decay* no lineales y evaluar su significación constituye un vacío importante en las herramientas disponibles para la Biogeografía y, en concreto, para el análisis de la diversidad beta. En esta tesis se ha estudiado el problema que genera la dependencia de los datos para el ajuste de modelos no lineales de *distance-decay* con el objetivo final de diseñar y estudiar el funcionamiento de un estadístico basado en la pseudo- R^2 (denominada así en esta tesis, en lugar de R^2 , por trabajar con datos heterocedásticos y con dependencia que no cumplen las asunciones de los modelos utilizados) de Modelos Lineales Generalizados (*Generalized Linear Models*, GLM) y en combinación con un contraste de significación adecuado, construido a partir de un modelo nulo en el cual similitud y distancia son independientes entre ellas, y por lo tanto no hay una relación de *distance-decay*, pero en el que se preserva la dependencia entre pares inherente a este tipo de datos. Para abordar este

problema, en esta tesis se ha desarrollado un contraste de significación para modelos exponenciales y *power-law* basado en permutaciones e, independientemente, un contraste de significación basado en el remuestreo por bloques de localidad, al que se ha denominado “remuestreo por bloques de localidad”. El “remuestreo por bloques de localidad” incorpora la dependencia de los datos al realizar el remuestreo a partir de conjuntos de datos que incluyen todas aquellas similitudes/distancias en las que interviene una localidad determinada. Este método es el que se propone como el más apropiado para el ajuste de modelos no lineales de *distance-decay* y el cálculo de su significación, y se muestra su aplicación a datos reales de presencia/ausencia de diferentes taxones de coleópteros en el continente europeo.

2 INTRODUCCIÓN

La ecología de comunidades y la biogeografía abordan el estudio de la distribución espacial de las especies y las comunidades biológicas, así como de los procesos subyacentes que gobiernan esas distribuciones (McGill, 2010). Una de las observaciones básicas sobre la diversidad biológica es que no se distribuye de manera uniforme en el planeta: las especies de seres vivos varían entre lugares y esta variación es diferente según la escala espacial que estemos considerando para nuestro estudio (Ricklefs, 2004; Willig et al., 2003). Las agrupaciones de seres vivos en un lugar específico constituyen lo que denominamos comunidades biológicas o ensamblajes de especies (*species assemblages*), y su composición puede variar enormemente entre localidades, desde ser prácticamente idénticas hasta ser completamente diferentes. Por tanto, en los estudios de diversidad biológica nos enfrentamos a dos preguntas básicas: “¿cómo medimos las diferencias entre comunidades biológicas o, en otras palabras, cómo caracterizamos los patrones de diversidad?” y, por otro lado, “¿cuáles son los procesos ecológicos, evolutivos o biogeográficos responsables de que existan dichos patrones de diversidad?”. No es posible abordar la pregunta sobre las causas sin haber establecido previamente los métodos para medir adecuadamente en qué medida las comunidades biológicas son diferentes.

Existen varias metodologías para medir la diversidad biológica, siendo la más simple e intuitiva el conteo de especies presentes en un área determinada. Esta sencilla medida nos proporciona una cuantificación inicial de la diversidad biológica, conocida como riqueza de especies. Una característica clave de la riqueza de especies es que depende de la escala espacial a la que se mide (Rahbek, 2005; Whittaker & Triantis, 2012; Whittaker et al., 2001) o, en otros términos, tiene dependencia espacial con la escala (*scale-dependence*). Es decir, la riqueza de especies tiende a aumentar a medida que ampliamos la escala geográfica de estudio, dando lugar a una de las leyes fundamentales en Ecología: la relación especies-área (*species-area relationship* (Rosenzweig, 1995)). Entonces, si nos centramos en un territorio concreto, al que denominaremos “región”, generalmente observaremos que la riqueza de especies que podemos encontrar en esa región es mayor que la riqueza de especies de una localidad cualquiera de dicha región. Esto nos lleva a una observación crucial: si la composición de especies de las comunidades de diferentes localidades de una región es idéntica en todas ellas, la diversidad

regional será la misma que la diversidad de cada localidad individual (diversidad local). Sin embargo, si la composición de especies es diferente entre localidades, entonces la diversidad regional será mayor que la diversidad local. En este marco teórico, denominamos diversidad gamma (γ) a la riqueza de especies en la región, diversidad alfa (α) a la riqueza promedio en las localidades y diversidad beta (β) a la relación que existe entre ambas (Whittaker, 1960). Por tanto, el hecho de que áreas geográficas más grandes alberguen una mayor diversidad de especies está intrínsecamente ligado al hecho de que diferentes localidades suelen tener diferentes especies.

La diversidad beta contempla, por tanto, dos aspectos fundamentales de la caracterización de los patrones de diversidad: (i) mide la variación en la composición de especies entre comunidades biológicas y (ii) expresa la relación entre la diversidad regional y la diversidad local (Figura 1). Aunque están íntimamente relacionadas desde el punto de vista conceptual, estas dos facetas suelen ser abordadas mediante aproximaciones metodológicas diferentes. Por un lado, la variación en la composición entre comunidades biológicas se puede medir con índices de disimilitud biótica, que cuantifican cómo de diferentes son dos comunidades teniendo en cuenta que la ausencia de una especie en ambas comunidades no es informativa y, por tanto, no debe ser tenida en cuenta en el cálculo, lo que también se denomina como “el problema de los dobles ceros” (Legendre & Legendre, 1998). Estos índices se calculan para pares de comunidades (*pairwise beta diversity* o *pairwise biotic dissimilarity*) a partir de datos de incidencia de especies (por ejemplo los índices de Jaccard (1912) o Sørensen (1948)) o datos de abundancia de especies (ej. índice de Bray-Curtis (Bray & Curtis, 1957)). Por otro lado, también es posible calcular un valor de diversidad beta global para múltiples localidades teniendo en cuenta la ratio entre la diversidad gamma y la diversidad alfa ($\beta = \gamma/\alpha$) en la región de estudio, tal y como propuso inicialmente Whittaker (1960). Cuanto mayor sea esta ratio, mayor será el número de especies que aparecen a escala regional en relación al número promedio de especies que podemos observar a escala local.

A pesar de ser un concepto muy intuitivo, la cuantificación de la diversidad beta ha suscitado un interesante debate a lo largo de los años. El desarrollo de los índices de disimilitud, también llamados coeficientes de asociación, fue anterior e independiente a la formulación de los conceptos de diversidad alfa, beta y gamma (por ejemplo, Jaccard, 1912; Simpson, 1943; Sørensen, 1948). El término diversidad beta fue introducido por primera vez por Whittaker (1960) y definido por él como “la magnitud del cambio en la composición de la comunidad” entre sitios. Para medir la magnitud de ese cambio de composición entre comunidades, la diversidad beta compara la diversidad en dos escalas diferentes (diversidad alfa y gamma), como se ha explicado anteriormente. Es necesario aclarar que esta comparación entre escalas se podría realizar usando la formulación multiplicativa clásica ($\beta = \gamma / \alpha$) o usando la partición aditiva de la diversidad ($\beta = \gamma - \alpha$), tal y como fue propuesta por otros autores (Lande, 1996; Veech & Crist, 2010). Sin embargo, estudios posteriores, como el de Jost (2007), demostraron que el tipo de partición elegida (ej. aditiva o multiplicativa) no es trivial si pretendemos que los componentes de diversidad alfa y diversidad beta que con él se calculan sean independientes.

En el caso de medidas de diversidad beta basadas en recuentos de especies, es decir, medidas basadas en la riqueza, el uso de la partición multiplicativa clásica es obligatorio (Baselga, 2010a) para garantizar la necesaria independencia entre la diversidad alfa y la beta (Wilson & Shmida, 1984).

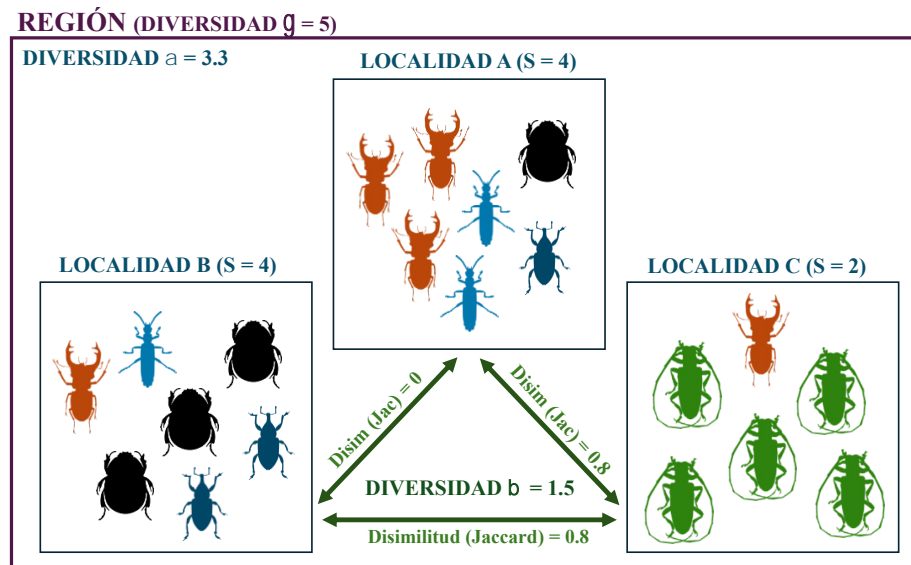


Figura 1. Ejemplo ilustrativo de la relación entre el cálculo de la diversidad beta global como la relación multiplicativa entre la diversidad gamma y la diversidad alfa ($\beta = \gamma/\alpha$) y la diversidad beta calculada como la disimilitud biótica entre pares de comunidades. En este ejemplo, la disimilitud biótica ha sido calculada con el índice de Jaccard, que tiene en cuenta el número de especies exclusivas de cada comunidad dividido por el número de especies totales. Nótese que si se usara un índice de disimilitud basado en abundancia (ej. Bray-Curtis), existiría un valor no nulo de disimilitud biótica entre la localidad A y la localidad B (Bray = 0.43) debido a las diferencias en abundancia de las especies, a pesar de que las especies en ambas localidades sean las mismas. (Crédito siluetas: phylopic.org, licencia CC0, CC BY-SA y CC BY 4.0)

Por otro lado, si consideramos la diversidad beta como una medida de la disimilitud entre sitios (Koleff et al., 2003), es importante conocer la relación matemática entre el cálculo de la diversidad beta basado en la partición multiplicativa y el cálculo basado en los índices de disimilitud biótica entre pares de comunidades biológicas. Estos puntos de vista han sido también denominados diversidad proporcional y diversidad de diferenciación, respectivamente (Jurasinski et al., 2009), pero esta distinción es más formal que conceptual porque los índices de disimilitud tradicionales son en realidad transformaciones monotónicas de la beta multiplicativa de Whittaker (es decir, $\beta = \gamma/\alpha$) calculada para dos sitios (Jost, 2007). En otras palabras, las medidas de diversidad de diferenciación son, de hecho, sólo casos especiales de medidas de diversidad proporcional para $N = 2$. El problema queda claramente ejemplificado por el hecho de que los índices de disimilitud por pares pueden generalizarse y aplicarse a situaciones que involucran múltiples sitios, resultando en medidas de disimilitud de múltiples sitios (*multiple-site dissimilarity*) que, nuevamente, son transformaciones monotónicas de la diversidad beta de Whittaker ($\beta = \gamma/\alpha$) para cualquier número de sitios (Diserud & Ødegaard, 2007). Por lo tanto, estos puntos de vista aparentemente divergentes sobre la diversidad beta son en realidad consistentes con un concepto único de diversidad beta.

Una de las preguntas básicas en ecología de comunidades es si las diferencias en la composición de las comunidades se pueden predecir en función de las diferencias en las condiciones ambientales que habitan y, por tanto, por la existencia de presiones selectivas diferentes entre localidades, o si existen otros procesos fundamentales que podrían explicar tal variación biótica (ver la revisión de la teoría de las comunidades ecológicas ofrecida por Vellend (2016)). En este sentido, es clave comprender el papel de la distancia espacial entre dichas comunidades ya que, cuanto más alejadas estén entre ellas, existirá una mayor tendencia a compartir un menor número de especies (Figura 2). Esta tendencia puede deberse a la simple incapacidad de ciertas especies para alcanzar territorios lejanos, porque a mayor distancia espacial menor es la probabilidad de que las especies que habitan una localidad hayan colonizado la otra, especialmente en los grupos biológicos que tienen una fuerte limitación intrínseca a la dispersión (por ejemplo, insectos no alados). Alternativamente, también puede deberse a que las condiciones ambientales presentan una estructura espacial, es decir, sitios próximos tienden a tener condiciones ambientales más parecidas que sitios lejanos (autocorrelación espacial). La autocorrelación espacial del ambiente puede derivar en una autocorrelación espacial de la diversidad biológica si la presencia o ausencia de las especies está determinada por esas condiciones ambientales (predominancia de procesos de selección). El análisis de cómo disminuye la similitud entre comunidades biológicas según aumenta la distancia espacial entre ellas se denomina *distance-decay of similarity* (descenso de similitud con distancia) (Nekola & White, 1999) y generalmente se usa el propio término en inglés: *distance-decay*. Nótese que la similitud biológica y la disimilitud biótica son complementarias: Similitud = (1 – Disimilitud).

El estudio del *distance-decay* constituye uno de los métodos fundamentales de análisis de la diversidad beta (Anderson et al., 2011; Morlon et al., 2008). De hecho, el patrón de *distance-decay* emerge como un patrón universal (Graco-Roza et al., 2022; Nekola & White, 1999; Soininen, Lennon, et al., 2007; Soininen, McDonald, et al., 2007) ya que se ha demostrado que existe un descenso de la similitud biológica con la distancia en diferentes grupos taxonómicos (Astorga et al., 2012; Poulin, 2003; Qian & Ricklefs, 2012; Saito et al., 2015; Svenning et al., 2011; Wetzel et al., 2012), hábitats (Chust et al., 2016; Jiménez-Valverde et al., 2010; Olden et al., 2010; Si et al., 2015; Tovo & Favretti, 2018), escalas espaciales (Arribas et al., 2020; Bell, 2010; Buckley & Jetz, 2008; Freijeiro & Baselga, 2016; König et al., 2017), y niveles de organización biológica (Baselga et al., 2013; Baselga et al., 2022; Baselga et al., 2015; Diniz-Filho & Bini, 2011; Gómez-Rodríguez et al., 2019). Por tanto, una de las preguntas fundamentales en ecología y biogeografía no es sólo si existe o no un patrón de *distance-decay* en las comunidades de estudio sino, sobre todo, cuál es la intensidad de dicho descenso y si es diferente entre grupos biológicos con rasgos morfológicos o funcionales diferentes. En este sentido, el análisis del *distance-decay* constituye una herramienta fundamental para entender la variación en la composición de las comunidades y los procesos que la generan (Gómez-Rodríguez & Baselga, 2018; Steinitz et al., 2006; Tornero et al., 2018, entre otros; Tuomisto et al., 2003). Además, desde el punto de vista aplicado, esta aproximación también se ha usado en estudios de conservación para identificarlos factores responsables de la variación en la composición de las comunidades en zonas amenazadas (Draper et al., 2019; Zhang et al., 2019).

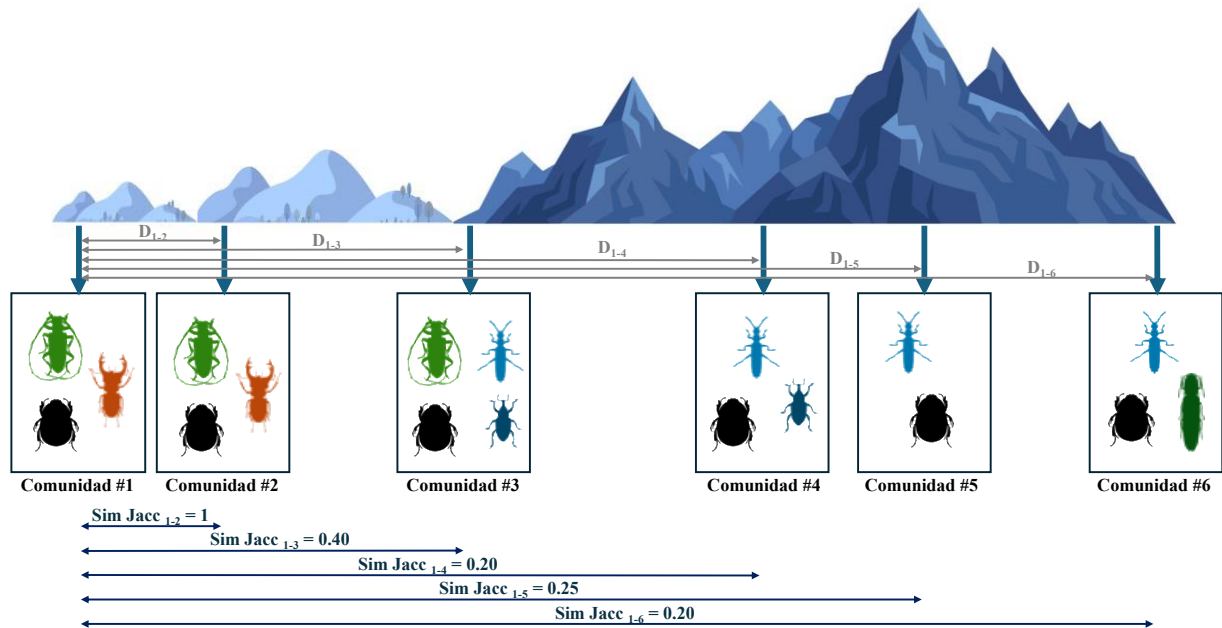


Figura 2. Ejemplo ilustrativo de cómo disminuye la similitud biótica entre la comunidad focal (comunidad #1) y el resto de las comunidades biológicas según se incrementa la distancia espacial entre ellas. Se muestra también cómo esa variación en la composición de las comunidades podría estar asociada a la variación en las condiciones ambientales, la cual también presentaría una estructura espacial al ser más similares las condiciones abióticas de los sitios de menor altitud en relación con los de mayor altitud. En un estudio de *distance-decay* es necesario analizar cómo varía la similitud biótica con la distancia espacial teniendo en cuenta todas las localidades, no sólo la comunidad focal que se usa como ejemplo aquí. La similitud biótica ha sido medida con el índice de Jaccard. (Crédito siluetas: phylopic.org (licencia CC0, CC BY-SA y CC BY 4.0) y pch.vector/Freepik)

La utilidad del *distance-decay* en el estudio de las causas de la estructura espacial de la diversidad biológica radica en que la disminución de la similitud biológica con la distancia espacial refleja procesos ecológicos subyacentes que operan a través de varias escalas espaciales. Comprender estos procesos es crucial para desentrañar los mecanismos responsables de la variación en la composición de las comunidades biológicas. Los procesos más importantes son los siguientes:

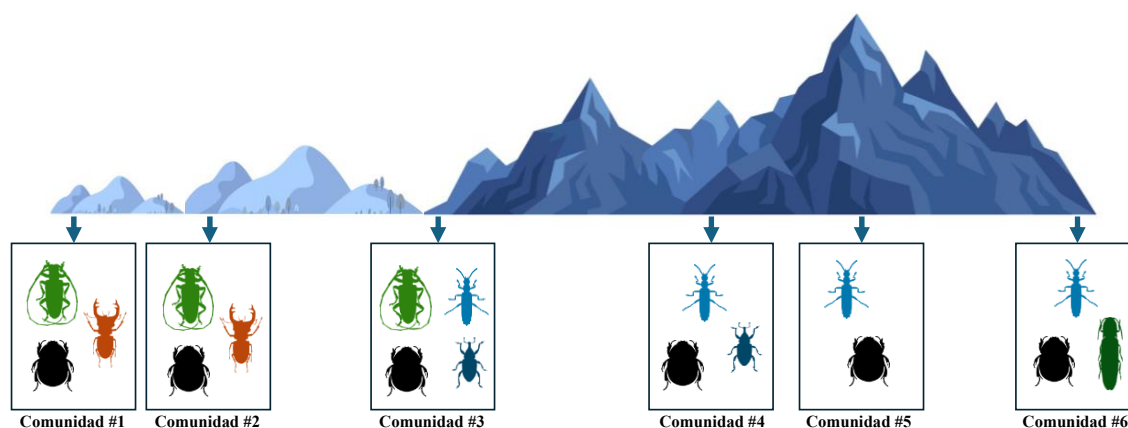
1. Limitación de la dispersión: La dispersión juega un papel fundamental en la configuración de la composición de las comunidades biológicas (Leibold & Chase, 2018; Leibold et al., 2004; Saito et al., 2015) y de los patrones biogeográficos (Svenning & Skov, 2007a). A medida que aumenta la distancia espacial entre dos localidades, la capacidad de las especies para dispersarse entre ellas disminuye, lo que conlleva una reducción de su similitud biótica. Es decir, tendrán menos especies en común y habrá más especies que aparezcan exclusivamente en una o en otra comunidad. La limitación a la dispersión puede surgir de la propia capacidad intrínseca de los organismos para desplazarse o expandir sus rangos de distribución (Gómez-Rodríguez & Baselga, 2018; Qian, 2009) o de factores extrínsecos, como las barreras físicas o la fragmentación del hábitat (Soininen, McDonald, et al., 2007).

2. Filtrado ambiental: El filtrado ambiental ocurre cuando ciertas especies son favorecidas o excluidas de determinados sitios en función de sus tolerancias fisiológicas, rasgos morfológicos o rasgos vitales (Chase & Myers, 2011; Kraft et al., 2015). Esto es debido a que, para cada especie, existe un rango de condiciones ambientales en las que puede tener un crecimiento poblacional positivo, lo que se denomina el nicho ecológico de la especie (Chase & Leibold, 2003; Hutchinson, 1957; Peterson et al., 2011). Como las condiciones ambientales varían en el espacio, también varía la idoneidad de dichas condiciones ambientales para las diferentes especies y, con ello, para la comunidad biológica en su conjunto, ya que esta estará conformada por distintas especies, cada una de ellas con su propio nicho ecológico. A medida que aumenta la distancia espacial entre localidades, las condiciones ambientales tienden a ser más diferentes entre ellas, como se ha explicado anteriormente, lo que puede resultar en diferencias en la composición de la comunidad si predominan los procesos de selección (*sensu* Vellend, 2016) y con ellos el filtrado diferencial de especies (Nekola & White, 1999; Soininen, McDonald, et al., 2007).

3. Contingencias históricas: La historia de una región, incluido su pasado geológico, climático y evolutivo, puede condicionar la composición de las comunidades que observamos actualmente (Baselga, Gómez-Rodríguez, et al., 2012; Graham et al., 2012; Hawkins & Porter, 2003; Montoya et al., 2007; Ricklefs & Schluter, 1993). Las denominadas contingencias históricas, tales como las fluctuaciones climáticas del pasado o los eventos geológicos, generan patrones biogeográficos reconocibles hoy en día asociados a la contracción, expansión o fragmentación de los rangos de distribución de las especies (Condamine et al., 2013; Dias et al., 2014; Dynesius & Jansson, 2000; Jansson, 2003). Estos potenciales legados históricos han de ser tenidos en cuenta a la hora de interpretar las causas de la similitud (o disimilitud) entre comunidades biológicas, particularmente a grandes escalas espaciales, en las que la limitación a la dispersión puede jugar un papel fundamental a la hora de explicar qué lugares han sido o no recolonizados tras el evento histórico que alteró la composición de la comunidad (Fitzpatrick et al., 2013; Gómez-Rodríguez & Baselga, 2018; Martín-Devasa et al., 2024; Qian et al., 2009). Un ejemplo del papel de estas contingencias históricas son los patrones de distribución de algunos animales y plantas en el continente europeo, en el cual se observa una recolonización incompleta de las latitudes más septentrionales tras las últimas glaciaciones (Gómez-Rodríguez & Baselga, 2018; Svenning & Skov, 2007b).

4. Interacciones bióticas: Las interacciones entre especies, incluidas la competencia, la depredación, el mutualismo o la facilitación, tienen un papel fundamental dentro de los procesos de selección que determinan la estructura y composición de las comunidades (Callaway et al., 2002; Chase & Myers, 2011; Graham et al., 2009). Al igual que en el caso de las condiciones abióticas, la intensidad y la frecuencia de estas interacciones puede variar entre localidades, lo que conlleva cambios en la composición de la comunidad. Las diferencias en el tipo, intensidad y frecuencia de las interacciones tienden a ser mayores según se incrementa la distancia entre las comunidades biológicas, pudiendo generar por tanto un patrón de *distance-decay* (Trøjelsgaard et al., 2015).

La aproximación más común en el estudio de los patrones de *distance-decay* es el ajuste de un modelo estadístico de la relación entre dos matrices de datos pareados: la similitud entre comunidades y la distancia espacial (Figura 3). Dicho modelo estadístico permite la estimación de parámetros con relevancia biológica, tales como la proporción de varianza de la similitud entre comunidades que es explicada por la distancia geográfica, o la tasa según la cual esta similitud desciende con la distancia.



MATRIZ DE SIMILITUD BIÓTICA

	Com#1	Com#2	Com#3	Com#4	Com#5	Com#6
Com#1	1.00					
Com#2	1.00	1.00				
Com#3	0.40	0.40	1.00			
Com#4	0.20	0.20	0.75	1.00		
Com#5	0.25	0.25	0.50	0.67	1.00	
Com#6	0.20	0.20	0.40	0.50	0.67	1.00

MATRIZ DE DISTANCIA ESPACIAL

	Com#1	Com#2	Com#3	Com#4	Com#5	Com#6
Com#1	0.0					
Com#2	1.0	0.0				
Com#3	2.5	1.5	0.0			
Com#4	4.0	3.0	1.5	0.0		
Com#5	5.0	4.0	2.5	1.0	0.0	
Com#6	6.5	5.5	4.0	2.5	1.5	0.0

Figura 3. Ejemplo ilustrativo de la construcción de matrices de similitud biótica (izquierda) y distancia espacial (derecha) a partir de los datos de composición de comunidades y su disposición en el espacio. Nótese que en este caso se incluye la diagonal de las matrices, con valor 1 en el caso de la similitud biótica (es decir, comunidades idénticas, ya que son la misma) y valor 0 en el caso de la distancia espacial. En el cálculo de los modelos de *distance-decay* no se tienen en cuenta estos valores, pero se muestran en la figura por su valor ilustrativo. La similitud biótica ha sido medida con el índice de Jaccard y la distancia espacial es distancia euclídea. (Crédito siluetas: phylopic.org (licencia CC0, CC BY-SA y CC BY 4.0) y pch.vector/Freepik)

La búsqueda de un modelo estadístico adecuado que capture el *distance-decay* empezó con el artículo fundacional de Nekola & White (1999), en el cual mostraban que un modelo exponencial (la regresión de la similitud transformada logarítmicamente frente a la distancia espacial) era el mejor ajuste para sus datos. Desde entonces, los patrones de *distance-decay* han sido principalmente estudiados con modelos lineales (*Linear Models*, LM) con o sin transformar las variables logarítmicamente (por ejemplo, Nekola & White, 1999); o bien con modelos lineales generalizados (*Generalized Linear Models*, GLM) (por ejemplo, Millar et al.,

2011; Soininen, McDonald, et al., 2007). Más recientemente, Nekola & McGill (2014) emplearon datos simulados para demostrar que la forma funcional de la relación en los *distance-decay* puede ser una función exponencial negativa, generalmente para datos recogidos en escalas espaciales amplias, o bien una función *power-law*, si la escala espacial es pequeña. No obstante, la principal dificultad de la construcción de modelos *distance-decay* no reside en la identificación de la mejor función de ajuste, sino en cómo se determina la significación del modelo ya que, al estar construido a partir de matrices triangulares de datos, se viola la asunción de independencia de los datos. Es decir, la similitud entre la comunidad #1 y la comunidad #3 de la Figura 4 no es independiente de la similitud entre la comunidad #3 y la comunidad #6, ya que los datos de composición biótica de la comunidad #3 han sido utilizados para calcular ambas similitudes.

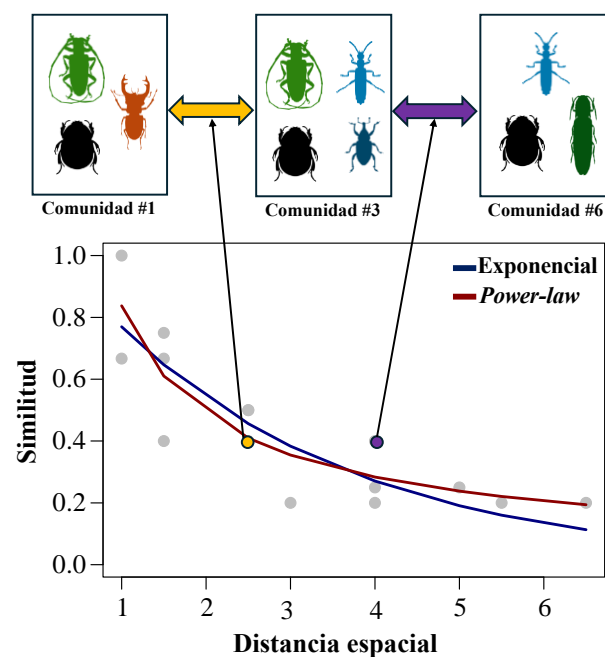


Figura 4. Representación gráfica del descenso en la similitud biótica con la distancia (*distance-decay*) de las comunidades y región representadas en las Figuras 2 y 3. Se muestra el ajuste de la función exponencial y la función *power-law*. Además, se han resaltado los casos de la comparación entre la comunidad #1 y la comunidad #3 (amarillo) e, independientemente, de la comunidad #3 y la comunidad #6 (violeta). De esta forma se muestra la violación de la asunción de independencia de los casos (cada uno de los puntos del gráfico), al estar la comunidad #3 involucrada en el cálculo de ambos casos (y de todos los pares que incluyan a la comunidad #3). La similitud biótica ha sido medida con el índice de Jaccard y la distancia espacial es distancia euclídea. (Crédito siluetas: phylopic.org (licencia CC0, CC BY-SA y CC BY 4.0))

La evaluación de la significación de los modelos *distance-decay* debe tener en cuenta la dependencia por pares inherente a los datos de similitud biótica y distancia espacial. Por ello, los contrastes de significación que se usan rutinariamente en los LMs y GLMs (por ejemplo, el test de la F , ver Hastie & Pregibon, 1993) no deberían aplicarse a datos de *distance-decay*, ya que se viola la asunción de independencia de estos contrastes. Esta dependencia en las variables de similitud y distancia deriva de su construcción a partir de comparaciones entre todos los

posibles pares de comunidades (Legendre et al., 1994; Smouse et al., 1986), tal y como se muestra en la Figura 3. Por lo tanto, cada observación independiente (p. ej. una comunidad biológica) está implicada en el cálculo de múltiples valores de similitud/distancia y por ello los casos usados para construir el modelo (las similitudes) no son independientes entre sí (ver Figura 4). En los modelos lineales, el sesgo estadístico causado por esta dependencia a pares en matrices de distancia ha sido tenido en cuenta en el diseño del test de Mantel (Mantel, 1967), que se usa como contraste de significación para este tipo de modelos. El test de Mantel evalúa si dos variables de distancia (o similitud) son independientes comparando un estadístico lineal (el coeficiente de correlación de Pearson) o un estadístico no paramétrico (coeficientes de correlación de rangos de Spearman o Kendall) con la distribución de dicha correlación bajo la hipótesis nula de independencia de las matrices (calibración del contraste), obtenida usando permutaciones de los sitios de muestreo en una de las matrices. Es decir, el test de Mantel es adecuado para evaluar la hipótesis de una relación lineal entre los datos, pero no está diseñado para testar otro tipo de relaciones, por lo menos de forma paramétrica. Si se utilizan los estadísticos no paramétricos de correlación, el test de Mantel sí es capaz de acomodar la relación no lineal que se espera entre la similitud de las comunidades y la distancia que las separa (Dietz, 1983). Sin embargo, siempre que sea posible es preferible utilizar modelos paramétricos no lineales (p.ej. exponencial o *power-law*) antes que recurrir al uso de contrastes no paramétricos. Además de tener mayor poder estadístico, los modelos paramétricos tienen la ventaja de que permiten la comparación de parámetros (por ejemplo, la pendiente o el intercepto) entre modelos de *distance-decay* ajustados a datos de distintos taxones y/o regiones. Estas comparaciones son clave para inferir los procesos que han generado los patrones de *distance-decay* (Gómez-Rodríguez & Baselga, 2018; Peguero et al., 2021; Saito et al., 2015; Soininen, McDonald, et al., 2007; Yang et al., 2021), que es el objetivo último de la ecología de comunidades o la biogeografía.

La ausencia de independencia entre los casos de las matrices de distancia y similitud debida a su estructura a pares supone el primer reto para el diseño de un contraste de significación aplicable a modelos no lineales de *distance-decay*. Este contraste debe comparar el patrón no lineal ajustado por el modelo con el esperado si la similitud y la distancia fueran independientes entre ellas (hipótesis nula del contraste) y debe diseñarse teniendo en cuenta la dependencia por pares. Cabe recordar que la mayoría de los contrastes de significación (p.ej. el test de la F) presentan la asunción básica de que las observaciones de una misma variable son independientes entre ellas. Cuando se viola esta asunción, como es el caso de los modelos *distance-decay*, se causa una inflación artificial de los grados de libertad del contraste conocida como pseudoreplicación, que aumenta a medida que lo hace el tamaño muestral (Hurlbert, 1984) (Figura 5).

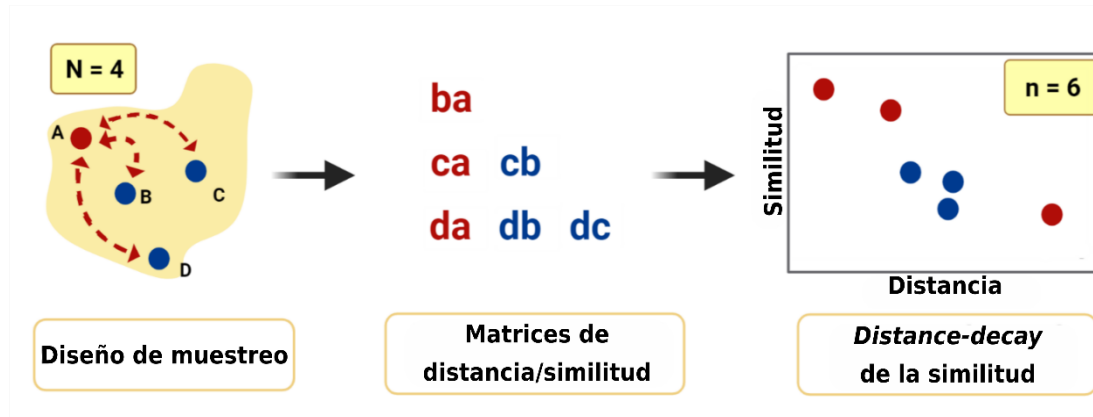


Figura 5. Ejemplo ilustrativo de la pseudoreplicación en datos de distancia y similitud. El tamaño muestral aumenta de forma artificial tras calcular las matrices de similitud y distancia, ya que de cuatro localidades se obtienen seis valores de similitud (ya que para N localidades tenemos $N*(N - 1)/2$ valores de similitud y distancia). Otra evidencia de la existencia de pseudoreplicación y dependencia a pares viene dada por el hecho de que la localidad A está implicada en el cálculo de todos los valores de similitud y distancia marcados en rojo ($N - 1$ valores). Figura creada en BioRender.com.

La pseudoreplicación puede aumentar el error tipo I de los contrastes de significación, provocando que se considere significativo el modelo cuando esto no es cierto (falsos positivos). En el contexto particular de los modelos de *distance-decay*, la pseudoreplicación se genera porque a partir de N sitios se obtienen $N * (N - 1)/2$ similitudes/distancias. Para evitar esta inflación del error de tipo I, el test de Mantel estima la distribución nula del coeficiente de correlación entre similitud y distancia a través de permutaciones de los sitios de muestreo en la matriz de similitud. Después compara la correlación observada con esta distribución nula y calcula el p-valor del contraste como la proporción de casos en los que la correlación nula es mayor que la observada. A pesar de su diseño específico para el tipo de datos usados en los *distance-decay*, es necesario recordar que no es apropiado aplicar el test de Mantel a este tipo de modelos ya que es un contraste de significación robusto para modelos lineales, pero no para el tipo de relaciones no lineales (p.ej. exponencial negativa o *power-law*) que se observan en la relación entre la similitud biótica y la distancia espacial (Nekola & McGill, 2014). Además puede presentar sesgo si no se considera la estructura espacial de los datos (Guillot & Rousset, 2013; Legendre et al., 2015; Legendre & Fortin, 2010). Autores previos han mostrado que cuando la relación entre las variables se aleja de la linealidad, la habilidad del test de Mantel para rechazar la hipótesis nula cuando es falsa puede reducirse (baja potencia estadística) (Diniz-Filho et al., 2013; Zeller et al., 2016), lo que limita su aplicación a los modelos de *distance-decay*, en los que se espera que su forma funcional no sea lineal (Nekola & McGill, 2014). En resumen, no se puede aplicar el test de Mantel para testar la significación de un modelo no lineal de *distance-decay* debido a que no compara su hipótesis nula con el modelo ajustado, sino con un coeficiente de correlación lineal o no paramétrico.

La detección de patrones no lineales (alta potencia estadística o bajo error de tipo II) es el segundo reto en el diseño de un contraste de significación para modelos no lineales de *distance-decay*. Una posible aproximación es el uso de un estadístico pseudo- R^2 en el contexto de un GLM, el cual estima qué proporción de la variación presente en los datos es explicada por el modelo ajustado (McCullagh & Nelder, 1989). Así, una vez ajustado un GLM con la relación funcional de interés (por ejemplo, exponencial negativa o *power-law*), la pseudo- R^2 compara el ajuste del modelo con un modelo nulo en el cual similitud y distancia son independientes entre ellas y por lo tanto no hay una relación de *distance-decay*. Si seguimos el razonamiento del test de Mantel y combinamos un estadístico basado en la pseudo- R^2 con permutaciones de los sitios de muestreo en la matriz respuesta (similitud), sería posible estimar la distribución nula de dicho estadístico bajo condiciones de dependencia a pares en las matrices. Otra posible opción es utilizar un método de remuestreo que considere la estructura pareada de los datos para generar remuestras independientes (con reemplazamiento) de las matrices de similitud y distancia. En este caso, el objetivo es mantener la estructura pareada concreta de los datos. Para abordar este problema, en esta tesis se propone que las remuestras sean tomadas a partir de conjuntos de datos que incluyan todas aquellas similitudes/distancias en las que intervenga una localidad determinada. Este método se ha denominado “remuestreo por bloques de localidad”. Aunque en el caso de un contraste de significación ambos métodos probablemente obtengan resultados semejantes, un remuestreo por bloques de localidad adaptado a datos de *distance-decay* podría ser extendido a otras preguntas basadas en modelos que se ven afectados por la dependencia a pares, tales como la estimación de la varianza de los parámetros del modelo o el contraste de las diferencias entre parámetros de dos modelos de *distance-decay* para distintos datos (por ejemplo, dos grupos biológicos o dos regiones diferentes). El valor de la pseudo- R^2 ha sido previamente utilizado para contrastes de significación en contextos similares; por ejemplo en combinación con aleatorizaciones que no preservan la estructura a pares de los datos (Gómez-Rodríguez & Baselga, 2018), en regresiones no lineales de matrices de distancias (Lichstein, 2007), o bien empleado a la par como contraste de significación y herramienta para la selección de modelos de *distance-decay* no lineal (*Generalized Dissimilarity Modelling*; Ferrier et al., 2007). Sin embargo, hasta donde sabemos ningún estudio ha evaluado de manera formal el comportamiento de un contraste de significación con permutaciones o, del propuesto en esta tesis como remuestreo por bloques de localidad, para modelos de *distance-decay*.

Los modelos de *distance-decay* han sido utilizados de forma extensiva en ecología de comunidades y biogeografía (ver, por ejemplo, Nekola & White, 1999; Sojininen et al., 2007, y los cientos de artículos que los citan). A priori, sabemos que la relación entre la similitud entre comunidades y la distancia espacial (o ambiental) que las separa debe ser no lineal (Nekola & McGill, 2014). Por lo tanto, la ausencia de un procedimiento estandarizado para evaluar su significación constituye un vacío importante en las herramientas disponibles para la biogeografía y la ecología de comunidades y, en concreto, para el análisis de la diversidad beta. El objetivo global de esta tesis es diseñar y estudiar el funcionamiento de un estadístico basado en la pseudo- R^2 en combinación con permutaciones o remuestreo por bloques de localidad como un contraste de significación para modelos exponenciales y *power-law* de *distance-decay*.

En esta tesis se abordan tres aspectos relevantes sobre los modelos *distance-decay*, que se desarrollan en secciones paralelas de la Metodología (Capítulo 4) y los Resultados (Capítulo 5). Además, antes del desarrollo metodológico formal, se realizan comprobaciones previas de los atributos de los patrones de *distance-decay* presentes en los datos empíricos de coleópteros europeos usados en esta tesis. Esta sección se denomina, de forma abreviada, “**Estudio no paramétrico**”.

El primer aspecto metodológico, que se aborda en los respectivos subcapítulos de la Metodología y los Resultados identificado como “**Estructura del error**”, explora la utilización de modelos GLM para ajustar los modelos de *distance-decay*, analizando si la estructura del error seleccionada (Gaussiana o Gamma) influye en los parámetros y la varianza explicada por el modelo y, de ser así, qué familia de modelos es más apropiada.

A continuación, dentro de la sección identificada como “**Métodos bootstrap**”, en sus respectivos apartados de Metodología y Resultados, se explora la citada técnica combinada con un contraste de significación del modelo basado en la *deviance*, y se implementa un estudio de simulación para comprobar si dicho procedimiento permite una buena estimación de la significación de los modelos.

Finalmente, en la última sección (“**Implementación de un método para contraste de hipótesis en modelos no lineales de distance-decay basado en permutaciones y remuestreo por bloques de localidad**”), se proponen dos estrategias alternativas para el contraste de la significación de los modelos no lineales de *distance-decay*, que consideran la estructura pareada de los datos de *distance-decay*, utilizando para la estimación de la distribución nula del estadístico pseudo- R^2 basado en la *deviance*. Estas dos alternativas se construyen utilizando, o bien (i) permutaciones de los sitios de muestreo en la matriz de similitud (equivalente a las permutaciones aplicadas en el test de Mantel, pero utilizando el estadístico pseudo- R^2 de un modelo no lineal) o bien, alternativamente, (ii) un método de remuestreo por bloques de localidad, que se introduce de forma novedosa en esta tesis. Posteriormente, se evalúa mediante simulaciones la capacidad del contraste de significación de la pseudo- R^2 para detectar la ausencia de *distance-decay* (error de tipo I) y su presencia (potencia estadística, equivalente a $1 - \text{error de tipo II}$) y, se aplica el contraste a datos reales. Estas nuevas metodologías constituyen una alternativa apropiada y preferible al test de Mantel para modelos paramétricos no lineales, dado que este trabaja exclusivamente con relaciones lineales o no paramétricas.

3 OBJETIVOS

El objetivo global de esta tesis es diseñar y estudiar el funcionamiento de un estadístico y un contraste de significación para modelos de descenso de la similitud biótica con la distancia espacial, también denominados modelos *distance-decay*. Los modelos *distance-decay* se basan en regresiones no lineales entre matrices de similitud/distancia, en las que existe una pseudorreplicación de los datos causada por la dependencia a pares de los mismos, lo que impide la aplicación de modelos GLM estándar y hace necesario desarrollar un contraste de significación que incorpore de forma explícita la dependencia inherente a este tipo de datos.

Los objetivos parciales son:

- Analizar la estructura del error de los modelos *distance-decay* ajustados mediante Modelos Lineales Generalizados (*Generalized Linear Models*, GLMs) para determinar su influencia en la parametrización y, con ello, identificar qué familia de modelos (Gaussiana o Gamma) es más apropiada para los modelos *distance-decay*.
- Determinar si un contraste de significación basado en la *deviance* y técnicas de *bootstrap* desarrolladas por autores previos son apropiadas para la estimación de la significación de los modelos no lineales de *distance-decay*.
- Desarrollar y aplicar estrategias novedosas para la estimación de la significación de los modelos no lineales de *distance-decay* que tengan en cuenta la estructura pareada inherente a este tipo de datos. Se proponen dos técnicas alternativas según el método usado para la estimación de la distribución nula del estadístico pseudo- R^2 :
 - Permutaciones de los sitios de muestreo en la matriz de similitud, de forma análoga al test de Mantel, pero utilizando el estadístico pseudo- R^2 de un modelo no lineal.
 - Remuestreo por bloques de localidad, que tiene en cuenta la dependencia de los datos al realizar el muestreo de los mismos, ya que los bloques que se remuestran son los que incluyen todas aquellas similitudes/distancias en las que interviene una localidad determinada
- Implementar los métodos estadísticos desarrollados en la librería “betapart” de R, para garantizar que estén accesibles para todos los investigadores en este campo de estudio.

4 MÉTODOLÓGÍA

En esta tesis se analiza cuál es la metodología apropiada para el ajuste de modelos no lineales de *distance-decay*. Para ello, se estudian y aplican los métodos propuestos a datos reales de distribución de especies, los cuales son descritos a continuación. Asimismo se incluye un apartado sobre los aspectos comunes del ajuste de los modelos de *distance-decay* utilizados en esta tesis y que están relacionados principalmente con el cálculo de las matrices de similitud biótica para los datos empíricos utilizados.

4.1.DATOS DE DISTRIBUCIÓN DE COLEÓPTEROS EN EUROPA

El conjunto de datos empíricos utilizado en esta tesis incluye tablas de presencia/ausencia para 21 taxones supraespecíficos de coleópteros (grandes géneros, tribus, subfamilias o familias; 4648 especies en total) en 34 territorios europeos (extensión: 11° W – 60° E de longitud y 36° N – 72° N de latitud). Las características principales de este conjunto de datos fueron descritas en Baselga, Lobo et al. (2012), Gómez-Rodríguez et al. (2015) y Gómez-Rodríguez & Baselga (2018). Los 21 taxones son representativos de los principales linajes de escarabajos (superfamilias Caraboidea, Staphylinoidea, Scarabaeoidea, Chrysomeloidea y Curculionoidea) e incluyen un amplio espectro de estrategias ecológicas y capacidades de dispersión.

Las tablas de presencia/ausencia se compilaron a partir de Löbl & Smetana (2003, 2004, 2006, 2010) y la web de Fauna Europaea (Fauna Europaea versión 1.1, disponible en línea en <http://www.faunaeur.org>). Los territorios generalmente corresponden a países europeos, pero se excluyeron todas las islas y los países de menor tamaño (< 2000 km², es decir, Andorra, Liechtenstein, San Marino, Mónaco y Ciudad del Vaticano) para evitar diferencias extremas en áreas y patrones biogeográficos derivados de la insularidad. Además, Croacia y Bosnia-Herzegovina se agruparon en una sola unidad porque el primer país rodea al segundo, y la región europea de Rusia se dividió en tres unidades [norte, centro y sur] dada su área extremadamente grande. Luxemburgo y Moldavia fueron excluidos de los análisis debido a la ausencia de datos para algunos taxones, lo que plantea problemas para el cálculo de la similitud biótica y la comparación estadística con los otros territorios. Por la misma razón, dos taxones incluidos en un artículo anterior (Baselga, Lobo, et al., 2012), Baridinae y Nanophyidae, no fueron considerados en este estudio (ya que carecen de datos en Albania, Eslovenia y Macedonia). El conjunto de datos final comprende los siguientes taxones: *Amara*, Apionidae, *Carabus*, Cerambycinae, *Chaetocnema*, Chrysomelinae, Cryptocephalinae, Hydraenini, Lamiinae, Lepturinae, Limnebiini, Ochtebiini, *Otiorhynchus*, Phyllobiini, Polydrusini, Pselaphinae, *Pterostichus*, Scarabaeidae, Scolytinae, Silphidae y *Trechus*.

El número de especies de los conjuntos de datos oscila entre 40 (familia Silphidae) y 817 (subfamilia Pselaphinae). Se incluyen grupos con muy baja capacidad de dispersión (por ejemplo, el género *Otiorhynchus*, con 698 especies, todas ellas ápteras) y grupos con alta capacidad de dispersión, como la subfamilia Scolytinae, en el que no hay especies ápteras y se ha observado que algunos individuos pueden llegar a dispersarse hasta 100 km en un día mediante vuelos a elevada altitud (Jackson et al., 2008). El hábito trófico predominante es la herbivoría, siendo la mayoría de los grupos fitófagos (n =11, por ejemplo la subfamilia Chrysomelinae), y el menos representado es la coprofagia (familia Scarabaeidae) y la negrofagia (familia Silphidae). En la Figura 6 se muestran algunas especies representativas de los grupos mencionados.

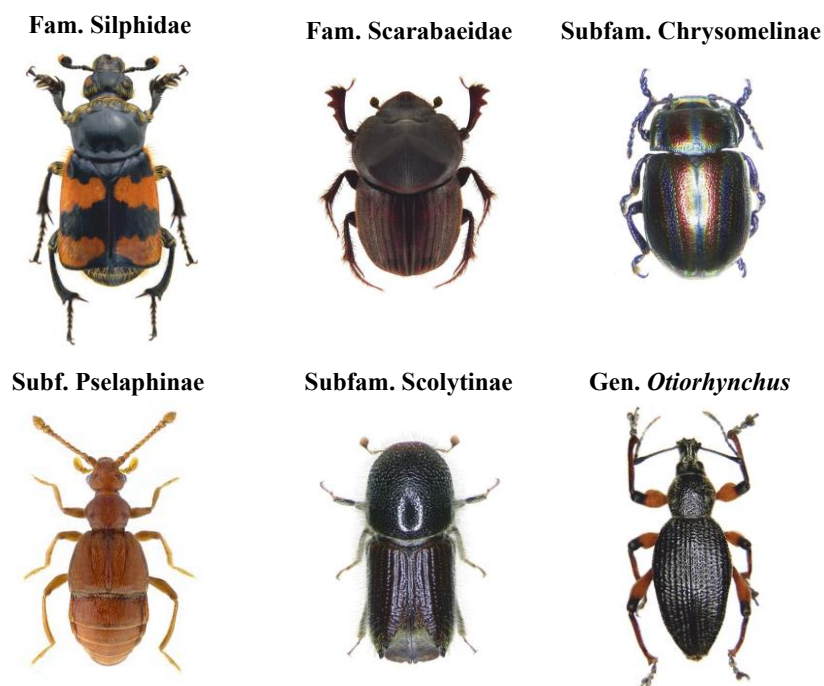


Figura 6. Especímenes representativos de algunos de los taxones utilizados en esta tesis (fuente: wikimedia commons, autor: Udo Schmidt, licencia Creative Commons Attribution-Share Alike 2.0).

Es necesario tener en cuenta que trabajos previos han mostrado que el patrón de *distance-decay* de los coleópteros es diferente en el sur y norte de Europa (Gómez-Rodríguez & Baselga, 2018). Esto puede ser atribuido a las distintas historias biogeográficas de estas dos regiones y, en concreto, al efecto de las glaciaciones y los refugios glaciares en la distribución de las especies. Estos modelos previos de *distance-decay* no fueron diseñados teniendo en cuenta la dependencia por pares para el contraste de la significación. De hecho, en esta tesis se realiza una valoración previa, basada en una aproximación no paramétrica para determinar si existe realmente una diferencia entre los patrones del norte y del sur y que, por tanto, las diferencias previamente observadas no son resultado del propio modelo elegido (ver sección 4.2). Es decir, se estudia si existe una diferencia en los conjuntos de datos evitando ningún tipo de parametrización.

En los análisis comparando el patrón de *distance-decay* en el norte y sur de Europa, se han considerado dos conjuntos de datos para cada taxón: uno incluye los países del norte de Europa (número de países, $n = 18$) y otro los del sur de Europa ($n = 16$).

Países del sur de Europa: Albania (AL), Austria (AT), Bulgaria (BG), Croacia-Bosnia-Herzegovina (CBH), Suiza (CH), España (ES), Francia (FR), Grecia (GR), Hungría (HU), Italia (IT), Macedonia (MK), Portugal (PT), Rumanía (RO), Eslovenia (SL), Serbia (SS), Turquía-Europea (TR).

Países del norte de Europa: Bélgica (BE), Bielorrusia (BY), República Checa (CZ), Alemania (DE), Dinamarca (DK), Estonia (EO), Finlandia (FI), Lituania (LT), Letonia (LV), Países Bajos (NL), Noruega (NO), Polonia (PL), Rusia-centro (RSC), Rusia-norte (RSN), Rusia-sur (RSS), Suecia (SE), Eslovaquia (SK), Ucrania (UK).

4.1.1 Cálculo de la similitud entre faunas y distancias entre territorios europeos

A partir de las tablas de presencia/ausencia de coleópteros en territorios europeos se han calculado las matrices de similitud biótica entre las faunas de los diferentes territorios para cada taxón, diferenciando entre territorios del norte y del sur de Europa. Igualmente, se ha calculado una matriz de distancia geográfica entre los diferentes territorios. De esta forma, se dispone de dos matrices de distancia geográfica entre territorios, una para el norte y otra para el sur, y de 42 matrices de similitud biótica, dos para cada taxón para el que se estudiará su patrón de *distance-decay* en el norte e, independientemente, en el sur de Europa.

La matriz de distancia geográfica se ha obtenido calculando, para cada par de territorios, la distancia euclídea entre los centroides con la función `vegdist()` de la librería `vegan` (Oksanen et al., 2017) de R (R Development Core Team, 2022). El resultado es una semimatriz de datos de distancia espacial que constituye la variable independiente (variable X) de los modelos de *distance-decay*. La matriz de similitud biótica entre territorios se ha calculado usando el índice de similitud de Simpson (β_{sim} , ecuación 4.1) (Simpson, 1943, 1960) aplicado a los datos de presencia/ausencia de especies de cada territorio. El conjunto de especies presente es un determinado país, es decir, la fauna de dicho territorio, se considerará en esta tesis equivalente a una comunidad biológica para reducir el número de términos utilizados en la descripción de los patrones de *distance-decay*, si bien el término comunidad biológica no es estrictamente aplicable a escalas espaciales tan amplias.

La fórmula del índice de disimilitud de Simpson es

$$\beta_{sim} = \frac{\min(b,c)}{a+\min(b,c)} \quad (\text{Ecuación 4.1})$$

donde a es el número de especies presentes en ambos sitios, b es el número de especies que están presentes en el primer sitio pero no en el segundo y c es el número de especies que están presentes en el segundo sitio pero no en el primero. El índice de disimilitud de Simpson (β_{sim}) calcula la disimilitud biótica entre comunidades eliminando el efecto de la diferencia en riqueza entre ellas. Por tanto, informa de la sustitución o reemplazamiento de especies entre

comunidades y no está afectado por los patrones de anidamiento de las comunidades, en los que las especies que aparecen en la comunidad de menor riqueza son subconjuntos de los que aparecen en las comunidades de mayor riqueza (es decir, comunidades empobrecidas, ver Figura 7). Baselga (2010b) ofrece una discusión detallada sobre las ventajas del uso del índice de Simpson y, en general, de la partición de la diversidad beta en componentes que midan de manera separada los patrones de reemplazamiento y anidamiento. Se ha calculado una semimatriz de disimilitud biótica de cada taxón usando la función `beta.pair()` especificando la familia “Sørensen” de índices de la librería “betapart” (Baselga et al., 2023) en R. Esta semimatriz de disimilitud biótica se ha convertido una matriz de similitud biótica ($1 - \beta_{sim}$) para cada taxón, la cual se ha usado como variable respuesta (variable Y) de los diferentes modelos de *distance-decay* construídos.

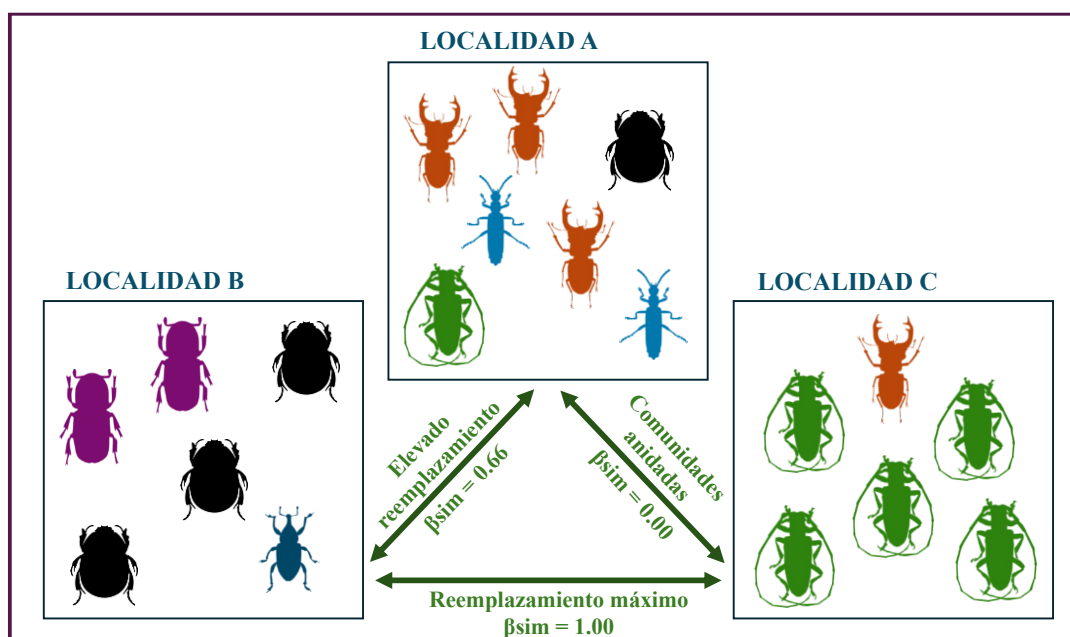


Figura 7. Ejemplo ilustrativo de los valores que adopta el índice de disimilitud de Simpson (β_{sim}) ante patrones de reemplazamiento (comparación A-B y B-C) y patrones anidados (comparación A-C). Como se muestra en la figura, la disimilitud entre las localidades A y C es máxima ya que, a pesar de que la localidad C sólo tiene dos especies de las que aparecen en la localidad A, estas son un subconjunto de las especies de localidad A, lo que da lugar a un patrón de anidamiento. (Crédito siluetas: phylopic.org (licencia CC0, CC BY-SA y CC BY 4.0))

4.2 ESTUDIO NO PARAMÉTRICO DE LAS DIFERENCIAS NORTE-SUR EN LOS PATRONES DE *DISTANCE-DECAY* DE COLEÓPTEROS EUROPEOS

Como análisis preliminar para determinar qué conjunto de datos es el más apropiado para el estudio de los patrones *distance-decay* de coleópteros europeos, se ha realizado un estudio no paramétrico de los patrones en el norte y el sur de Europa, con el objetivo de verificar si las diferencias previamente reportadas en Gómez-Rodríguez & Baselga (2018) se mantienen cuando se aplican modelos sin constricciones paramétricas. De esta forma, se ha confirmado que las diferencias previamente observadas no son resultado del propio modelo elegido y, por tanto, es oportuno mantener dicha subdivisión de los datos empíricos en el resto de los análisis realizados en esta tesis.

Tal y como se ha descrito previamente, se han separado los países de la zona norte (18) y la zona sur (16), y los datos de presencia/ausencia asociados y se han calculado las matrices de distancia (X) y similitud (Y) para cada uno, de tal forma que en las matrices del grupo norte hay 153 casos (pares de países) y en las del grupo sur 120. A continuación, para estudiar si el patrón de descenso de la similitud biológica con la distancia geográfica es semejante en ambas zonas (norte vs. sur), se han utilizado tests no paramétricos, evitando así la asunción de una forma funcional determinada en la modelización de los datos. Para ello, se han realizado dos contrastes basados ambos en un ANCOVA no paramétrico: uno para comprobar si las curvas de regresión ajustadas son iguales y otro para comprobar si son paralelas (Young & Bowman, 1995). Se ha utilizado la función `sm.ancova()` de la librería “sm” en R (Bowman & Azzalini, 2018). Estos análisis se han realizado para cada taxón de estudio, de forma que es posible determinar si la diferencia norte-sur existe en todos, en la mayoría o en prácticamente ningún conjunto de datos.

Los contrastes realizados permiten determinar si el efecto de una variable explicativa X sobre una variable respuesta Y es el mismo (sin restricciones paramétricas) considerando distintos grupos de datos, uno de ellos en términos de si dichos efectos son iguales (contraste de igualdad, ecuación 4.2) y el otro en términos de si son paralelos (contraste de paralelismo, ecuación 4.3). El estadístico del contraste de igualdad es

$$TS = \frac{\sum_{i=1}^P \sum_{j=1}^{n_i} \{\hat{g}_i(x_{ij}) - \hat{g}(x_{ij})\}^2}{\hat{\sigma}^2} \quad (\text{Ecuación 4.2})$$

y el estadístico del contraste de paralelismo es

$$TS = \frac{\sum_{i=1}^P \sum_{j=1}^{n_i} \{\hat{\alpha}_i + \hat{g}(x_{ij}) - \hat{g}_i(x_{ij})\}^2}{\hat{\sigma}^2} \quad (\text{Ecuación 4.3})$$

en donde las estimaciones individuales de cada curva ($\hat{g}_i(x_{ij})$) se contrastan con una curva común estimada bajo la suposición de igualdad ($\hat{g}(x_{ij})$). En el caso del paralelismo se incluye $\hat{\alpha}_i$ que es un intercepto estimado por mínimos cuadrados. Las curvas son estimadas mediante el estimador de tipo núcleo de Nadaraya-Watson (Nadaraya, 1964). Esta comparación se hace en todos los puntos x_{ij} (observaciones de X) y en el denominador aparece la varianza del error (Young & Bowman, 1995).

Cabe indicar que este contraste se plantea sobre datos independientes en cada uno de los grupos, una hipótesis que no se cumple en los datos de similitud y distancia, por lo que los resultados obtenidos deben ser analizados con cautela y siempre teniendo en cuenta que el objetivo de este apartado es valorar la oportunidad de dividir los datos en subconjuntos “norte” y “sur” en los apartados siguientes. No se ha encontrado en la literatura una modificación de este contraste que permita tener en cuenta la dependencia particular de este tipo de datos.

A los p-valores devueltos por los contrastes se les ha aplicado la corrección de Benjamini & Yekutieli (2001). Se ha elegido dicha corrección por la posible correlación positiva entre los elementos de las variables debida la situación espacial de las observaciones. Este método controla la tasa de falsos descubrimientos (*false discovery rate*, FDR) y es más potente que los métodos que controlan el error FWER (*familywise error*) si se espera rechazar muchas de las hipótesis nulas (Goeman & Solari, 2014). Se ha considerado una significación del 5%.

4.3 ESTRUCTURA DEL ERROR EN MODELOS PARAMÉTRICOS (GLM) DE *DISTANCE-DECAY*

Tal como se ha expuesto en la Introducción, el descenso de la similitud entre comunidades biológicas a medida que aumenta la distancia que las separa (*distance-decay of similarity*) es un patrón macroecológico universal. Su estudio y modelado ha sido un elemento fundamental para entender la variación de las comunidades biológicas en el espacio, y los procesos que gobiernan esa variación (Gómez-Rodríguez & Baselga, 2018; Soininen, McDonald, et al., 2007; Tornero et al., 2018; Tuomisto et al., 2003, entre otros). A la hora de analizar los patrones de *distance-decay*, es necesario ajustar modelos estadísticos paramétricos, con los que se cuantifique la intensidad del descenso de la similitud con la distancia espacial y en qué medida la variación que observamos en los valores de similitud biótica se explica sólo por la distancia espacial o podría estar determinada también por otros factores. Esta cuantificación, a través de los parámetros estadísticos del modelo, nos permite comparar los patrones de *distance-decay* entre diferentes organismos y regiones. Los elementos principales de interés son, por tanto, (i) la tasa según la cual esta similitud desciende con la distancia y (ii) la proporción de varianza de la similitud entre comunidades que es explicada por la distancia geográfica.

A pesar de su interés ecológico, no existe todavía una estrategia canónica sobre cómo deben construirse los modelos *distance-decay*, ya que presentan una serie de peculiaridades y retos que no han sido resueltos en su totalidad. Por ejemplo, aunque muchos trabajos previos utilizan Modelos Lineales Generalizados (*Generalized Linear Model*, GLM) para ajustar este tipo de patrones (Gómez-Rodríguez & Baselga, 2018), no existen estudios previos que analicen qué tipo de estructura del error es la más apropiada en estos casos. Los GLMs son excelentes candidatos para modelizar los patrones de *distance-decay* porque, por un lado, permiten ajustar relaciones no lineales mediante el uso de diferentes links que determinen cómo se relacionan las variables de estudio (por ejemplo, logarítmicamente) y, por otro, permiten utilizar diferentes estructuras del error, que no necesariamente tienen por qué ser normales (Gaussianas), ampliando así el número de opciones disponibles.

En esta sección utilizamos los datos de presencia/ausencia de 21 taxones de coleópteros en países del norte e, independientemente, del sur de Europa para analizar, con datos empíricos, si una distribución Gamma de la estructura del error permitiría un mejor ajuste de los modelos, comparada con una estructura del error Gaussiana. Se ha considerado la función Gamma debido a que es una función continua que puede adoptar una gran variedad de formas. Por otro, es necesario destacar que el conjunto de datos utilizado se considera idóneo para este tipo de análisis debido a la gran variabilidad de estrategias vitales entre los diferentes taxones y, con ello, la gran variabilidad en los patrones de *distance-decay* (Gómez-Rodríguez & Baselga, 2018) que se consideran en el estudio de la distribución del error.

Para cada taxón de coleópteros y área geográfica (norte / sur de Europa), se han ajustado, de forma independiente, dos GLMs con link logarítmico, uno de ellos con error de estructura Gaussiana, y el otro con error de distribución Gamma. En dichos modelos, la variable dependiente es el índice de similitud de Simpson (Baselga, 2010b; Koleff et al., 2003) entre

pares de países europeos, y la variable explicativa es la distancia espacial entre los centroides de dichos países. Para cada uno de los conjuntos de datos, se han comparado los parámetros del modelo (intercepto y pendiente) y la varianza explicada (pseudo- R^2) entre el modelo con error Gaussiano y el modelo con error Gamma. Para ello se compararon empíricamente las distribuciones de dichos estadísticos (intercepto, pendiente y pseudo- R^2), estimadas mediante *bootstrap* (500 remuestras de los datos de similitud). Las diferencias entre las distribuciones de los estadísticos de los modelos con error Gaussiano y los modelos con error Gamma se evaluaron mediante pruebas pareadas de Wilcoxon y se les aplicó la corrección de Benjamini & Yekutieli (2001).

4.4 MÉTODOS *BOOTSTRAP*: EXPLORACIÓN Y APROXIMACIÓN INICIAL

El principal objetivo de esta sección es desarrollar un método sencillo para ajustar un modelo estadístico paramétrico y no lineal a los patrones de *distance-decay*, para que se pueda comparar el descenso de la similitud con la distancia entre distintos taxones mediante la comparación de sus parámetros.

Los principales aspectos a tener en cuenta en la formulación del modelo son:

- Los patrones de diversidad resultantes de algunos procesos biogeográficos son heterocedásticos (su varianza aumenta con la distancia geográfica). Esta circunstancia es razonable, ya que al aumentar la separación entre dos zonas de muestreo es más probable que los factores biológicos que afectan a las especies que en ellos viven sean distintos. Por lo tanto, la heterocedasticidad refleja información biológica relevante y no debe ser eliminada.
- El índice de similitud de Simpson está comprendido entre 0 y 1, exclusivamente.
- Es previsible la existencia de una estructura compleja de dependencia en los datos, difícil de modelizar, debida tanto a la estructura espacial del muestreo como a la construcción a pares de las matrices de similitud/distancia.

Este apartado se ha centrado por tanto en realizar una primera aproximación estadística que permita modelizar este tipo de datos de forma sencilla, interpretable y robusta frente a esta estructura intrínseca de dependencia entre los valores de similitud que incluyen a un mismo punto de muestreo, por ser la violación más grave de las condiciones de los modelos habituales.

4.4.1 Modelo teórico inicial

Se ha asumido *a priori* que la relación entre los datos de similitud (Y) y distancia (X) utilizados en este estudio es de forma exponencial (ecuación 4.4) siguiendo las recomendaciones del trabajo realizado por Nekola y McGill (2014), que postula que esta forma funcional es la apropiada cuando la escala espacial es amplia, tal y como ocurre en el continente europeo. Sin embargo, y con el objetivo de considerar un número más amplio de escenarios potenciales, se contempla también la posibilidad de que la forma funcional que mejor se ajuste en algunos casos sea de tipo *power-law* (ecuación punto 4.5), debido a la distinta capacidad de dispersión de las especies de cada conjunto de datos y las diferencias en las características geográficas entre el norte y el sur de Europa. Por lo tanto, la metodología que se presenta a continuación se ha desarrollado en base a un modelo exponencial, pero a la hora de comparar modelos aplicados a los datos reales se ha considerado también la función *power-law*.

$$Y = ae^{bX} \quad (\text{Ecuación 4.4})$$

$$Y = aX^b \quad (\text{Ecuación 4.5})$$

La ecuación de la curva exponencial es linealizable, y por tanto se ha utilizado un GLM como modelo inicial para los datos en el cual la función link es el logaritmo (McCullagh & Nelder, 1989). Se ha elegido la normal como distribución de Y ya que es continua y tiene soporte para todo el rango de la variable (similitud de Simpson). No obstante, cuando alguno de sus valores es 0 deben indicarse valores iniciales para la estimación de los coeficientes de regresión al utilizar la función `glm()` de R (R Development Core Team, 2022) para ajustar el modelo. Se ha comprobado que con distintos valores absolutos cercanos a 0 (del orden de 10^{-5} o 10^{-4}) o incluso el 0 mismo no varía el resultado. Sin embargo, si se dan valores iniciales demasiado grandes la función no converge. Finalmente, el modelo teórico para los datos se recoge en la ecuación 4.7. Además, para el ajuste de un GLM correcto los datos deben ser independientes entre ellos e Y debe ser homocedástica.

$$E(Y|X) = g^{-1}(\beta X) = e^{\beta X} \leftrightarrow g(E(Y|X)) = \ln(E(Y|X)) = \beta X \quad \text{con} \quad Y \sim N(\mu, \sigma) \quad (\text{Ecuación 4.6})$$

$$Y = e^{\beta X} + \varepsilon = e^{\beta_0} e^{\beta_1 X} + \varepsilon = a e^{bX} + \varepsilon \quad \text{con} \quad Y \sim N(\mu, \sigma), Y \in [0, 1] \quad (\text{Ecuación 4.7})$$

Sin embargo, tal y como se ha explicado previamente, los valores de las variables han sido obtenidos a partir de muestras tomadas en distintos países lo suficientemente cercanos como para que exista dependencia espacial entre ellos y, además, las variables de estudio son una semimatriz de distancia euclídea (X) y una semimatriz de similitud biótica entre cada par de países (Y). Es importante destacar que el efecto de la proximidad entre países es precisamente el patrón de interés que se modeliza (*distance-decay*) y, por tanto, forma parte del objetivo de estudio. Sin embargo, la dependencia asociada a la construcción por pares sí constituye una violación indeseada de la asunción de independencia de las observaciones que se usan para ajustar el modelo. La dependencia espacial en sí misma, también denominada autocorrelación espacial, ha sido ampliamente estudiada en contextos ecológicos en los que, por ejemplo, mediante el uso de variogramas o covariogramas se estudia el comportamiento espacial de variables simples (Legendre & Legendre, 1998). Sin embargo, es precisamente la naturaleza multivariante de los datos (datos de similitud calculados a partir de la presencia/ausencia de múltiples especies, consideradas variables independientes entre sí) y, por tanto, de la dependencia a pares inherente de la construcción de matrices de similitud, el principal reto estadístico de la modelización de estos datos. Este reto ha sido resuelto para los modelos lineales (Mantel, 1967), pero en la actualidad no existe una vía para el ajuste de modelos no lineales de *distance-decay*. En esta tesis se ha optado por explorar cómo se debe modificar el proceso de ajuste de un GLM, un método utilizado habitualmente en este contexto biológico, para adaptarlo a la falta de independencia de los datos.

Los tipos de dependencia descritos en el párrafo anterior tendrán el efecto de que los datos van a ser más semejantes entre sí de lo esperado. Por ejemplo, es lógico suponer que los valores de la similitud entre las faunas de los Países Bajos y Bélgica serán parecidos a la similitud de las faunas entre Bélgica y Alemania, tanto por su cercanía como porque ambos valores de similitud han sido calculados a partir de un mismo elemento (la fauna de Bélgica está siendo usada en ambas comparaciones). En este contexto, es previsible que la varianza real de los estimadores sea más pequeña que la estimada por el método clásico de IRWLS (*Iteratively Reweighted Least*

Squares (Faraway, 2006)), que es el utilizado por la función `glm()` de R. También que los *p*-valores devueltos por las aproximaciones asintóticas de los contrastes basados en el modelo no sean correctos. Además, algunos de los conjuntos de datos presentan una fuerte heterocedasticidad (más información disponible en sección 5.4.2). Sin embargo, dentro de esta tesis sólo se aborda el problema de la dependencia, y se deja el de la heterocedasticidad para futuros estudios ya que, además, se considera una propiedad intrínseca e inevitable de los datos.

El objetivo principal de este apartado es explorar los métodos de *bootstrap* y, en concreto los denominados “*bootstrap* por bloques”, propuestos por autores previos como métodos para evitar los efectos de la dependencia en la estimación de los parámetros del modelo y sus varianzas en los contrastes de significación que se les puedan aplicar y en el proceso de selección del mejor modelo mediante AIC. En las siguientes secciones de este apartado se explican los métodos *bootstrap* utilizados y se estudian las primeras aproximaciones a esas tres cuestiones mediante un estudio de simulación.

4.4.2 Métodos *bootstrap* por bloques propuestos por autores previos

Bootstrap de bloques móviles según Künsch (1989) y Liu & Singh (1992)

El *bootstrap* por bloques o *bootstrap* de bloques móviles (*Moving Blocks Bootstrap*, MBB) fue propuesto de forma independiente por Künsch (1989) y por Liu & Singh (1992). Este método *bootstrap* genera remuestras a partir de la distribución muestral empírica considerando en ella la existencia de dependencia. Se consigue mediante la definición de bloques de datos que se remuestran conjuntamente, ya que si existe autocorrelación de algún tipo en el mecanismo generador de datos quiere decir que es más probable que una observación de la muestra original sea más semejante a las adyacentes a ella que a las más lejanas, y que su propio valor haya influido en el de esas observaciones adyacentes. Por lo tanto, no sería correcto suponer que en una remuestra las probabilidades de aparición de observaciones adyacentes van a ser independientes entre ellas, porque las frecuencias de esos valores están ligadas en la población y por tanto en las muestras. Este método permite, por tanto, tener en cuenta la dependencia espacial de los datos, pero no la construcción por pares de las matrices de similitud/distancia.

Las remuestras por bloques se obtendrían de la siguiente forma:

1. Se elige el tamaño del bloque, bl , que debe ser un número positivo y entero. A partir de bl se obtiene k como el menor número entero que sea igual o mayor que $\frac{n}{bl}$.
2. Se define $q = n - bl + 1$ bloques a partir de los índices de los valores de la muestra $1, 2, \dots, n$ tomando bl valores consecutivos empezando en el valor i , es decir: B_i para $i = 1, 2, \dots, q$. Se almacenan en forma de columnas de una matriz $bl \times q$.
3. Se obtienen k observaciones de bloques completos de entre los q posibles bloques, con distribución equiprobable. De esta forma, se forma un vector I con los k bloques seleccionados de longitud $bl \times k$, obteniéndolos de la matriz de bloques. Los valores de este vector son índices de los valores muestrales de X e Y .

4. Se definen los vectores remuestra X^* e Y^* seleccionando en la muestra los valores de X y Y que coinciden con las n primeras componentes de I .

Bootstrap por bloques basado en la propuesta de Sherman & le Cessie (1997)

Además, dada la estructura particular de las matrices de distancia, se ha explorado otro tipo de *bootstrap* por bloques siguiendo la idea propuesta por Sherman & le Cessie (1997). Estos autores propusieron un método *bootstrap* para trabajar con datos que se pudieran agrupar con algún factor de bloqueo (por ejemplo, individuos agrupados en familias) y, por tanto, consiste en el remuestreo de los bloques formados por ese factor. Lo utilizaron para construir intervalos de confianza de los coeficientes de regresión.

Los componentes de un vector proveniente de una matriz de distancia que recoge todos los pares de distancia posibles entre una serie de lugares presentan una estructura compleja, en la que observaciones lejanas pueden estar igualmente relacionadas. Por ejemplo, si se han recogido datos de presencia/ausencia de diversas especies en los puntos A, B, C y D, la matriz de distancia tendrá la siguiente forma:

$$\begin{bmatrix} ba \\ ca & cb \\ da & db & dc \end{bmatrix}$$

Y el vector obtenido a partir de ella será $x = ba, ca, da, cb, db, dc$, en el cual se observa que, independientemente de la disposición espacial en la que se encuentren las localidades A, B, C y D, los valores da y dc están relacionados entre sí, ya que para el cálculo de ambos se ha utilizado información de la localidad D. Sin embargo, esta dependencia no es capturada por el método de bloques móviles estándar ya que, al no estar próximas, nunca irían en un mismo bloque. Por lo tanto, se ha querido comprobar si al elegir manualmente los bloques como todos los valores que proceden de una misma localización se podrían obtener mejores resultados que con el método de bloques móviles estándar.

Para obtener remuestras en el método de bloques manuales se han seguido estos pasos:

1. Siendo s el número de localizaciones muestreadas (países), se definen s bloques seleccionando las observaciones de la fila i y la columna i de las matrices de distancia (X) y similitud (Y) ($i = 1, 2, \dots, s$). Estos s bloques (vectores) se almacenan como columnas de una matriz de bloques, de tal forma que se tienen dos matrices de bloques con las observaciones de las variables, una para X y otra para Y .
2. Se arrojan s observaciones de los índices de los bloques de entre los s posibles bloques, con distribución equiprobable.
3. Se forman dos vectores juntando las columnas de las matrices de bloques de X e Y correspondientes a estos índices. Estos dos vectores son los vectores remuestra X^* e Y^*

Este tipo de *bootstrap* permite por tanto obtener estimaciones y distribuciones de estadísticos que normalmente exigen ausencia de correlación de los datos bajo condiciones de dependencia.

En esta aproximación inicial se ha probado el uso de estos dos métodos de *bootstrap* por bloques (móviles y manuales) junto a datos simulados para aproximar la desviación típica y el sesgo de los coeficientes de regresión (siguiendo a Fitzenberger, 1998, que trabajó con la regresión lineal), la distribución de un contraste de significación basado en la *deviance* de los modelos y la frecuencia de selección de distintos modelos según su AIC. Para evaluar su comportamiento con los datos provenientes de una matriz de distancia se ha realizado un estudio de simulación, que se expone a continuación en este capítulo. El código de R correspondiente a todo el estudio de simulación se encuentra en la sección 9.3.

4.4.3 Estudio de simulación

Se han simulado datos con R (R Development Core Team, 2022) a partir del modelo teórico (ecuaciones 4.6 y 4.7) e incluyendo en X la estructura propia de las matrices de distancia, para mantener la correlación espacial y la debida a la construcción de dicha matriz. Los valores del intercepto (a) y la pendiente (b) se han tomado de aplicar la función `glm()` sobre los datos de *ptero* para la zona sur, *distgeo.s* y *ptero.s* (Figura 8). Han sido $a = 0.7575$ y $b = -0.0003564$.

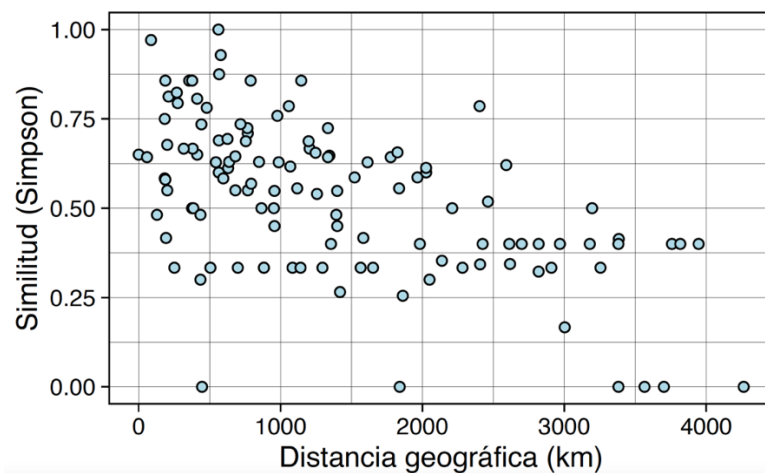


Figura 8: Representación de la similitud frente a la distancia geográfica para el conjunto de datos *ptero.s* (género *Pterostichus* en el sur de Europa).

En el proceso para simular los datos se han seguido los siguientes pasos:

1. Se han generado una serie de localizaciones de muestreo (N) tomando una muestra aleatoria del rango de coordenadas (x, y) presentes en las muestras originales. A partir de esas N coordenadas se ha calculado la semimatriz de distancia euclídea, que es la muestra simulada de X y tiene un tamaño de $n = \frac{N^2}{2} - \frac{N}{2}$.

2. Una vez que se obtiene X , se ha generado una semimatriz de errores normales con media 0 y desviación típica 0.15. Esta desviación se ha elegido por tanteo para aproximar la de *ptero.s*.
3. Finalmente, se ha generado Y a partir de X , la matriz de errores, a y b siguiendo la ecuación 4.6.

En ocasiones los datos simulados son menores que 0 y mayores que 1 (rango del índice de similitud de Simpson) pero se acepta esta desviación porque no es una diferencia muy grande y no debería afectar excesivamente al ajuste del GLM ni al comportamiento del método *bootstrap*.

Se han simulado muestras de $N = 16$ y $N = 20$ y con cuatro tamaños de bloque aplicables al *bootstrap* de bloques móviles. Para cada tamaño de bloque móvil se ha simulado una muestra de $N = 16$ y otra de $N = 20$, y en las ocho muestras resultantes se ha aplicado también el método *bootstrap* de bloques manuales para los análisis. En todos los casos se indican valores de inicio de 0.005 y -0.00001 en la función `glm()` para la estimación de los coeficientes de regresión.

4.4.4 Contraste de significación del modelo basado en la *deviance*

La *deviance* en el contexto de un GLM es una medida que, en el caso de que los datos sean independientes y pertenezcan a la familia de la distribución exponencial, equivale al doble de la diferencia entre la log-verosimilitud del modelo saturado y la log-verosimilitud del modelo ajustado. El modelo saturado es aquel en el que los datos se explican de forma exacta, con tantos parámetros como observaciones haya, y por tanto no resulta informativo; pero se toma de referencia ideal como el mejor ajuste posible. Es decir, la *deviance* en sí es una medida de lo bien que el modelo ajustado explica los datos, considerando previamente que se trata de un modelo válido, y cuanto menor sea su valor, mejor será el ajuste del modelo. Se puede obtener un estadístico que mide la diferencia entre la *deviance* de dos modelos anidados, y sirve para contrastar si dicha diferencia en el ajuste es significativa. En el caso de un GLM en el que la distribución de Y es de tipo normal y los datos son independientes este estadístico tiene una distribución F bajo la hipótesis nula y se rechaza en la cola de la derecha (Faraway, 2006).

Este contraste puede aplicarse también a la comparación entre el modelo nulo (modelo que consiste tan sólo en el intercepto, que es la media de Y) y el modelo ajustado. De esta forma se contrasta si el modelo ajustado, en caso de ser válido, supone realmente una diferencia frente a la ausencia de modelo. Es decir, es un test de significación del modelo. En R este contraste lo realiza la función `anova()` (R Development Core Team, 2022).

Se ha evaluado error de tipo I (la probabilidad de rechazar la hipótesis nula cuando es cierta) para ocho casos de estudio con diferentes combinaciones de número de sitios (N) y tamaño de bloque (bl), con 1000 simulaciones y 2500 réplicas *bootstrap*, todos bajo la hipótesis nula de que los datos de Y no siguen ningún tipo de patrón relacionado con X sino que Y es una variable aleatoria e independiente de X . Para ello se han aleatorizado los datos de Y en cada muestra simulada. Dicho error se ha comprobado en el caso de usar la distribución F asintótica y en el

de utilizar la distribución *bootstrap* del estadístico bajo la hipótesis nula, ambos con un nivel crítico de 0.05. Para construir los tests *bootstrap* se han estimado sus distribuciones bajo la hipótesis nula mediante la aleatorización de las observaciones dentro de los bloques remuestreados de Y , no toda la remuestra. Es decir, se considera que la hipótesis nula es que al margen de compartir la misma estructura espacial y la debida a la construcción de la matriz de distancia, X no ofrece información sobre Y . Si se aleatorizara toda la remuestra, la hipótesis nula sería que X e Y no comparten absolutamente ningún tipo de patrón o estructura, y entonces el patrón causado en X por los dos tipos de correlación de los datos (que debe ser semejante en ambas variables) podría identificarse como explicativo del patrón de Y , oscureciendo el verdadero aporte del modelo.

Para realizar un contraste de significación de los modelos de *distance-decay* con el método *bootstrap* expuesto anteriormente, se han calculado los p-valores como la proporción de veces que el estadístico *bootstrap* ha superado el valor del estadístico de la muestra.

4.4.5 Selección de modelos mediante AIC

El AIC o Criterio de Información de Akaike (ecuación 4.8) se utiliza para seleccionar el modelo (o modelos) que mejor explica los datos de entre un conjunto de modelos candidatos, es decir, elige el mejor modelo en relación al resto:

$$AIC = -2\log(L(\hat{\theta})) + 2K \quad (\text{Ecuación 4.8})$$

siendo $\hat{\theta}$ las estimaciones de los parámetros del modelo y K el número de parámetros estimados del modelo. El primer término del AIC refleja la verosimilitud del modelo y el segundo es una penalización por el número de parámetros que se han estimado para ajustarlo; cuando uno disminuye, el otro aumenta. Cuanto menor sea el AIC, mejor ajuste tendrá el modelo (Burnham & Anderson, 2002).

Si bien es un criterio aplicable a cualquier tipo de modelo, en el contexto de datos dependientes presenta la dificultad de que no se conoce la forma de la verosimilitud. En este estudio de simulación se ha probado a estimar las frecuencias *bootstrap* de selección de dos modelos: el real, el GLM original de tipo exponencial, y el alternativo, un GLM que devuelve una curva *power-law* al introducir la variable distancia en forma logarítmica y mantener como link el logaritmo. Se ha seguido la metodología de Burnham & Anderson (2002) pero sustituyendo el *bootstrap naive* por el de bloques móviles y bloques manuales y utilizando 10000 remuestras *bootstrap*.

4.5 IMPLEMENTACIÓN DE UN MÉTODO PARA CONTRASTE DE HIPÓTESIS EN MODELOS NO LINEALES DE *DISTANCE-DECAY* BASADO EN PERMUTACIONES Y REMUESTREO POR BLOQUES DE LOCALIDAD¹

4.5.1 Una nueva familia de simulaciones con dependencia por pares

En este apartado se propone un contraste de significación basado en un estadístico pseudo- R^2 para modelos exponenciales y *power-law* de *distance-decay* de la similitud entre comunidades y se evalúa su funcionamiento. Para ello se han realizado simulaciones que permiten estimar (1) su error de tipo I, es decir, la probabilidad de rechazar la hipótesis nula cuando es cierta; y (2) su potencia estadística, es decir, la probabilidad de rechazar la hipótesis nula cuando es falsa. Las simulaciones de datos de comunidades biológicas en el espacio se han realizado bajo dos hipótesis (Figura 9):

1. **Nula (H_0):** ausencia de relación entre la similitud de las comunidades y la distancia que las separa (no hay patrón de *distance-decay*). Todas las especies tienen la misma probabilidad de estar presentes en cualquier sitio, y la composición específica de una comunidad no depende de su localización espacial. Por lo tanto, la distancia espacial y la similitud biológica son independientes y tanto el modelo nulo como el de *distance-decay* presentan el mismo ajuste de los datos.
2. **Alternativa (H_1):** existe un *distance-decay* no lineal de la similitud entre comunidades. Las especies están agregadas espacialmente y la composición específica de una comunidad se correlaciona con la de otras comunidades cercanas. Por lo tanto, la distancia espacial es un predictor de la similitud entre comunidades y el modelo de *distance-decay* se ajusta mejor a los datos que el modelo nulo.

Como control del efecto del tamaño muestral de las simulaciones en el error del contraste pseudo- R^2 , se han simulado comunidades en N sitios ($N = \{10, 50, 100\}$), con un total de $S = 150$ especies.

Para la *simulación de comunidades bajo la hipótesis nula*, se han generado N coordenadas espaciales aleatorias a partir de una distribución uniforme con un rango entre 0 y 9000. Después, independientemente de estas coordenadas, se han simulado matrices de presencia/ausencia de especies con dimensiones $N \times S$ (N sitios y S especies) usando una variable aleatoria binomial con diferente probabilidad de éxito para cada especie. Esta probabilidad fue tomada de las probabilidades de una simulación de distribución lognormal para asegurar la existencia de una alta proporción de especies de distribución reducida y pocas de distribución amplia, simulando así un patrón habitual en comunidades naturales (Gaston, 1996).

¹ Una versión adaptada y en inglés de este capítulo ha sido publicada como: Martínez-Santalla S¹, Martín-Devasa R^{1,2}, Gómez-Rodríguez C¹, Crujeiras RM³, Baselga A. 2022¹. Assessing the non-linear decay of community similarity: permutation and site-block resampling significance tests. *Journal of Biogeography* (WILEY, Online ISSN:1365-2699) 49: 968-978. <https://doi.org/10.1111/jbi.14351>

¹ CRETUS, Universidade de Santiago de Compostela; ² University of Helsinki, ³ CITMaga, Universidade de Santiago de Compostela

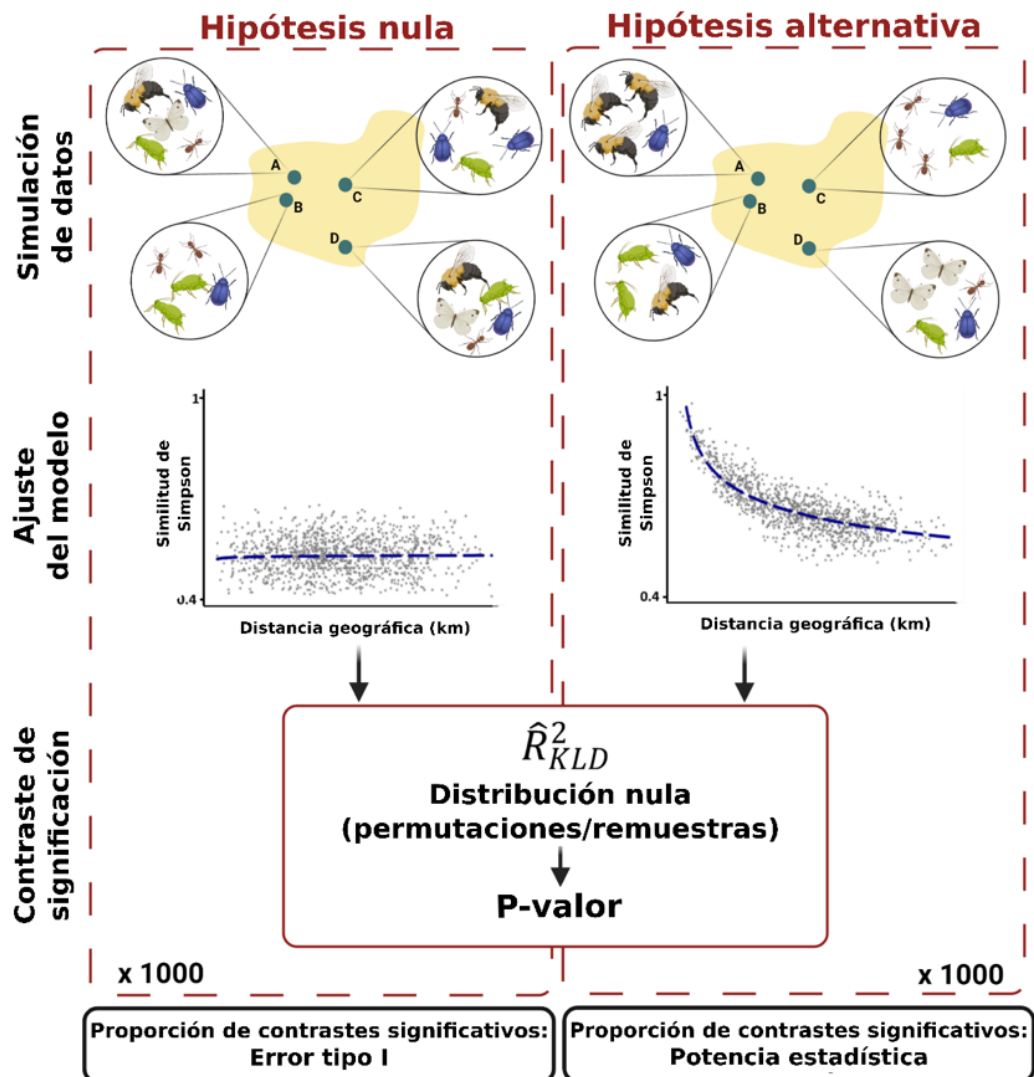


Figura 9. Esquema de las simulaciones empleadas en este estudio para evaluar el funcionamiento del contraste de la pseudo- R^2 (R_{KLD}^2) como contraste de significación para modelos exponenciales y *power-law* del *distance-decay* de la similitud. Imagen creada con BioRender.com, con siluetas del programa.

Para la *simulación de comunidades bajo la hipótesis alternativa*, hemos generado campos aleatorios gaussianos (*Gaussian Random Fields*, GRF) usando la función `grf()` de la librería “`geoR`” v.1.7-5.2.2 (Ribeiro Jr & Diggle, 2020) en R con una función de covarianza exponencial (ecuación 4.9). La función `grf()` nos permitió simular S especies como variables gaussianas espacialmente correlacionadas con valores en N puntos espaciales aleatorios. La covarianza exponencial del GRF fuerza el descenso de la semejanza entre variables gaussianas (especies) observadas en dos puntos de muestreo a medida que aumenta la distancia que las separa. Esta función de covarianza viene dada por:



$$C(h) = \sigma^2 e^{\left(-\frac{h}{\varphi}\right)} \quad (\text{Ecuación 4.9})$$

En donde φ es el parámetro de rango que controla la intensidad de la autocorrelación espacial, σ^2 es la meseta parcial, h es la distancia entre dos puntos y $C(h)$ es la covarianza entre las variables en esos dos puntos. Para valores altos de φ , la similitud entre la composición de las comunidades descenderá más lentamente con el aumento de la distancia y viceversa. Para conseguir matrices de presencia/ausencia de especies $N \times S$, que son las requeridas por numerosos índices de similitud, como el índice de Simpson, hemos transformado las variables gaussianas de especie a datos binomiales usando diferentes probabilidades de éxito para cada especie. Dicha probabilidad fue tomada de los valores de una distribución lognormal simulada, para asegurar, al igual que en el caso de la hipótesis nula, la existencia de una alta proporción de especies de distribución reducida y pocas especies de distribución amplia. Bajo esta hipótesis se han simulado tres escenarios con diferente grado de autocorrelación espacial en la composición específica de los sitios muestreados ($\varphi = \{300, 5000, 40000\}$), y $\sigma^2 = 1$.

4.5.2 Generación de matrices de similitud/distancia y ajuste de modelos *distance-decay*

Una vez se han obtenido las tablas de datos de presencia/ausencia y las coordenadas espaciales para cada conjunto de datos (datos simulados o datos reales de distribución de coleópteros europeos), se han calculado las matrices triangulares (semimatrices) de la distancia espacial y la similitud entre comunidades. Estas matrices se han usado posteriormente para ajustar los modelos de *distance-decay* y testar su significación usando el contraste de la pseudo- R^2 .

Para las semimatrices de distancia (D_x), se han calculado las distancias euclídeas entre las coordenadas (o centroides) de los sitios, tal y como se explica en el apartado 4.1.1. Las semimatrices de similitud (Sim_y) se han obtenido aplicando el índice de Simpson con la función `beta.pair()` de la librería “betapart” de R (Baselga & Orme, 2012; Baselga et al., 2023).

Como se ha explicado anteriormente, se ha optado por modelos GLMs, en vez de LMs con variables transformadas, para ajustar los modelos de *distance-decay* ya que presentan la ventaja de que sus parámetros de regresión se estiman considerando la similitud sin transformarla logarítmicamente. Por lo tanto, los residuos utilizados para calcular el estadístico pseudo- R^2 en un GLM están sin transformar y se evita el posible sesgo causado por una transformación directa de la similitud y la modificación y/o eliminación de los valores 0 de similitud que sería necesaria para la transformación logarítmica de dicha similitud (Millar et al., 2011; O'Hara & Kotze, 2010). Por tanto, se han ajustado GLMs exponenciales eligiendo el logaritmo como función link y errores gaussianos e, independientemente, *GLMs power-law* en los cuales se mantiene la función link y se transforma logarítmicamente la distancia geográfica. Para cuantificar la devianza total, se han construido GLMs nulos mediante la regresión de la similitud sobre su propia media. Todos los modelos fueron calculados usando la función `glm2()` de la librería ‘glm2’ de R, v.1.2.1 (Marschner, 2011).

4.5.3 El contraste de la pseudo-R²: R_{KLD}^2

Para testar la significación de modelos exponenciales y *power-law* de *distance-decay*, se ha usado la medida pseudo-R² basada en la divergencia de Kullback-Leibler descrita por Cameron & Windmeijer (1997). La elección de un estadístico pseudo-R² para este contraste de significación se basa en que (1) su hipótesis nula es la independencia entre similitud y distancia, (2) se puede calcular para modelos exponenciales y *power-law* y (3) estima la proporción de la *deviance* del modelo nulo que es explicada por el modelo ajustado, sirviendo así como medida de varianza explicada. Su fórmula es:

$$R_{KLD}^2 = \frac{D(y, \hat{\mu}_0) - D(y, \hat{\mu}_1)}{D(y, \hat{\mu}_0)} \quad (\text{Ecuación 4.10})$$

En donde $D(y, \hat{\mu}_0)$ es la *deviance* del modelo nulo (modelo compuesto sólo por el intercepto de la similitud con la distancia) y $D(y, \hat{\mu}_1)$ es la *deviance* del modelo de *distance-decay*. En el contexto de un GLM, la *deviance* de un modelo es una medida de cuán bien se ajusta el modelo a los datos y , por tanto, un valor menor de *deviance* implica un mejor ajuste del modelo (McCullagh & Nelder, 1989), como se ha explicado anteriormente. Por lo tanto, un valor mayor de R_{KLD}^2 implica un ajuste mejor a los datos por parte del modelo de *distance-decay* respecto al modelo nulo. Para estimar el valor del estadístico en nuestros datos, \hat{R}_{KLD}^2 , calculamos los valores de *deviance* de los modelos ajustados usando la función *deviance* de la librería ‘stats’ v.3.5.0 (R Development Core Team, 2022).

Para datos en los que no hay dependencia a pares, este contraste es equivalente a realizar un test de la F (Hastie & Pregibon, 1993). Sin embargo, para datos de similitud y distancia, la distribución del estadístico bajo la hipótesis nula no se conoce porque las observaciones de las variables no son independientes (Legendre et al., 1994). Por consiguiente, para obtener el p-valor del contraste R_{KLD}^2 se ha estimado su distribución bajo la hipótesis nula empleando métodos de remuestreo. La clave para incorporar la dependencia a pares presente en las variables de similitud y distancia en estos métodos es considerar todos los valores de similitud/distancia relacionados con un sitio particular como la unidad básica del remuestreo, como se ha explicado anteriormente. Estos conjuntos de datos (todos los valores de similitud/distancia en los que una comunidad/sitio determinado está implicado) serán denominados en esta tesis “bloques de localidad”. Así, se ha estimado la distribución nula de R_{KLD}^2 (Figura 9) usando (i) permutaciones por bloques de localidad, que permutan conjuntamente todos los valores de similitud relacionados con una sola localidad (bloque de localidad), y (ii) remuestreo por bloques de localidad, que toma muestras con reemplazamiento de bloques de localidad en las matrices de similitud y distancia y luego aleatoriza los valores dentro de cada bloque de la matriz de similitud (Figura 10).

Permutaciones de bloques de localidad

Se ha empleado el siguiente procedimiento para obtener permutaciones de los sitios de muestreo en la matriz de similitud (Sim_y), lo cual hace que similitud y distancia (D_x) sean independientes entre sí (Figura 10):

1. Reordenamos Sim_y por filas y columnas (Sim_y^*) sin modificar D_x .
2. Ajustamos un GLM a Sim_y^* y D_x , y un GLM nulo a Sim_y^* ; después calculamos ambos valores de *deviance* y R_{KLD}^2 .
3. Repetimos 1-2 para B permutaciones.

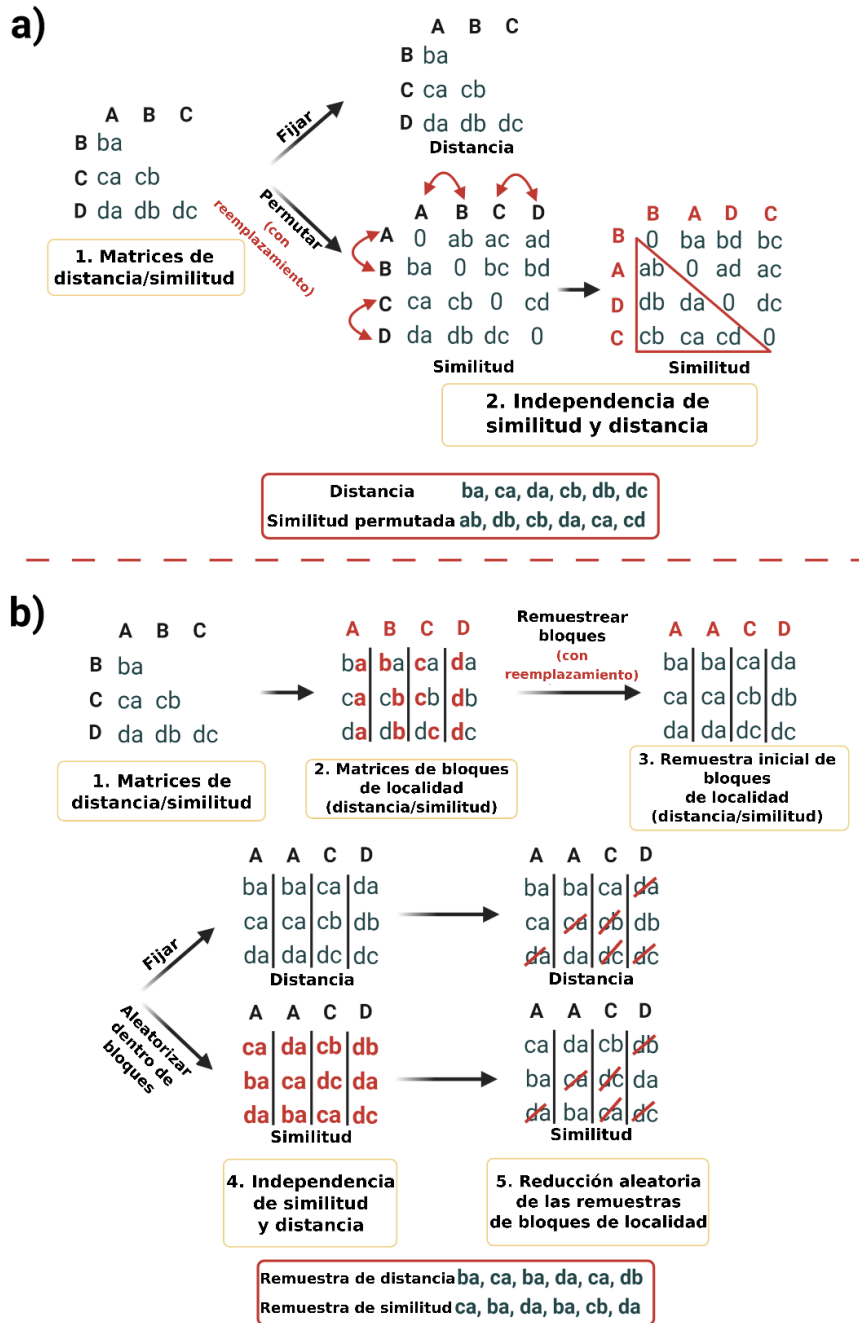


Figura 10. Representación esquemática de los procedimientos de permutación (a) y remuestreo por bloques de localidad (b) para estimar la distribución nula del estadístico pseudo- R^2 (R_{KLD}^2) incorporando la estructura a pares de los datos de *distance-decay*. Creado con BioRender.com



Remuestreo de bloques de localidad

Aquí, en contraste con el procedimiento de permutación, se han obtenido con reemplazamiento las remuestras de los bloques de localidad (todas las similitudes o distancias construidas a partir de un sitio particular) y se han guardado en una matriz de bloques de similitud o distancia. Este método es un desarrollo del *bootstrap* por bloques manuales presentado en el apartado 4.4.2, asegurando la preservación del tamaño de las matrices de remuestreo, el cual no se había contemplado previamente y es fundamental para el cálculo del p-valor.

Para obtener remuestras bajo la hipótesis nula, se han aleatorizado los valores dentro de cada bloque en la matriz de similitud. El procedimiento es el siguiente (Figura 10b):

1. Para N localidades, se han definido bloques de localidad seleccionando los valores de D_x y Sim_y en los que estuviese implicada la localidad dada. Se han almacenado los N bloques como las columnas de dos matrices de bloques, una por variable (D_x y Sim_y).
2. Se han obtenido N remuestras de bloques (columnas de las matrices de bloques) con reemplazamiento, asegurando que los mismos bloques se muestreasen para D_x y Sim_y , y se han almacenado en dos matrices de remuestras.
3. Para forzar que Sim_y sea independiente de D_x (hipótesis nula del contraste) sin cambiar la estructura de bloques, se ha aleatorizado la matriz de remuestras de Sim_y por columnas (se aleatorizaron los valores de cada bloque remuestreado en Sim_y).
4. Inicialmente las matrices de remuestras se construyeron con observaciones duplicadas ($N * [N - 1]$ valores) respecto a las matrices originales de similitud y distancia ($N * [N - 1]/2$ valores) debido a que cada bloque de localidades incluye $N - 1$ valores. En este paso se ha reducido la dimensión de las matrices de remuestras a la original de las variables muestreando $N * [N - 1]/2$ valores de ambas matrices de remuestras. Dichos valores se han almacenado en los vectores de remuestras D_x^* y Sim_y^* .
5. Se ha ajustado un GLM a D_x^* y Sim_y^* , y un GLM nulo a Sim_y^* , y a continuación se ha calculado los valores de *deviance* y R_{KLD}^2 .*.
6. Se han repetido los pasos 1 – 5 B veces.

Una vez se hubo generado la distribución nula estimada por cualquiera de estos dos métodos, se ha obtenido el p-valor del contraste como la proporción de veces en las que R_{KLD}^2 * fue mayor que \hat{R}_{KLD}^2 (valor del estadístico en la muestra original).

4.5.4 Evaluación del funcionamiento del contraste de significación

Para evaluar el desempeño de R_{KLD}^2 (ecuación 4.10) en combinación con los métodos de permutaciones y remuestreo por bloques de localidad para testar la significación de patrones de *distance-decay*, se han utilizado datos simulados (tal como se explica en la sección 4.5.1) para estimar su probabilidad de (i) rechazar la hipótesis nula cuando es cierta (error tipo I) y (ii) rechazar la hipótesis nula cuando es falsa (potencia estadística = 1 – error de tipo II). Para todos

los escenarios de simulación el nivel de significación elegido fue de 0.05. Para comparar el contraste propuesto con otros contrastes de referencia, también se estimó el error de tipo I del test de Mantel utilizando la función `mantel()` de la librería ‘vegan’ (Oksanen et al., 2017) y el del test de la F aplicándolo como un ANOVA de tipo I mediante la función `anova()` de la librería ‘stats’. El test de Mantel se utiliza para evaluar la significación de la correlación entre matrices de datos cuando la relación entre ellas es lineal y, por tanto, no sería a priori idóneo para modelos *distance-decay*, ya que autores previos han mostrado que la relación entre la similitud biótica y la distancia no suele ser lineal (Nekola & McGill, 2014). Por su parte, el test de la F asume la independencia de los valores de disimilitud, por lo que esperamos que tenga un error de tipo I inflado. Estas comparaciones han permitido, por tanto, evaluar si los procedimientos de permutaciones o remuestreo por bloques que aquí se proponen son capaces de corregir la inflación del error de tipo I del test de la F . Por otro lado, también se ha comparado la potencia estadística del test de significación propuesto (basado en permutaciones o remuestreo por bloques) con las del test de la F o el test de Mantel. Dado que el test propuesto permite la utilización de modelos no lineales, esperamos una mayor potencia que para el test de Mantel.

Para estimar el error de tipo I y la potencia estadística ($1 - \text{error de tipo II}$), se han simulado datos bajo las hipótesis nula y alternativa, respectivamente. A continuación, se ha aplicado el contraste de la pseudo- R^2 (R_{KLD}^2) a esos datos, utilizando (i) 2500 permutaciones/remuestras de bloques de localidad, (ii) el test de Mantel con 2500 permutaciones y (iii) el test de la F . Este procedimiento de simulación se ha repetido 1000 veces (Figura 9). El error de tipo I esperado fue el mismo que el nivel de significación, 0.05, lo cual indicaría la correcta calibración del estadístico. En el caso de la potencia estadística, se espera que la adecuación del modelo elegido a la relación de *distance-decay* simulada condicione el rendimiento de R_{KLD}^2 . En los escenarios de $\varphi = 5000$ y $\varphi = 40000$, la disminución de la similitud ocurre a lo largo de todo el rango de distancia espacial simulado y debería ser fácilmente detectada por todos los modelos estudiados (exponencial, *power-law* y el modelo lineal del test de Mantel). En cambio, en el escenario con una correlación espacial más débil ($\varphi = 300$) existe un descenso inicial muy rápido de la similitud, pasando de valores de 0.9 a 0.4 pero, sin embargo, en la mayor parte del rango de distancias (el 95%, aproximadamente) el patrón es similar a la hipótesis nula. Por lo tanto, debería ser más difícil de detectar con el contraste de la pseudo- R^2 que los otros dos escenarios de *distance-decay* simulados y, por tanto, permite comprobar el funcionamiento del contraste con patrones de *distance-decay* más débiles (ver patrones de *distance-decay* de ejemplo en Figura 11).

Todas las simulaciones y el análisis estadístico se han realizado en R v.3.5.0 (R Development Core Team, 2022). Las simulaciones se han ejecutado en el superordenador FINISTERRAE, un servicio facilitado por el Centro de Supercomputación de Galicia (CESGA).

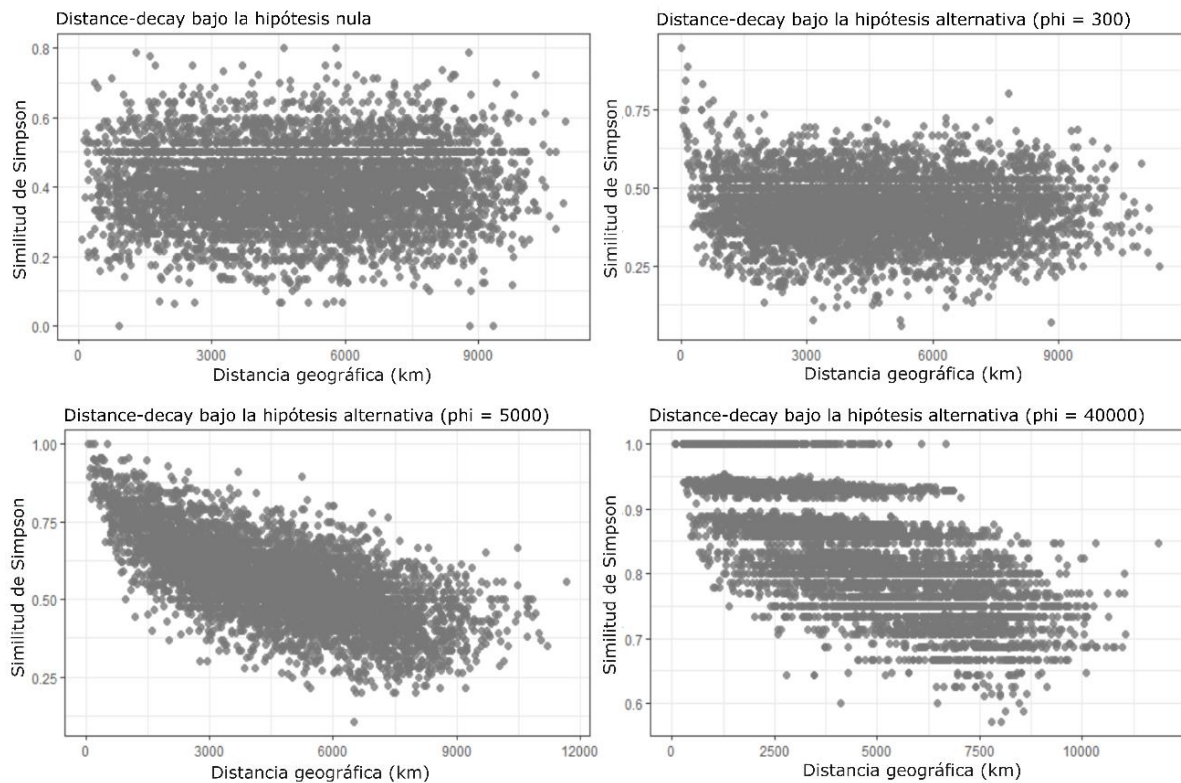


Figura 11. Datos simulados para 100 localidades ($N = 100$) bajo la hipótesis nula (no hay *distance-decay*) y distintas hipótesis alternativas, variando entre autocorrelación espacial leve a fuerte en la composición específica de las comunidades ($\varphi = 300, 5000, 40000$).

4.5.5 Ejemplo de aplicación a datos reales

Finalmente, se ha ilustrado la aplicación del contraste de la pseudo- R^2 asociada a permutaciones en los conjuntos de datos reales de coleópteros europeos descritos en la sección 4.1. Para estos conjuntos de datos, y siempre separando las regiones del sur y norte de Europa, se han comparado los ajustes de modelos exponenciales y *power-law* de *distance-decay*, usando el valor estimado de R_{KLD}^2 para mostrar las diferencias entre ellos. Además, se ha incluido en el análisis una nueva función sigmoideal propuesta recientemente para el modelado de los patrones de *distance-decay*: la función Gompertz (Martín-Devasa et al., 2022b). Esta función es idónea cuando similitud no desciende marcadamente con la distancia entre localidades próximas, dando lugar a una pequeña meseta inicial en los patrones de *distance-decay*. La fórmula de la función Gompertz (Gompertz, 1833) es, en sí misma, una generalización del modelo exponencial negativo:

$$f(x) = ae^{-be^{-cx}} \quad (\text{Ecuación 4.11})$$

donde a es la asíntota superior, b es un parámetro de posición, c la tasa de cambio y x la distancia entre sitios.

Los modelos de *distance-decay* se ajustaron mediante regresiones no lineales y se estimó la significación de los modelos usando el método de permutaciones por bloques descrito en las secciones previas. La similitud en la composición de especies entre pares de sitios se calculó con el índice de similitud por pares de Simpson, $1-\beta_{\text{sim}}$ (Baselga, 2010b; Koleff et al., 2003), utilizando la función `beta.pair()` de la librería “betapart” (Baselga & Orme, 2012; Baselga et al., 2023). Las distancias espaciales se calcularon como la distancia euclídea entre pares de sitios. Las funciones exponencial negativa, *power-law* y Gompertz (Martín-Devasa et al., 2022b) se ajustaron utilizando la función `nlsLM()` de la librería “minpack.lm” (Elzhov et al., 2016), que realiza una regresión de mínimos cuadrados no lineal utilizando el algoritmo de Levenberg-Marquardt (Levenberg, 1944; Marquardt, 1963). Se ha optado aquí por las regresiones no lineales mediante `nlsLM()` en vez de GLM para poder ajustar también modelos Gompertz, ya que recientemente se ha mostrado que es una función adecuada para ajustar modelos de *distance-decay* cuando los tamaños de las distribuciones de las especies son muy amplios en relación al área de estudio (Martín-Devasa et al., 2022b). El ajuste de cada tipo modelo se comparó empíricamente en base al valor de R_{KLD}^2 . El ajuste de los modelos y el test de permutaciones por bloques de localidad (resultado de esta tesis) han sido implementados en la función `decay.model()` de la librería “betapart” (Baselga & Orme, 2012; Baselga et al., 2023) de R. A los p-valores, calculados con 1000 permutaciones, se les ha aplicado la corrección de Benjamini & Yekutieli (2001). La representación gráfica de los modelos se realizó con la función `plot.decay()` de la citada librería.

5 RESULTADOS

5.1 ESTUDIO NO PARAMÉTRICO DE LAS DIFERENCIAS NORTE-SUR EN LOS PATRONES DE *DISTANCE-DECAY* DE COLEÓPTEROS EUROPEOS

El estudio no paramétrico de los patrones de *distance-decay* en el norte y sur de Europa mostró diferencias en los patrones de *distance-decay* en la mayoría de los taxones estudiados, en consonancia con las observaciones realizadas anteriormente por Gómez-Rodríguez & Baselga (2018). De esta forma, se verifica que la división de datos utilizada en esta tesis es apropiada y que los resultados de estudios previos que apoyaban esta división no estaban sesgados por el propio modelo.

En la Tabla 1 y la Figura 12 se recogen los resultados de este análisis, que muestran que sólo se puede asumir que la hipótesis nula de igualdad de las curvas de regresión es cierta para ochte¹ (p-valor = 1). Respecto al contraste de paralelismo, se acepta que las curvas ajustadas para cada región son paralelas en los casos de apion, ceram, lamin, leptu, ochte, otior, phyll, polyd y ptero (9 de 21 conjuntos de datos). Es decir, en estos casos la similitud descende de la misma forma en ambas regiones, aunque parte de valores iniciales de similitud distintos (salvo para ochte).

Por lo tanto, se observa que la similitud entre las faunas de países suele ser menor en el sur, y su descenso con la distancia mucho más marcado en la mayoría de los casos. No obstante, aparecen irregularidades en algunos conjuntos de datos: en algún caso la curva estimada aumenta bruscamente en distancias mayores (como en chaet e hydra). Al margen del posible error de muestreo, esta circunstancia podría reflejar el resultado de la historia biogeográfica de la distribución de las especies y comunidades (por ejemplo, las ubicaciones comparadas han sido refugios climáticos y por eso algunas de las especies que en ellas viven son las mismas) y/o características del hábitat (hábitats similares podrían albergar a las mismas especies si estas logran colonizar ambas zonas). Esta explicación también se podría aplicar a los valores aislados de similitud 0 que aparecen en scoly y silph. Dichos puntos parecen tener gran influencia en la estimación de la curva, de tal forma que parece que si se eliminase el que aparece en los datos de scoly, las curvas de la zona norte y la zona sur coincidirían. Sin embargo, se ha decidido no eliminar ningún dato en el análisis para evitar cualquier posible pérdida de información, por lo que la división entre la zona norte y la zona sur se ha mantenido para ambos conjuntos de datos. Por tanto, se ha decidido analizar por separado los grupos norte y sur en todos los casos.

¹ Para facilitar la lectura, se usarán los siguientes acrónimos de los taxones en esta tesis: *Amara* (amara), Apionidae (apion), *Carabus* (carab), Cerambycinae (ceram), *Chaetocnema* (chaet), Chrysomelinae (chrys), Cryptocephalinae (crypt), Hydraenini (hydra), Lamiinae (lamin), Lepturinae (leptu), Limnebiini (limne), Ochtebiini (ochte), *Otiorhynchus* (otior), Phyllobiini (phyll), Polydrusini (polyd), Pselaphinae (psela), *Pterostichus* (ptero), Scarabaeidae (scara), Scolytinae (scoly), Silphidae (silph) y *Trechus* (trech).

Tabla 1. P-valores para el ANCOVA no paramétrico del índice de similitud de Simpson sobre la distancia euclídea. Corregidos según el método de Benjamini & Yekutieli (2001).

Taxón (acrónimo)	Contraste de igualdad	Contraste de paralelismo
amara	0	0
apion	0	1
carab	0	0
ceram	0	0.0836
chaet	0	0.0015
chrys	0	0
crypt	0	0
hydra	0	0
lamin	0	0.2589
leptu	0	0.4765
limne	0	0.0001
ochte	1	1
otior	0	0.3711
phyll	0	0.3258
polyd	0	0.3784
psela	0	0
ptero	0	0.1284
scara	0	0
scoly	0.0077	0.0028
silph	0	0
trech	0	0.0002

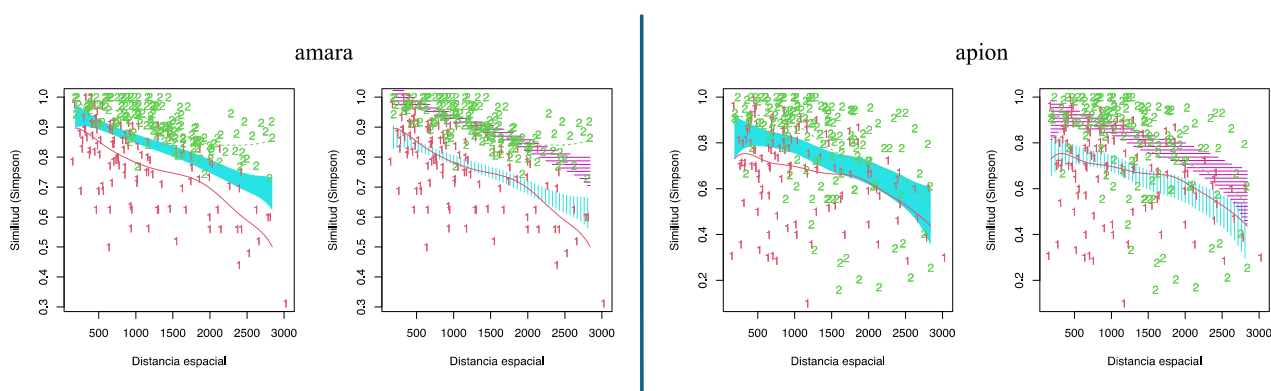


Figura 12. Bandas de referencia para los contrastes de igualdad y paralelismo (primera y segunda gráfica de cada par, respectivamente) de las curvas de regresión de la similitud biológica, medida con el índice de Simpson, y la distancia geográfica. 1: datos del sur de Europa; 2: datos del norte de Europa.

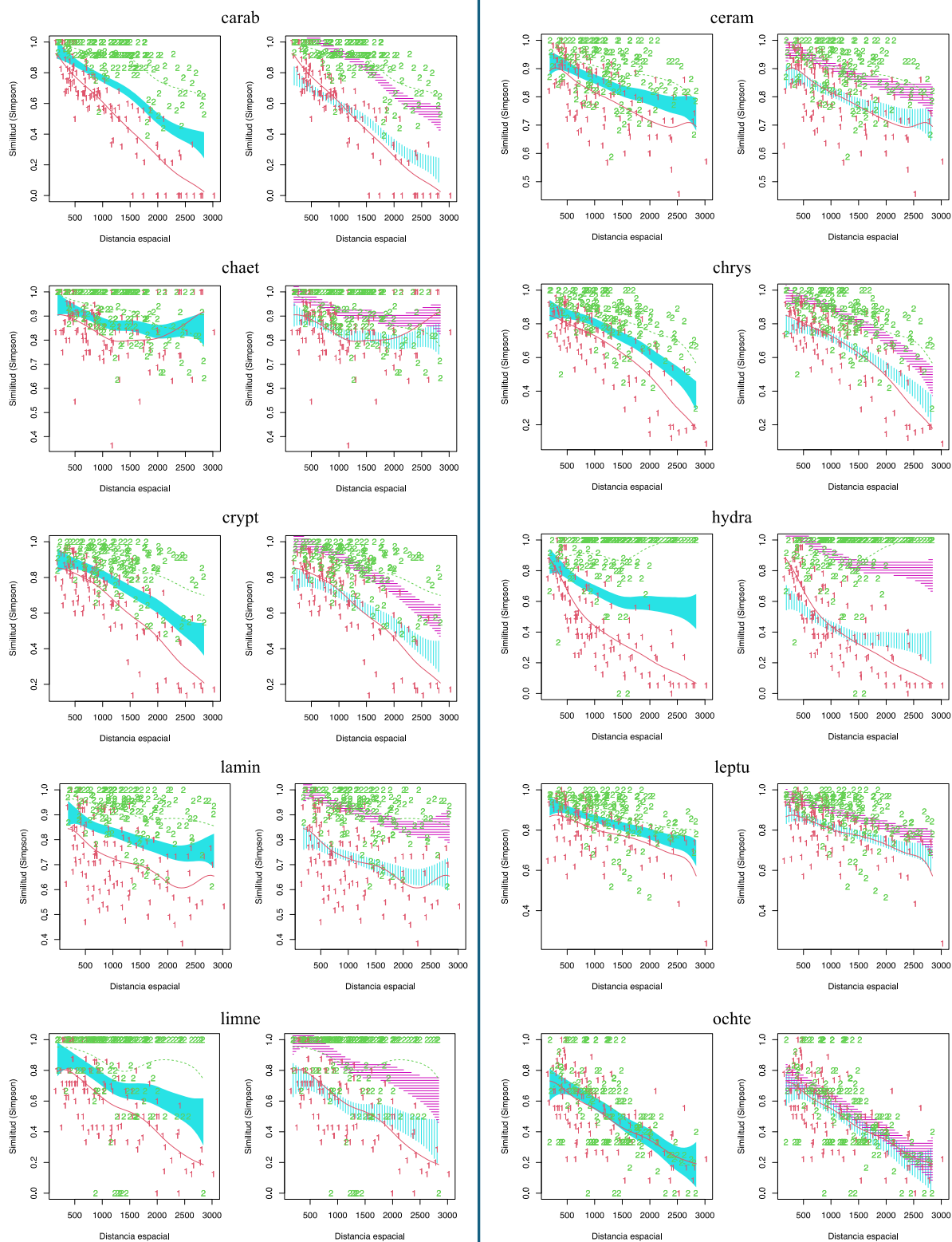


Figura 12 (cont.). Bandas de referencia para los contrastes de igualdad y paralelismo (primera y segunda gráfica de cada par, respectivamente) de las curvas de regresión de la similitud biológica, medida con el índice de Simpson, y la distancia geográfica. 1: datos del sur de Europa; 2: datos del norte de Europa.

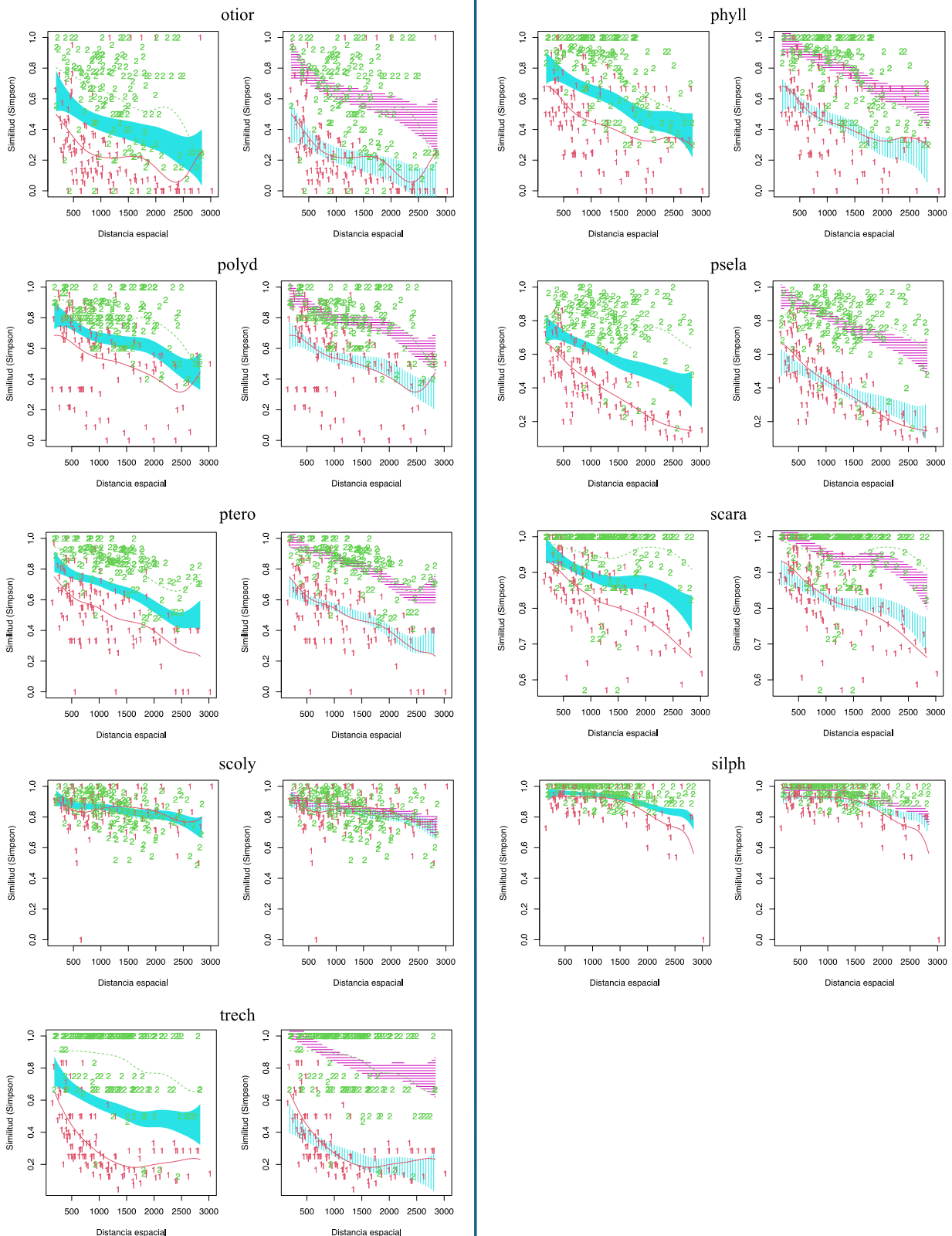


Figura 12 (cont.). Bandas de referencia para los contrastes de igualdad y paralelismo (primera y segunda gráfica de cada par, respectivamente) de las curvas de regresión de la similitud biológica, medida con el índice de Simpson, y la distancia geográfica. 1: datos del sur de Europa; 2: datos del norte de Europa.

5.2 ESTRUCTURA DEL ERROR EN MODELOS PARAMÉTRICOS (GLM) DE *DISTANCE-DECAY*

Una vez determinado que, incluso sin asunciones paramétricas, los patrones de *distance-decay* son diferentes entre en norte y el sur de Europa, se exploró si una estructura del error con distribución Gamma en los GLMs usados para ajustar dichos patrones ofrecía ventajas sobre una estructura del error de tipo Gaussiano. Se ha considerado la función Gamma como una posible alternativa a la distribución Gaussiana debido a la diversidad de formas que puede tomar y que es una función continua.

Por tanto, para cada taxón de coleópteros y área geográfica (norte / sur de Europa) por separado, se ajustaron dos GLMs con link logarítmico, pero uno de ellos con error de estructura Gaussiana, y el otro con error de distribución Gamma y, posteriormente, se compararon empíricamente las distribuciones de dichos estadísticos (intercepto, pendiente y pseudo- R^2).

En la mayoría de los conjuntos de datos no se observaron diferencias significativas en los parámetros del modelo (intercepto, pendiente y pseudo- R^2) entre los modelos gaussiano y gamma (Figuras 13a y 13b). La divergencia entre el ajuste de ambos modelos parece ser mayor cuando la similitud disminuye más rápido, como se puede ver en las Figuras 13a y 13b. Sólo un conjunto de datos, *Carabus* en el sur de Europa, mostró diferencias significativas en el intercepto, la pendiente y la varianza explicada ($p = 0.011$, $p = 0.029$ y $p < 0.005$, respectivamente). En todo caso, incluso aunque las diferencias en pseudo- R^2 no son significativas en el resto de las comparaciones, la varianza explicada siempre es mayor en el modelo con estructura Gaussiana en todos los conjuntos de datos (Tabla 2, Figuras 13a y 13b). Además, el modelo gamma puede proporcionar resultados no deseados, con valores de intercepto grandes que superan notablemente el valor 1 como se observa, por ejemplo, en *Ochtebiini* en el norte de Europa (Figura 13a).

Podemos, por tanto, concluir que los GLMs con estructura de error Gamma no ofrecen ventajas significativas, sino, al contrario, los modelos con estructura de error Gaussiana tienden a tener mejores ajustes a los datos, por lo que resultan preferibles.

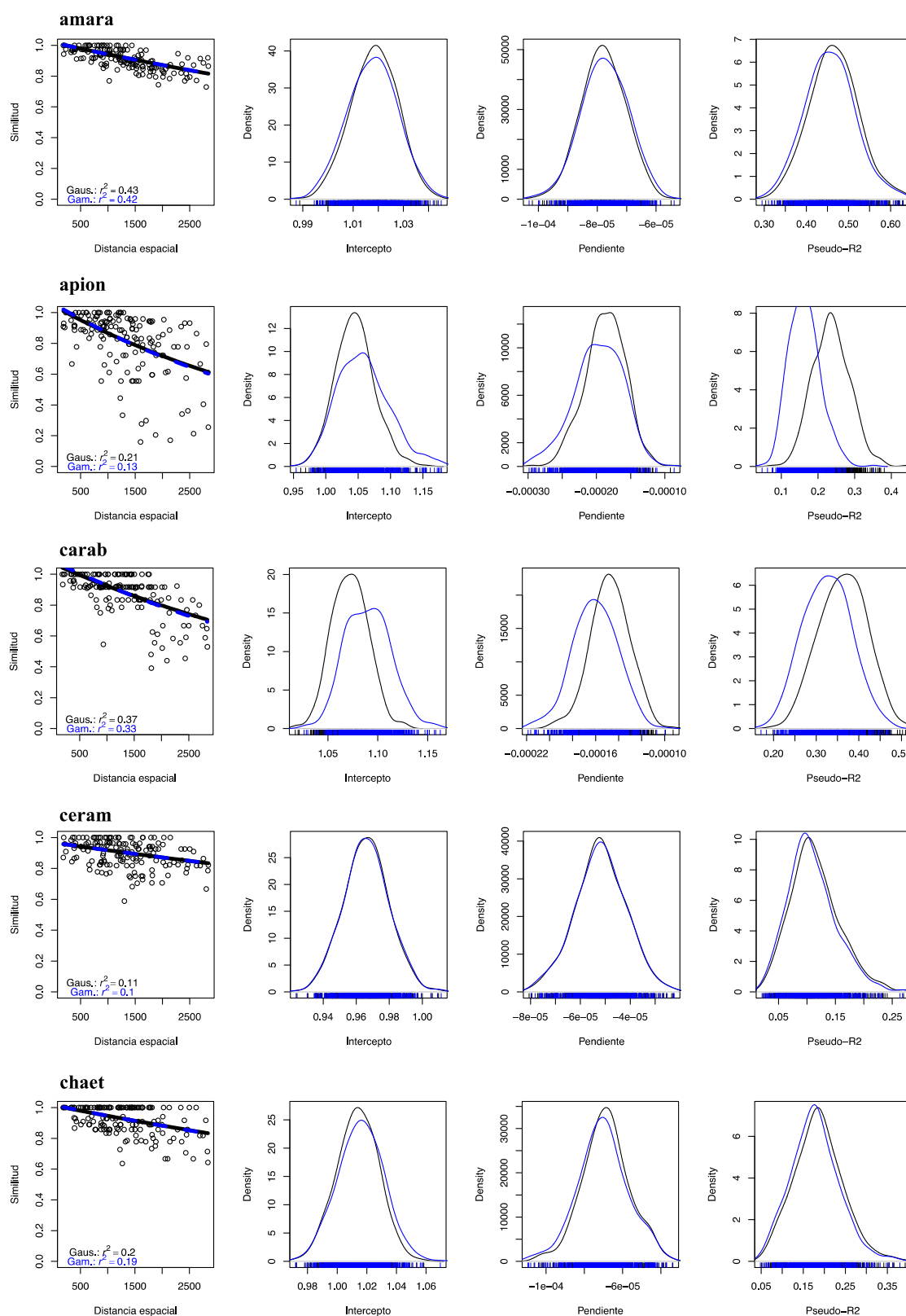


Figura 13a. Para cada taxón de estudio en el norte de Europa, se muestran (i) las curvas de ajuste de la relación entre la distancia geográfica y la similitud biótica considerando un error de distribución Gaussiana (negro) y un error de distribución Gamma (azul), (ii) la curva de distribución de los interceptos, (iii) la curva de distribución de las pendientes y (iv) la distribución de valores de pseudo- R^2 estimadas mediante *bootstraps* de 500 remuestras.

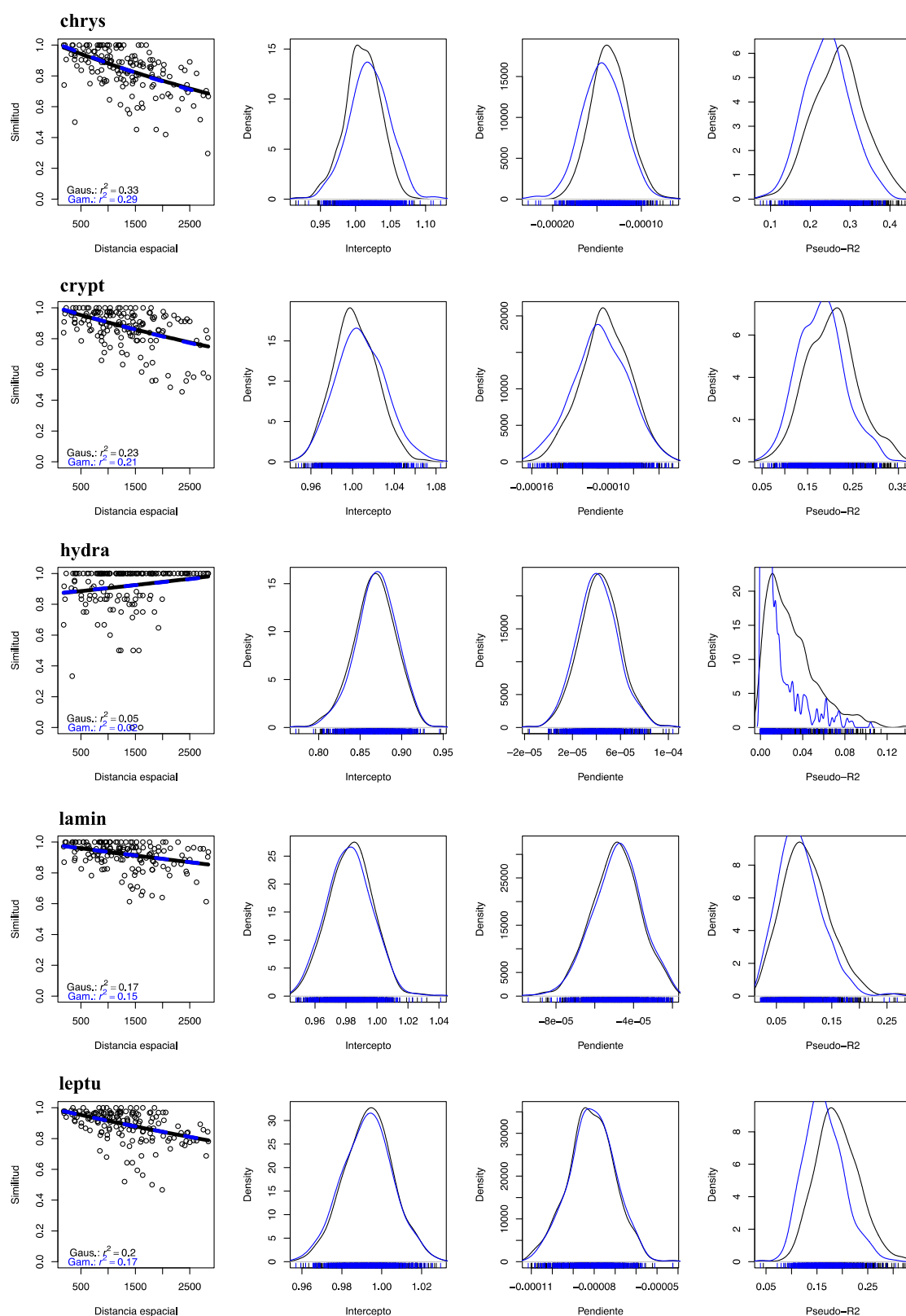


Figura 13a (cont.). Para cada taxón de estudio en el norte de Europa, se muestran (i) las curvas de ajuste de la relación entre la distancia geográfica y la similitud biótica considerando un error de distribución Gaussiana (negro) y un error de distribución Gamma (azul), (ii) la curva de distribución de los interceptos, (iii) la curva de distribución de las pendientes y (iv) la distribución de valores de pseudo- R^2 estimadas mediante *bootstraps* de 500 remuestras.

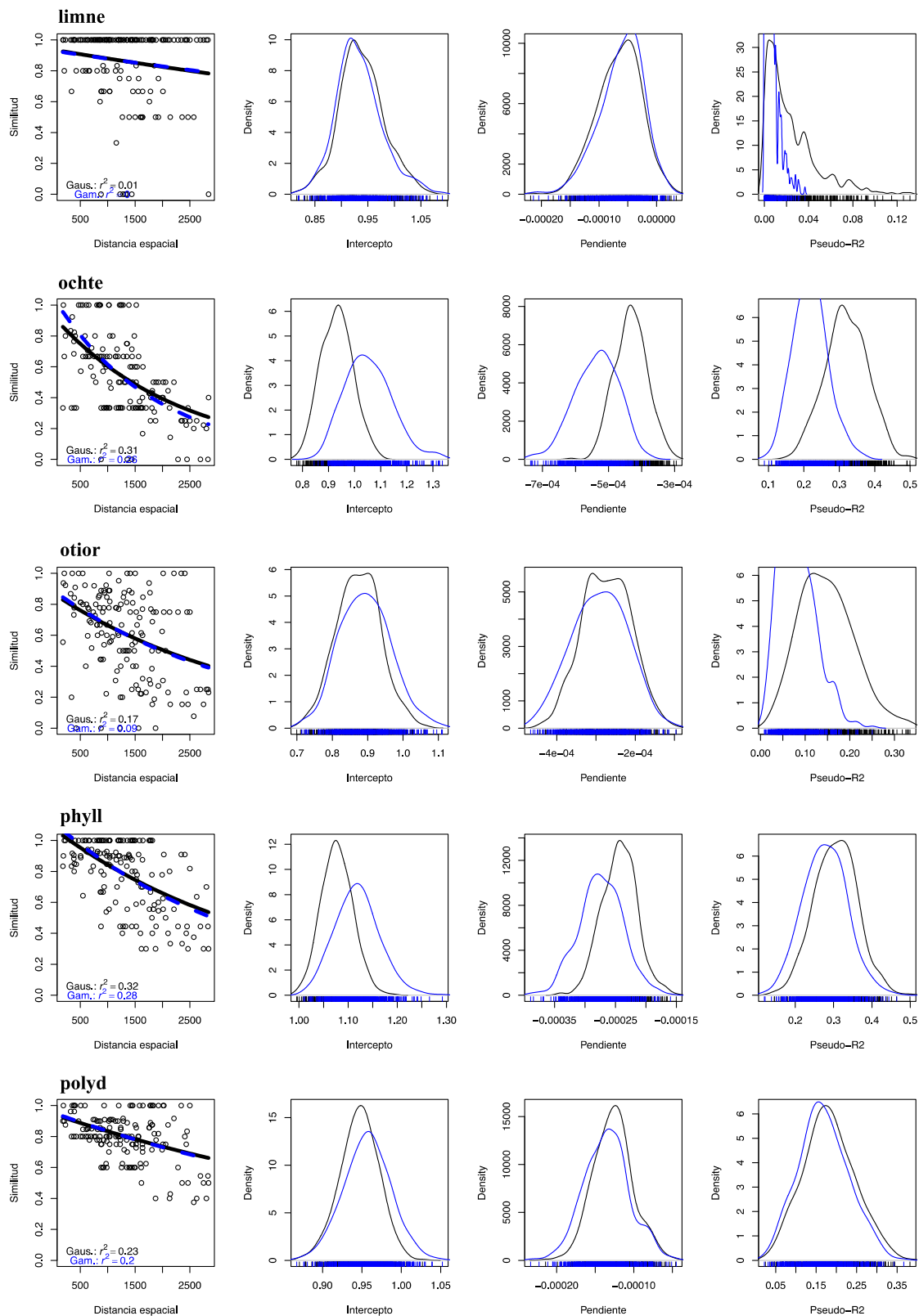


Figura 13a (cont.). Para cada taxón de estudio en el norte de Europa, se muestran (i) las curvas de ajuste de la relación entre la distancia geográfica y la similitud biótica considerando un error de distribución Gaussiana (negro) y un error de distribución Gamma (azul), (ii) la curva de distribución de los interceptos, (iii) la curva de distribución de las pendientes y (iv) la distribución de valores de pseudo- R^2 estimadas mediante *bootstraps* de 500 remuestras.

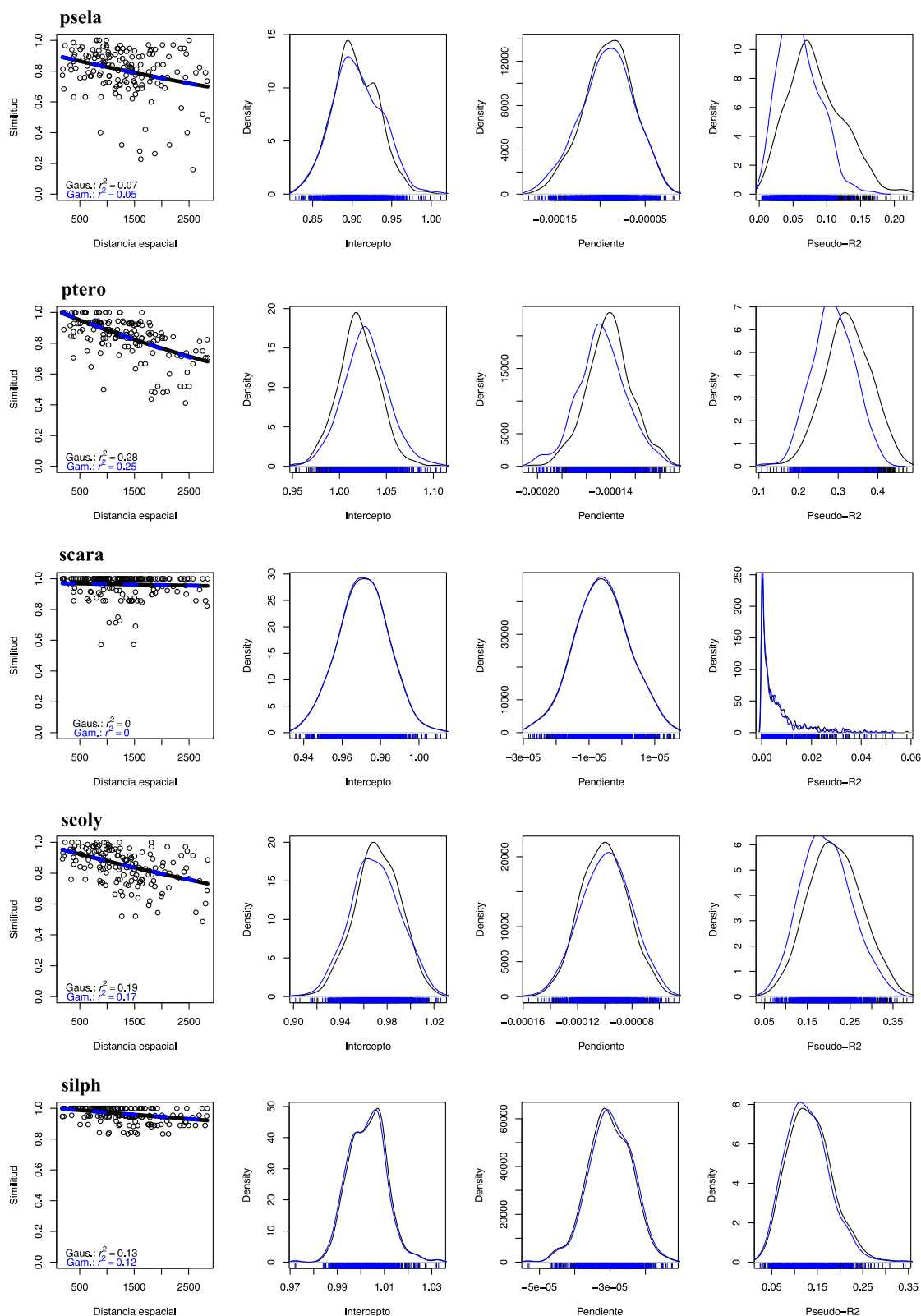


Figura 13a (cont.). Para cada taxón de estudio en el norte de Europa, se muestran (i) las curvas de ajuste de la relación entre la distancia geográfica y la similitud biótica considerando un error de distribución Gaussiana (negro) y un error de distribución Gamma (azul), (ii) la curva de distribución de los interceptos, (iii) la curva de distribución de las pendientes y (iv) la distribución de valores de pseudo-R² estimadas mediante *bootstraps* de 500 remuestras.

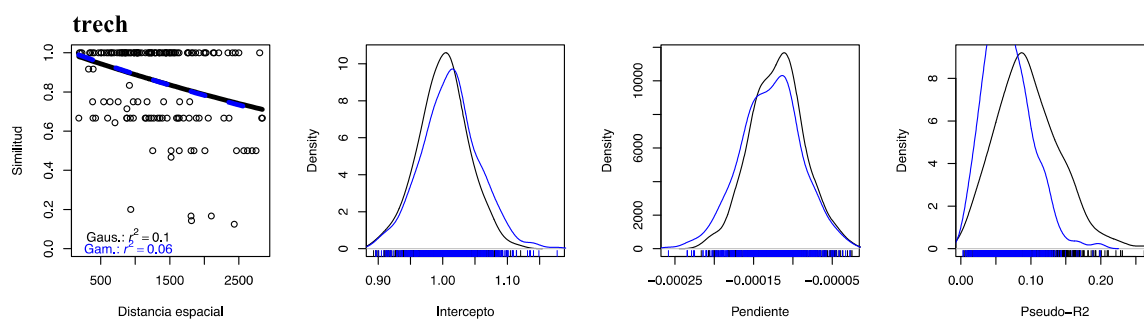


Figura 13a (cont.). Para cada taxón de estudio en el norte de Europa, se muestran (i) las curvas de ajuste de la relación entre la distancia geográfica y la similitud biótica considerando un error de distribución Gaussiana (negro) y un error de distribución Gamma (azul), (ii) la curva de distribución de los interceptos, (iii) la curva de distribución de las pendientes y (iv) la distribución de valores de pseudo-R² estimadas mediante *bootstraps* de 500 remuestras.

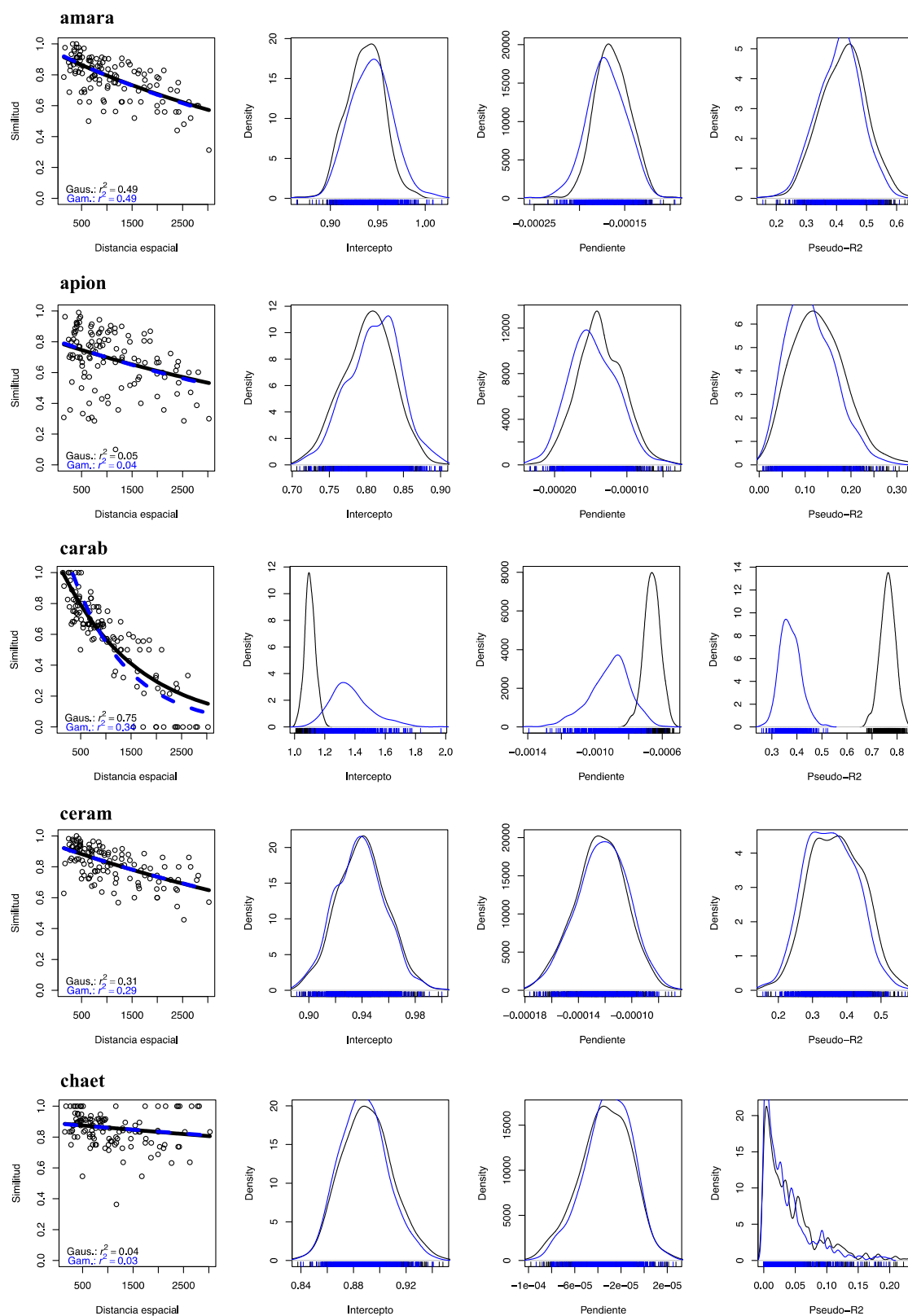


Figura 13b. Para cada taxón de estudio en el sur de Europa, se muestran (i) las curvas de ajuste de la relación entre la distancia geográfica y la similitud biótica considerando un error de distribución Gaussiana (negro) y un error de distribución Gamma (azul), (ii) la curva de distribución de los interceptos, (iii) la curva de distribución de las pendientes y (iv) la distribución de valores de pseudo- R^2 estimadas mediante *bootstraps* de 500 remuestras.

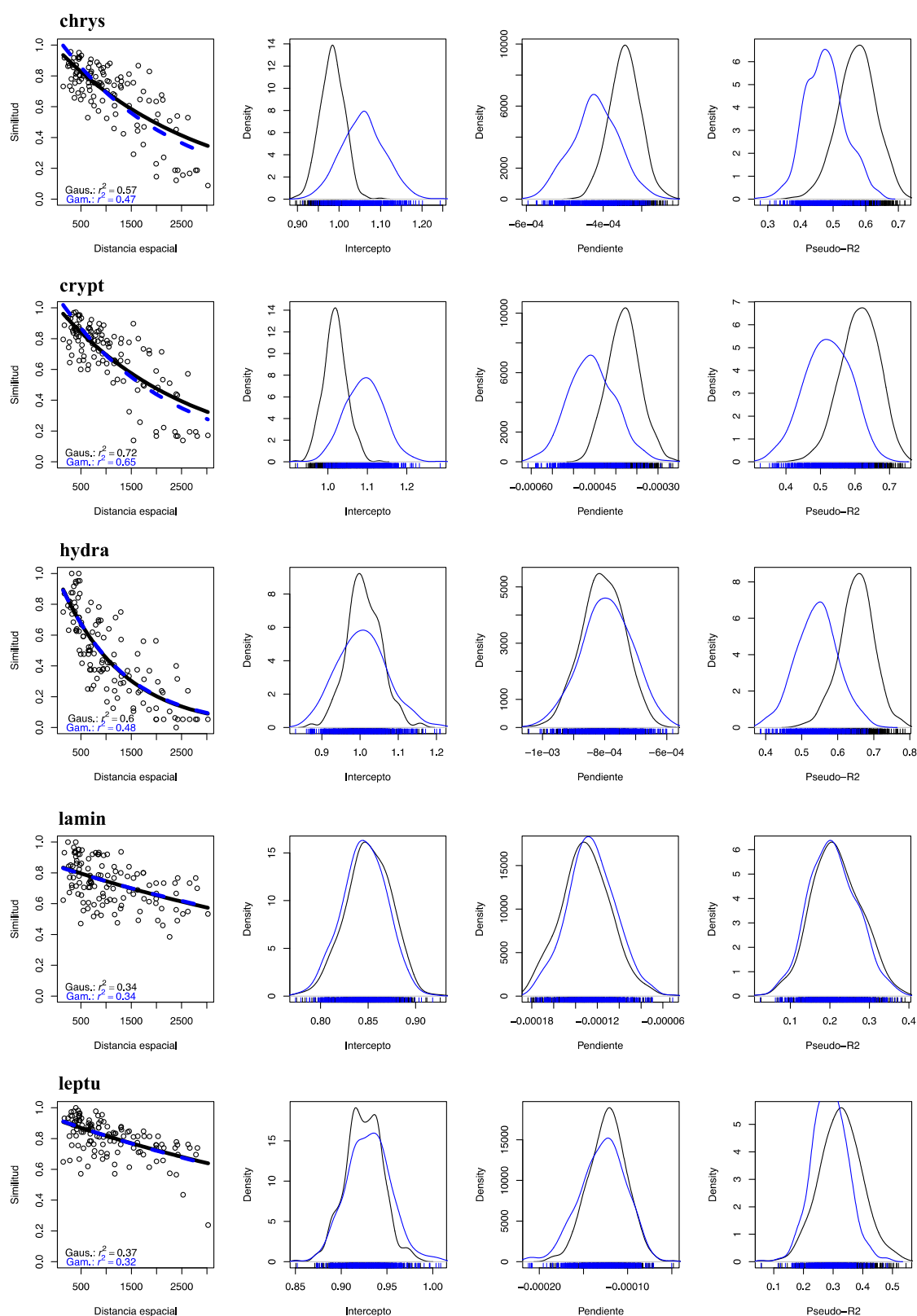


Figura 13b (cont). Para cada taxón de estudio en el sur de Europa, se muestran (i) las curvas de ajuste de la relación entre la distancia geográfica y la similitud biótica considerando un error de distribución Gaussiana (negro) y un error de distribución Gamma (azul), (ii) la curva de distribución de los interceptos, (iii) la curva de distribución de las pendientes y (iv) la distribución de valores de pseudo- R^2 estimadas mediante *bootstraps* de 500 remuestras.

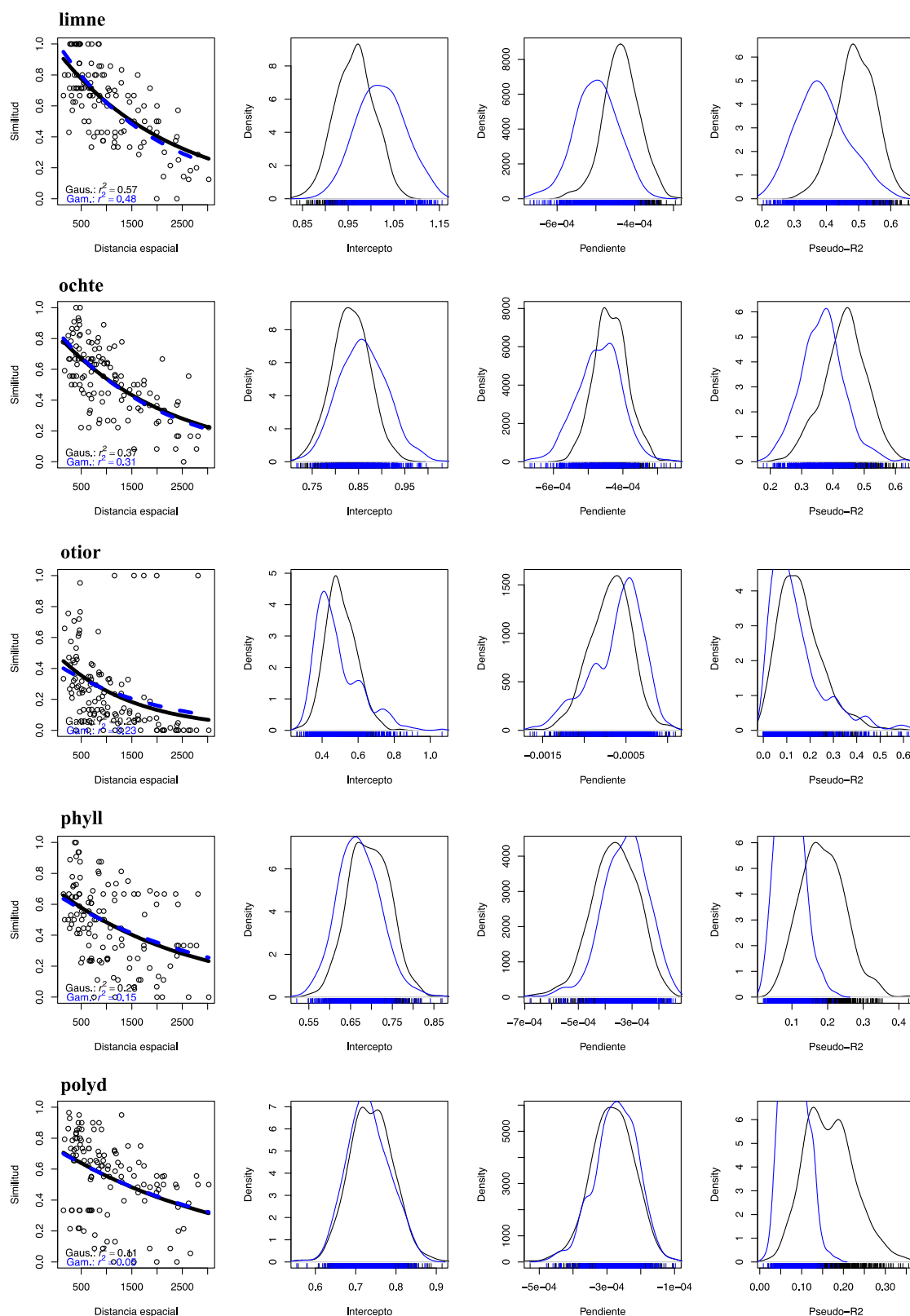


Figura 13b (cont). Para cada taxón de estudio en el sur de Europa, se muestran (i) las curvas de ajuste de la relación entre la distancia geográfica y la similitud biótica considerando un error de distribución Gaussiana (negro) y un error de distribución Gamma (azul), (ii) la curva de distribución de los interceptos, (iii) la curva de distribución de las pendientes y (iv) la distribución de valores de pseudo- R^2 estimadas mediante *bootstraps* de 500 remuestras.

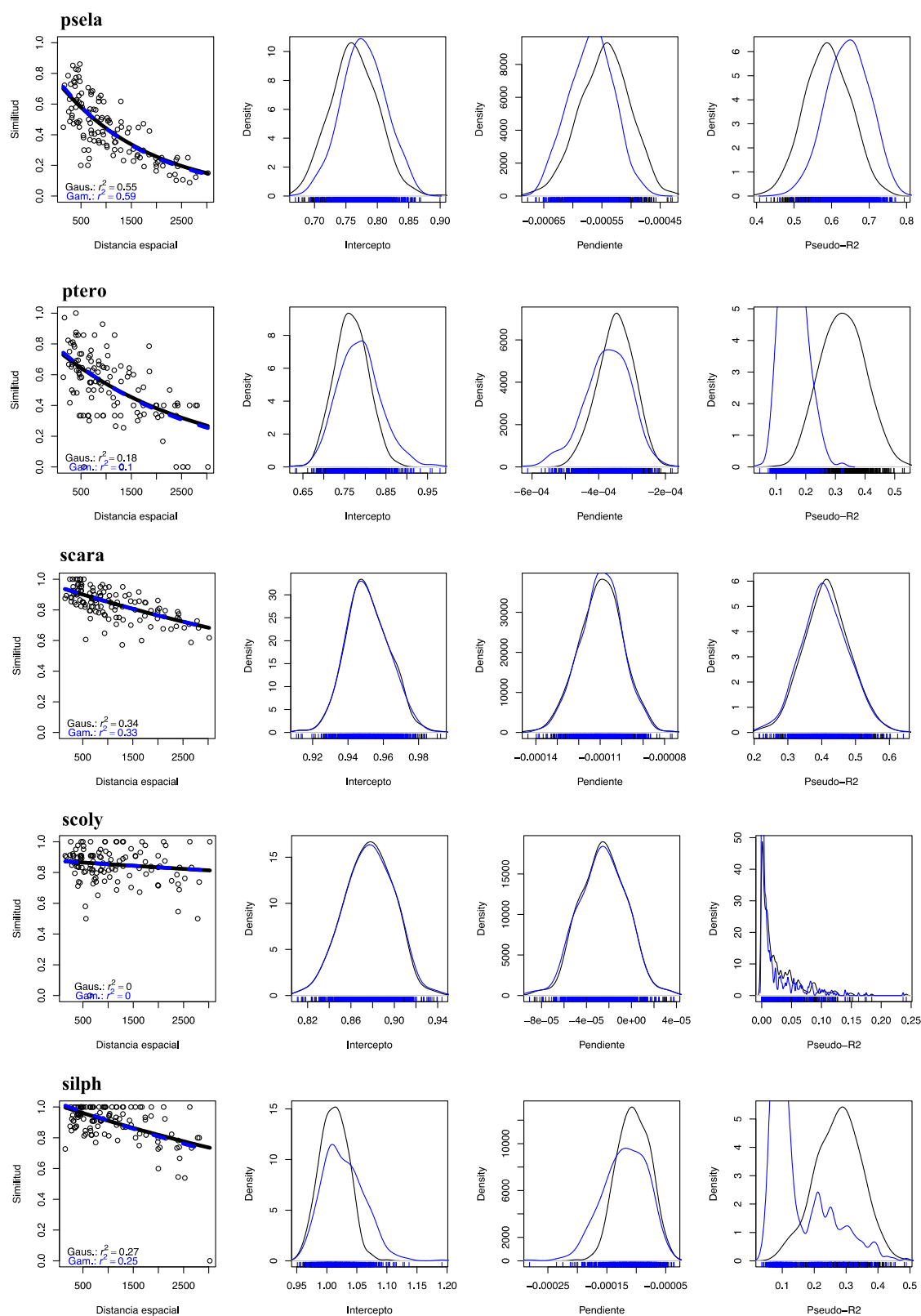


Figura 13b (cont). Para cada taxón de estudio en el sur de Europa, se muestran (i) las curvas de ajuste de la relación entre la distancia geográfica y la similitud biótica considerando un error de distribución Gaussiana (negro) y un error de distribución Gamma (azul), (ii) la curva de distribución de los interceptos, (iii) la curva de distribución de las pendientes y (iv) la distribución de valores de pseudo- R^2 estimadas mediante *bootstraps* de 500 remuestras.

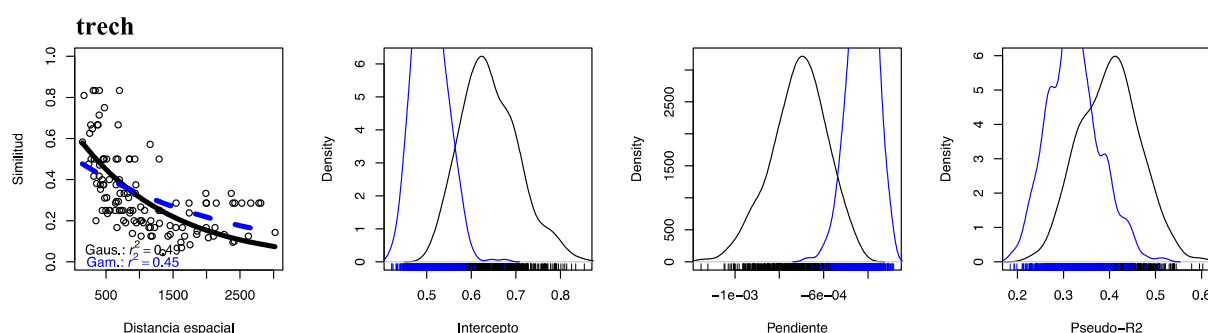


Figura 13b (cont). Para cada taxón de estudio en el sur de Europa, se muestran (i) las curvas de ajuste de la relación entre la distancia geográfica y la similitud biótica considerando un error de distribución Gaussiana (negro) y un error de distribución Gamma (azul), (ii) la curva de distribución de los interceptos, (iii) la curva de distribución de las pendientes y (iv) la distribución de valores de pseudo-R² estimadas mediante *bootstraps* de 500 remuestras.

Tabla 2. Valores estimados (media de la distribución *bootstrap*) para el intercepto (a), pendiente (b) y varianza explicada (pseudo-R²) de modelos con estructura del error Gaussiana o Gamma, para el norte de Europa. Ninguna diferencia entre los parámetros de los modelos (Gaussiana vs. Gamma) del norte de Europa fue significativa. $\alpha < 0.05$; contraste de pruebas pareadas de Wilcoxon; corrección de Benjamini-Yekutieli para p-valores.

Taxón	GAUSSIANA			GAMMA		
	Intercepto	Pendiente	pseudo-R ²	Intercepto	Pendiente	pseudo-R ²
amara	1.02	-0.00008	0.46	1.02	-0.00008	0.45
apion	1.04	-0.00019	0.24	1.06	-0.00020	0.17
carab	1.07	-0.00015	0.36	1.09	-0.00016	0.33
ceram	0.97	-0.00005	0.12	0.97	-0.00005	0.11
chaet	1.01	-0.00007	0.19	1.02	-0.00007	0.18
chrys	1.01	-0.00014	0.28	1.02	-0.00014	0.25
crypt	1.00	-0.00010	0.21	1.01	-0.00011	0.19
hydra	0.87	0.00004	0.03	0.87	0.00004	0.01
lamin	0.98	-0.00005	0.11	0.98	-0.00005	0.09
leptu	0.99	-0.00008	0.19	0.99	-0.00008	0.16
limne	0.94	-0.00007	0.02	0.94	-0.00006	0.01
ochte	0.93	-0.00044	0.32	1.06	-0.00055	0.22
otior	0.88	-0.00027	0.15	0.90	-0.00029	0.08
phyll	1.08	-0.00025	0.31	1.12	-0.00028	0.28
polyd	0.94	-0.00013	0.18	0.95	-0.00013	0.17
psela	0.91	-0.00009	0.09	0.91	-0.00009	0.06
ptero	1.02	-0.00014	0.32	1.03	-0.00015	0.28
scara	0.97	-0.00001	0.01	0.97	-0.00001	0.01
scoly	0.97	-0.00010	0.21	0.97	-0.00010	0.19
silph	1.00	-0.00003	0.14	1.00	-0.00003	0.13
trech	1.00	-0.00012	0.10	1.02	-0.00013	0.07

Tabla 2 (cont). Valores estimados (media de la distribución *bootstrap*) para el intercepto (a), pendiente (b) y varianza explicada (pseudo-R²) de modelos con estructura del error Gaussiana o Gamma, para el sur de Europa. Solo el género *Carabus* mostró diferencias significativas entre los parámetros de los modelos (Gaussiana vs. Gamma), y se indica en negrita.

Taxón	GAUSSIANA			GAMMA		
	Intercepto	Pendiente	pseudo-R ²	Intercepto	Pendiente	pseudo-R ²
amara	0.94	-0.00016	0.42	0.94	-0.00017	0.41
apion	0.80	-0.00013	0.12	0.81	-0.00014	0.11
carab	1.10	-0.00066	0.76	1.37	-0.00091	0.37
ceram	0.94	-0.00012	0.36	0.94	-0.00012	0.35
chaet	0.89	-0.00004	0.05	0.89	-0.00003	0.04
chrys	0.98	-0.00035	0.58	1.06	-0.00043	0.48
crypt	1.02	-0.00038	0.61	1.09	-0.00046	0.52
hydra	1.01	-0.00080	0.65	1.01	-0.00079	0.55
lamin	0.85	-0.00013	0.21	0.84	-0.00012	0.21
leptu	0.92	-0.00012	0.32	0.93	-0.00013	0.28
limne	0.97	-0.00044	0.50	1.03	-0.00050	0.39
ochte	0.83	-0.00043	0.44	0.86	-0.00047	0.37
otior	0.50	-0.00067	0.15	0.47	-0.00057	0.12
phyll	0.70	-0.00036	0.19	0.67	-0.00032	0.09
polyd	0.74	-0.00028	0.17	0.73	-0.00028	0.08
psela	0.76	-0.00054	0.59	0.78	-0.00057	0.64
ptero	0.77	-0.00035	0.33	0.79	-0.00038	0.16
scara	0.95	-0.00011	0.41	0.95	-0.00011	0.41
scoly	0.87	-0.00002	0.03	0.88	-0.00003	0.02
silph	1.01	-0.00011	0.28	1.03	-0.00012	0.15
trech	0.64	-0.00072	0.40	0.51	-0.00043	0.32

5.3 MÉTODOS *BOOTSTRAP*: EXPLORACIÓN Y APROXIMACIÓN INICIAL

5.3.1 Estimación de los parámetros del modelo y su desviación típica mediante métodos de *bootstrap* propuestos por autores previos

El principal interés de conocer la verdadera desviación típica de los coeficientes de regresión es la construcción de intervalos de confianza y la realización de contrastes sobre dichos coeficientes, enfocados tanto a ajustar un buen modelo a los datos como a la comparación de las curvas de descenso de la similitud biológica con la distancia entre distintos conjuntos de datos.

Para cada muestra simulada (8 en total) se han estimado las distribuciones de los coeficientes de regresión a y b mediante ambos métodos *bootstrap* aplicando la función `glm()` con el parámetro `family = gaussian(link = "log")` a cada remuestra de X e Y . Esta función estima los parámetros del modelo mediante IRWLS. A partir de dichas distribuciones se ha obtenido su media muestral (estimación puntual de los coeficientes) y su desviación típica muestral. A continuación se ha calculado el sesgo de los estimadores. También se han almacenado las estimaciones que devuelve la función `glm()` y se ha calculado su sesgo. Los resultados de estas simulaciones se encuentran en las Tablas 3-6 y en la Figura 14.

Tabla 3. Estimación del intercepto (a), su sesgo y su desviación típica en los modelos *distance-decay* de las muestras simuladas ($n = 10000$ remuestras) y 16 sitios (N). Se comparan el método GLM estándar, sin *bootstrap*, y los métodos de simulación basados en bloques móviles y bloques manuales. En la tabla, bl es el tamaño de los bloques móviles. Valor simulado de $a = 0.7575$.

Simulación		Estimación a	Sesgo a	Desviación típica a
N = 16, $bl = 10$	GLM (sin <i>bootstrap</i>)	0.7543	-0.0032	0.0517
	Bloques móviles	0.7558	-0.0017	0.0385
	Bloques manuales	0.7551	-0.0025	0.0286
N = 16, $bl = 20$	GLM (sin <i>bootstrap</i>)	0.7838	0.0262	0.0521
	Bloques móviles	0.8052	0.0477	0.0372
	Bloques manuales	0.7842	0.0266	0.0266
N = 16, $bl = 40$	GLM (sin <i>bootstrap</i>)	0.8522	0.0946	0.0585
	Bloques móviles	0.8856	0.1280	0.0450
	Bloques manuales	0.8527	0.0951	0.0361
N = 16, $bl = 80$	GLM (sin <i>bootstrap</i>)	0.6942	-0.0634	0.0612
	Bloques móviles	0.6817	-0.0758	0.0156
	Bloques manuales	0.6953	-0.0623	0.0295

Tabla 4. Estimación de la pendiente (b), su sesgo y su desviación típica en los modelos *distance-decay* de las muestras simuladas ($n = 10000$ remuestras) y 16 sitios (N). Se comparan el método GLM estándar, sin *bootstrap*, y los métodos de simulación basados en bloques móviles y bloques manuales. En la tabla, bl es el tamaño de los bloques móviles. Valor simulado de $b = -0.0003564$.

Simulación	Estimación b	Sesgo b	Desviación típica b
N =16, $bl = 10$	GLM (sin <i>bootstrap</i>)	-3.360e-04	2.036e-05
	Bloques móviles	-3.370e-04	1.940e-05
	Bloques manuales	-3.360e-04	2.032e-05
N =16, $bl = 20$	GLM (sin <i>bootstrap</i>)	-3.369e-04	-1.284e-05
	Bloques móviles	-3.804e-04	-2.404e-05
	Bloques manuales	-3.689e-04	-1.252e-05
N =16, $bl = 40$	GLM (sin <i>bootstrap</i>)	-4.457e-04	-8.939e-05
	Bloques móviles	-4.819e-04	-1.255e-05
	Bloques manuales	-4.459e-04	-8.955e-05
N =16, $bl = 80$	GLM (sin <i>bootstrap</i>)	-2.941e-04	6.222e-05
	Bloques móviles	-2.936e-04	6.277e-05
	Bloques manuales	-2.945e-04	6.192e-05

Tabla 5. Estimación del intercepto (a), su sesgo y su desviación típica en los modelos *distance-decay* de las muestras simuladas ($n = 10000$ remuestras) y 20 sitios (N). Se comparan el método GLM estándar, sin *bootstrap*, y los métodos de simulación basados en bloques móviles y bloques manuales. En la tabla, bl es el tamaño de los bloques móviles. Valor simulado de $a = 0.7575$.

Simulación	Estimación a	Sesgo a	Desviación típica a
N =20, $bl = 10$	GLM (sin <i>bootstrap</i>)	0.7647	0.0071
	Bloques móviles	0.7537	-0.0017
	Bloques manuales	0.7657	0.0082
N =20, $bl = 20$	GLM (sin <i>bootstrap</i>)	0.7602	0.0027
	Bloques móviles	0.7723	0.0147
	Bloques manuales	0.7602	0.0026
N =20, $bl = 40$	GLM (sin <i>bootstrap</i>)	0.7128	-0.0447
	Bloques móviles	0.7228	-0.0347
	Bloques manuales	0.7133	-0.0443
N =20, $bl = 80$	GLM (sin <i>bootstrap</i>)	0.7796	0.0221
	Bloques móviles	0.7979	0.0404
	Bloques manuales	0.7816	0.0241

Tabla 6. Estimación de la pendiente (b), su sesgo y su desviación típica en los modelos *distance-decay* de las muestras simuladas ($n = 10000$ remuestras) y 20 sitios (N). Se comparan el método GLM estándar, sin *bootstrap*, y los métodos de simulación basados en bloques móviles y bloques manuales. En la tabla, bl es el tamaño de los bloques móviles. Valor simulado de $b = -0.0003564$.

Simulación		Estimación b	Sesgo b	Desviación típica b
N =20, $bl = 10$	GLM (sin <i>bootstrap</i>)	-3.695e-04	-1.312e-05	3.131e-05
	Bloques móviles	-3.617e-04	-5.264e-06	2.670e-05
	Bloques manuales	-3.700e-04	-1.364e-05	2.094e-05
N =20, $bl = 20$	GLM (sin <i>bootstrap</i>)	-3.782e-04	-1.635e-05	3.661e-05
	Bloques móviles	-3.791e-04	-2.274e-05	3.017e-05
	Bloques manuales	-3.724e-04	-1.598e-05	1.895e-05
N =20, $bl = 40$	GLM (sin <i>bootstrap</i>)	-3.322e-04	2.418e-05	3.496e-05
	Bloques móviles	-3.367e-04	1.974e-05	2.194e-05
	Bloques manuales	-3.320e-04	2.439e-05	2.711e-05
N =20, $bl = 80$	GLM (sin <i>bootstrap</i>)	-3.551e-04	1.343e-06	3.242e-05
	Bloques móviles	-3.657e-04	-9.270e-06	3.471e-05
	Bloques manuales	-3.551e-04	1.307e-06	2.366e-05

En primer lugar, se aprecia gran variedad en los sesgos para todos los métodos dependiendo de la propia muestra, el número de sitios y, para el método *bootstrap* de bloques móviles, del tamaño de bloques considerado (Tablas 3-6). Por ejemplo, para $N = 16$, $bl = 10$ el sesgo de a y de b alcanza sus valores mínimos en el conjunto de los tres métodos mientras que en la muestra $N = 16$, $bl = 40$ alcanza sus valores máximos, muy superiores a los anteriores, de tal forma que el valor real apenas si forma parte de cualquiera de las distribuciones *bootstrap*. Concretamente para $N = 16$, en valor absoluto el sesgo de a se encuentra entre 0.0017 y 0.1280, y el de b entre $1.252 \cdot 10^{-5}$ y $1.255 \cdot 10^{-4}$. Al aumentar el tamaño de muestra a $N = 20$, la fluctuación entre los sesgos en relación al conjunto de la muestra ya no es tan extrema: en valor absoluto, el sesgo de a se encuentra entre 0.0038 y 0.0447, y el de b entre $1.307 \cdot 10^{-6}$ y $2.439 \cdot 10^{-5}$.

En segundo lugar, las distribuciones *bootstrap* de a y b son en general bastante simétricas (Figura 14), especialmente para $N = 20$, con lo cual utilizar la media como estimación es acertado. De la misma forma, aunque el ajuste de los modelos obtenidos por los dos métodos *bootstrap* y el devuelto por la función `glm()` diverge considerablemente para $N = 16$, $bl = 40$ y $N = 16$, $bl = 80$, cuando $N = 20$ los tres métodos generan ajustes muy superpuestos (Figura 14).

Respecto al sesgo, el *bootstrap* de bloques manuales devuelve valores equivalentes a los de la función `glm()` tanto para a como para b (Tablas 3-6). Sin embargo, el *bootstrap* de bloques móviles en la mayoría de las ocasiones los supera a ambos e incluso los duplica; y en una minoría de ellos es menor o incluso la mitad. En el caso de la desviación típica, las estimaciones *bootstrap* devuelven valores sensiblemente más pequeños que la función `glm()`, salvo para $N = 16$, $bl = 10$ y $N = 20$, $bl = 80$ en los que el *bootstrap* de bloques móviles los supera. Por el contrario, las estimaciones del *bootstrap* de bloques manuales siempre son menores que las de `glm()` y en la mayoría de ocasiones también son menores que las del método de bloques móviles.

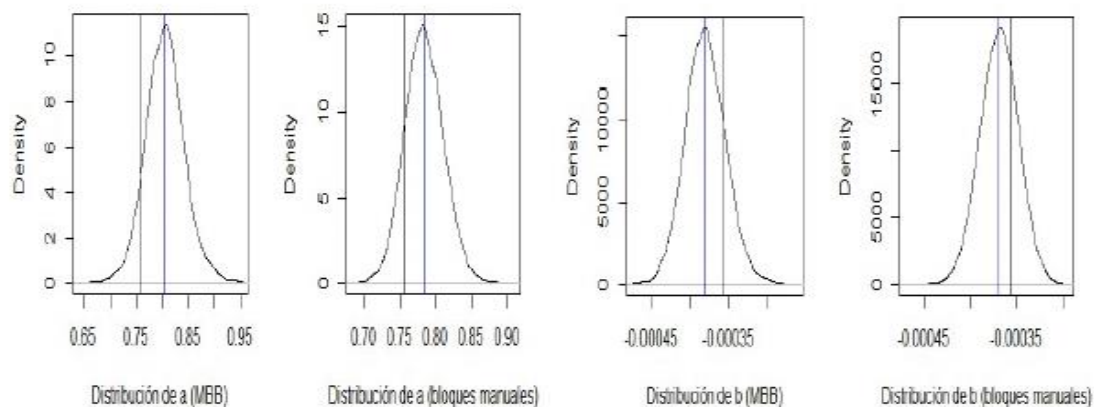
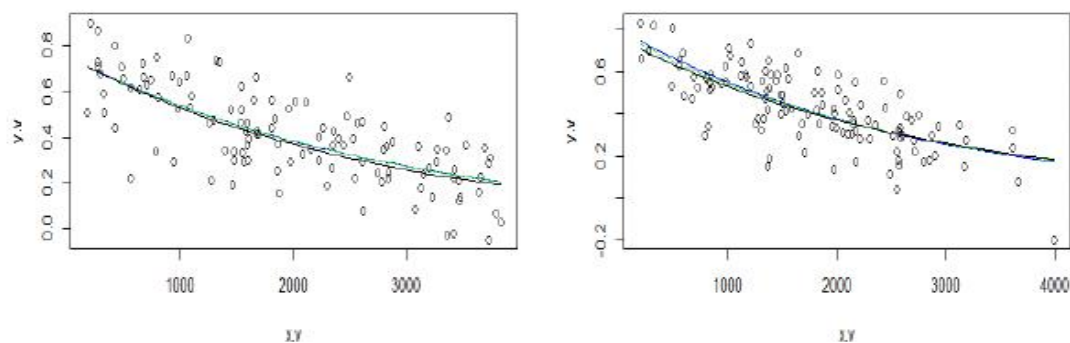
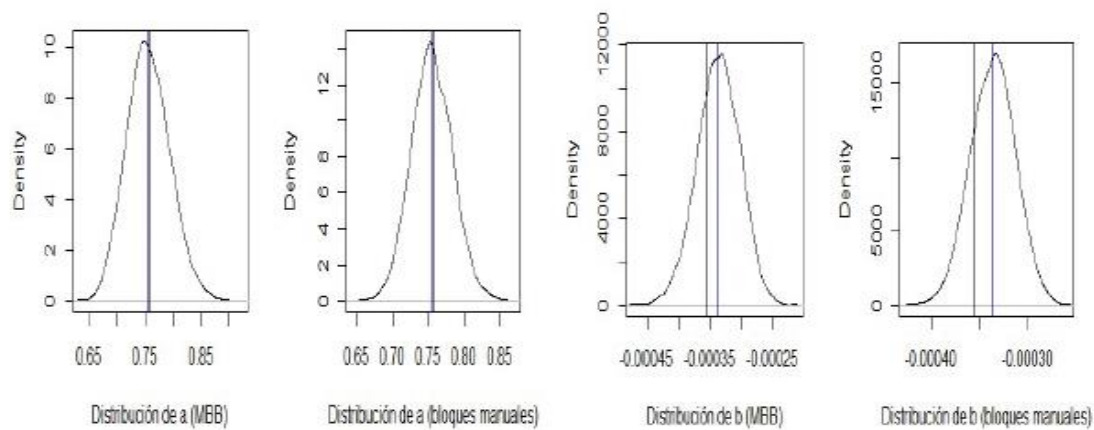


Figura 14. Distribuciones del intercepto (a) y el coeficiente de regresión (b) estimados mediante *bootstrap* (10000 réplicas). En las gráficas de las distribuciones (a, b, e, f), la línea vertical azul es la media muestral de la distribución, usada como estimador del parámetro, y la línea vertical negra es el verdadero valor del parámetro. En las gráficas *distance-decay* (c, d), la curva negra se corresponde con el modelo estimado mediante la función `glm()` *bootstrap*; la curva azul se corresponde con el ajuste hecho a partir de los parámetros estimados mediante *bootstrap* de bloques móviles (MBB); y la curva verde con el ajuste hecho a partir de los parámetros obtenidos con el *bootstrap* de bloques manuales. N es el número de sitios, bl es el tamaño de los bloques.

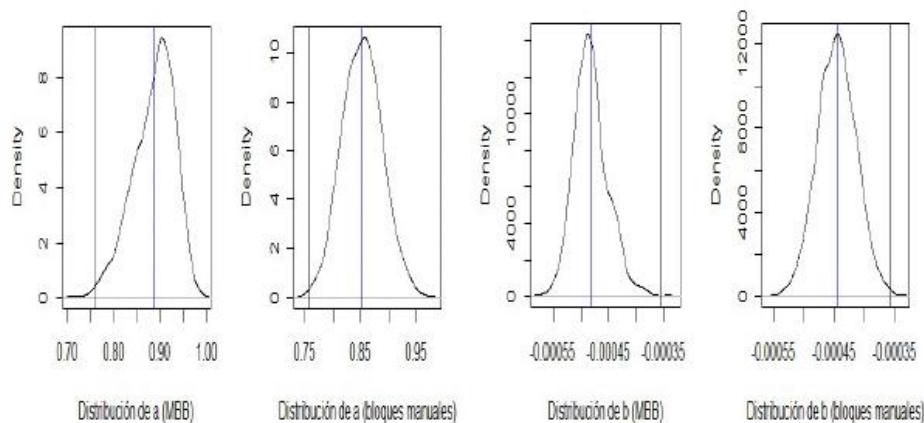
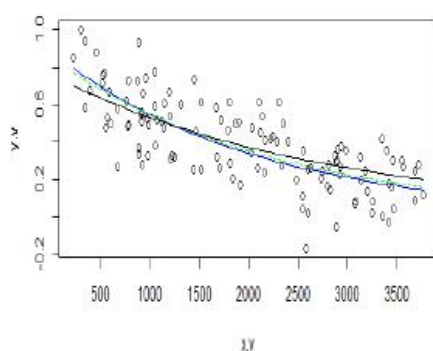
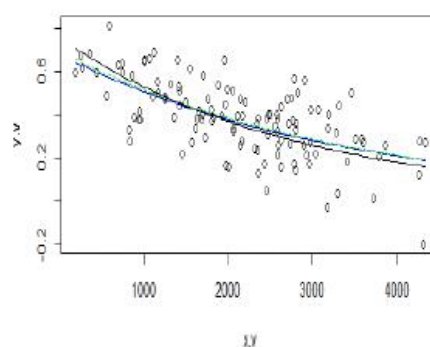
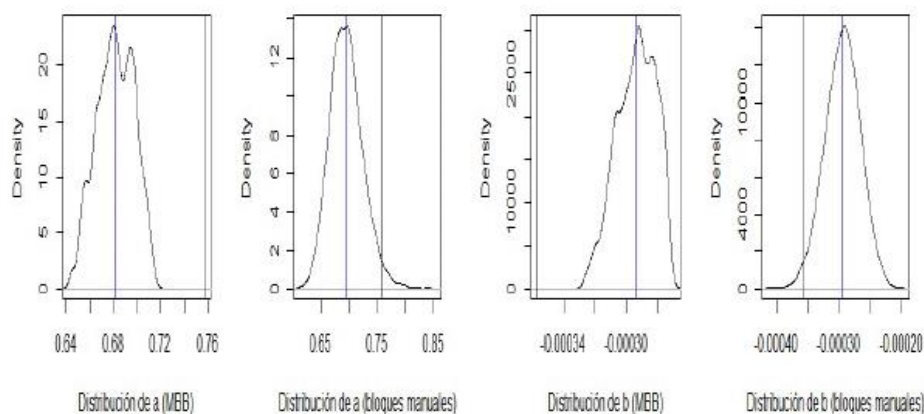
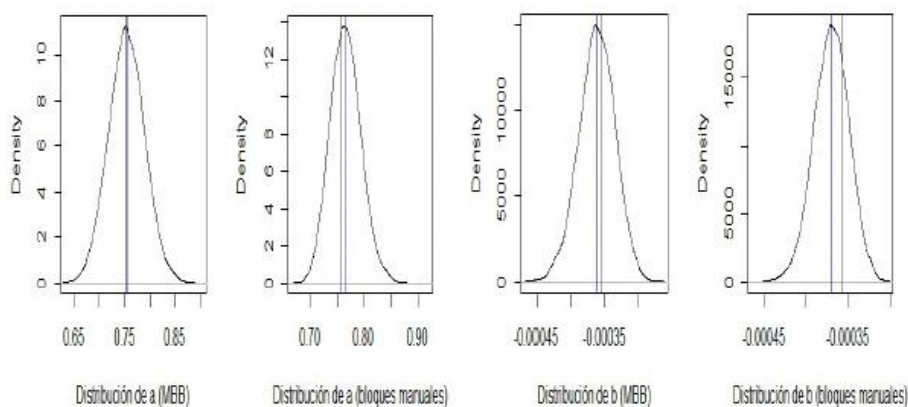
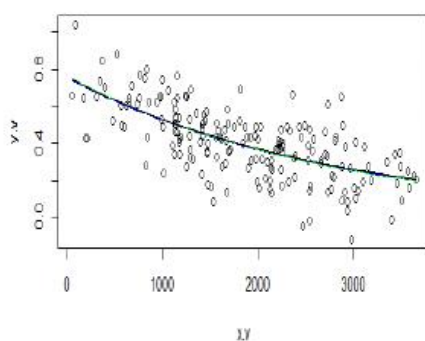
(g) $N=16$, $bl=40$, distribución de a (h) $N=16$, $bl=40$, distribución de b (i) $N=16$, $bl=40$, ajuste de los modelos(j) $N=16$, $bl=80$, ajuste de los modelos(k) $N=16$, $bl=80$, distribución de a (l) $N=16$, $bl=80$, distribución de b

Figura 14 (cont.). Distribuciones del intercepto (a) y el coeficiente de regresión (b) estimados mediante *bootstrap* (10000 réplicas). En las gráficas de las distribuciones (a , b , e , f), la línea vertical azul es la media muestral de la distribución, usada como estimador del parámetro, y la línea vertical negra es el verdadero valor del parámetro. En las gráficas *distance-decay* (c , d), la curva negra se corresponde con el modelo estimado mediante la función *glm()* sin *bootstrap*; la curva azul se corresponde con el ajuste hecho a partir de los parámetros estimados mediante *bootstrap* de bloques móviles (MBB); y la curva verde con el ajuste hecho a partir de los parámetros obtenidos con el *bootstrap* de bloques manuales. N es el número de sitios, bl es el tamaño de los bloques.

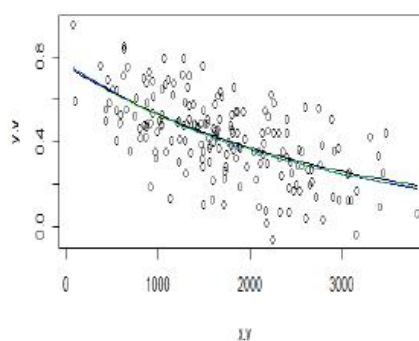


(m) $N=20$, $bl=10$, distribución de a

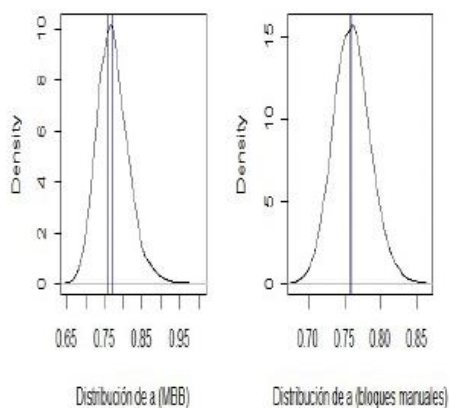
(n) $N=20$, $bl=10$, distribución de b



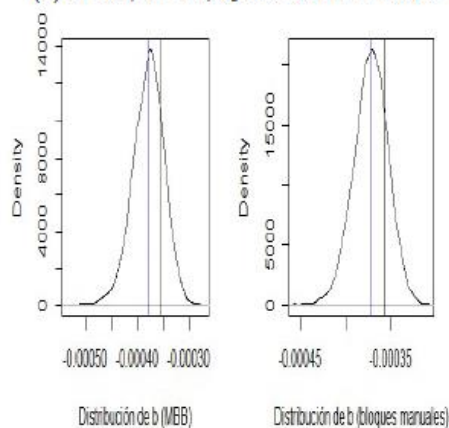
(ñ) $N=20$, $bl=10$, ajuste de los modelos



(o) $N=20$, $bl=20$, ajuste de los modelos



(p) $N=20$, $bl=20$, distribución de a



(q) $N=20$, $bl=20$, distribución de b

Figura 14 (cont.). Distribuciones del intercepto (a) y el coeficiente de regresión (b) estimados mediante *bootstrap* (10000 réplicas). En las gráficas de las distribuciones (a, b, e, f), la línea vertical azul es la media muestral de la distribución, usada como estimador del parámetro, y la línea vertical negra es el verdadero valor del parámetro. En las gráficas *distance-decay* (c, d), la curva negra se corresponde con el modelo estimado mediante la función `glm()` sin *bootstrap*; la curva azul se corresponde con el ajuste hecho a partir de los parámetros estimados mediante *bootstrap* de bloques móviles (MBB); y la curva verde con el ajuste hecho a partir de los parámetros obtenidos con el *bootstrap* de bloques manuales. N es el número de sitios, bl es el tamaño de los bloques.

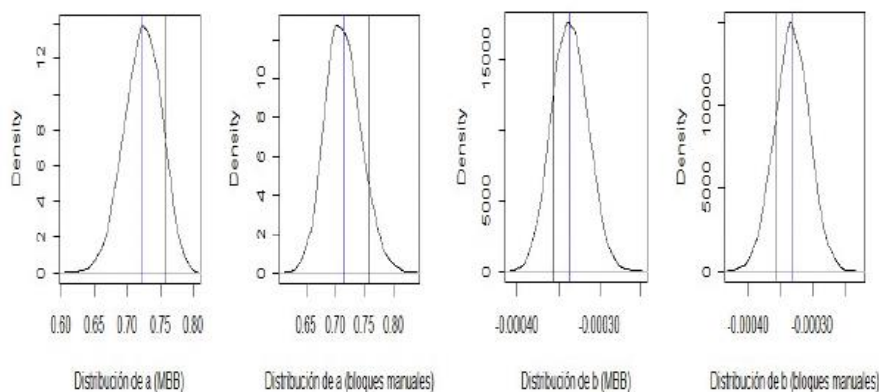
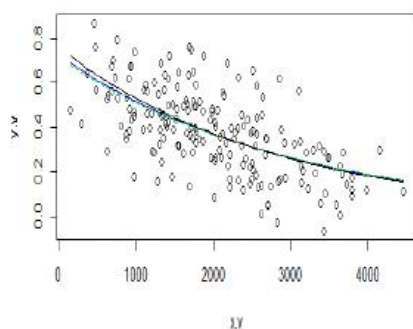
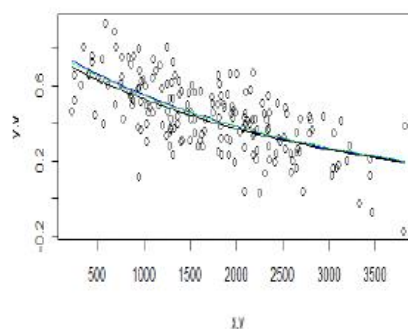
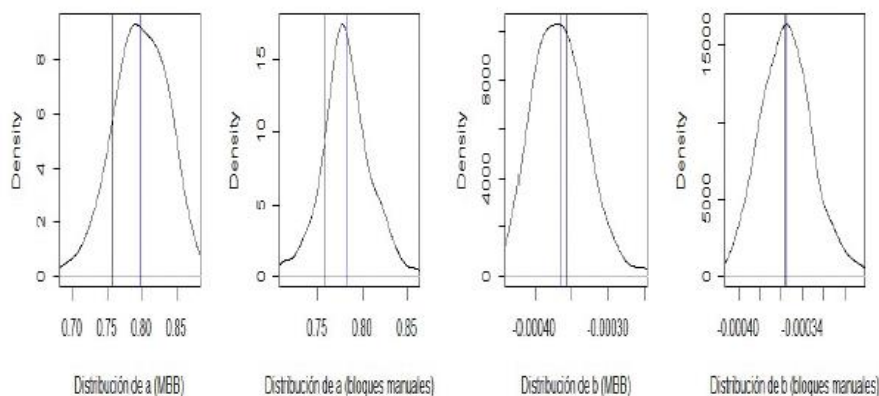
(r) $N=20$, $bl=40$, distribución de a (s) $N=20$, $bl=40$, distribución de b (t) $N=20$, $bl=40$, ajuste de los modelos(u) $N=20$, $bl=80$, ajuste de los modelos(v) $N=20$, $bl=80$, distribución de a (w) $N=20$, $bl=80$, distribución de b

Figura 14 (cont.). Distribuciones del intercepto (a) y el coeficiente de regresión (b) estimados mediante *bootstrap* (10000 réplicas). En las gráficas de las distribuciones (a, b, e, f), la línea vertical azul es la media muestral de la distribución, usada como estimador del parámetro, y la línea vertical negra es el verdadero valor del parámetro. En las gráficas *distance-decay* (c, d), la curva negra se corresponde con el modelo estimado mediante la función `glm()` sin *bootstrap*; la curva azul se corresponde con el ajuste hecho a partir de los parámetros estimados mediante *bootstrap* de bloques móviles (MBB); y la curva verde con el ajuste hecho a partir de los parámetros obtenidos con el *bootstrap* de bloques manuales. N es el número de sitios, bl es el tamaño de los bloques.

Por lo tanto, parece que el *bootstrap* de bloques móviles es inestable con este tipo de datos y este tamaño muestral. En algunos casos puntuales devuelve los mejores resultados, y en otros su sesgo y/o su estimación de la desviación típica es mayor que para los otros métodos. Esta inestabilidad en función de la muestra no permite diferenciar cual es realmente el mejor tamaño de bloque. En cambio, el método de bloques manuales es más coherente a lo largo de todas las muestras: no aumenta el sesgo de los coeficientes respecto al `glm()` y sus estimaciones de la desviación típica son menores. Este resultado coincide con lo esperado: en condiciones de dependencia de los datos la varianza real de los estimadores de los coeficientes de regresión será menor que la obtenida mediante IRWLS. De todas formas, sería conveniente calcular el sesgo de las propias estimaciones de la desviación típica para determinar si efectivamente el método de bloques manuales funciona correctamente. Por todo ello, se considera que se debe profundizar en la obtención de una aproximación basada en la creación de bloques forzados para la construcción de modelos de *distance-decay* no lineales, como se realiza en el apartado 5.4 de esta tesis, descartando los otros dos métodos explorados en este capítulo.

5.3.2 Contraste de significación del modelo basado en la *deviance*

Para determinar la significación del modelo, se ha realizado una aproximación basada en la *deviance* y la comparación del modelo nulo con el modelo ajustado. La *deviance* es una medida de lo bien que el modelo ajustado explica los datos, considerando previamente que se trata de un modelo válido, y cuanto menor sea el valor de la *deviance*, mejor será el ajuste del modelo. En el contexto de la comparación del modelo nulo y el modelo ajustado, se contrasta si el modelo ajustado, en caso de ser válido, supone realmente una diferencia frente a la ausencia de modelo (modelo nulo). Para ello se han calculado los p-valores como la proporción de veces que el estadístico *bootstrap* ha superado el valor del estadístico observado en los datos originales. Los resultados se muestran en la Tabla 7.

Tabla 7. Nivel (error de tipo I) del contraste de la *deviance* entre el modelo ajustado y el modelo nulo. Se han realizado 1000 simulaciones y 2500 réplicas *bootstrap* utilizando el GLM estándar sin *bootstrap*, y los métodos de simulación basados en bloques móviles y bloques manuales. El valor crítico es $\alpha = 0.05$. N es el número de sitios, y *bl* el tamaño de los bloques móviles.

	N = 16			
	<i>bl</i> = 10	<i>bl</i> = 20	<i>bl</i> = 40	<i>bl</i> = 80
GLM (sin <i>bootstrap</i>)	0.047	0.059	0.052	0.046
Bloques móviles	0.031	0.053	0.049	0.040
Bloques manuales	0.032	0.044	0.038	0.025
	N = 20			
	<i>bl</i> = 10	<i>bl</i> = 20	<i>bl</i> = 40	<i>bl</i> = 80
GLM (sin <i>bootstrap</i>)	0.056	0.059	0.049	0.048
Bloques móviles	0.041	0.047	0.046	0.047
Bloques manuales	0.040	0.034	0.034	0.029

El contraste usando la distribución F asintótica (GLM sin *bootstrap*) sobrepasa en más de 0.005 el nivel de 0.05 en tres de los casos ($N = 16$ y $bl = 20$, $N = 20$ y $bl = 10$, $N = 20$ y $bl = 20$) y en el resto se aproxima bastante, por lo que parece ser ligeramente liberal dependiendo de la muestra (Tabla 7). No se aprecia diferencia en el patrón entre los casos con $N = 16$ y con $N = 20$. El test *bootstrap* de bloques móviles en ocasiones alcanza al nivel de 0.05 y en otras es inferior. Al aumentar el tamaño de muestra se acerca más al nivel nominal. Los mejores resultados para ambos tamaños de muestra se han conseguido con los tamaños de bloque 20 y 40. El test *bootstrap* de bloques manuales es demasiado conservador, rechaza demasiado poco la hipótesis nula y nunca llega al nivel de significación, por lo que sus resultados no serán fiables. Los otros dos contrastes, especialmente el de bloques móviles, sí que proporcionan resultados aceptables para el análisis de datos a pesar de no estar considerando la dependencia por pares de los datos y, por tanto, no ser teóricamente adecuados para los modelos no lineales de *distance-decay*. Esto sugiere que quizás la simulación de la hipótesis nula no esté verdaderamente reproduciendo la dependencia por pares de los datos y que sea este aspecto el que provoca que el método *bootstrap* con bloques manuales parezca demasiado conservador. Este aspecto es el que se explora en la sección 5.4 de esta tesis.

5.3.3 Selección de modelos mediante AIC

Para determinar la distribución de la verosimilitud en modelos de datos dependientes, como es el caso de los *distance-decay*, se han estimado las frecuencias *bootstrap* de selección de dos modelos: uno que se corresponde con el modelo que se usó para generar los datos simulados (modelo exponencial), y un modelo alternativo distinto (*power-law*) al utilizado para generar los datos, con lo cual se espera que, basándonos en el AIC, nunca se seleccione el modelo alternativo como el modelo que mejor se ajusta a los datos. Este estudio de simulación es importante para determinar si el AIC es un método fiable para seleccionar el modelo (o modelos) que mejor explique los datos de entre un conjunto de modelos candidatos, cuando se usa para datos empíricos.

Se han ejecutado 10000 réplicas *bootstrap* y los resultados se muestran en la Tabla 8. Se observa que cualquiera de los dos métodos *bootstrap* ha funcionado correctamente, seleccionando el modelo verdadero (GLM exponencial) en la amplia mayoría de los casos. Además, en 10 de las 16 simulaciones, la proporción de réplicas *bootstrap* en las que se selecciona el modelo verdadero es superior a 0.9.

Tabla 8. Frecuencias *bootstrap* de selección de los modelos mediante AIC. Se han realizado 10000 réplicas *bootstrap* usando el método de simulación basados en bloques móviles e, independientemente, el basado en bloques manuales. N es el número de sitios, y bl el tamaño de los bloques.

Simulación		GLM exponencial	GLM <i>power-law</i>
N =16, bl = 10	Bloques móviles	0.85	0.23
	Bloques manuales	0.7551	0.15
N =16, bl = 20	Bloques móviles	1	0
	Bloques manuales	1	0
N =16, bl = 40	Bloques móviles	0.88	0.12
	Bloques manuales	0.76	0.24
N =16, bl = 80	Bloques móviles	0.90	0.10
	Bloques manuales	0.79	0.21
N =20, bl = 10	Bloques móviles	0.91	0.09
	Bloques manuales	0.98	0.02
N =20, bl = 20	Bloques móviles	0.98	0.02
	Bloques manuales	0.97	0.03
N =20, bl = 40	Bloques móviles	0.73	0.27
	Bloques manuales	0.95	0.05
N =20, bl = 80	Bloques móviles	1	0
	Bloques manuales	1	0

5.4. IMPLEMENTACIÓN DE UN MÉTODO PARA CONTRASTE DE HIPÓTESIS EN MODELOS NO LINEALES DE *DISTANCE-DECAY* BASADO EN PERMUTACIONES Y REMUESTRO POR BLOQUES DE LOCALIDAD²

5.4.1 Simulaciones para la estimación de los errores de tipo I y II

Como se esperaba, el test de la F falló en devolver un error de tipo I adecuado y además su inflación se incrementó con el tamaño muestral (N), es decir, el número de comunidades comparadas (Figura 15). Por el contrario, tanto el método de permutaciones por bloques como el de remuestro por bloques de localidad usados para estimar la distribución nula de R_{KLD}^2 evitaron la inflación del error de tipo I. La estimación de su error de tipo I fue cercana a 0.05 y no aumentó con el número de localidades en la matriz de similitud (Figura 16). El test de Mantel también presentó un error de tipo I correcto. En las simulaciones de la hipótesis nula, todos los estadísticos (R_{KLD}^2 exponencial, R_{KLD}^2 *power-law* y la r de Mantel) presentaron distribuciones de media cercana a 0 (Figura 16); lo que refleja la equivalencia en este escenario entre los modelos ajustados y el modelo nulo.

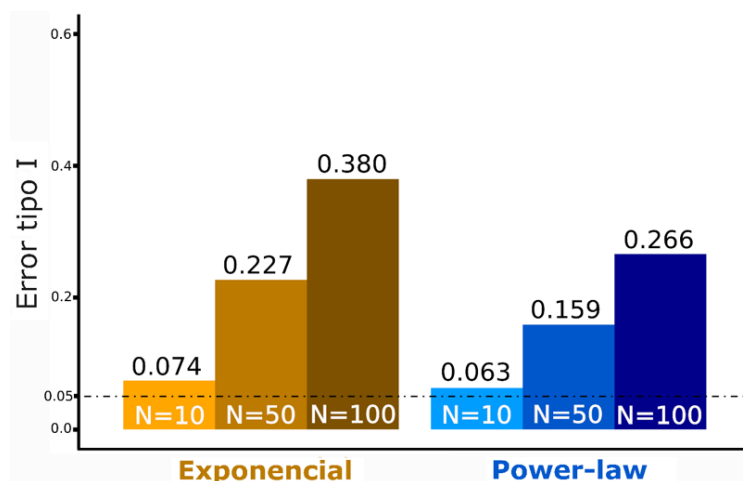


Figura 15. Inflación del error de tipo I para el test de la F aplicado a los modelos exponencial negativo y *power-law* de *distance-decay*. N es el número de localidades. El error de tipo I se estimó simulando datos 1000 veces bajo la hipótesis nula, para a continuación ajustar un modelo exponencial negativo o *power-law* en cada simulación y calcular la significación con la función `anova()` en R. La proporción de contrastes significativos es el error de tipo I.

En las simulaciones de la hipótesis alternativa para los escenarios con alta autocorrelación espacial en la composición de las comunidades ($\varphi = 5000$ y $\varphi = 40000$), la mayoría de los valores estimados de R_{KLD}^2 fueron superiores a 0.25; por lo que el modelo exponencial y el *power-law* explicaron una proporción considerable de la variabilidad en la similitud entre

² Una versión adaptada y en inglés de este capítulo ha sido publicada como:

Martínez-Santalla S¹, Martín-Devasa R^{1,2}, Gómez-Rodríguez C¹, Crujeiras RM³, Baselga A. 2022¹. Assessing the non-linear decay of community similarity: permutation and site-block resampling significance tests. *Journal of Biogeography* (WILEY, Online ISSN:1365-2699) 49: 968-978. <https://doi.org/10.1111/jbi.14351>

¹ CRETUS, Universidade de Santiago de Compostela; ² University of Helsinki, ³ CITMAga, Universidade de Santiago de Compostela

comunidades en función de la distancia espacial entre ellas. Sin embargo, en el escenario de baja autocorrelación espacial ($\varphi = 300$), los valores de \hat{R}_{KLD}^2 se mantuvieron de forma general por debajo de 0.05; lo cual implica que los modelos exponencial y *power-law* no explicaron una proporción mayor de la variabilidad de la que explica el modelo nulo (Figuras 16-18).

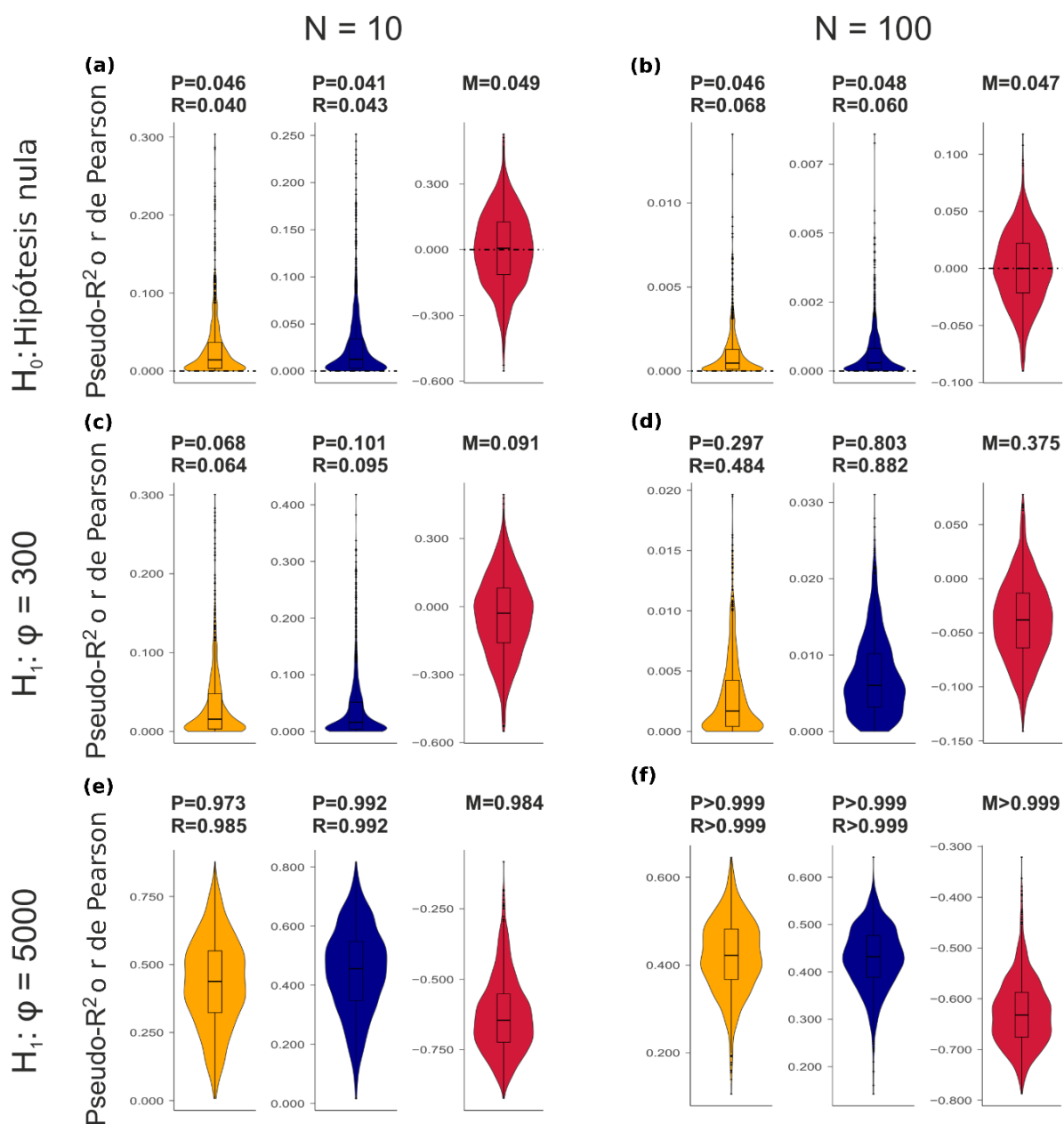


Figura 16. Proporción de contrastes significativos en 1000 simulaciones para las permutaciones por bloques de localidad (P), el remuestreo por bloques de localidad (R) y el test de Mantel (M). Las simulaciones se corrieron bajo la hipótesis nula (a, b) y bajo la alternativa (c, d para $\varphi = 300$; y e, f para $\varphi = 5000$), con un número bajo ($N = 10$ en a, c, e) y alto ($N = 100$ en b, d, f) de localidades. La proporción de contrastes significativos en simulaciones bajo la hipótesis nula estima el error de tipo I; mientras que bajo la hipótesis alternativa estima la potencia estadística. La distribución estimada de los estadísticos (R_{KLD}^2 o la r de Pearson), y las estimaciones del error de tipo I y la potencia estadística se muestran para modelos exponenciales negativos (amarillo), *power-law* (azul) y modelos lineales para Mantel (rojo).

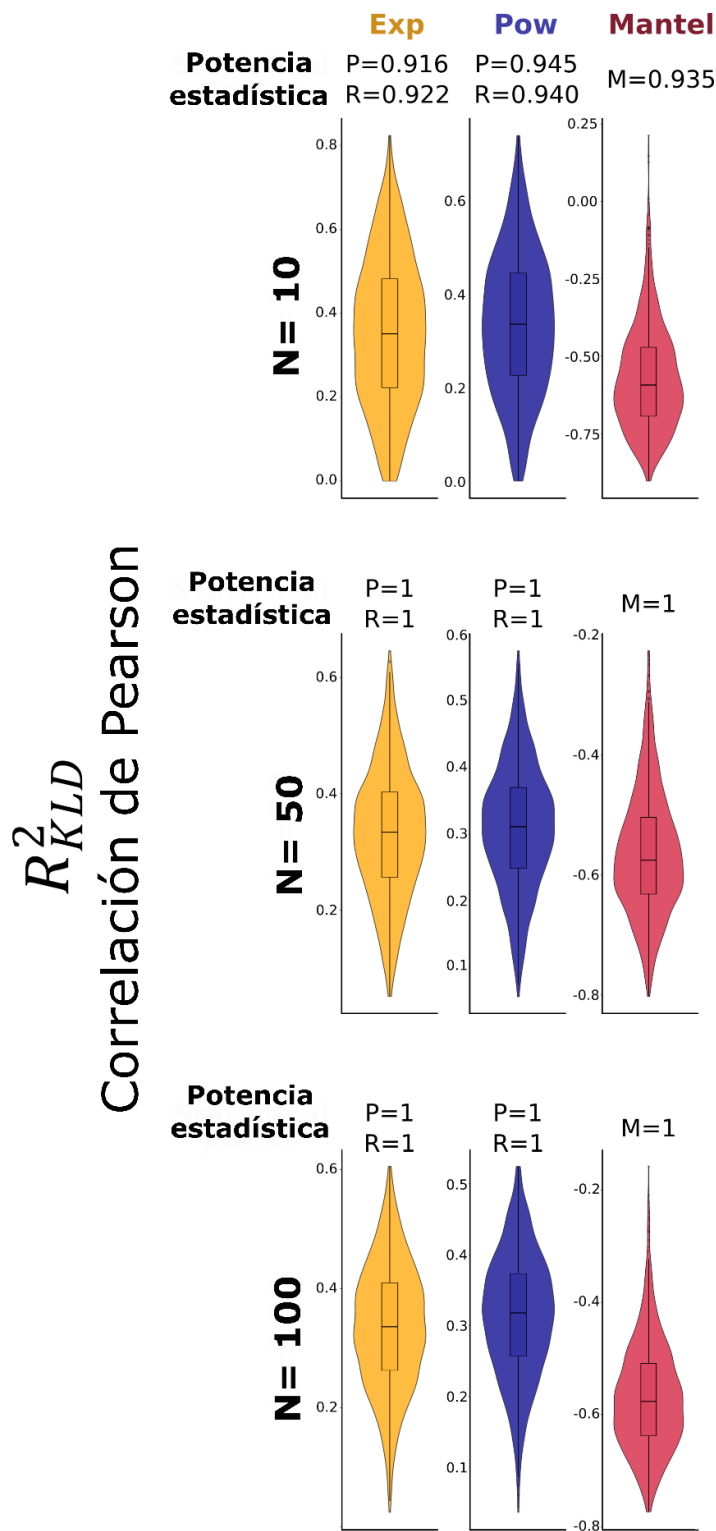


Figura 17. Proporción de contrastes significativos en 1000 simulaciones bajo $\varphi = 40000$ para distintos tamaños muestrales (N). La proporción de pruebas significativas en simulaciones bajo la hipótesis alternativa constituye la potencia estadística del contraste. Simulaciones realizadas con permutaciones por bloques de localidad (P), remuestreo por bloques de localidad (R), y el test de Mantel (M). La distribución estimada de los estadísticos (R^2_{KLD} o r de Pearson en modelos lineales) y la potencia estadística se muestran para los modelos exponencial negativo (amarillo), *power-law* (azul) y lineal de Mantel (rojo).

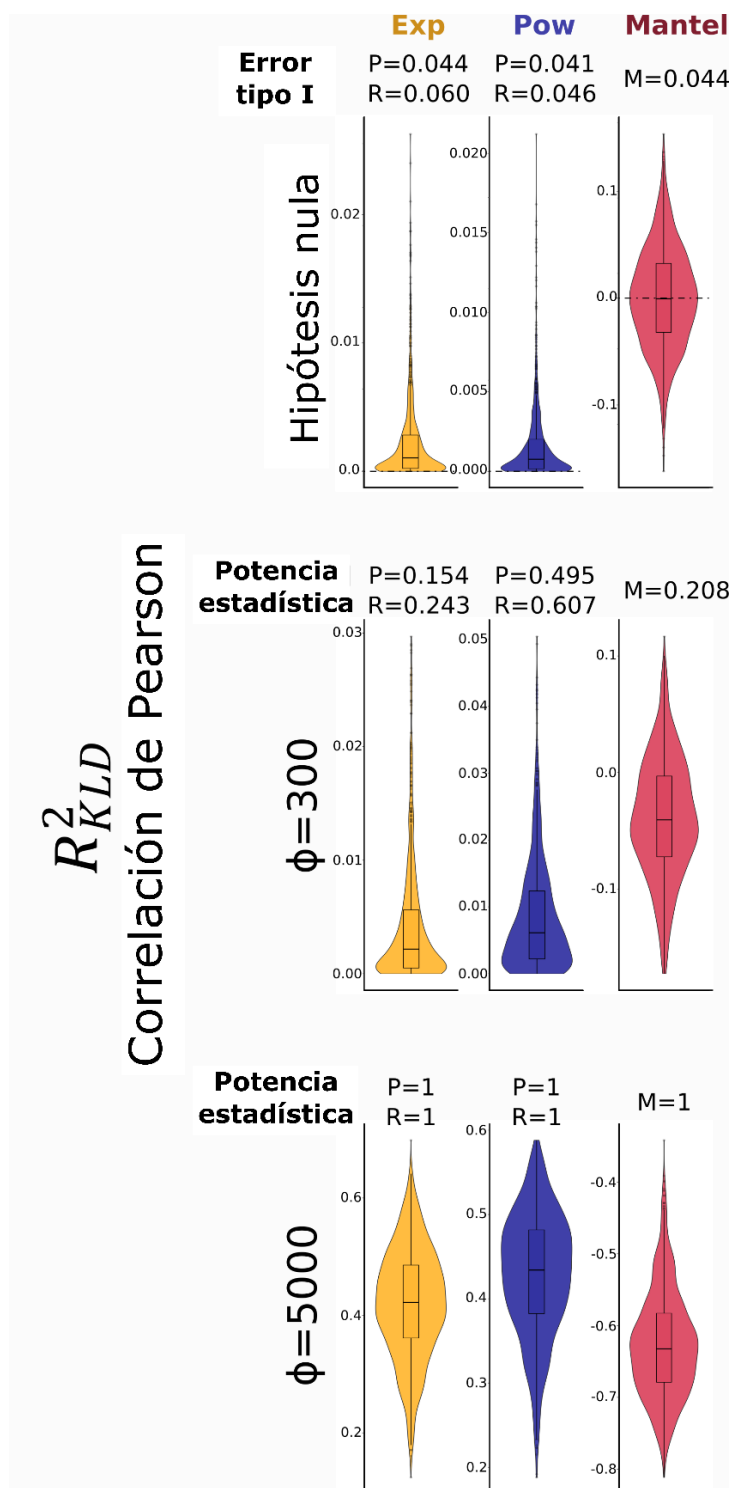


Figura 18. Proporción de contrastes significativos en 1000 simulaciones para las permutaciones por bloques de localidad (P), el remuestreo por bloques de localidad (R) y el test de Mantel (M). Las simulaciones se hicieron para 50 localidades ($N = 50$) bajo la hipótesis nula (error tipo I) y las hipótesis alternativas (potencia estadística para $\phi = 300$ y $\phi = 5000$). La proporción de pruebas significativas en simulaciones bajo la hipótesis nula constituye la estimación del error de tipo I; mientras que la misma proporción calculada para las simulaciones bajo hipótesis alternativas es la potencia estadística del contraste. La distribución estimada de los estadísticos (R_{KLD}^2 o r de Pearson en modelos lineales), el error de tipo I y la potencia estadística se muestran para los modelos exponencial negativo (amarillo), *power-law* (azul) y lineal de Mantel (rojo).

Esto afectó a la potencia estadística, ya que este escenario ($\varphi = 300$) es muy similar a la hipótesis nula. En consecuencia, para simulaciones bajo la hipótesis alternativa el contraste de significación R_{KLD}^2 , ya sea basado en permutaciones o remuestreo por bloques de localidad, fue capaz de forma general de rechazar la hipótesis nula para todas las combinaciones de autocorrelación espacial y tamaño muestral (potencia estadística > 0.9); salvo para el escenario de *distance-decay* con la menor autocorrelación espacial ($\varphi = 300$) (Figuras 16-18). Por el contrario, en $\varphi = 300$ la potencia estadística estimada aumentó con el tamaño muestral, particularmente para el R_{KLD}^2 del modelo *power-law*; aunque todos los contrastes se comportaron de forma pobre para tamaños muestrales extremadamente pequeños ($N = 10$), como se esperaba. El R_{KLD}^2 del modelo exponencial y el test de Mantel, los cuales no pueden acomodar la rápida caída inicial de la similitud, siempre presentaron menor potencia estadística que el R_{KLD}^2 del modelo *power-law* en este escenario.

5.4.2 Aplicación a datos empíricos: coleópteros europeos

El contraste R_{KLD}^2 detectó patrones significativos ($\alpha = 0.05$) de *distance-decay* en la mayoría de los casos de datos reales (Tablas 9 y 10), independientemente de si el modelo aplicado era exponencial, *power-law*, o Gompertz (Figuras 5-6). Los únicos casos en los que el contraste R_{KLD}^2 no rechazó la hipótesis nula fueron Hydraenini, Limnebiini y Scarabaeidae en el norte de Europa, y *Chaetocnema* y Scolytinae en el sur de Europa. En general, hemos encontrado que el modelo exponencial negativo, el *power-law* y el Gompertz no son equivalentes en su ajuste a nuestros datos reales. También se observaron diferencias entre el norte y el sur de Europa. En el norte, el modelo Gompertz fue el que mejor se ajustó a la mayoría de los datos, aunque siempre con valores de R_{KLD}^2 muy similares a los del modelo exponencial, que presentó ajustes muy parecidos. En el sur, observamos mucha mayor variabilidad, con diferentes modelos ofreciendo el mejor ajuste en diferentes taxones. Por ejemplo, el modelo *power-law* presentó mejor ajuste a los datos (mayores valores de \hat{R}_{KLD}^2) en *Chaetocnema*, *Otiorhynchus*, Phyllobiini y *Trechus*. En el resto de los taxones, los modelos exponencial y Gompertz presentaron mejores ajustes, siempre con valores de R_{KLD}^2 similares, aunque en algunos taxones el ajuste del modelo exponencial fue el mejor (Cerambycinae e Hydraenini), y en otros el ajuste de los modelos Gompertz fue mejor (*Amara*, *Carabus*, Chrysomelinae, Cryptocephalinae, Lepturinae, Limnebiini y Silphidae).

Es reseñable que los gráficos de residuos (Figuras 19 y 20) muestran que una gran parte de los conjuntos de datos son heterocedásticos. Esto es particularmente cierto en los casos del norte (todos parecen tener cierto grado de heterocedasticidad salvo Cerambycinae). En cambio, la heterocedasticidad parece ser mucho más baja en los datos del sur de Europa, donde solo algunos casos como Silphidae y *Trechus* presentan una heterocedasticidad acusada. Considerando la alta frecuencia con la que la heterocedasticidad aparece en los diferentes modelos de *distance-decay*, este es un aspecto que debería explorarse en un futuro en el contexto del desarrollo de nuevos métodos de análisis del descenso de la similitud biológica con la distancia geográfica.

Tabla 9. Valores de varianza explicada (R_{KLD}^2) y valores de significación, tanto brutos como resultantes tras la aplicación de corrección de Benjamini & Yekutieli (2001) [BY], para los modelos exponenciales, *power-law* y Gompertz en 21 taxones de coleópteros europeos en el norte de Europa.

	Exponencial			Power-law			Gompertz		
	R_{KLD}^2	p	p [BY]	R_{KLD}^2	p	p [BY]	R_{KLD}^2	p	p [BY]
amara	0.46	0.001	0.005	0.39	0.001	0.005	0.44	0.001	0.007
apion	0.24	0.001	0.005	0.18	0.001	0.005	0.25	0.001	0.007
carab	0.36	0.001	0.005	0.25	0.001	0.005	0.41	0.001	0.007
ceram	0.11	0.001	0.005	0.10	0.001	0.005	0.11	0.003	0.015
chaet	0.18	0.001	0.005	0.15	0.001	0.005	0.19	0.002	0.012
chrys	0.27	0.001	0.005	0.19	0.001	0.005	0.29	0.001	0.007
crypt	0.21	0.001	0.005	0.15	0.001	0.005	0.22	0.001	0.007
hydra	0.02	0.100	0.403	0.01	0.171	0.655	0.02	0.139	0.560
lamin	0.10	0.003	0.014	0.10	0.001	0.005	0.10	0.004	0.019
leptu	0.18	0.001	0.005	0.17	0.001	0.005	0.18	0.001	0.007
limne	0.02	0.252	0.965	0.03	0.151	0.608	0.01	0.275	1.000
ochte	0.32	0.001	0.005	0.22	0.001	0.005	0.34	0.001	0.007
otior	0.14	0.001	0.005	0.12	0.003	0.015	0.14	0.002	0.012
phyll	0.31	0.001	0.005	0.22	0.001	0.005	0.33	0.001	0.007
polyd	0.17	0.001	0.005	0.15	0.001	0.005	0.18	0.001	0.007
psela	0.08	0.036	0.153	0.06	0.056	0.238	0.08	0.030	0.128
ptero	0.32	0.001	0.005	0.24	0.001	0.005	0.34	0.001	0.007
scara	0.00	0.562	1.000	0.01	0.348	1.000	0.00	0.608	1.000
scoly	0.21	0.001	0.005	0.19	0.001	0.005	0.20	0.001	0.007
silph	0.13	0.011	0.050	0.13	0.005	0.024	0.11	0.021	0.095
trech	0.09	0.003	0.014	0.06	0.014	0.063	0.10	0.003	0.015

Tabla 10. Valores de varianza explicada (R_{KLD}^2) y valores de significación, tanto brutos como resultantes tras la aplicación de corrección de Benjamini & Yekutieli (2001) [BY], para los modelos exponenciales, *power-law* y Gompertz en 21 taxones de coleópteros europeos en el sur de Europa.

	Exponencial			Power-law			Gompertz		
	R_{KLD}^2	p	p [BY]	R_{KLD}^2	p	p [BY]	R_{KLD}^2	p	p [BY]
amara	0.42	0.001	0.005	0.36	0.001	0.005	0.43	0.001	0.005
apion	0.12	0.020	0.081	0.08	0.047	0.189	0.12	0.024	0.097
carab	0.76	0.001	0.005	0.62	0.001	0.005	0.78	0.001	0.005
ceram	0.36	0.001	0.005	0.31	0.001	0.005	0.35	0.001	0.005
chaet	0.03	0.176	0.674	0.06	0.054	0.207	0.03	0.221	0.846
chrys	0.58	0.001	0.005	0.43	0.001	0.005	0.61	0.001	0.005
crypt	0.61	0.001	0.005	0.46	0.001	0.005	0.65	0.001	0.005
hydra	0.65	0.001	0.005	0.59	0.001	0.005	0.64	0.001	0.005
lamin	0.21	0.005	0.023	0.21	0.003	0.014	0.21	0.005	0.023
leptu	0.32	0.001	0.005	0.25	0.001	0.005	0.33	0.001	0.005
limne	0.49	0.001	0.005	0.38	0.001	0.005	0.50	0.001	0.005
ochte	0.44	0.001	0.005	0.37	0.001	0.005	0.44	0.001	0.005
otior	0.14	0.003	0.014	0.15	0.002	0.010	0.13	0.003	0.014
phyll	0.18	0.001	0.005	0.19	0.001	0.005	0.18	0.001	0.005
polyd	0.16	0.013	0.055	0.14	0.014	0.060	0.16	0.018	0.077
psela	0.59	0.001	0.005	0.50	0.001	0.005	0.59	0.001	0.005
ptero	0.33	0.001	0.005	0.29	0.001	0.005	0.33	0.001	0.005
scara	0.41	0.001	0.005	0.38	0.001	0.005	0.41	0.001	0.005
scoly	0.01	0.285	1.000	0.01	0.233	0.849	0.01	0.263	0.959
silph	0.27	0.001	0.005	0.15	0.002	0.010	0.34	0.001	0.005
trech	0.40	0.001	0.005	0.47	0.001	0.005	0.39	0.001	0.005

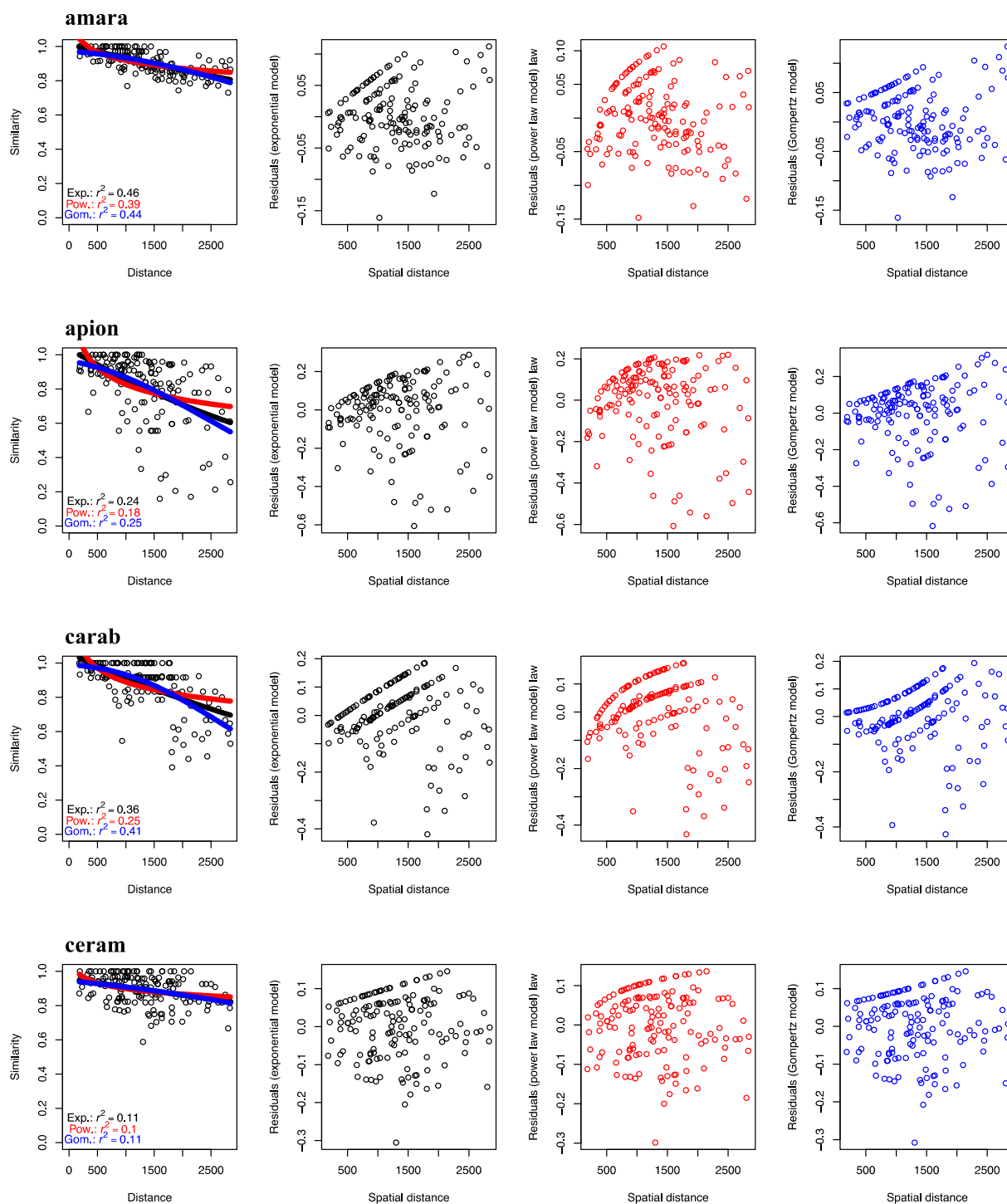


Figura 19. Modelos de *distance-decay* para 21 taxones de coleópteros del norte de Europa. Se muestra el ajuste de las funciones exponencial negativa (negro), *power-law* (rojo) y Gompertz (azul), y sus respectivos valores de varianza explicada (R^2_{KLD}). Para cada uno de esos modelos, se muestra la distribución de los residuos en función de la distancia geográfica.

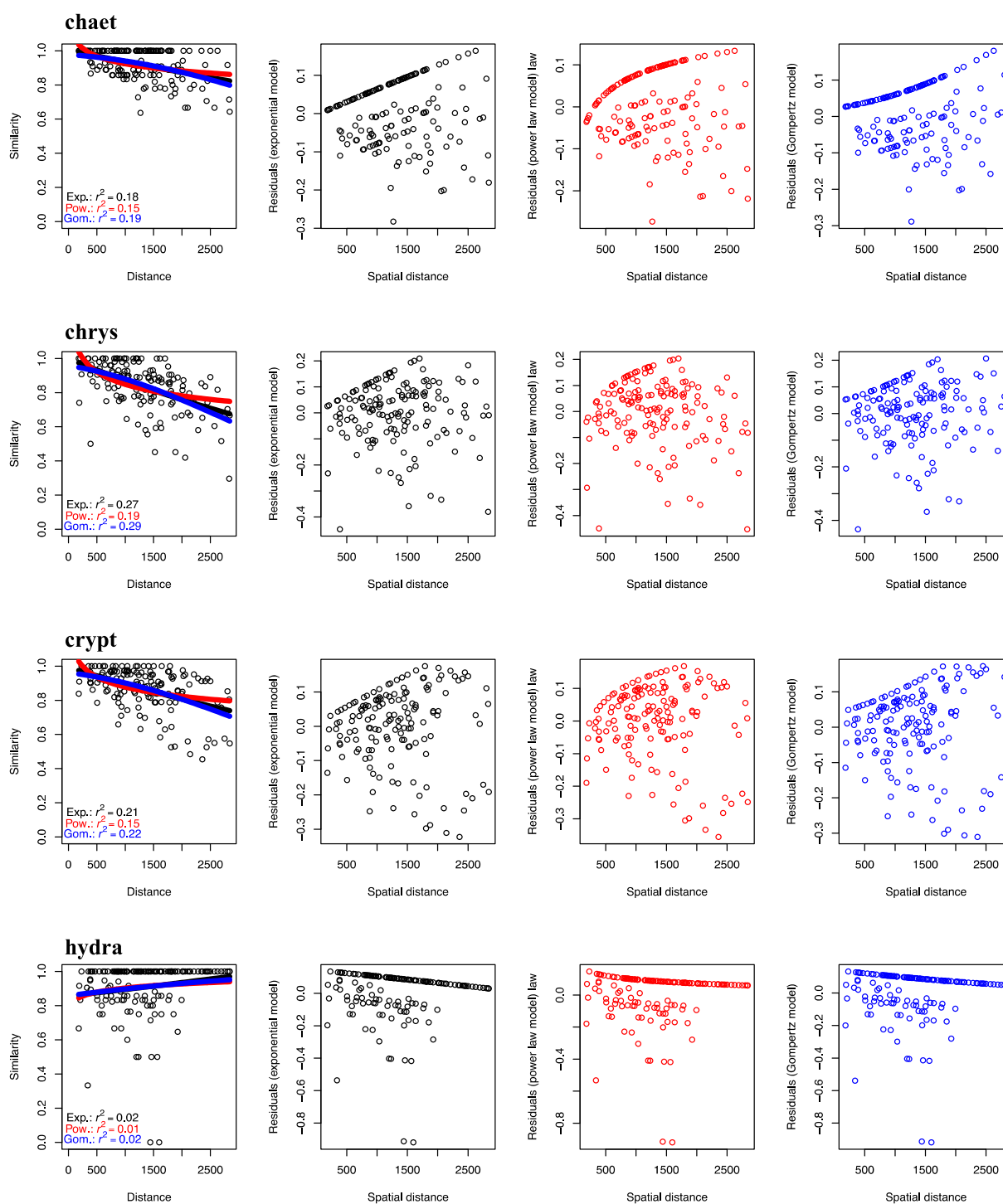


Figura 19 (cont.). Modelos de *distance-decay* para 21 taxones de coleópteros del norte de Europa. Se muestra el ajuste de las funciones exponencial negativa (negro), *power-law* (rojo) y Gompertz (azul), y sus respectivos valores de varianza explicada (R^2_{KLD}). Para cada uno de esos modelos, se muestra la distribución de los residuos en función de la distancia geográfica.

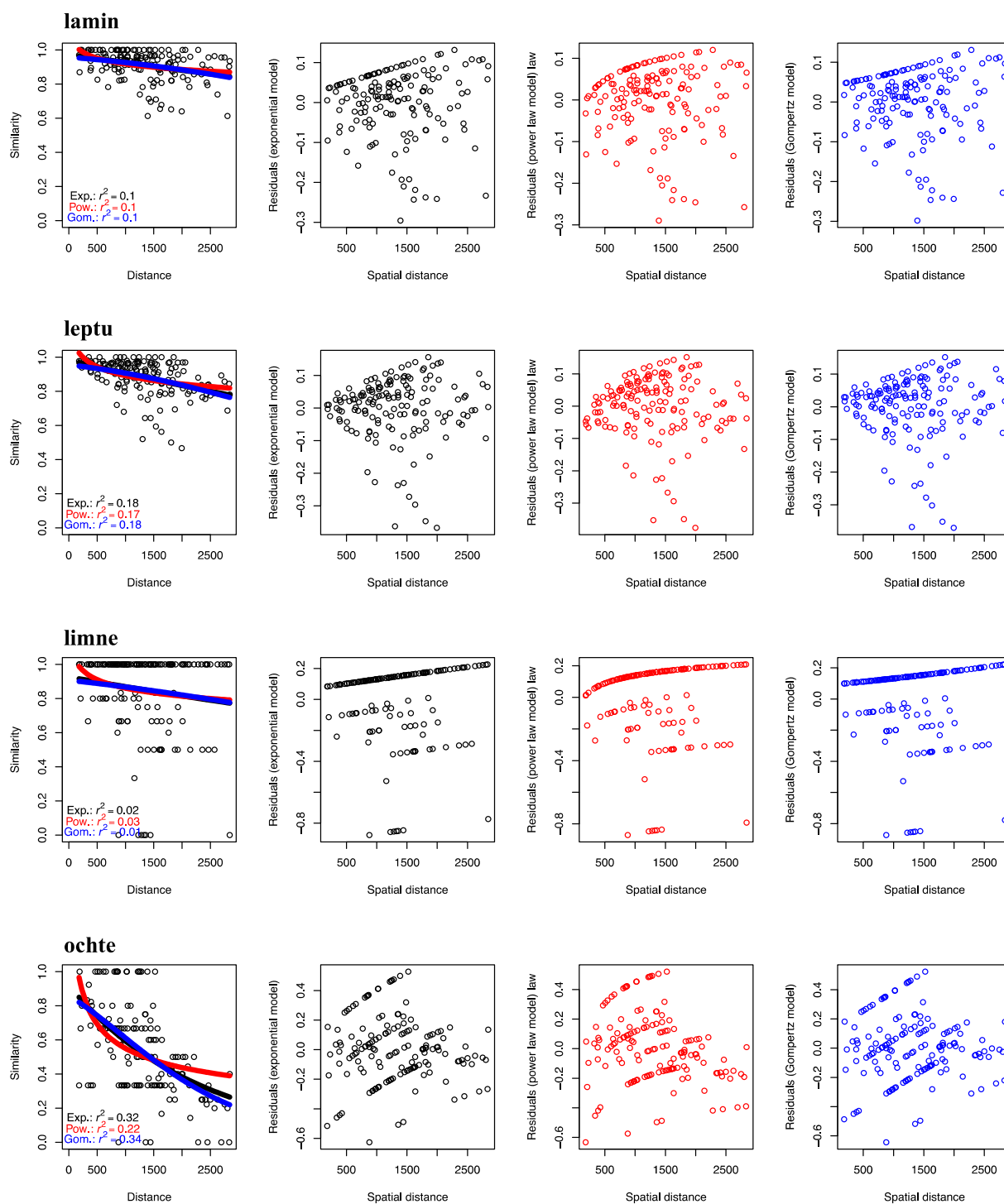


Figura 19 (cont.). Modelos de *distance-decay* para 21 taxones de coleópteros del norte de Europa. Se muestra el ajuste de las funciones exponencial negativa (negro), *power-law* (rojo) y Gompertz (azul), y sus respectivos valores de varianza explicada (R^2_{KLD}). Para cada uno de esos modelos, se muestra la distribución de los residuos en función de la distancia geográfica.

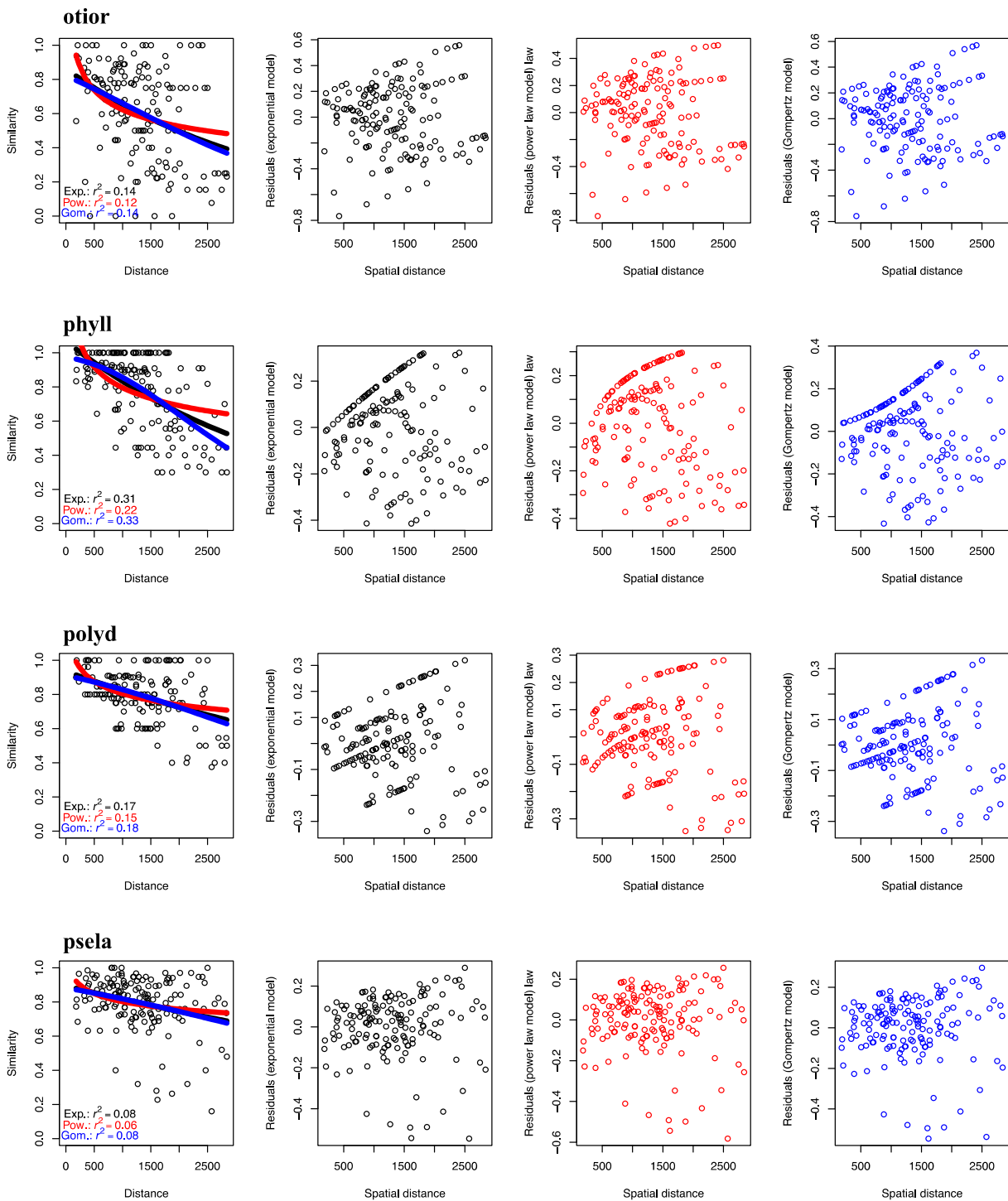


Figura 19 (cont.). Modelos de *distance-decay* para 21 taxones de coleópteros del norte de Europa. Se muestra el ajuste de las funciones exponencial negativa (negro), *power-law* (rojo) y Gompertz (azul), y sus respectivos valores de varianza explicada (R^2_{KLD}). Para cada uno de esos modelos, se muestra la distribución de los residuos en función de la distancia geográfica.

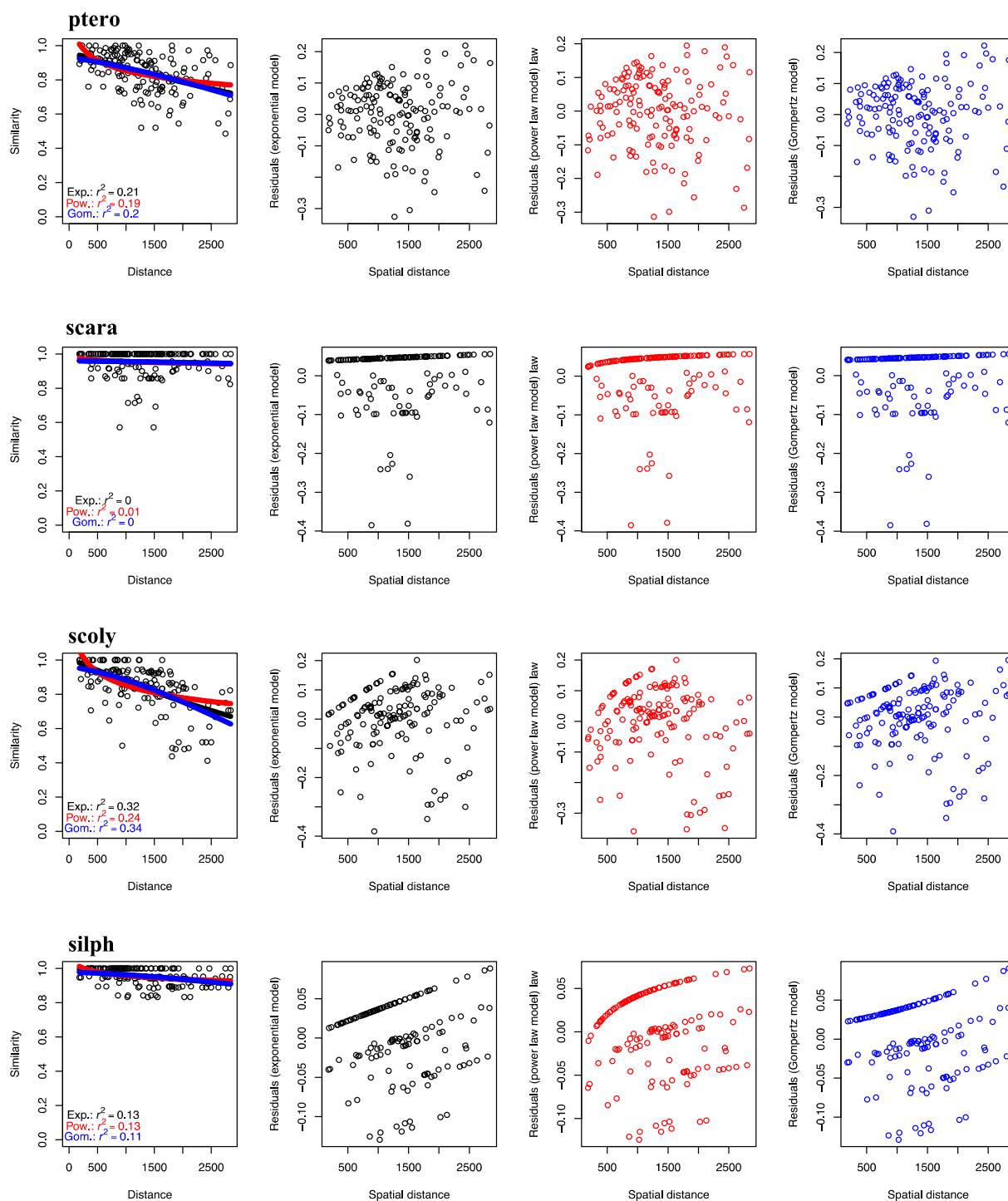


Figura 19 (cont.). Modelos de *distance-decay* para 21 taxones de coleópteros del norte de Europa. Se muestra el ajuste de las funciones exponencial negativa (negro), *power-law* (rojo) y Gompertz (azul), y sus respectivos valores de varianza explicada (R^2_{KLD}). Para cada uno de esos modelos, se muestra la distribución de los residuos en función de la distancia geográfica.

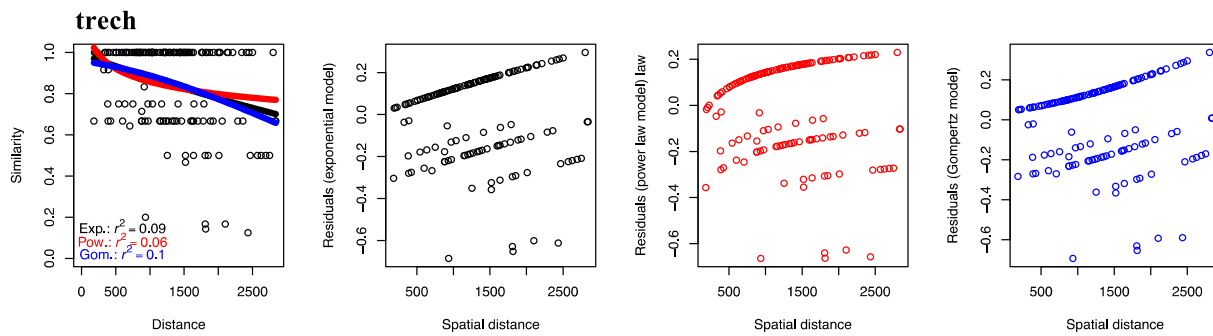


Figura 19 (cont.). Modelos de *distance-decay* para 21 taxones de coleópteros del norte de Europa. Se muestra el ajuste de las funciones exponencial negativa (negro), *power-law* (rojo) y Gompertz (azul), y sus respectivos valores de varianza explicada (R^2_{KLD}). Para cada uno de esos modelos, se muestra la distribución de los residuos en función de la distancia geográfica.

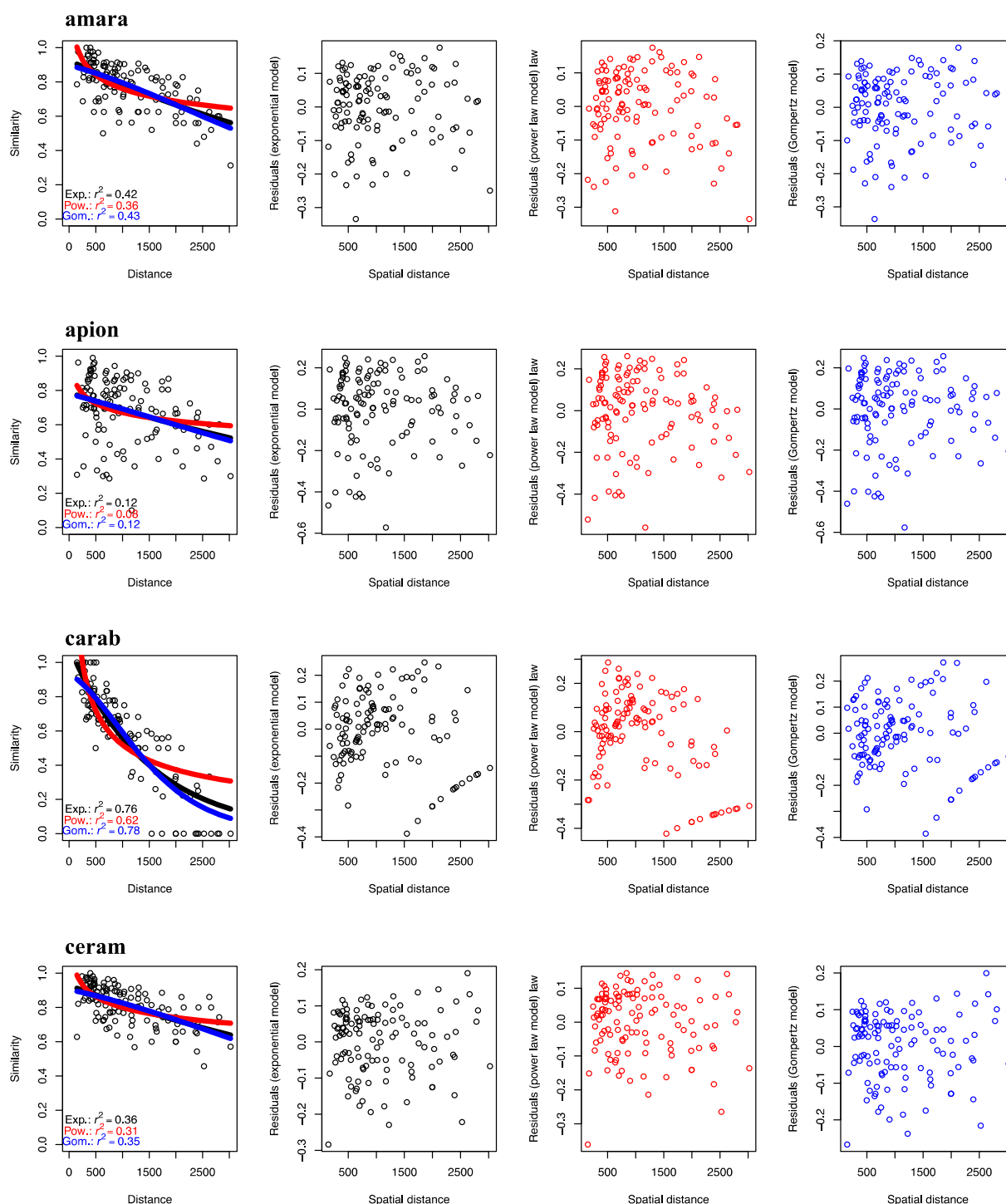


Figura 20. Modelos de *distance-decay* para 21 taxones de coleópteros del sur de Europa. Se muestra el ajuste de las funciones exponencial negativa (negro), *power-law* (rojo) y Gompertz (azul), y sus respectivos valores de varianza explicada (R^2_{KLD}). Para cada uno de esos modelos, se muestra la distribución de los residuos en función de la distancia geográfica.

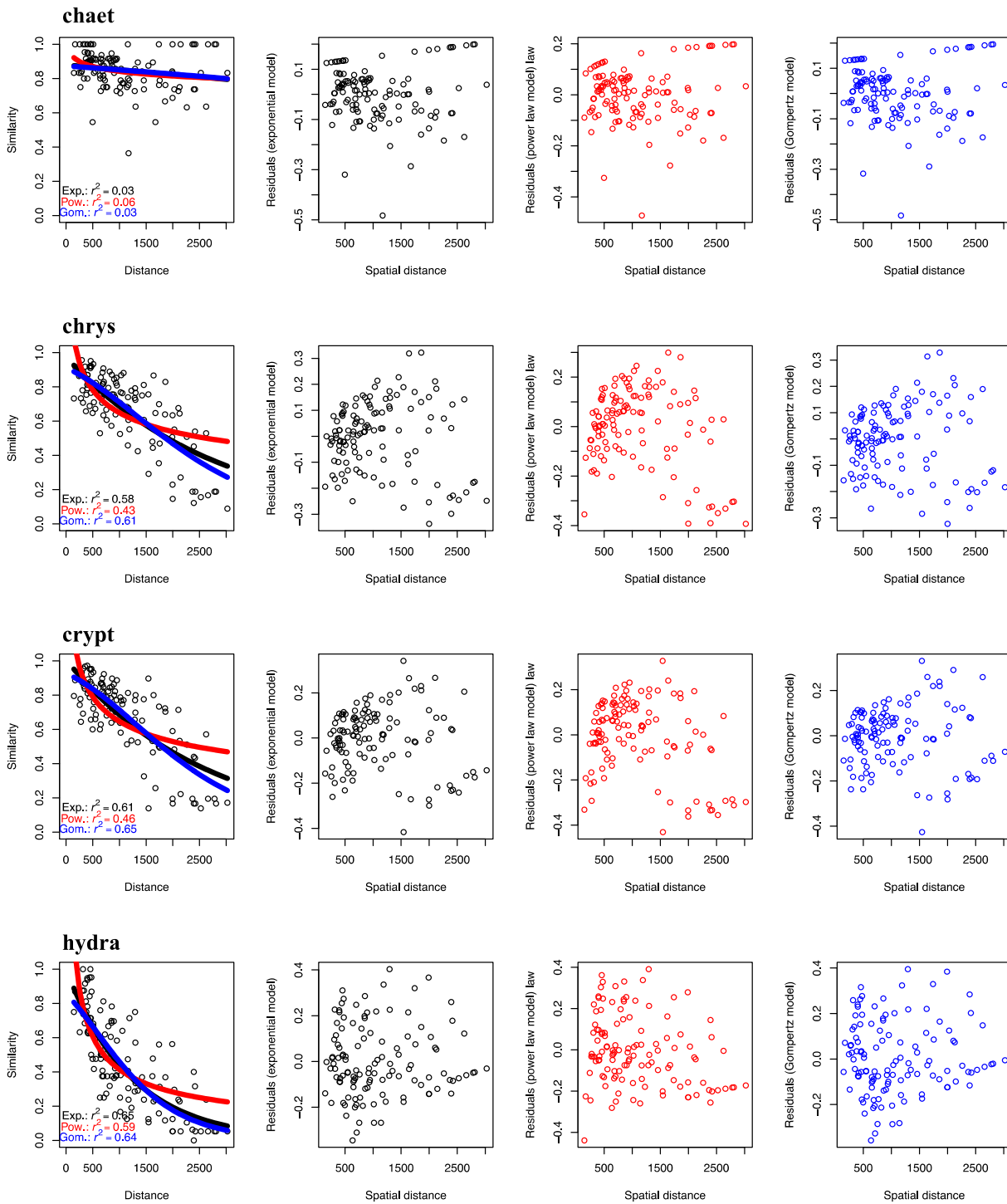


Figura 20 (cont.). Modelos de *distance-decay* para 21 taxones de coleópteros del sur de Europa. Se muestra el ajuste de las funciones exponencial negativa (negro), *power-law* (rojo) y Gompertz (azul), y sus respectivos valores de varianza explicada (R^2_{KLD}). Para cada uno de esos modelos, se muestra la distribución de los residuos en función de la distancia geográfica.

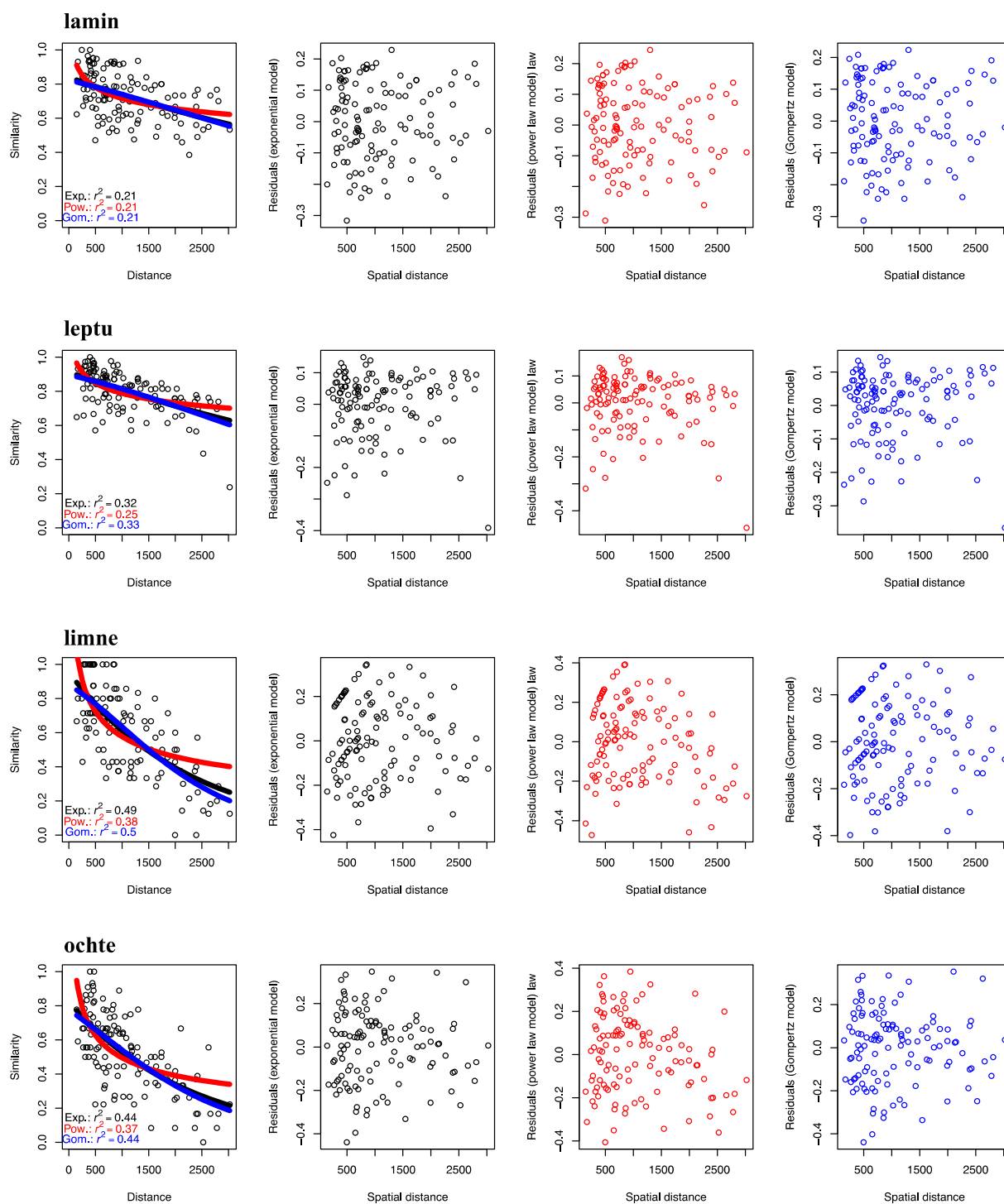


Figura 20 (cont.). Modelos de *distance-decay* para 21 taxones de coleópteros del sur de Europa. Se muestra el ajuste de las funciones exponencial negativa (negro), *power-law* (rojo) y Gompertz (azul), y sus respectivos valores de varianza explicada (R_{KLD}^2). Para cada uno de esos modelos, se muestra la distribución de los residuos en función de la distancia geográfica.

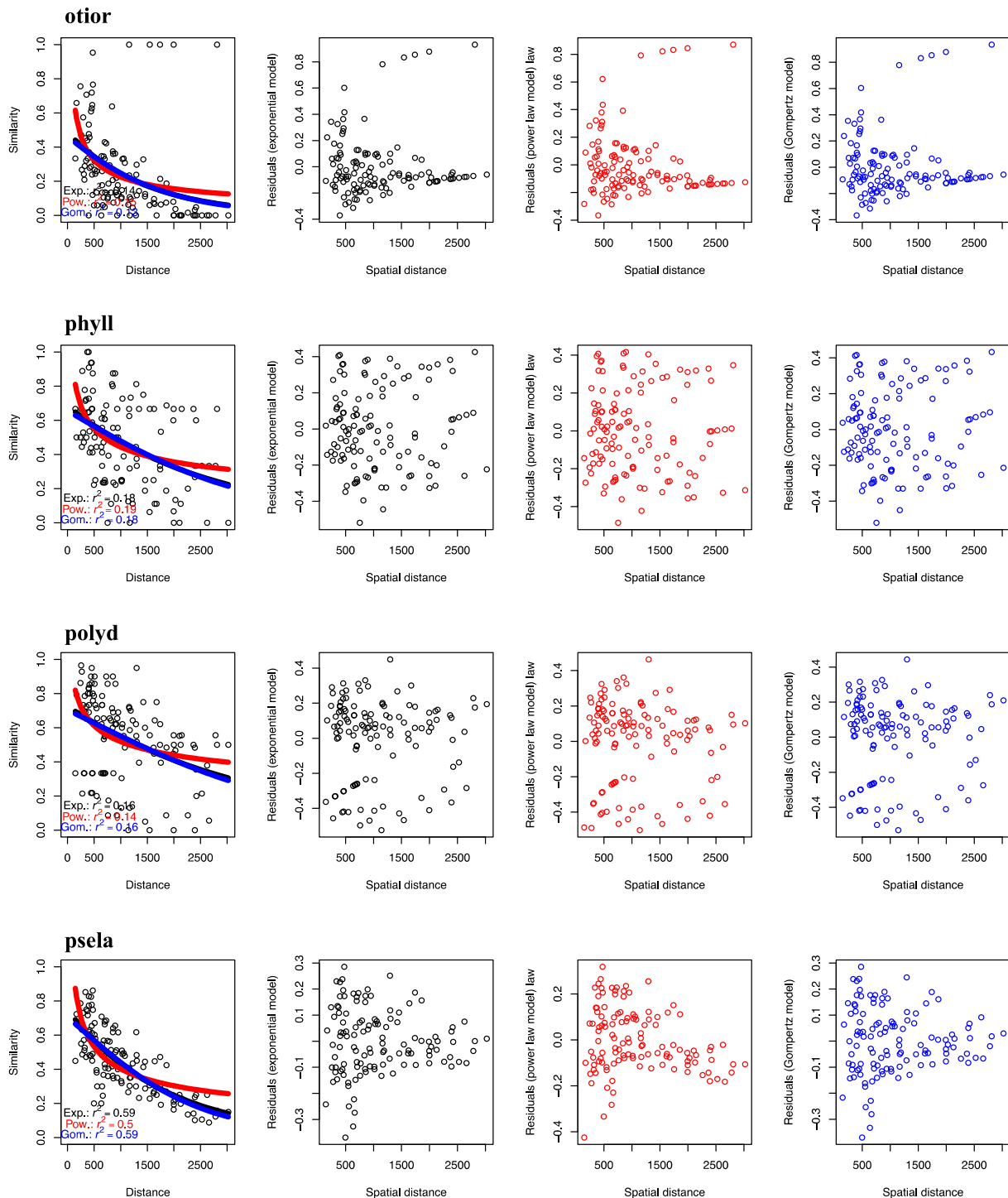


Figura 20 (cont.). Modelos de *distance-decay* para 21 taxones de coleópteros del sur de Europa. Se muestra el ajuste de las funciones exponencial negativa (negro), *power-law* (rojo) y Gompertz (azul), y sus respectivos valores de varianza explicada (R^2_{KLD}). Para cada uno de esos modelos, se muestra la distribución de los residuos en función de la distancia geográfica.

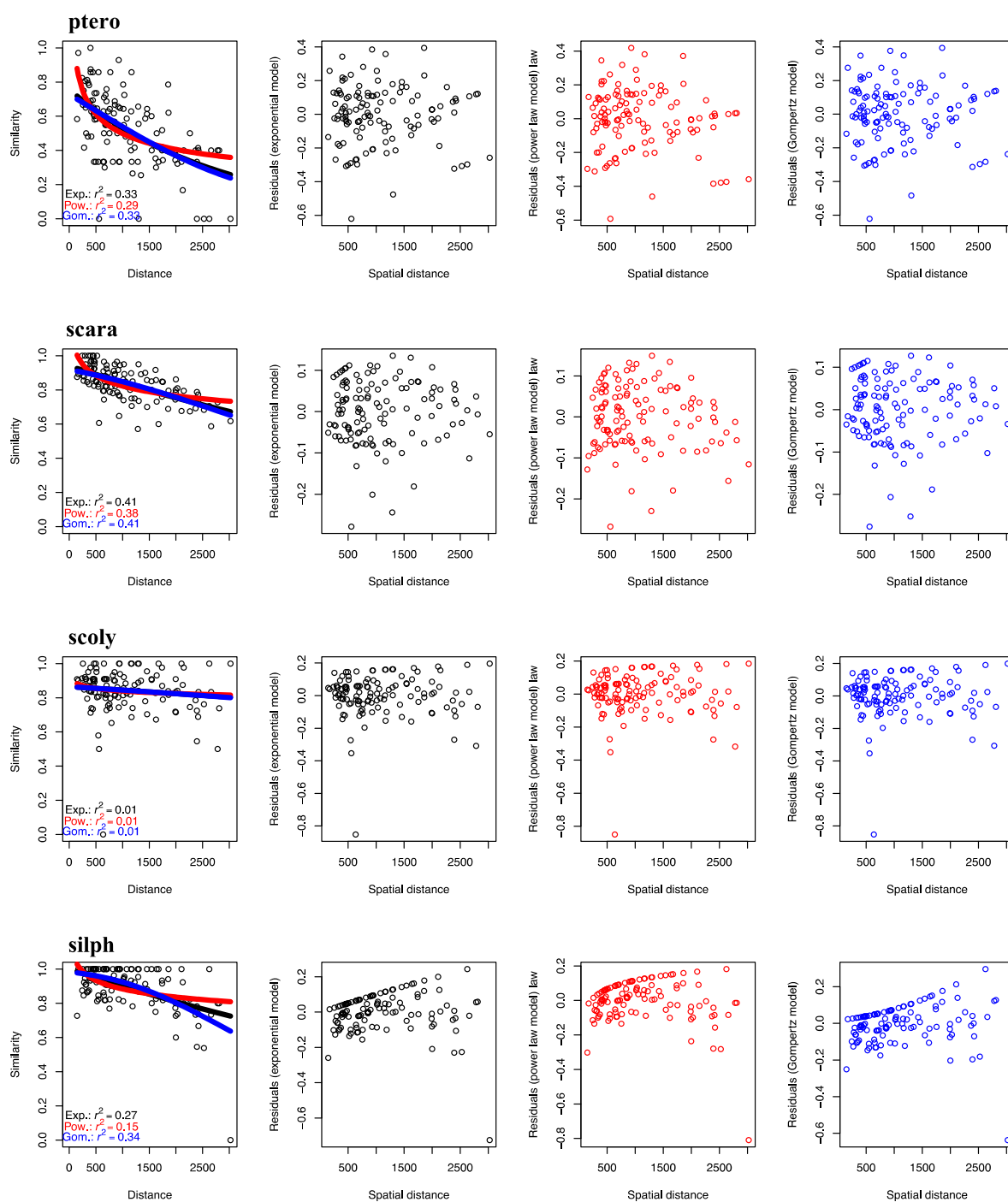


Figura 20 (cont.). Modelos de *distance-decay* para 21 taxones de coleópteros del sur de Europa. Se muestra el ajuste de las funciones exponencial negativa (negro), *power-law* (rojo) y Gompertz (azul), y sus respectivos valores de varianza explicada (R_{KLD}^2). Para cada uno de esos modelos, se muestra la distribución de los residuos en función de la distancia geográfica.

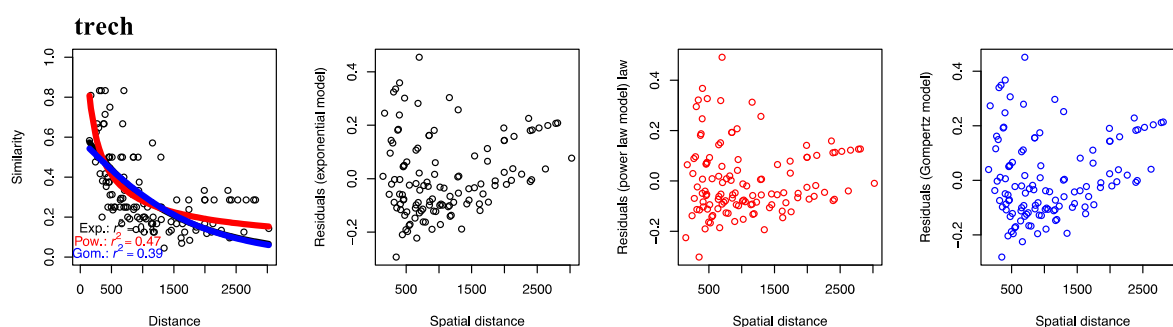


Figura 20 (cont.). Modelos de *distance-decay* para 21 taxones de coleópteros del sur de Europa. Se muestra el ajuste de las funciones exponencial negativa (negro), *power-law* (rojo) y Gompertz (azul), y sus respectivos valores de varianza explicada (R_{KLD}^2). Para cada uno de esos modelos, se muestra la distribución de los residuos en función de la distancia geográfica.

5.4.3 Implementación del contraste estadístico en la librería “betapart” de R

Los métodos de permutaciones y de remuestreo por bloques han sido implementados en la librería *betapart* (Baselga & Orme, 2012), en su versión 1.6 (Baselga et al., 2023), que está disponible en CRAN (<https://cran.r-project.org/package=betapart>):

Package ‘betapart’

March 13, 2023

Type Package

Title Partitioning Beta Diversity into Turnover and Nestedness Components

Version 1.6

Date 2023-03-13

Imports ape, fastmatch, geometry, picante, reddy, doSNOW, foreach, snow, itertools, minpack.lm

Suggests vegan

Description Functions to compute pair-wise dissimilarities (distance matrices) and multiple-site dissimilarities, separating the turnover and nestedness-resultant components of taxonomic (incidence and abundance based), functional and phylogenetic beta diversity.

License GPL (>= 2)

NeedsCompilation no

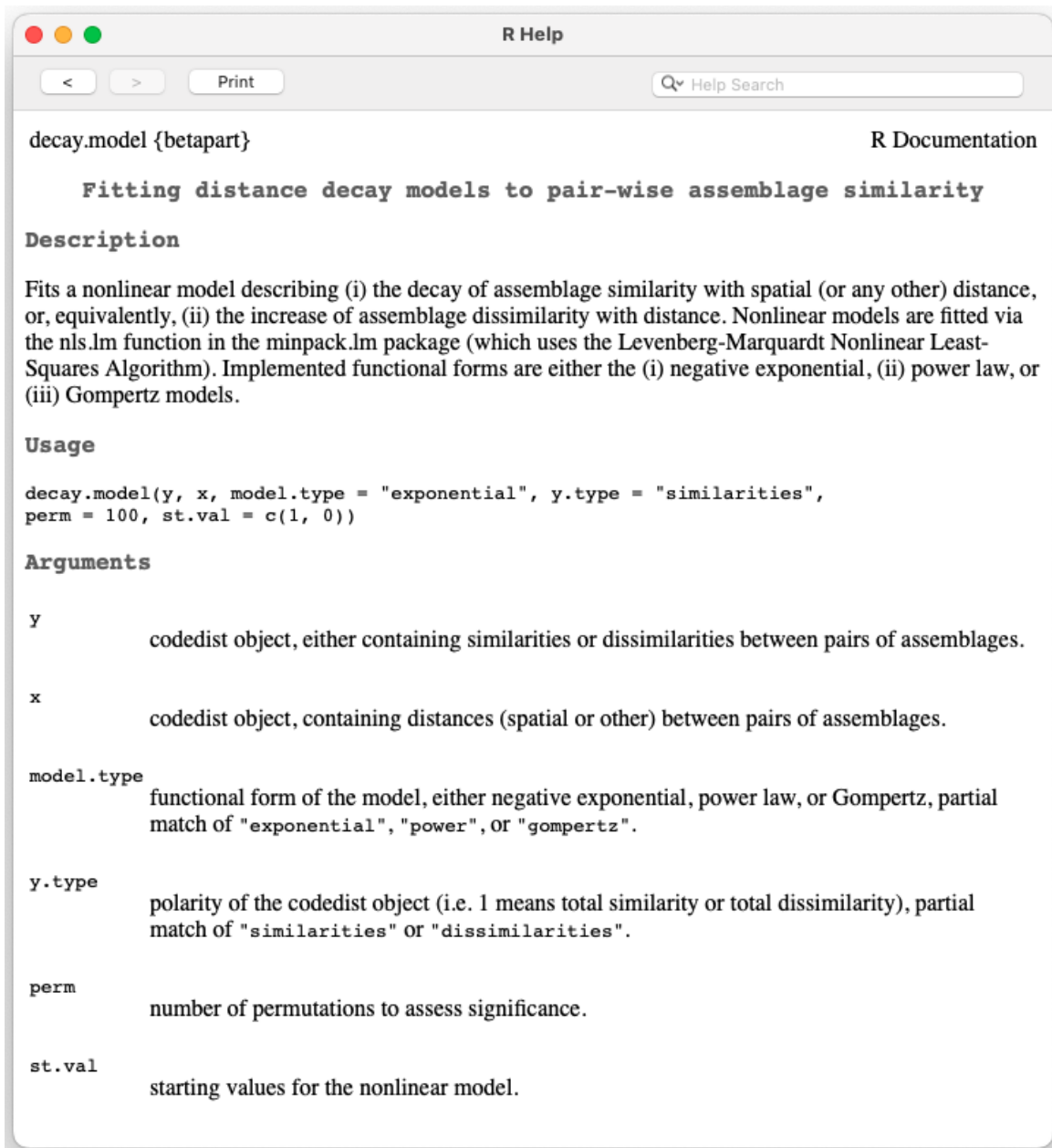
Maintainer Andres Baselga <andres.baselga@usc.es>

Author Andres Baselga [aut, cre],
David Orme [aut],
Sebastien Villegier [aut],
Julien De Bortoli [aut],
Fabien Leprieur [aut],
Maxime Logez [aut],
Sara Martinez-Santalla [aut],
Ramiro Martin-Devasa [aut],
Carola Gomez-Rodriguez [aut],
Rosa M. Crujeiras [aut],
Renato Henriques-Silva [ctb]

Repository CRAN

Date/Publication 2023-03-13 17:10:15 UTC

El método de permutaciones por bloques se utiliza dentro de la función `decay.model()` para estimar la significación de los modelos de *distance-decay*, ya sean exponenciales, *power-law* o Gompertz:



The image shows a screenshot of the R Help window for the `decay.model` function. The window title is "R Help" and it contains navigation buttons (back, forward, print) and a search bar. The main content is the documentation for `decay.model {betapart}`, which includes a title, a description, usage instructions, and a list of arguments with their descriptions.

decay.model {betapart} R Documentation

Fitting distance decay models to pair-wise assemblage similarity

Description

Fits a nonlinear model describing (i) the decay of assemblage similarity with spatial (or any other) distance, or, equivalently, (ii) the increase of assemblage dissimilarity with distance. Nonlinear models are fitted via the `nls.lm` function in the `minpack.lm` package (which uses the Levenberg-Marquardt Nonlinear Least-Squares Algorithm). Implemented functional forms are either the (i) negative exponential, (ii) power law, or (iii) Gompertz models.

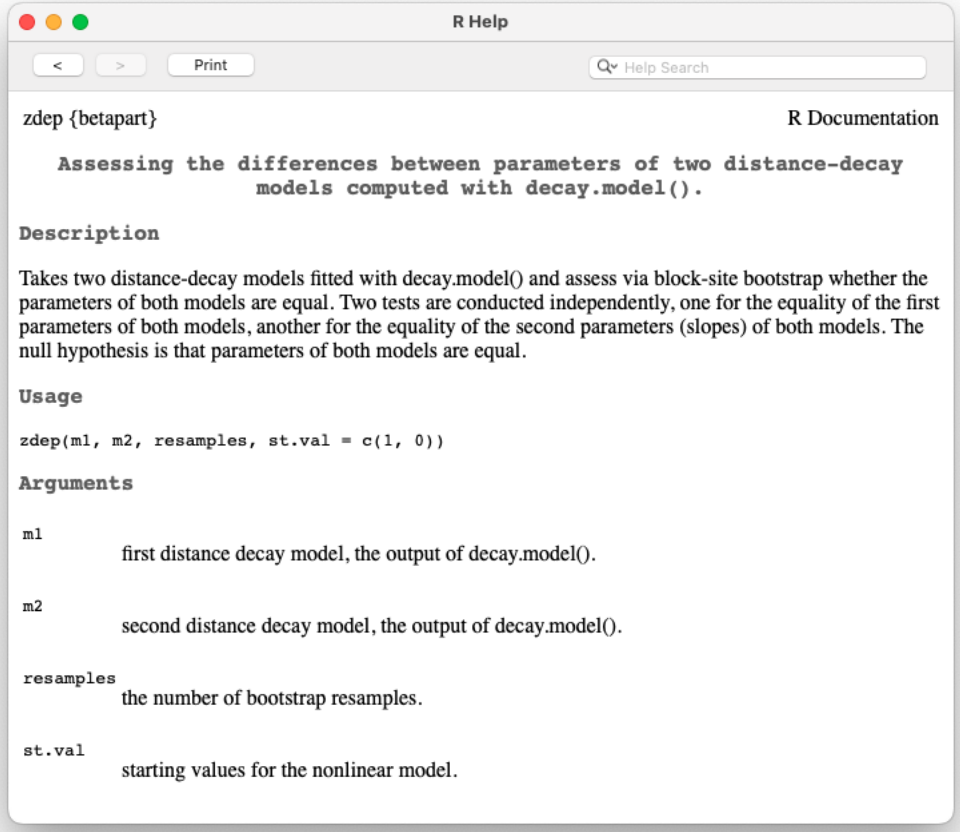
Usage

```
decay.model(y, x, model.type = "exponential", y.type = "similarities",
  perm = 100, st.val = c(1, 0))
```

Arguments

y	codedist object, either containing similarities or dissimilarities between pairs of assemblages.
x	codedist object, containing distances (spatial or other) between pairs of assemblages.
model.type	functional form of the model, either negative exponential, power law, or Gompertz, partial match of "exponential", "power", or "gompertz".
y.type	polarity of the codedist object (i.e. 1 means total similarity or total dissimilarity), partial match of "similarities" or "dissimilarities".
perm	number of permutations to assess significance.
st.val	starting values for the nonlinear model.

El método de remuestreo por bloques se utiliza dentro de la función `zdep()` (Martín-Devasa et al., 2022a) para estimar la significación de la diferencia entre parámetros de dos modelos de *distance-decay*:



The image shows a screenshot of the R Help window for the `zdep` function. The window title is "R Help" and it includes a search bar and navigation buttons. The content is as follows:

zdep {betapart} R Documentation

Assessing the differences between parameters of two distance-decay models computed with `decay.model()`.

Description

Takes two distance-decay models fitted with `decay.model()` and assess via block-site bootstrap whether the parameters of both models are equal. Two tests are conducted independently, one for the equality of the first parameters of both models, another for the equality of the second parameters (slopes) of both models. The null hypothesis is that parameters of both models are equal.

Usage

```
zdep(m1, m2, resamples, st.val = c(1, 0))
```

Arguments

m1 first distance decay model, the output of `decay.model()`.

m2 second distance decay model, the output of `decay.model()`.

resamples the number of bootstrap resamples.

st.val starting values for the nonlinear model.

6 DISCUSIÓN

Esta tesis gira en torno a los modelos de descenso de la similitud biótica con la distancia (modelos *distance-decay*), abarcando tres aspectos fundamentales: la estructura del error de modelos GLM aplicados a patrones *distance-decay*, la exploración de técnicas *bootstrap* para la estimación de la significación en modelos con dependencia por pares, y finalmente la utilización de técnicas de permutación y remuestreo por bloques de localidad para implementar un test de contraste de significación basado en una medida pseudo- R^2 (R_{KLD}^2) para modelos exponenciales y *power-law* del *distance-decay* de la similitud. Además, se muestra también la aplicación de esta nueva aproximación a modelos de *distance-decay* ajustados con una función recientemente propuesta: la función Gompertz (Martín-Devasa et al., 2022b). Estas aportaciones metodológicas de carácter novedoso en el estudio de los patrones *distance-decay* constituyen en sí mismas una solución a uno de los problemas clave en el estudio del *distance-decay* y, en general, de los patrones de diversidad beta inferidos a partir de comparaciones de matrices de disimilitud/distancia: la dependencia por pares de las observaciones. Si tenemos en cuenta que las matrices de (di)similitud biótica son una de las herramientas clave para la cuantificación de las diferencias entre comunidades y, por tanto, para determinar en qué medida esas diferencias contribuyen a la diversidad gamma de una región, esta tesis aporta soluciones metodológicas en uno de los campos fundamentales de estudio del análisis de la diversidad beta.

La primera cuestión que se aborda es qué tipo de estructura del error es más apropiada para modelos GLM aplicados a patrones de *distance-decay*. A pesar de que, desde un punto de vista teórico, la función Gamma podría permitir un mejor ajuste de los modelos de *distance-decay*, se observa que, para la mayoría de los taxones analizados, ajustar un GLM con errores gaussianos y uno con errores Gamma no proporciona resultados diferentes, es decir, los modelos son robustos a diferentes estructuras de error, como también lo muestran Steinbauer et al. (2012) para GLMs de error binomial y gaussiano. Para los taxones donde el ajuste de los modelos difiere, el modelo con errores gaussianos proporciona un pseudo- R^2 mejor y también valores de intercepto biológicamente más realistas (es decir, cercanos a 1). Estos resultados se suman al hecho de que un modelo con errores Gamma no se puede ajustar a datos con valores 0, mientras que el modelo de error gaussiano sí. Por tanto, es recomendable usar una estructura de errores Gaussiana al ajustar un GLM a patrones de *distance-decay*.

Es importante aclarar que, para realizar el análisis de la estructura del error se ha utilizado un amplio conjunto de datos empíricos: la distribución de más de 20 taxones diferentes de coleópteros en territorios europeos, que ha permitido ajustar una gran variedad de patrones de *distance-decay* debido a la gran variedad de tolerancias fisiológicas, rasgos morfológicos y rasgos vitales existentes entre los coleópteros. Esta gran variedad de patrones de *distance-decay*, especialmente cuando se diferencia la región norte y sur y de Europa, había sido

mostrada previamente por otros autores (Gómez-Rodríguez & Baselga, 2018). Sin embargo, existía la posibilidad de que estas diferencias observadas previamente estuvieran determinadas por el tipo de modelo ajustado, el cual no estaba diseñado para tener en cuenta la dependencia por pares en el contraste de la significación. Por ello, y para evitar que los análisis de la tesis se basaran en conjuntos de datos con menor variabilidad de la esperada inicialmente, se realizó un análisis no paramétrico preliminar que confirmó que existían diferencias marcadas entre taxones y entre la región del norte y del sur de Europa. Por ello, todos los análisis de esta tesis han considerado la separación norte-sur de los datos empíricos.

La siguiente sección puede considerarse un primer intento de resolver la estimación de la significación de los modelos no lineales de *distance-decay*, que están afectados por la pseudo-replicación de los casos (e inflación de los grados de libertad) debido a que a partir de N comunidades biológicas se obtienen $N * (N-1)$ valores de similitud (o distancia), que son los casos utilizados en los modelos *distance-decay*. Es evidente, entonces, que los modelos incluyen muchos más casos [$N * (N-1)$] que observaciones independientes [N]. Por ello, era necesario explorar la utilidad de las técnicas estadísticas propuestas por autores previos para corregir los problemas asociados a la dependencia de los datos en los modelos de *distance-decay*. En primer lugar, se valoró el método *bootstrap* de bloques móviles según Künsch (1989) y Liu & Singh (1992), el cual fue diseñado para corregir la autocorrelación de las observaciones. La autocorrelación se produce cuando es probable que una observación de la muestra original sea más semejante a las adyacentes a ella que a las más lejanas y que su propio valor haya influido en el de esas observaciones adyacentes. Esta aproximación no corrige por tanto la dependencia por pares que existe en los datos de *distance-decay*, ya que los valores de similitud biótica tienen dependencia entre ellos no sólo porque las comunidades a partir de las que fueron calculados puedan estar próximas entre sí, sino, sobre todo, porque la misma comunidad está involucrada en el cálculo de varios índices de similitud. Es decir, los mismos datos de presencia/ausencia o de abundancia de especies en una comunidad determinada se usan para el cálculo de múltiples valores de similitud biótica, uno por cada comunidad con la que son comparados. Por tanto, no se puede considerar que su propio valor haya influido sólo en el de esas observaciones adyacentes, ya que los valores de similitud calculados también con las comunidades más alejadas tienen dependencia con los valores de similitud calculados con las comunidades más próximas. En segundo lugar, se exploró el *bootstrap* por bloques siguiendo la idea propuesta por Sherman & Cessie (1997), en la que los datos se agrupan según algún factor de bloqueo que pueda ser responsable de la dependencia por pares, siendo por tanto un método en principio prometedor.

Las técnicas de *bootstrap* propuestas por autores previos se han explorado asociadas a un test de contraste basado en la *deviance* de los modelos. En esta aproximación inicial, no se pudo concluir que el *bootstrap* por bloques, ya sean móviles o manuales, funcione correctamente en el proceso de estimación de un GLM, ya que la implementación que se realizó no mostraba que la dependencia que presentan los datos de similitud biológica y distancia geográfica se corrigiese. Se observó que el método de *bootstrap* con bloques móviles no corrige la inflación del error tipo I, que también se observaba al usar contrastes que no pretendían corregir la

dependencia en los datos, como el contraste estándar de significación (test de la F). Esto se debe, como se ha comentado anteriormente, a que los bloques móviles están diseñados para acomodar la dependencia espacial de los datos, pero no la dependencia entre pares que es intrínseca a las matrices de similitud y distancia. Por el contrario, las simulaciones diseñadas en esta exploración inicial sugirieron que, en contra de lo predicho por el efecto de la dependencia por pares, el test de la F , y el *bootstrap* de bloques móviles produjeron errores de tipo I más próximos al valor nominal ($\alpha = 0.05$) que el *bootstrap* con bloques manuales, que se mostró demasiado conservador (error de Tipo I < 0.05) pese a ser, desde un punto de vista conceptual, más apropiado. Fue por ello necesario revisar el procedimiento de simulación de datos y estudiar cuál es el sesgo de esta estimación de error de tipo I en datos de *distance-decay*, que es lo que se resolvió en la última sección de esta tesis. Dentro del estudio de las técnicas de *bootstrap* propuestas por autores previos también se pudo determinar que, para elegir si se ajusta mejor a los datos de similitud biológica y distancia geográfica un GLM exponencial o uno *power-law* es indistinto utilizar los valores de AIC del GLM ajustado sin *bootstrap* o las frecuencias de selección que devuelven cualquiera de los dos métodos de *bootstrap* propuestos. Por tanto, se puede concluir que el diseño de simulaciones apropiadas es clave para determinar la significación de los modelos, pero no para la selección del mejor modelo.

En la última sección se evaluó el funcionamiento de un contraste de significación basado en una medida de pseudo- R^2 (R_{KLD}^2) para modelos exponenciales y *power-law* del *distance-decay* de la similitud, utilizando permutaciones por bloques o remuestreos por bloques de localidad para estimar la distribución de la pseudo- R^2 bajo la hipótesis nula, y de esa manera poder estimar la significación del modelo. Para ello, se realizaron simulaciones de datos de comunidades biológicas en el espacio bajo la hipótesis nula y la hipótesis alternativa. En la hipótesis nula (no hay patrón de *distance-decay*) todas las especies tienen la misma probabilidad de estar presentes en cualquier sitio, y la composición específica de una comunidad no depende de su localización espacial. En la hipótesis alternativa (existe un *distance-decay* no lineal de la similitud entre comunidades) las especies están agregadas espacialmente y la composición específica de una comunidad se correlaciona, en mayor o menor grado, con la de otras comunidades cercanas. Además, en ambos casos se tuvo en cuenta la variabilidad en los rangos de distribución de las especies en condiciones reales, asegurando que las simulaciones generaran comunidades biológicas en las que había una alta proporción de especies de distribución reducida y pocas de distribución amplia. Además, para testar la significación de los modelos, también fue necesario simular la distribución bajo la hipótesis nula del estadístico utilizado, la pseudo- R^2 basada en la divergencia de Kullback-Leibler descrita por Cameron & Windmeijer (1997). La distribución de este estadístico es desconocida en un contexto de *distance-decay* debido a la dependencia por pares de los datos y, en esta tesis se ha generado mediante métodos de permutación o remuestreo por bloques de localidad. La clave ha sido, por tanto, incorporar la dependencia a pares presente en las variables de similitud y distancia al considerar todos los valores de similitud/distancia relacionados con un sitio particular como la unidad básica del remuestreo. Estos conjuntos de datos (todos los valores de similitud/distancia

en los que una comunidad/sitio determinado está implicado) han sido denominados en esta tesis “bloques de localidad”.

Los resultados mostraron que el contraste de significación R_{KLD}^2 funciona bien para patrones no lineales de *distance-decay*, complementado con el uso de permutaciones o remuestreo por bloques de localidad para estimar la distribución de R_{KLD}^2 bajo la hipótesis nula. En otras palabras, tanto las permutaciones como el remuestreo por bloques de localidad incorporaron la dependencia a pares de los valores de similitud y distancia en sus estimaciones, corrigiendo la inflación del error de tipo I que aparecería si se hubieran usado contrastes estándar de significación, como se mostró en la sección anterior. Por tanto, en esta tesis se propone un método novedoso que posee importantes ventajas en el estudio de los patrones de *distance-decay*: (i) proporciona una medida de bondad de ajuste, el estadístico pseudo- R^2 (R_{KLD}^2), que se puede interpretar de forma sencilla como la proporción de variación en la similitud entre comunidades explicada por la distancia espacial (o predictores alternativos con estructura pareada, como medidas de distancia climática); (ii) proporciona una estimación robusta de la significación estadística del modelo usado, con errores de tipo I y potencia estadística apropiados; y (iii) permite la implementación de modelos lineales generalizados (GLMs) para ajustar relaciones de *distance-decay*, con todas las ventajas derivadas de dichos GLMs tales como la comparación de modelos mediante AIC o la estimación de parámetros de relevancia biológica (p.ej. la pendiente del *distance-decay*), que pueden ser comparados entre distintas regiones o grupos taxonómicos.

El contraste de significación R_{KLD}^2 tuvo, por tanto, un rendimiento óptimo bajo condiciones de ausencia y presencia de *distance-decay*, independientemente de la estrategia de calibración empleada para estimar su hipótesis nula (permutaciones o remuestreo por bloques de localidad). Es decir, su error de tipo I fue bajo y su potencia estadística alta, ya fuera usando permutaciones o remuestreo por bloques de localidad. El test de Mantel funcionó correctamente para detectar patrones de *distance-decay* cuando dicho patrón estaba claramente presente a lo largo de todo el rango de distancia y podía aproximarse a un modelo lineal. Sin embargo, el test de Mantel presentó una potencia estadística reducida cuando el patrón era curvo y menos marcado (datos simulados con $\varphi = 300$), como se espera en los patrones de *distance-decay*, ya que autores previos han mostrado que la relación funcional que mejor los captura es la exponencial-negativa, la *power-law* o la Gompertz (Martín-Devasa et al., 2022b; Nekola & McGill, 2014). Nos gustaría remarcar que, además, se sabe *a priori* que los modelos lineales no son apropiados para índices de similitud/disimilitud, ya que están definidos entre 0 y 1 y esto lleva a que su relación con distancias espaciales o ambientales sean no lineales y asintóticas. Por esta razón, el test de Mantel puede fallar en la detección de una relación de *distance-decay* si esta se aleja de la linealidad (Diniz-Filho et al., 2013; Zeller et al., 2016). Esta limitación del test de Mantel se ejemplifica en el escenario de *distance-decay* simulado con la autocorrelación espacial más débil ($\varphi = 300$). En este escenario, la similitud desciende rápidamente a distancias muy cortas y llega pronto a su asíntota inferior (el patrón es fuertemente no lineal). Como se esperaba, el R_{KLD}^2 del modelo *power-law* detecta de forma más eficaz el patrón de *distance-decay* (mayor potencia estadística) que el R_{KLD}^2 del modelo exponencial o el modelo lineal (test de Mantel)

debido a que la función *power-law* puede acomodar un *distance-decay* acusado desde el inicio. Las debilidades del test de Mantel como contraste de significación para patrones de *distance-decay* marcadamente no lineales se pueden observar también en el análisis de los datos empíricos. Por ejemplo, en el ejemplo más pronunciado y menos lineal de los casos reales de *distance-decay* estudiados (*Trechus*), se encuentran diferencias en el ajuste de los distintos modelos, siendo el modelo *power-law* el que presentó el valor más alto de *deviance* explicada. Estos dos resultados muestran la importancia de incorporar a las herramientas de ecología de comunidades y biogeografía un contraste de significación para modelos no lineales de *distance-decay* ya que el uso de un contraste de significación lineal para un modelo no lineal podría llevar a no detectar el patrón de *distance-decay* que realmente estamos modelando.

Los nuevos métodos que se proponen en esta tesis están sujetos a algunas limitaciones. En primer lugar, es importante reconocer de forma explícita que un valor pseudo- R^2 no sustituye las pruebas diagnósticas necesarias sobre las asunciones del modelo (tales como homocedasticidad o normalidad de los residuos). Su uso como medida de varianza explicada debe estar condicionado, por tanto, a la validación estadística de estas asunciones o al conocimiento de que el modelo es adecuado para los datos de interés y la pregunta de investigación (Nekola & McGill, 2014), como se recomienda en el trabajo de Nakagawa & Schielzeth (2013). Adicionalmente, el estadístico pseudo- R^2 puede ser demasiado optimista como estimación de la varianza explicada para tamaños muestrales pequeños y en estos casos su valor debería corregirse (Heinzel et al., 2005).

En segundo lugar, la estructura espacial de otras diferencias en las características de las comunidades o de las localidades (p.ej. la distancia climática entre ellas) también podría afectar al funcionamiento del contraste. En este sentido, cabe destacar la versatilidad del uso de R_{KLD}^2 en combinación con permutaciones o remuestreo por bloques de localidad, el cual podría ser aplicado como contraste de significación en otro tipo de modelos no lineales de *distance-decay*, por ejemplo cuando los predictores no son distancias en el espacio geográfico sino entre variables climáticas (Astorga et al., 2012; Saito et al., 2015), en el tiempo (Matsuoka et al., 2016), etc. En estudios equivalentes hechos con modelos lineales, el uso del test de Mantel con variables de distancia con estructura espacial (p.ej. una variable de distancia climática) ha sido criticado por varios autores porque ni el test original ni el test de Mantel parcial (Smouse et al., 1986) son capaces de acomodar el sesgo introducido por esta estructura espacial (Guillot & Rousset, 2013; Legendre et al., 2015). Por ello, aunque es en principio viable, debemos suponer que si se utiliza el R_{KLD}^2 en modelos construidos con variables predictoras estructuradas que no sean distancias espaciales, seguramente presente el mismo sesgo. Una posible solución para variables con estructura espacial podría ser la modificación del contraste siguiendo el trabajo de Crabot et al. (2019) para el test de Mantel. Estos autores han propuesto un método basado en aleatorizaciones espacialmente restringidas para corregir la inflación de su error de tipo I debida a la autocorrelación espacial presente en las variables. Finalmente, el remuestreo por bloques de localidad que se describe en esta tesis también podría ser utilizado para abordar otras cuestiones relacionadas con el ajuste de los modelos de *distance-decay* y que puedan verse afectadas por la estructura pareada de los datos. Algunos ejemplos de aplicaciones serían la

estimación de la varianza de parámetros de regresión para contrastes de comparación de modelos o la estimación del AIC y la incertidumbre en la selección de modelos al tiempo que se considera la estructura real de los datos de *distance-decay* (Fieberg et al., 2020).

En conclusión, esta tesis ha desarrollado una metodología de ajuste y contraste de significación de modelos de *distance-decay* utilizando una medida de pseudo- R^2 basada en la *deviance*, R_{KLD}^2 , en combinación con permutaciones o remuestreo por bloques de localidad. La metodología que aquí se propone constituye una opción no lineal que debería considerarse como una alternativa preferible al uso del test de Mantel, sea lineal o no paramétrico. Adicionalmente, tanto el estadístico R_{KLD}^2 como el método de remuestreo por bloques de localidad propuesto tienen potencial para ser adaptados a distintas preguntas y aproximaciones sobre el estudio de patrones no lineales de *distance-decay*, tales como la evaluación del efecto de factores no espaciales (p.ej. variables climáticas) en la similitud entre comunidades o la estimación de la varianza de parámetros de los modelos (p.ej. la tasa a la cual desciende la similitud a medida que aumenta la distancia). Los métodos desarrollados han sido implementados en la librería “betapart” de R (Baselga et al., 2023), para que estén disponibles para toda la comunidad científica.

7 CONCLUSIONES

1. Un análisis preliminar (ANCOVA no paramétrico) de los datos de similitud entre las faunas de países europeos y su relación con la distancia espacial para 21 taxones de coleópteros mostró que la relación entre la similitud biológica y la distancia geográfica es marcadamente distinta entre la región del norte y del sur de Europa, tanto en términos de igualdad o paralelismo de los patrones, como incluso en términos de heterocedasticidad. Por tanto, los patrones de *distance-decay* de coleópteros europeos deben analizarse separadamente en el norte y el sur de Europa.
2. Para la mayoría de los casos de estudio analizados (patrones de *distance-decay* en 21 taxones de coleópteros, analizando de manera independiente el patrón en el norte y el sur de Europa), el ajuste de un GLM con errores gaussianos y uno con errores gamma no proporciona resultados diferentes, es decir, los modelos son robustos a diferentes estructuras de error.
3. Para los casos de estudio donde el ajuste de los modelos difiere, el modelo con errores gaussianos proporciona un mejor ajuste a los datos (mayor pseudo- R^2) y también interceptos biológicamente más realistas (es decir, cercanos a 1).
4. Además, los modelos GLM con errores gamma no se pueden ajustar a datos con valores 0, que son posibles en datos de similitud, mientras que el modelo de error gaussiano sí. Por tanto, es preferible que los modelos de *distance-decay* sean ajustados con GLMs con estructura gaussiana en el error.
5. En la exploración inicial de los métodos *bootstrap* disponibles en la literatura el método de *bootstrap* con bloques móviles no parece corregir la inflación del error tipo I, que también se observa al usar contrastes estándar de significación (test de la F). Esto podría deberse a que los bloques móviles están diseñados para acomodar la dependencia espacial de los datos, pero no la dependencia entre pares que es intrínseca a los matrices de similitud y distancia.
6. Las simulaciones diseñadas en el análisis inicial mostraron que, en contra de lo predicho por el efecto de la dependencia por pares, el test de la F , y el *bootstrap* de bloques móviles produjeron errores de tipo I más próximos al valor nominal ($\alpha = 0.05$) que el *bootstrap* con bloques manuales, que se mostró demasiado conservador (error de tipo I < 0.05). Esto sugirió la necesidad de desarrollar un nuevo tipo de simulaciones que reprodujesen de manera precisa la dependencia por pares de las similitudes.

7. El desarrollo de un nuevo método de simulación hipótesis, de manera que se asegure que los datos de similitud presentan dependencia por pares similar a la de datos empíricos, permitió desarrollar a su vez un contraste de significación basado en la pseudo- R^2 para modelos no lineales de descenso de la similitud biótica. Esta combinación de métodos demostró que esta aproximación para el contraste de la significación es adecuada para modelos exponenciales y *power-law*, y preferible al test de Mantel, que está diseñado para relaciones lineales.
8. La estimación de la significación basada bien en permutaciones por bloques de localidad, o bien en remuestreo por bloques de localidad, produce niveles de error de tipo I muy cercanos al alfa nominal, y valores de error de tipo II muy bajos.
9. Además de ser adecuado para la modelización de los patrones de *distance-decay*, este contraste de significación podría aplicarse a otros tipos de regresión no lineal que se basen en distancias, por ejemplo, la relación entre la distancia genética entre individuos de diferentes poblaciones y la distancia espacial que separa dichas poblaciones.
10. El nuevo método de remuestreo por bloques de localidad también podría aplicarse a otras preguntas estadísticas sobre modelos de *distance-decay* que no pueden ser respondidas usando permutaciones, como la estimación de la varianza de parámetros del modelo. Esta estimación es clave para poder desarrollar tests de igualdad entre los parámetros de dos modelos de *distance-decay* (para dos taxones diferentes, por ejemplo).
11. Los métodos desarrollados en esta tesis han sido implementados en la librería “betapart” de R, mostrando su adecuación a modelos no lineales con diferentes formas (exponencial negativa, *power-law*, o Gompertz).

8 BIBLIOGRAFÍA

- Anderson, M. J., Crist, T. O., Chase, J. M., Vellend, M., Inouye, B. D., Freestone, A. L., Sanders, N. J., Cornell, H. V., Comita, L. S., Davies, K. F., Harrison, S. P., Kraft, N. J. B., Stegen, J. C., & Swenson, N. G. (2011). Navigating the multiple meanings of beta diversity: a roadmap for the practicing ecologist. *Ecology Letters*, *14*(1), 19-28.
- Arribas, P., Andújar, C., Salces-Castellano, A., Emerson, B. C., & Vogler, A. P. (2020). The limited spatial scale of dispersal in soil arthropods revealed with whole-community haplotype-level metabarcoding. *Molecular Ecology*, *30*, 48-61.
- Astorga, A., Oksanen, J., Luoto, M., Soininen, J., Virtanen, R., & Muotka, T. (2012). Distance decay of similarity in freshwater communities: do macro- and microorganisms follow the same rules? *Global Ecology and Biogeography*, *21*(3), 365-375.
- Baselga, A. (2010a). Multiplicative partition of true diversity yields independent alpha and beta components, additive partition does not. *Ecology*, *91*(7), 1974-1981.
- Baselga, A. (2010b). Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, *19*(1), 134-143.
- Baselga, A., Fujisawa, T., Crampton-Platt, A., Bergsten, J., Foster, P. G., Monaghan, M. T., & Vogler, A. P. (2013). Whole-community DNA barcoding reveals a spatio-temporal continuum of biodiversity at species and genetic levels. *Nature Communications*, *4*, 1892.
- Baselga, A., Gómez-Rodríguez, C., Araújo, M. B., Castro-Insua, A., Arenas, M., Posada, D., & Vogler, A. P. (2022). Joint analysis of species and genetic variation to quantify the role of dispersal and environmental constraints in community turnover. *Ecography*, *2022*(5), e05808.
- Baselga, A., Gómez-Rodríguez, C., & Lobo, J. M. (2012). Historical legacies in world amphibian diversity revealed by the turnover and nestedness components of beta diversity. *PLOS ONE*, *7*(2), e32341.
- Baselga, A., Gómez-Rodríguez, C., & Vogler, A. P. (2015). Multi-hierarchical macroecology at species and genetic levels to discern neutral and non-neutral processes. *Global Ecology and Biogeography*, *24*, 873-882.
- Baselga, A., Lobo, J. M., Svenning, J. C., Aragón, P., & Araújo, M. B. (2012). Dispersal ability modulates the strength of the latitudinal richness gradient in European beetles. *Global Ecology and Biogeography*, *21*, 1106-1113.
- Baselga, A., & Orme, C. D. L. (2012). betapart: an R package for the study of beta diversity. *Methods in Ecology and Evolution*, *3*, 808-812.
- Baselga, A., Orme, C. D. L., Villéger, S., De Bortoli, J., Leprieur, F., Logez, M., Martínez-Santalla, S., Martín-Devasa, R., Gómez-Rodríguez, C., & Crujeiras, R. M. (2023). *betapart: Partitioning beta diversity into turnover and nestedness components. R package version 1.6 Available at: <http://CRAN.R-project.org/package=betapart>.*

- Bell, T. (2010). Experimental tests of the bacterial distance–decay relationship. *The ISME Journal*, 4(11), 1357-1365.
- Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4), 1165-1188.
- Bowman, A., & Azzalini, A. (2018). *sm: Smoothing Methods for Nonparametric Regression and Density Estimation*. R package version 2.2-5.5 <https://cran.r-project.org/web/packages/sm/sm.pdf>. Accedido 4 de septiembre de 2018.
- Bray, J. R., & Curtis, J. T. (1957). An ordination of the upland forest communities of Southern Wisconsin. *Ecological Monographs*, 27(4), 325-349.
- Buckley, L. B., & Jetz, W. (2008). Linking global turnover of species and environments. *Proceedings of the National Academy of Sciences of the United States of America*, 105(46), 17836-17841.
- Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Inference: A Practical Information-Theoretic Approach*. Springer-Verlag.
- Callaway, R. M., Brooker, R. W., Choler, P., Kikvidze, Z., Lortie, C. J., Michalet, R., Paolini, L., Pugnaire, F. I., Newingham, B., Aschehoug, E. T., Armas, C., Kikodze, D., & Cook, B. J. (2002). Positive interactions among alpine plants increase with stress [Article]. *Nature*, 417(6891), 844-848.
- Cameron, C. A., & Windmeijer, F. A. G. (1997). An R-squared measure of goodness of fit for some common nonlinear regression models. *Journal of Econometrics*, 77(2), 329-342.
- Chase, J. M., & Leibold, M. A. (2003). *Ecological Niches: Linking Classical and Contemporary Approaches*. University of Chicago Press.
- Chase, J. M., & Myers, J. A. (2011). Disentangling the importance of ecological niches from stochastic processes across scales. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1576), 2351-2363.
- Chust, G., Villarino, E., Chenuil, A., Irigoien, X., Bizsel, N., Bode, A., Broms, C., Claus, S., Fernández de Puelles, M. L., Fonda-Umani, S., Hoarau, G., Mazzocchi, M. G., Mozetič, P., Vandepitte, L., Veríssimo, H., Zervoudaki, S., & Borja, A. (2016). Dispersal similarly shapes both population genetics and community patterns in the marine realm. *Scientific Reports*, 6(1), 28730.
- Condamine, F. L., Soldati, L., Clamens, A. L., Rasplus, J. Y., & Kergoat, G. J. (2013). Diversification patterns and processes of wingless endemic insects in the Mediterranean Basin: historical biogeography of the genus *Blaps* (Coleoptera: Tenebrionidae). *Journal of Biogeography*, 40(10), 1899-1913.
- Crabot, J., Clappe, S., Dray, S., & Datry, T. (2019). Testing the Mantel statistic with a spatially-constrained permutation procedure. *Methods in Ecology and Evolution*, 10(4), 532-540.
- Dias, M., Oberdorff, T., Huguény, B., Leprieur, F., Jézéquel, C., Cornu, J.-F., Brosse, S., Grenouillet, G., & Tedesco, P. A. (2014). Global imprint of historical connectivity on freshwater fish biodiversity. *Ecology Letters*, 17, 1130-1140.
- Dietz, E. J. (1983). Permutation tests for association between two distance matrices. *Systematic Biology*, 32(1), 21-26.
- Diniz-Filho, J. A., Soares, T. N., Lima, J. S., Dobrovolski, R., Landeiro, V. L., de Campos Telles, M. P., Rangel, T. F., & Bini, L. M. (2013). Mantel test in population genetics. *Genetics and Molecular Biology*, 36(4), 475-485.

- Diniz-Filho, J. A. F., & Bini, L. M. (2011). Geographical patterns in biodiversity: towards an integration of concepts and methods from genes to species diversity. *Natureza & Conservação*, 9(2), 179-187.
- Diserud, O. H., & Ødegaard, F. (2007). A multiple-site similarity measure. *Biology Letters*, 3(1), 20-22.
- Draper, F. C., Baraloto, C., Brodrick, P. G., Phillips, O. L., Martinez, R. V., Honorio Coronado, E. N., Baker, T. R., Zárate Gómez, R., Amasifuen Guerra, C. A., Flores, M., Garcia Villacorta, R., V. A. Fine, P., Freitas, L., Monteagudo-Mendoza, A., J. W Brienens, R., & Asner, G. P. (2019). Imaging spectroscopy predicts variable distance decay across contrasting Amazonian tree communities. *Journal of Ecology*, 107(2), 696-710.
- Dynesius, M., & Jansson, R. (2000). Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *Proceedings of the National Academy of Sciences of the United States of America*, 97(16), 9115-9120.
- Elzhov, T. V., Mullen, K. M., Spiess, A.-N., & Bolker, B. (2016). *minpack.lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds*. <https://CRAN.R-project.org/package=minpack.lm>.
- Faraway, J. J. (2006). Generalized Linear Models. In J. J. Faraway (Ed.), *Extending the Linear Model with R. Generalized Linear, Mixed effects and Nonparametric Regression Models* (pp. 126-135). Chapman and Hall/CRC.
- Fieberg, J. R., Vitense, K., & Johnson, D. H. (2020). Resampling-based methods for biologists. *PeerJ*, 8, e9089.
- Fitzenberger, B. (1998). The moving blocks bootstrap and robust inference for linear least squares and quantile regressions. *Journal of Econometrics*, 82(2), 235-287.
- Fitzpatrick, M. C., Sanders, N. J., Normand, S., Svenning, J.-C., Ferrier, S., Gove, A. D., & Dunn, R. R. (2013). Environmental and historical imprints on beta diversity: insights from variation in rates of species turnover along gradients. *Proceedings of the Royal Society B: Biological Sciences*, 280(1768), 20131201.
- Freijeiro, A., & Baselga, A. (2016). Spatial and environmental correlates of species richness and turnover patterns in European cryptocephaline and chrysomeline beetles. *Zookeys*, 597, 81-99.
- Gaston, K. J. (1996). Species-range-size distributions: patterns, mechanisms and implications. *Trends in Ecology & Evolution*, 11(5), 197-201.
- Goeman, J. J., & Solari, A. (2014). Multiple hypothesis testing in genomics. *Statistics in Medicine*, 33(11), 1946-1978.
- Gómez-Rodríguez, C., & Baselga, A. (2018). Variation among European beetle taxa in patterns of distance decay of similarity suggests a major role of dispersal processes. *Ecography*, 41, 1825-1834.
- Gómez-Rodríguez, C., Freijeiro, A., & Baselga, A. (2015). Dispersal and ecological traits explain differences in beta diversity patterns of European beetles. *Journal of Biogeography*, 42, 1526-1537.
- Gómez-Rodríguez, C., Miller, K. E., Castillejo, J., Iglesias-Piñeiro, J., & Baselga, A. (2019). Understanding dispersal limitation through the assessment of diversity patterns across phylogenetic scales below the species level. *Global Ecology and Biogeography*, 28, 353-364.

- Gompertz, B. (1833). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London*, 2, 252-253.
- Graco-Roza, C., Aarnio, S., Abrego, N., Acosta, A. T. R., Alahuhta, J., Altman, J., Angiolini, C., Aroviita, J., Attorre, F., Baastrup-Spohr, L., Barrera-Alba, J. J., Belmaker, J., Biurrun, I., Bonari, G., Bruelheide, H., Burrascano, S., Carboni, M., Cardoso, P., Carvalho, J. C., . . . Soininen, J. (2022). Distance decay 2.0 – A global synthesis of taxonomic and functional turnover in ecological communities. *Global Ecology and Biogeography*, 31(7), 1399-1421.
- Graham, C. H., Parra, J. L., Rahbek, C., & McGuire, J. A. (2009). Phylogenetic structure in tropical hummingbird communities. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 19673-19678.
- Graham, C. H., Parra, J. L., Tinoco, B. A., Stiles, F. G., & McGuire, J. A. (2012). Untangling the influence of ecological and evolutionary factors on trait variation across hummingbird assemblages. *Ecology*, 93(8), S99-S111.
- Guillot, G., & Rousset, F. (2013). Dismantling the Mantel tests. *Methods in Ecology and Evolution*, 4(4), 336-344.
- Hastie, T. J., & Pregibon, D. (1993). Generalized Linear Models. In J. M. Chambers & T. J. Hastie (Eds.), *Statistical Models in S* (pp. 195-249). Chapman & Hall.
- Hawkins, B. A., & Porter, E. E. (2003). Relative influences of current and historical factors on mammal and bird diversity patterns in deglaciated North America. *Global Ecology and Biogeography*, 12(6), 475-481.
- Heinzel, H., Waldhör, T., & Mittlböck, M. (2005). Careful use of pseudo R-squared measures in epidemiological studies. *Statistics in Medicine*, 24(18), 2867-2872.
- Hurlbert, S. H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecological Monographs*, 54(2), 187-211.
- Hutchinson, G. E. (1957). Population studies - Animal ecology and demography - Concluding remarks [Article]. *Cold Spring Harbor Symposia on Quantitative Biology*, 22, 415-427.
- Jaccard, P. (1912). The distribution of the flora in the alpine zone. *New Phytologist*, 11, 37-50.
- Jackson, P. L., Straussfogel, D., Lindgren, B. S., Mitchell, S., & Murphy, B. (2008). Radar observation and aerial capture of mountain pine beetle, *Dendroctonus ponderosae* Hopk. (Coleoptera: Scolytidae) in flight above the forest canopy. *Canadian Journal of Forest Research*, 38(8), 2313-2327.
- Jansson, R. (2003). Global patterns in endemism explained by past climatic change. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 270(1515), 583-590.
- Jiménez-Valverde, A., Baselga, A., Melic, A., & Txasko, N. (2010). Climate and regional beta-diversity gradients in spiders: dispersal capacity has nothing to say? *Insect Conservation and Diversity*, 3(1), 51-60.
- Jost, L. (2007). Partitioning diversity into independent alpha and beta components. *Ecology*, 88(10), 2427-2439.
- Jurasinski, G., Retzer, V., & Beierkuhnlein, C. (2009). Inventory, differentiation, and proportional diversity: a consistent terminology for quantifying species diversity. *Oecologia*, 159(1), 15-26.
- Koleff, P., Gaston, K. J., & Lennon, J. K. (2003). Measuring beta diversity for presence-absence data. *Journal of Animal Ecology*, 72, 367-382.

- König, C., Weigelt, P., & Kreft, H. (2017). Dissecting global turnover in vascular plants. *Global Ecology and Biogeography*, 26(2), 228-242.
- Kraft, N. J. B., Adler, P. B., Godoy, O., James, E. C., Fuller, S., & Levine, J. M. (2015). Community assembly, coexistence and the environmental filtering metaphor. *Functional Ecology*, 29(5), 592-599.
- Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3), 1217-1241.
- Lande, R. (1996). Statistics and partitioning of species diversity, and similarity among multiple communities. *Oikos*, 76(1), 5-13.
- Legendre, P., Fortin, M.-J., & Borcard, D. (2015). Should the Mantel test be used in spatial analysis? *Methods in Ecology and Evolution*, 6(11), 1239-1247.
- Legendre, P., & Fortin, M. J. (2010). Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, 10(5), 831-844.
- Legendre, P., Lapointe, F. J., & Casgrain, P. (1994). Modeling brain evolution from behavior - a permutational regression approach. *Evolution*, 48(5), 1487-1499.
- Legendre, P., & Legendre, L. (1998). *Numerical Ecology*, 2nd ed. Elsevier.
- Leibold, M. A., & Chase, J. M. (2018). *Metacommunity Ecology, Volume 59* (S. A. Levin & H. S. Horn, Eds.). Princeton University Press.
- Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., Holt, R. D., Shurin, J. B., Law, R., Tilman, D., Loreau, M., & Gonzalez, A. (2004). The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters*, 7(7), 601-613.
- Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2(2), 164-168.
- Liu, R. Y., & Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In R. LePage & L. Billard (Eds.), *Exploring the limits of bootstrap* (pp. 225-248). Wiley.
- Löbl, I., & Smetana, A. (2003). *Catalogue of Palaearctic Coleoptera. Vol. 1: Archostemata-Myxophaga-Adephaga* (1 ed.). Apollo Books.
- Löbl, I., & Smetana, A. (2004). *Catalogue of Palaearctic Coleoptera. Vol. 2: Hydrophiloidea-Staphylinoidea* (1 ed.). Apollo Books.
- Löbl, I., & Smetana, A. (2006). *Catalogue of Palaearctic Coleoptera. Vol. 3: Scarabaeoidea, Scirtoidea, Dascilloidea, Buprestoidea and Byrrhoidea* (1 ed.). Apollo Books.
- Löbl, I., & Smetana, A. (2010). *Catalogue of Palaearctic Coleoptera. Vol. 6: Chrysomeloidea* (1 ed.). Apollo Books.
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27(2), 209-220.
- Marquardt, D. W. (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2), 431-441.
- Marschner, I. C. (2011). *glm2: Fitting generalized linear models with convergence problems*. Paquete de R versión 1.2.1. <https://cran.r-project.org/package=glm2>.

- Martín-Devasa, R., Jiménez-Valverde, A., Leprieur, F., Baselga, A., & Gómez-Rodríguez, C. (2024). Dispersal limitation shapes distance-decay patterns of European spiders at the continental scale. *Global Ecology and Biogeography*, *33*, e13810.
- Martín-Devasa, R., Martínez-Santalla, S., Gómez-Rodríguez, C., Crujeiras, R. M., & Baselga, A. (2022a). Comparing distance-decay parameters: A novel test under pairwise dependence. *Ecological Informatics*, *72*, 101894.
- Martín-Devasa, R., Martínez-Santalla, S., Gómez-Rodríguez, C., Crujeiras, R. M., & Baselga, A. (2022b). Species range size shapes distance-decay in community similarity. *Diversity and Distributions*, *28*(7), 1348-1357.
- Matsuoka, S., Kawaguchi, E., & Osono, T. (2016). Temporal distance decay of similarity of ectomycorrhizal fungal community composition in a subtropical evergreen forest in Japan. *FEMS Microbiology Ecology*, *92*(5), fiw061.
- McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models. Monographs on Statistics and Applied Probability 37 (2^a ed.)*. Chapman & Hall.
- McGill, B. J. (2010). Towards a unification of unified theories of biodiversity. *Ecology Letters*, *13*(5), 627-642.
- Millar, R. B., Anderson, M. J., & Tolimieri, N. (2011). Much ado about nothings: using zero similarity points in distance-decay curves. *Ecology*, *92*(9), 1717-1722.
- Montoya, D., Rodríguez, M. A., Zavala, M. A., & Hawkins, B. A. (2007). Contemporary richness of holarctic trees and the historical pattern of glacial retreat. *Ecography*, *30*(2), 173-182.
- Morlon, H., Chuyong, G., Condit, R., Hubbell, S., Kenfack, D., Thomas, D., Valencia, R., & Green, J. L. (2008). A general framework for the distance–decay of similarity in ecological communities. *Ecology Letters*, *11*(9), 904-917.
- Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications*, *9*(1), 141-142.
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133-142.
- Nekola, J. C., & McGill, B. J. (2014). Scale dependency in the functional form of the distance decay relationship. *Ecography*, *37*(4), 309-320.
- Nekola, J. C., & White, P. S. (1999). The distance decay of similarity in biogeography and ecology. *Journal of Biogeography*, *26*(4), 867-878.
- O'Hara, R., & Kotze, J. (2010). Do not log-transform count data. *Nature Precedings*.
- Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., . . . Weedon, J. (2017). *vegan: Community Ecology Package*. R package version 2.6-4. <https://CRAN.R-project.org/package=vegan>.
- Olden, J. D., Kennard, M. J., Leprieur, F., Tedesco, P. A., Winemiller, K. O., & Garcia-Berthou, E. (2010). Conservation biogeography of freshwater fishes: recent progress and future challenges. *Diversity and Distributions*, *16*(3), 496-513.
- Peguero, G., Ferrín, M., Sardans, J., Verbruggen, E., Ramírez-Rojas, I., Van Langenhove, L., Verryckt, L. T., Muriene, J., Iribar, A., Zinger, L., Grau, O., Orivel, J., Stahl, C., Courtois, E. A., Asensio, D., Gargallo-Garriga, A., Llusà, J., Margalef, O., Ogaya, R., . . . Peñuelas,

- J. (2021). Decay of similarity across tropical forest communities: integrating spatial distance with soil nutrients. *Ecology* 103(2), e03599.
- Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M., & Araújo, M. B. (2011). *Ecological Niches and Geographic Distributions*. Princeton University Press.
- Poulin, R. (2003). The decay of similarity with geographical distance in parasite communities of vertebrate hosts. *Journal of Biogeography*, 30(10), 1609-1615.
- Qian, H. (2009). Beta diversity in relation to dispersal ability for vascular plants in North America. *Global Ecology and Biogeography*, 18(3), 327-332.
- Qian, H., Badgley, C., & Fox, D. L. (2009). The latitudinal gradient of beta diversity in relation to climate and topography for mammals in North America. *Global Ecology and Biogeography*, 18(1), 111-122.
- Qian, H., & Ricklefs, R. E. (2012). Disentangling the effects of geographic distance and environmental dissimilarity on global patterns of species turnover. *Global Ecology and Biogeography*, 21(3), 341-351.
- R Development Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at <https://www.R-project.org/>.
- Rahbek, C. (2005). The role of spatial scale and the perception of large-scale species-richness patterns. *Ecology Letters*, 8(2), 224-239.
- Ribeiro Jr, P. J., & Diggle, P. J. (2020). *geoR: Analysis of Geostatistical Data*. Paquete de R versión 1.7-5.2.2. <https://CRAN.R-project.org/package=geoR>.
- Ricklefs, R. E. (2004). A comprehensive framework for global patterns in biodiversity. *Ecology Letters*, 7(1), 1-15.
- Ricklefs, R. E., & Schluter, D. (1993). Species diversity: regional and historical influences. In R. E. Ricklefs & D. Schluter (Eds.), *Species Diversity in Ecological Communities: Historical and Geographical Perspectives* (pp. 350–364). Chicago University Press.
- Rosenzweig, M. L. (1995). *Species Diversity in Space and Time*. Cambridge University Press.
- Saito, V. S., Soininen, J., Fonseca-Gessner, A. A., & Siqueira, T. (2015). Dispersal traits drive the phylogenetic distance decay of similarity in Neotropical stream metacommunities. *Journal of Biogeography*, 42(11), 2101-2111.
- Sherman, M., & Cessie, S. I. (1997). A comparison between bootstrap methods and generalized estimating equations for correlated outcomes in generalized linear models. *Communications in Statistics - Simulation and Computation*, 26(3), 901-925.
- Si, X. F., Baselga, A., & Ding, P. (2015). Revealing beta-diversity patterns of breeding bird and lizard communities on inundated land-bridge islands by separating the turnover and nestedness components. *PLOS ONE*, 10(5).
- Simpson, G. G. (1943). Mammals and the nature of continents. *American Journal of Science*, 241, 1-31.
- Simpson, G. G. (1960). Notes on the measurement of faunal resemblance. *American Journal of Science*, 258, 300-311.
- Smouse, P. E., Long, J. C., & Sokal, R. R. (1986). Multiple-regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, 35(4), 627-632.

- Soininen, J., Lennon, J. J., & Hillebrand, H. (2007). A multivariate analysis of beta diversity across organisms and environments. *Ecology*, *88*(11), 2830-2838.
- Soininen, J., McDonald, R., & Hillebrand, H. (2007). The distance decay of similarity in ecological communities. *Ecography*, *30*(1), 3-12.
- Sørensen, T. A. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species content, and its application to analyses of the vegetation on Danish commons. *Kongelige Danske Videnskabernes Selskabs Biologiske Skrifter*, *5*, 1-34.
- Steinbauer, M. J., Dolos, K., Reineking, B., & Beierkuhnlein, C. (2012). Current measures for distance decay in similarity of species composition are influenced by study extent and grain size. *Global Ecology and Biogeography*, *21*(12), 1203-1212.
- Steinitz, O., Heller, J., Tsoar, A., Rotem, D., & Kadmon, R. (2006). Environment, dispersal and patterns of species similarity. *Journal of Biogeography*, *33*(6), 1044-1054.
- Svenning, J. C., Fløjgaard, C., & Baselga, A. (2011). Climate, history and neutrality as drivers of mammal beta diversity in Europe: insights from multiscale deconstruction. *Journal of Animal Ecology*, *80*(2), 393-402.
- Svenning, J. C., & Skov, F. (2007a). Could the tree diversity pattern in Europe be generated by postglacial dispersal limitation? *Ecology Letters*, *10*(6), 453-460.
- Svenning, J. C., & Skov, F. (2007b). Ice age legacies in the geographical distribution of tree species richness in Europe. *Global Ecology and Biogeography*, *16*(2), 234-245.
- Tornero, I., Boix, D., Bagella, S., Pinto-Cruz, C., Caria, M. C., Belo, A., Lumbreras, A., Sala, J., Compte, J., & Gascón, S. (2018). Dispersal mode and spatial extent influence distance-decay patterns in pond metacommunities. *PLOS ONE*, *13*(8), e0203119.
- Tovo, A., & Favretti, M. (2018). The distance decay of similarity in tropical rainforests. A spatial point processes analytical formulation. *Theoretical Population Biology*, *120*, 78-89.
- Trøjelsgaard, K., Jordano, P., Carstensen, D. W., & Olesen, J. M. (2015). Geographical variation in mutualistic networks: similarity, turnover and partner fidelity. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1802), 20142925.
- Tuomisto, H., Ruokolainen, K., & Yli-Halla, M. (2003). Dispersal, environment, and floristic variation of western Amazonian forests. *Science*, *299*(5604), 241-244.
- Veech, J. A., & Crist, T. O. (2010). Diversity partitioning without statistical independence of alpha and beta. *Ecology*, *91*, 1964-1969.
- Vellend, M. (2016). *The Theory of Ecological Communities*. Princeton University Press.
- Wetzel, C. E., Bicudo, D. d. C., Ector, L., Lobo, E. A., Soininen, J., Landeiro, V. L., & Bini, L. M. (2012). Distance decay of similarity in Neotropical diatom communities. *PLOS ONE*, *7*(9), e45071.
- Whittaker, R. H. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, *30*(3), 280-338.
- Whittaker, R. J., & Triantis, K. A. (2012). The species-area relationship: an exploration of that 'most general, yet protean pattern'. *Journal of Biogeography*, *39*(4), 623-626.
- Whittaker, R. J., Willis, K. J., & Field, R. (2001). Scale and species richness: towards a general, hierarchical theory of species diversity. *Journal of Biogeography*, *28*(4), 453-470.
- Willig, M. R., Kaufman, D. M., & Stevens, R. D. (2003). Latitudinal gradients of biodiversity: Pattern, process, scale, and synthesis. *Annual Review of Ecology Evolution and Systematics*, *34*, 273-309.

- Wilson, M. V., & Shmida, A. (1984). Measuring beta diversity with presence absence data. *Journal of Ecology*, 72(3), 1055-1064.
- Yang, Q., Weigelt, P., Fristoe, T. S., Zhang, Z., Kreft, H., Stein, A., Seebens, H., Dawson, W., Essl, F., König, C., Lenzner, B., Pergl, J., Pouteau, R., Pyšek, P., Winter, M., Ebel, A. L., Fuentes, N., Giehl, E. L. H., Kartesz, J., . . . van Kleunen, M. (2021). The global loss of floristic uniqueness. *Nature Communications*, 12(1), 7290.
- Young, S. G., & Bowman, A. W. (1995). Non-parametric analysis of covariance. *Biometrics*, 51(3), 920-931.
- Zeller, K. A., Creech, T. G., Millette, K. L., Crowhurst, R. S., Long, R. A., Wagner, H. H., Balkenhol, N., & Landguth, E. L. (2016). Using simulations to evaluate Mantel-based methods for assessing landscape resistance to gene flow. *Ecology and Evolution*, 6(12), 4115-4128.
- Zhang, W., Lei, M., Li, Y., Wang, P., Wang, C., Gao, Y., Wu, H., Xu, C., Niu, L., Wang, L., & Zhang, H. (2019). Determination of vertical and horizontal assemblage drivers of bacterial community in a heavily polluted urban river. *Water Research*, 161, 98-107.

9 APÉNDICES

9.1 CÓDIGO DE R PARA COMPARACIÓN NO PARAMÉTRICA DE LOS PATRONES DE *DISTANCE-DECAY* DE COLEÓPTEROS EN EL NORTE Y EL SUR DE EUROPA

```
# - . # - . # - . # - . # - . # - . # - . # - . # - . # - . # - . # - . # - . #
# ESTUDIO NO PARAMÉTRICO DE DIFERENCIAS EN PATRONES DISTANCE-DECAY
# - . # - . # - . # - . # - . # - . # - . # - . # - . # - . # - . # - . # - . #

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# 1. Librerías
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

library(sm)
library(vegan)
library(betapart)

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# 2. Datos
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

# Lectura de tablas de datos de presencia/ausencia de coleópteros

amara <- read.table("./datos/amara.txt", header = T)
apion <- read.table("./datos/apionidae.txt", header = T)
carab <- read.table("./datos/carabus.txt", header = T)
ceram <- read.table("./datos/cerambycinae.txt", header = T)
chaet <- read.table("./datos/chaetocnema.txt", header = T)
chrys <- read.table("./datos/chrys.txt", header = T)
crypt <- read.table("./datos/crypt.txt", header = T)
hydra <- read.table("./datos/hydraenini.txt", header = T)
lamin <- read.table("./datos/lamiinae.txt", header = T)
leptu <- read.table("./datos/lepturinae.txt", header = T)
limne <- read.table("./datos/limnebiini.txt", header = T)
ochte <- read.table("./datos/ochtebiini.txt", header = T)
otior <- read.table("./datos/otiorhynchus.txt", header = T)
phyll <- read.table("./datos/phylllobini.txt", header = T)
polyd <- read.table("./datos/polydrosini.txt", header = T)
psela <- read.table("./datos/pselaphinae.txt", header = T)
ptero <- read.table("./datos/pterostichus.txt", header = T)
scara <- read.table("./datos/scarabaeidae.txt", header = T)
scoly <- read.table("./datos/scolytinae.txt", header = T)
```

```

silph <- read.table("../datos/silphidae.txt", header = T)
trech <- read.table("../datos/trechus.txt", header = T)

# Lectura de tablas de datos con información sobre territorios europeos

vars <- read.table("../datos/col-eur-vars.txt", header = T, row.names = 1)

# Generación de tablas independientes para países del norte (1) y del sur
(0) de Europa

amara.n <- subset(amara, vars$nortesur == 1)
apion.n <- subset(apion, vars$nortesur == 1)
carab.n <- subset(carab, vars$nortesur == 1)
ceram.n <- subset(ceram, vars$nortesur == 1)
chaet.n <- subset(chaet, vars$nortesur == 1)
chrys.n <- subset(chrys, vars$nortesur == 1)
crypt.n <- subset(crypt, vars$nortesur == 1)
hydra.n <- subset(hydra, vars$nortesur == 1)
lamin.n <- subset(lamin, vars$nortesur == 1)
leptu.n <- subset(leptu, vars$nortesur == 1)
limne.n <- subset(limne, vars$nortesur == 1)
ochte.n <- subset(ochte, vars$nortesur == 1)
otior.n <- subset(otior, vars$nortesur == 1)
phyll.n <- subset(phyll, vars$nortesur == 1)
polyd.n <- subset(polyd, vars$nortesur == 1)
psela.n <- subset(psela, vars$nortesur == 1)
ptero.n <- subset(ptero, vars$nortesur == 1)
scara.n <- subset(scara, vars$nortesur == 1)
scoly.n <- subset(scoly, vars$nortesur == 1)
silph.n <- subset(silph, vars$nortesur == 1)
trech.n <- subset(trech, vars$nortesur == 1)

amara.s <- subset(amara, vars$nortesur == 0)
apion.s <- subset(apion, vars$nortesur == 0)
carab.s <- subset(carab, vars$nortesur == 0)
ceram.s <- subset(ceram, vars$nortesur == 0)
chaet.s <- subset(chaet, vars$nortesur == 0)
chrys.s <- subset(chrys, vars$nortesur == 0)
crypt.s <- subset(crypt, vars$nortesur == 0)
hydra.s <- subset(hydra, vars$nortesur == 0)
lamin.s <- subset(lamin, vars$nortesur == 0)
leptu.s <- subset(leptu, vars$nortesur == 0)
limne.s <- subset(limne, vars$nortesur == 0)
ochte.s <- subset(ochte, vars$nortesur == 0)
otior.s <- subset(otior, vars$nortesur == 0)
phyll.s <- subset(phyll, vars$nortesur == 0)
polyd.s <- subset(polyd, vars$nortesur == 0)
psela.s <- subset(psela, vars$nortesur == 0)
ptero.s <- subset(ptero, vars$nortesur == 0)
scara.s <- subset(scara, vars$nortesur == 0)
scoly.s <- subset(scoly, vars$nortesur == 0)
silph.s <- subset(silph, vars$nortesur == 0)
trech.s <- subset(trech, vars$nortesur == 0)

```

```

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# 3. Comparación no paramétrica de patrones distance-decay
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

# Se realiza la comparación y se exporta la representación gráfica

pdf(file = "amara-contrast.pdf", height = 5, width = 9)
par(mfrow = c(1, 2))
sm.ancova(distgeo.v, amara.sim.v, n.s, model = "equal", xlab = "Distancia
  espacial (km)", ylab = "Similitud (Simpson)")
sm.ancova(distgeo.v, amara.sim.v, n.s, model = "parallel", xlab =
  "Distancia espacial (km)", ylab = "Similitud (Simpson)")
dev.off()

pdf(file = "apion-contrast.pdf", height = 5, width = 9)
par(mfrow = c(1, 2))
sm.ancova(distgeo.v, apion.sim.v, n.s, model = "equal", xlab = "Distancia
  espacial (km)", ylab = "Similitud (Simpson)")
sm.ancova(distgeo.v, apion.sim.v, n.s, model = "parallel", xlab =
  "Distancia espacial (km)", ylab = "Similitud (Simpson)")
dev.off()

pdf(file = "carab-contrast.pdf", height = 5, width = 9)
par(mfrow = c(1, 2))
sm.ancova(distgeo.v, carab.sim.v, n.s, model = "equal", xlab = "Distancia
  espacial (km)", ylab = "Similitud (Simpson)")
sm.ancova(distgeo.v, carab.sim.v, n.s, model = "parallel", xlab =
  "Distancia espacial (km)", ylab = "Similitud (Simpson)")
dev.off()

pdf(file = "ceram-contrast.pdf", height = 5, width = 9)
par(mfrow = c(1, 2))
sm.ancova(distgeo.v, ceram.sim.v, n.s, model = "equal", xlab = "Distancia
  espacial (km)", ylab = "Similitud (Simpson)")
sm.ancova(distgeo.v, ceram.sim.v, n.s, model = "parallel", xlab =
  "Distancia espacial (km)", ylab = "Similitud (Simpson)")
dev.off()

pdf(file = "chaet-contrast.pdf", height = 5, width = 9)
par(mfrow = c(1, 2))
sm.ancova(distgeo.v, chaet.sim.v, n.s, model = "equal", xlab = "Distancia
  espacial (km)", ylab = "Similitud (Simpson)")
sm.ancova(distgeo.v, chaet.sim.v, n.s, model = "parallel", xlab =
  "Distancia espacial (km)", ylab = "Similitud (Simpson)")
dev.off()

pdf(file = "chrys-contrast.pdf", height = 5, width = 9)
par(mfrow = c(1, 2))

```

```

sm.ancova(distgeo.v, chrys.sim.v, n.s, model = "equal", xlab = "Distancia
  espacial (km)", ylab = "Similitud (Simpson)")
sm.ancova(distgeo.v, chrys.sim.v, n.s, model = "parallel", xlab =
  "Distancia espacial (km)", ylab = "Similitud (Simpson)")
dev.off()

```

```

pdf(file = "crypt-contrast.pdf", height = 5, width = 9)
par(mfrow = c(1, 2))
sm.ancova(distgeo.v, crypt.sim.v, n.s, model = "equal", xlab = "Distancia
  espacial (km)", ylab = "Similitud (Simpson)")
sm.ancova(distgeo.v, crypt.sim.v, n.s, model = "parallel", xlab =
  "Distancia espacial (km)", ylab = "Similitud (Simpson)")
dev.off()

```

```

pdf(file = "hydra-contrast.pdf", height = 5, width = 9)
par(mfrow = c(1, 2))
sm.ancova(distgeo.v, hydra.sim.v, n.s, model = "equal", xlab = "Distancia
  espacial (km)", ylab = "Similitud (Simpson)")
sm.ancova(distgeo.v, hydra.sim.v, n.s, model = "parallel", xlab =
  "Distancia espacial (km)", ylab = "Similitud (Simpson)")
dev.off()

```

```

pdf(file = "lamin-contrast.pdf", height = 5, width = 9)
par(mfrow = c(1, 2))
sm.ancova(distgeo.v, lamin.sim.v, n.s, model = "equal", xlab = "Distancia
  espacial (km)", ylab = "Similitud (Simpson)")
sm.ancova(distgeo.v, lamin.sim.v, n.s, model = "parallel", xlab =
  "Distancia espacial (km)", ylab = "Similitud (Simpson)")
dev.off()

```

```

pdf(file = "leptu-contrast.pdf", height = 5, width = 9)
par(mfrow = c(1, 2))
sm.ancova(distgeo.v, leptu.sim.v, n.s, model = "equal", xlab = "Distancia
  espacial (km)", ylab = "Similitud (Simpson)")
sm.ancova(distgeo.v, leptu.sim.v, n.s, model = "parallel", xlab =
  "Distancia espacial (km)", ylab = "Similitud (Simpson)")
dev.off()

```

```

pdf(file = "limne-contrast.pdf", height = 5, width = 9)
par(mfrow = c(1, 2))
sm.ancova(distgeo.v, limne.sim.v, n.s, model = "equal", xlab = "Distancia
  espacial (km)", ylab = "Similitud (Simpson)")
sm.ancova(distgeo.v, limne.sim.v, n.s, model = "parallel", xlab =
  "Distancia espacial (km)", ylab = "Similitud (Simpson)")
dev.off()

```

```

pdf(file = "ochte-contrast.pdf", height = 5, width = 9)
par(mfrow = c(1, 2))
sm.ancova(distgeo.v, ochte.sim.v, n.s, model = "equal", xlab = "Distancia
  espacial (km)", ylab = "Similitud (Simpson)")
sm.ancova(distgeo.v, ochte.sim.v, n.s, model = "parallel", xlab =
  "Distancia espacial (km)", ylab = "Similitud (Simpson)")
dev.off()

```

```

pdf(file = "otior-contrast.pdf", height = 5, width = 9)

```

```

par(mfrow = c(1, 2))
sm.ancova(distgeo.v, otior.sim.v, n.s, model = "equal", xlab = "Distancia
  espacial (km)", ylab = "Similitud (Simpson)")
sm.ancova(distgeo.v, otior.sim.v, n.s, model = "parallel", xlab =
  "Distancia espacial (km)", ylab = "Similitud (Simpson)")
dev.off()

```

```

pdf(file = "phyll-contrast.pdf", height = 5, width = 9)
par(mfrow = c(1, 2))
sm.ancova(distgeo.v, phyll.sim.v, n.s, model = "equal", xlab = "Distancia
  espacial (km)", ylab = "Similitud (Simpson)")
sm.ancova(distgeo.v, phyll.sim.v, n.s, model = "parallel", xlab =
  "Distancia espacial (km)", ylab = "Similitud (Simpson)")
dev.off()

```

```

pdf(file = "polyd-contrast.pdf", height = 5, width = 9)
par(mfrow = c(1, 2))
sm.ancova(distgeo.v, polyd.sim.v, n.s, model = "equal", xlab = "Distancia
  espacial (km)", ylab = "Similitud (Simpson)")
sm.ancova(distgeo.v, polyd.sim.v, n.s, model = "parallel", xlab =
  "Distancia espacial (km)", ylab = "Similitud (Simpson)")
dev.off()

```

```

pdf(file = "psela-contrast.pdf", height = 5, width = 9)
par(mfrow = c(1, 2))
sm.ancova(distgeo.v, psela.sim.v, n.s, model = "equal", xlab = "Distancia
  espacial (km)", ylab = "Similitud (Simpson)")
sm.ancova(distgeo.v, psela.sim.v, n.s, model = "parallel", xlab =
  "Distancia espacial (km)", ylab = "Similitud (Simpson)")
dev.off()

```

```

pdf(file = "ptero-contrast.pdf", height = 5, width = 9)
par(mfrow = c(1, 2))
sm.ancova(distgeo.v, ptero.sim.v, n.s, model = "equal", xlab = "Distancia
  espacial (km)", ylab = "Similitud (Simpson)")
sm.ancova(distgeo.v, ptero.sim.v, n.s, model = "parallel", xlab =
  "Distancia espacial (km)", ylab = "Similitud (Simpson)")
dev.off()

```

```

pdf(file = "scara-contrast.pdf", height = 5, width = 9)
par(mfrow = c(1, 2))
sm.ancova(distgeo.v, scara.sim.v, n.s, model = "equal", xlab = "Distancia
  espacial (km)", ylab = "Similitud (Simpson)")
sm.ancova(distgeo.v, scara.sim.v, n.s, model = "parallel", xlab =
  "Distancia espacial (km)", ylab = "Similitud (Simpson)")
dev.off()

```

```

pdf(file = "scoly-contrast.pdf", height = 5, width = 9)
par(mfrow = c(1, 2))
sm.ancova(distgeo.v, scoly.sim.v, n.s, model = "equal", xlab = "Distancia
  espacial (km)", ylab = "Similitud (Simpson)")
sm.ancova(distgeo.v, scoly.sim.v, n.s, model = "parallel", xlab =
  "Distancia espacial (km)", ylab = "Similitud (Simpson)")
dev.off()

```

```
pdf(file = "silph-contrast.pdf", height = 5, width = 9)
par(mfrow = c(1, 2))
sm.ancova(distgeo.v, silph.sim.v, n.s, model = "equal", xlab = "Distancia
  espacial (km)", ylab = "Similitud (Simpson)")
sm.ancova(distgeo.v, silph.sim.v, n.s, model = "parallel", xlab =
  "Distancia espacial (km)", ylab = "Similitud (Simpson)")
dev.off()
```

```
pdf(file = "trech-contrast.pdf", height = 5, width = 9)
par(mfrow = c(1, 2))
sm.ancova(distgeo.v, trech.sim.v, n.s, model = "equal", xlab = "Distancia
  espacial (km)", ylab = "Similitud (Simpson)")
sm.ancova(distgeo.v, trech.sim.v, n.s, model = "parallel", xlab =
  "Distancia espacial (km)", ylab = "Similitud (Simpson)")
dev.off()
```

9.2 CÓDIGO DE R PARA EL ESTUDIO DE LA ESTRUCTURA DEL ERROR

```

# - . # - . # - . # - . # - . # - . # - . # - . # - . # - . # - . #
# ESTUDIO PARA LA ESTRUCTURA DEL ERROR
# - . # - . # - . # - . # - . # - . # - . # - . # - . # - . # - . #

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# 1. Librerías
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

library(sm)
library(vegan)
library(betapart)

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# 2. Datos
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

# Lectura de tablas de datos de presencia/ausencia de coleópteros

amara <- read.table("./datos/amara.txt", header = T)
apion <- read.table("./datos/apionidae.txt", header = T)
carab <- read.table("./datos/carabus.txt", header = T)
ceram <- read.table("./datos/cerambycinae.txt", header = T)
chaet <- read.table("./datos/chaetocnema.txt", header = T)
chrys <- read.table("./datos/chrys.txt", header = T)
crypt <- read.table("./datos/crypt.txt", header = T)
hydra <- read.table("./datos/hydraenini.txt", header = T)
lamin <- read.table("./datos/lamiinae.txt", header = T)
leptu <- read.table("./datos/lepturinae.txt", header = T)
limne <- read.table("./datos/limnebiini.txt", header = T)
ochte <- read.table("./datos/ochtebiini.txt", header = T)
otior <- read.table("./datos/otiorhynchus.txt", header = T)
phyll <- read.table("./datos/phyllobini.txt", header = T)
polyd <- read.table("./datos/polydrosini.txt", header = T)
psela <- read.table("./datos/pselaphinae.txt", header = T)
ptero <- read.table("./datos/pterostichus.txt", header = T)
scara <- read.table("./datos/scarabaeidae.txt", header = T)
scoly <- read.table("./datos/scolytinae.txt", header = T)
silph <- read.table("./datos/silphidae.txt", header = T)
trech <- read.table("./datos/trechus.txt", header = T)

# Lectura de tablas de datos con información sobre territorios europeos

vars <- read.table("./datos/col-eur-vars.txt", header = T, row.names = 1)
# Generación de tablas independientes para países del norte (1) y del sur
(0) de Europa

amara.n <- subset(amara, vars$nortesur == 1)
apion.n <- subset(apion, vars$nortesur == 1)
carab.n <- subset(carab, vars$nortesur == 1)
ceram.n <- subset(ceram, vars$nortesur == 1)
chaet.n <- subset(chaet, vars$nortesur == 1)
chrys.n <- subset(chrys, vars$nortesur == 1)

```

```

crypt.n <- subset(crypt, vars$nortesur == 1)
hydra.n <- subset(hydra, vars$nortesur == 1)
lamin.n <- subset(lamin, vars$nortesur == 1)
leptu.n <- subset(leptu, vars$nortesur == 1)
limne.n <- subset(limne, vars$nortesur == 1)
ochte.n <- subset(ochte, vars$nortesur == 1)
otior.n <- subset(otior, vars$nortesur == 1)
phyll.n <- subset(phyll, vars$nortesur == 1)
polyd.n <- subset(polyd, vars$nortesur == 1)
psela.n <- subset(psela, vars$nortesur == 1)
ptero.n <- subset(ptero, vars$nortesur == 1)
scara.n <- subset(scara, vars$nortesur == 1)
scoly.n <- subset(scoly, vars$nortesur == 1)
silph.n <- subset(silph, vars$nortesur == 1)
trech.n <- subset(trech, vars$nortesur == 1)

amara.s <- subset(amara, vars$nortesur == 0)
apion.s <- subset(apion, vars$nortesur == 0)
carab.s <- subset(carab, vars$nortesur == 0)
ceram.s <- subset(ceram, vars$nortesur == 0)
chaet.s <- subset(chaet, vars$nortesur == 0)
chrys.s <- subset(chrys, vars$nortesur == 0)
crypt.s <- subset(crypt, vars$nortesur == 0)
hydra.s <- subset(hydra, vars$nortesur == 0)
lamin.s <- subset(lamin, vars$nortesur == 0)
leptu.s <- subset(leptu, vars$nortesur == 0)
limne.s <- subset(limne, vars$nortesur == 0)
ochte.s <- subset(ochte, vars$nortesur == 0)
otior.s <- subset(otior, vars$nortesur == 0)
phyll.s <- subset(phyll, vars$nortesur == 0)
polyd.s <- subset(polyd, vars$nortesur == 0)
psela.s <- subset(psela, vars$nortesur == 0)
ptero.s <- subset(ptero, vars$nortesur == 0)
scara.s <- subset(scara, vars$nortesur == 0)
scoly.s <- subset(scoly, vars$nortesur == 0)
silph.s <- subset(silph, vars$nortesur == 0)
trech.s <- subset(trech, vars$nortesur == 0)

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# 3. Matrices de similitud biótica y distancia espacial
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

# Cálculo de matrices de distancia espacial (distancia euclídea) entre los
centroides de los territorios europeos

distgeo <-dist(data.frame(vars$x.utm, vars$y.utm), method = "euclidean")
# Entre todos los países

distgeo.n <-dist(subset(data.frame(vars$x.utm, vars$y.utm), vars$nortesur
== 1), method = "euclidean") # Países del norte

distgeo.s <-dist(subset(data.frame(vars$x.utm, vars$y.utm), vars$nortesur ==
0), method = "euclidean") # Países del sur

```

```
# Cálculo de matrices de similitud biótica entre territorios europeos
(índice de similitud de Simpson)
```

```
amara.n.sim <- 1 - beta.pair(amara.n)$beta.sim
apion.n.sim <- 1 - beta.pair(apion.n)$beta.sim
carab.n.sim <- 1 - beta.pair(carab.n)$beta.sim
ceram.n.sim <- 1 - beta.pair(ceram.n)$beta.sim
chaet.n.sim <- 1 - beta.pair(chaet.n)$beta.sim
chrys.n.sim <- 1 - beta.pair(chrys.n)$beta.sim
crypt.n.sim <- 1 - beta.pair(crypt.n)$beta.sim
hydra.n.sim <- 1 - beta.pair(hydra.n)$beta.sim
lamin.n.sim <- 1 - beta.pair(lamin.n)$beta.sim
leptu.n.sim <- 1 - beta.pair(leptu.n)$beta.sim
limne.n.sim <- 1 - beta.pair(limne.n)$beta.sim
ochte.n.sim <- 1 - beta.pair(ochte.n)$beta.sim
otior.n.sim <- 1 - beta.pair(otior.n)$beta.sim
phyll.n.sim <- 1 - beta.pair(phyll.n)$beta.sim
polyd.n.sim <- 1 - beta.pair(polyd.n)$beta.sim
psela.n.sim <- 1 - beta.pair(psela.n)$beta.sim
ptero.n.sim <- 1 - beta.pair(ptero.n)$beta.sim
scara.n.sim <- 1 - beta.pair(scara.n)$beta.sim
scoly.n.sim <- 1 - beta.pair(scoly.n)$beta.sim
silph.n.sim <- 1 - beta.pair(silph.n)$beta.sim
trech.n.sim <- 1 - beta.pair(trech.n)$beta.sim
```

```
amara.s.sim <- 1 - beta.pair(amara.s)$beta.sim
apion.s.sim <- 1 - beta.pair(apion.s)$beta.sim
carab.s.sim <- 1 - beta.pair(carab.s)$beta.sim
ceram.s.sim <- 1 - beta.pair(ceram.s)$beta.sim
chaet.s.sim <- 1 - beta.pair(chaet.s)$beta.sim
chrys.s.sim <- 1 - beta.pair(chrys.s)$beta.sim
crypt.s.sim <- 1 - beta.pair(crypt.s)$beta.sim
hydra.s.sim <- 1 - beta.pair(hydra.s)$beta.sim
lamin.s.sim <- 1 - beta.pair(lamin.s)$beta.sim
leptu.s.sim <- 1 - beta.pair(leptu.s)$beta.sim
limne.s.sim <- 1 - beta.pair(limne.s)$beta.sim
ochte.s.sim <- 1 - beta.pair(ochte.s)$beta.sim
otior.s.sim <- 1 - beta.pair(otior.s)$beta.sim
phyll.s.sim <- 1 - beta.pair(phyll.s)$beta.sim
polyd.s.sim <- 1 - beta.pair(polyd.s)$beta.sim
psela.s.sim <- 1 - beta.pair(psela.s)$beta.sim
ptero.s.sim <- 1 - beta.pair(ptero.s)$beta.sim
scara.s.sim <- 1 - beta.pair(scara.s)$beta.sim
scoly.s.sim <- 1 - beta.pair(scoly.s)$beta.sim
silph.s.sim <- 1 - beta.pair(silph.s)$beta.sim
trech.s.sim <- 1 - beta.pair(trech.s)$beta.sim
```

```
# Conversión de matrices de distancia espacial en vectores y generación de
un único vector con datos de norte y el sur
```

```
distgeo.n.v <- as.vector(distgeo.n)
length(distgeo.n)
#153
```

```

distgeo.s.v <- as.vector(distgeo.s)
  length(distgeo.s)
  #120

distgeo.v <- c(distgeo.n.v, distgeo.s.v)
  # Generación de un factor para separar los grupos en el vector
  conjunto
  n.s <- c(rep(1, length(distgeo.n)), rep(0, length(distgeo.s))) #1 ==
  norte, 0 == sur.
  n.s <- as.factor(n.s)

# Conversión de matrices de similitud biótica en vectores y generación de
un único vector con datos del norte y del sur

amara.n.sim.v <- as.vector(amara.n.sim)
amara.s.sim.v <- as.vector(amara.s.sim)
amara.sim.v <- c(amara.n.sim.v, amara.s.sim.v)

apion.n.sim.v <- as.vector(apion.n.sim)
apion.s.sim.v <- as.vector(apion.s.sim)
apion.sim.v <- c(apion.n.sim.v, apion.s.sim.v)

carab.n.sim.v <- as.vector(carab.n.sim)
carab.s.sim.v <- as.vector(carab.s.sim)
carab.sim.v <- c(carab.n.sim.v, carab.s.sim.v)

ceram.n.sim.v <- as.vector(ceram.n.sim)
ceram.s.sim.v <- as.vector(ceram.s.sim)
ceram.sim.v <- c(ceram.n.sim.v, ceram.s.sim.v)

chaet.n.sim.v <- as.vector(chaet.n.sim)
chaet.s.sim.v <- as.vector(chaet.s.sim)
chaet.sim.v <- c(chaet.n.sim.v, chaet.s.sim.v)

chrys.n.sim.v <- as.vector(chrys.n.sim)
chrys.s.sim.v <- as.vector(chrys.s.sim)
chrys.sim.v <- c(chrys.n.sim.v, chrys.s.sim.v)

crypt.n.sim.v <- as.vector(crypt.n.sim)
crypt.s.sim.v <- as.vector(crypt.s.sim)
crypt.sim.v <- c(crypt.n.sim.v, crypt.s.sim.v)

hydra.n.sim.v <- as.vector(hydra.n.sim)
hydra.s.sim.v <- as.vector(hydra.s.sim)
hydra.sim.v <- c(hydra.n.sim.v, hydra.s.sim.v)

lamin.n.sim.v <- as.vector(lamin.n.sim)
lamin.s.sim.v <- as.vector(lamin.s.sim)
lamin.sim.v <- c(lamin.n.sim.v, lamin.s.sim.v)

leptu.n.sim.v <- as.vector(leptu.n.sim)
leptu.s.sim.v <- as.vector(leptu.s.sim)
leptu.sim.v <- c(leptu.n.sim.v, leptu.s.sim.v)

```

```

limne.n.sim.v <- as.vector(limne.n.sim)
limne.s.sim.v <- as.vector(limne.s.sim)
limne.sim.v <- c(limne.n.sim.v, limne.s.sim.v)

ochte.n.sim.v <- as.vector(ochte.n.sim)
ochte.s.sim.v <- as.vector(ochte.s.sim)
ochte.sim.v <- c(ochte.n.sim.v, ochte.s.sim.v)

otior.n.sim.v <- as.vector(otior.n.sim)
otior.s.sim.v <- as.vector(otior.s.sim)
otior.sim.v <- c(otior.n.sim.v, otior.s.sim.v)

phyll.n.sim.v <- as.vector(phyll.n.sim)
phyll.s.sim.v <- as.vector(phyll.s.sim)
phyll.sim.v <- c(phyll.n.sim.v, phyll.s.sim.v)

polyd.n.sim.v <- as.vector(polyd.n.sim)
polyd.s.sim.v <- as.vector(polyd.s.sim)
polyd.sim.v <- c(polyd.n.sim.v, polyd.s.sim.v)

psela.n.sim.v <- as.vector(psela.n.sim)
psela.s.sim.v <- as.vector(psela.s.sim)
psela.sim.v <- c(psela.n.sim.v, psela.s.sim.v)

ptero.n.sim.v <- as.vector(ptero.n.sim)
ptero.s.sim.v <- as.vector(ptero.s.sim)
ptero.sim.v <- c(ptero.n.sim.v, ptero.s.sim.v)

scara.n.sim.v <- as.vector(scara.n.sim)
scara.s.sim.v <- as.vector(scara.s.sim)
scara.sim.v <- c(scara.n.sim.v, scara.s.sim.v)

scoly.n.sim.v <- as.vector(scoly.n.sim)
scoly.s.sim.v <- as.vector(scoly.s.sim)
scoly.sim.v <- c(scoly.n.sim.v, scoly.s.sim.v)

silph.n.sim.v <- as.vector(silph.n.sim)
silph.s.sim.v <- as.vector(silph.s.sim)
silph.sim.v <- c(silph.n.sim.v, silph.s.sim.v)

trech.n.sim.v <- as.vector(trech.n.sim)
trech.s.sim.v <- as.vector(trech.s.sim)
trech.sim.v <- c(trech.n.sim.v, trech.s.sim.v)

# Generación de lista de datos con los vectores conjuntos de todos los
taxones

data.sim.complete <- list(amara.n.sim.v, apion.n.sim.v, carab.n.sim.v,
  ceram.n.sim.v, chaet.n.sim.v, chrys.n.sim.v, crypt.n.sim.v,
  hydra.n.sim.v, lamin.n.sim.v, leptu.n.sim.v, limne.n.sim.v,
  ochte.n.sim.v, otior.n.sim.v, phyll.n.sim.v, polyd.n.sim.v,
  psela.n.sim.v, ptero.n.sim.v, scara.n.sim.v, scoly.n.sim.v,
  silph.n.sim.v, trech.n.sim.v, amara.s.sim.v, apion.s.sim.v,
  carab.s.sim.v, ceram.s.sim.v, chaet.s.sim.v, chrys.s.sim.v,

```

```

crypt.s.sim.v, hydra.s.sim.v, lamin.s.sim.v, leptu.s.sim.v,
limne.s.sim.v, ochte.s.sim.v, otior.s.sim.v, phyll.s.sim.v,
polyd.s.sim.v, psela.s.sim.v, ptero.s.sim.v, scara.s.sim.v,
scoly.s.sim.v, silph.s.sim.v, trech.s.sim.v)

names(data.sim.complete)<- c("amara.n.sim.v", "apion.n.sim.v",
  "carab.n.sim.v", "ceram.n.sim.v", "chaet.n.sim.v", "chrys.n.sim.v",
  "crypt.n.sim.v", "hydra.n.sim.v", "lamin.n.sim.v", "leptu.n.sim.v",
  "limne.n.sim.v", "ochte.n.sim.v", "otior.n.sim.v", "phyll.n.sim.v",
  "polyd.n.sim.v", "psela.n.sim.v", "ptero.n.sim.v", "scara.n.sim.v",
  "scoly.n.sim.v", "silph.n.sim.v", "trech.n.sim.v", "amara.s.sim.v",
  "apion.s.sim.v", "carab.s.sim.v", "ceram.s.sim.v", "chaet.s.sim.v",
  "chrys.s.sim.v", "crypt.s.sim.v", "hydra.s.sim.v", "lamin.s.sim.v",
  "leptu.s.sim.v", "limne.s.sim.v", "ochte.s.sim.v", "otior.s.sim.v",
  "phyll.s.sim.v", "polyd.s.sim.v", "psela.s.sim.v", "ptero.s.sim.v",
  "scara.s.sim.v", "scoly.s.sim.v", "silph.s.sim.v", "trech.s.sim.v")

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# 4. Generación de función para comparación de los coeficientes de
regresión y representación de patrones distance-decay
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

## Definición de parámetros de la función

# B: bootstrap repeats
# sim: vector de similitud
# dist: vector de distancias euclideas

# glm 1: fam = gauss link = log; black
# glm 2: fam = gamma link = log; blue
  # a: intercept (exp(glm$coeff[1]))
  # b: slope (glm$coeff[2])

## Definición de la función

test.coef.Boot.density <- function(B, sim, dis, ...){
  a.1 <- numeric(B)      # Vector almacenamiento interceptos glm1 (gauss)
  b.1 <- numeric(B)      # Vector almacenamiento pendientes glm1 (gauss)

  a.2 <- numeric(B)      # Vector almacenamiento interceptos glm2 (gamma)
  b.2 <- numeric(B)      # Vector almacenamiento pendientes glm2 (gamma)

  pseudo.r.1 <- numeric(B) # Vector almacenamiento pseudoR2 glm1 (gauss)
  pseudo.r.2 <- numeric(B) # Vector almacenamiento pseudoR2 glm2 (gamma)

  v.ind <- 1:length(sim)
  # groups <- c(rep(0, B), rep(1, B))
  # groups <- factor(groups)

# Remuestreo de los datos y ajuste de los modelos para cada remuestra
for (i in 1:B){
  rem.ind <- sample(v.ind, size = length(sim), replace = T)

```

```

rem.sim <- sim[rem.ind]
rem.dist <- dis[rem.ind]

model.1 <- glm(rem.sim + 0.01 ~ rem.dist, family = gaussian(link =
  "log"))
model.2 <- glm(rem.sim + 0.01 ~ rem.dist, family = Gamma(link = "log"))

a.1[i] <- exp(model.1$coefficients[1])
b.1[i] <- model.1$coefficients[2]

a.2[i] <- exp(model.2$coefficients[1])
b.2[i] <- model.2$coefficients[2]

pseudo.r.1[i] = 1 - model.1$deviance/model.1$null.deviance
pseudo.r.2[i] = 1 - model.2$deviance/model.2$null.deviance

}

# Ajuste de los modelos distance-decay con datos observados (sin
remuestreo) con error gaussiano o gamma

glm.0.gauss <- glm(sim + 0.01 ~ dis, family = gaussian(link = "log"))
glm.0.gamma <- glm(sim + 0.01 ~ dis, family = Gamma(link = "log"))

# Representación gráfica

par(mfrow = c(1, 4))
plot(sim ~ dis, ylab = "Dissimilarity", xlab = "Spatial distance", ylim =
  c(0, 1))

points(glm.0.gauss$fitted.values[order(dis)] ~ sort(dis), col = "black",
  pch = 16, type = "l", lwd = 4)
points(glm.0.gamma$fitted.values[order(dis)] ~ sort(dis), col = "blue", pch
  = 16, type = "l", lwd = 4, lty = 2)

text(900, 0.07, bquote("Gaus.:" ~ italic(r)^2 == .(round(pseudo.r.1, digits
  = 2))))
text(900, 0.01, bquote("Gam.:" ~ italic(r)^2 == .(round(pseudo.r.2, digits =
  2))), col = "blue")

a <- c(a.1, a.2) # Vector con interceptos de modelos gaussiano y gamma
b <- c(b.1, b.2) # Vector con pendientes de modelos gaussiano y gamma

ps.R <- c(pseudo.r.1, pseudo.r.2) # Vector con pseudoR2 de modelos
gaussiano y gamma

ps.R.1.mu <- mean(pseudo.r.1)
ps.R.2.mu <- mean(pseudo.r.2)

plot(density(a.1, bw = "SJ"), col = "black", main = "", xlab = "Intercept",
  xlim = c(min(c(a)), max(c(a))))
lines(density(a.2, bw = "SJ"), col = "blue")
rug(a.1, col = "black")
rug(a.2, col = "blue")

```



```
# Norte
pdf(file = "amara.n.pdf", height = 3, width = 10)
amara.coeff.n <- test.coef.Boot.density(B = 500, sim = amara.n.sim.v, dis =
  distgeo.n.v)
dev.off()

pdf(file = "apion.n.pdf", height = 3, width = 10)
apion.coeff.n <- test.coef.Boot.density(B = 500, sim = apion.n.sim.v, dis =
  distgeo.n.v)
dev.off()

pdf(file = "carab.n.pdf", height = 3, width = 10)
carab.coeff.n <- test.coef.Boot.density(B = 500, sim = carab.n.sim.v, dis =
  distgeo.n.v)
dev.off()

pdf(file = "ceram.n.pdf", height = 3, width = 10)
ceram.coeff.n <- test.coef.Boot.density(B = 500, sim = ceram.n.sim.v, dis =
  distgeo.n.v)
dev.off()

pdf(file = "chaet.n.pdf", height = 3, width = 10)
chaet.coeff.n <- test.coef.Boot.density(B = 500, sim = chaet.n.sim.v, dis =
  distgeo.n.v)
dev.off()

pdf(file = "chrys.n.pdf", height = 3, width = 10)
chrys.coeff.n <- test.coef.Boot.density(B = 500, sim = chrys.n.sim.v, dis =
  distgeo.n.v)
dev.off()

pdf(file = "crypt.n.pdf", height = 3, width = 10)
crypt.coeff.n <- test.coef.Boot.density(B = 500, sim = crypt.n.sim.v, dis =
  distgeo.n.v)
dev.off()

pdf(file = "hydra.n.pdf", height = 3, width = 10)
hydra.coeff.n <- test.coef.Boot.density(B = 500, sim = hydra.n.sim.v, dis =
  distgeo.n.v)
dev.off()

pdf(file = "lamin.n.pdf", height = 3, width = 10)
lamin.coeff.n <- test.coef.Boot.density(B = 500, sim = lamin.n.sim.v, dis =
  distgeo.n.v)
dev.off()

pdf(file = "leptu.n.pdf", height = 3, width = 10)
leptu.coeff.n <- test.coef.Boot.density(B = 500, sim = leptu.n.sim.v, dis =
  distgeo.n.v)
dev.off()

pdf(file = "limne.n.pdf", height = 3, width = 10)
limne.coeff.n <- test.coef.Boot.density(B = 500, sim = limne.n.sim.v, dis =
  distgeo.n.v)
dev.off()
```

```

pdf(file = "ochte.n.pdf", height = 3, width = 10)
ochte.coeff.n <- test.coef.Boot.density(B = 500, sim = ochte.n.sim.v, dis =
  distgeo.n.v)
dev.off()

pdf(file = "otior.n.pdf", height = 3, width = 10)
otior.coeff.n <- test.coef.Boot.density(B = 500, sim = otior.n.sim.v, dis =
  distgeo.n.v)
dev.off()

pdf(file = "phyll.n.pdf", height = 3, width = 10)
phyll.coeff.n <- test.coef.Boot.density(B = 500, sim = phyll.n.sim.v, dis =
  distgeo.n.v)
dev.off()

pdf(file = "polyd.n.pdf", height = 3, width = 10)
polyd.coeff.n <- test.coef.Boot.density(B = 500, sim = polyd.n.sim.v, dis =
  distgeo.n.v)
dev.off()

pdf(file = "psela.n.pdf", height = 3, width = 10)
psela.coeff.n <- test.coef.Boot.density(B = 500, sim = psela.n.sim.v, dis =
  distgeo.n.v)
dev.off()

pdf(file = "ptero.n.pdf", height = 3, width = 10)
ptero.coeff.n <- test.coef.Boot.density(B = 500, sim = ptero.n.sim.v, dis =
  distgeo.n.v)
dev.off()

pdf(file = "scara.n.pdf", height = 3, width = 10)
scara.coeff.n <- test.coef.Boot.density(B = 500, sim = scara.n.sim.v, dis =
  distgeo.n.v)
dev.off()

pdf(file = "scoly.n.pdf", height = 3, width = 10)
scoly.coeff.n <- test.coef.Boot.density(B = 500, sim = scoly.n.sim.v, dis =
  distgeo.n.v)
dev.off()

pdf(file = "silph.n.pdf", height = 3, width = 10)
silph.coeff.n <- test.coef.Boot.density(B = 500, sim = silph.n.sim.v, dis =
  distgeo.n.v)
dev.off()

pdf(file = "trech.n.pdf", height = 3, width = 10)
trech.coeff.n <- test.coef.Boot.density(B = 500, sim = trech.n.sim.v, dis =
  distgeo.n.v)
dev.off()

# Generación de tabla de datos con los parámetros de los modelos
tabla.parametros.n<-round(rbind(apply(amara.coeff.n[[1]], 2,
mean),apply(apion.coeff.n[[1]], 2, mean),apply(carab.coeff.n[[1]], 2,

```

```
mean), apply(ceram.coeff.n[[1]], 2, mean), apply(chaet.coeff.n[[1]], 2,
mean), apply(chrys.coeff.n[[1]], 2, mean), apply(crypt.coeff.n[[1]], 2,
mean), apply(hydra.coeff.n[[1]], 2, mean), apply(lamin.coeff.n[[1]], 2,
mean), apply(leptu.coeff.n[[1]], 2, mean), apply(limne.coeff.n[[1]], 2,
mean), apply(ochte.coeff.n[[1]], 2, mean), apply(otior.coeff.n[[1]], 2,
mean), apply(phyll.coeff.n[[1]], 2, mean), apply(polyd.coeff.n[[1]], 2,
mean), apply(psela.coeff.n[[1]], 2, mean), apply(ptero.coeff.n[[1]], 2,
mean), apply(scara.coeff.n[[1]], 2, mean), apply(scoly.coeff.n[[1]], 2,
mean), apply(silph.coeff.n[[1]], 2, mean), apply(trech.coeff.n[[1]], 2,
mean)), 5)
```

```
colnames(tabla.parametros.n) <- c("a.gauss", "b.gauss", "r2.gauss",
"a.gamma", "b.gamma", "r2.gamma")
```

```
write.table(tabla.parametros.n, file="tabla-parametros-n.txt")
```

```
# -----
```

```
# Sur
```

```
pdf(file = "amara.s.pdf", height = 3, width = 10)
```

```
amara.coeff.s <- test.coef.Boot.density(B = 500, sim = amara.s.sim.v, dis =
distgeo.s.v)
```

```
dev.off()
```

```
pdf(file = "apion.s.pdf", height = 3, width = 10)
```

```
apion.coeff.s <- test.coef.Boot.density(B = 500, sim = apion.s.sim.v, dis =
distgeo.s.v)
```

```
dev.off()
```

```
pdf(file = "carab.s.pdf", height = 3, width = 10)
```

```
carab.coeff.s <- test.coef.Boot.density(B = 500, sim = carab.s.sim.v, dis =
distgeo.s.v)
```

```
dev.off()
```

```
pdf(file = "ceram.s.pdf", height = 3, width = 10)
```

```
ceram.coeff.s <- test.coef.Boot.density(B = 500, sim = ceram.s.sim.v, dis =
distgeo.s.v)
```

```
dev.off()
```

```
pdf(file = "chaet.s.pdf", height = 3, width = 10)
```

```
chaet.coeff.s <- test.coef.Boot.density(B = 500, sim = chaet.s.sim.v, dis =
distgeo.s.v)
```

```
dev.off()
```

```
pdf(file = "chrys.s.pdf", height = 3, width = 10)
```

```
chrys.coeff.s <- test.coef.Boot.density(B = 500, sim = chrys.s.sim.v, dis =
distgeo.s.v)
```

```
dev.off()
```

```
pdf(file = "crypt.s.pdf", height = 3, width = 10)
```

```
crypt.coeff.s <- test.coef.Boot.density(B = 500, sim = crypt.s.sim.v, dis =
distgeo.s.v)
```

```
dev.off()
```

```
pdf(file = "hydra.s.pdf", height = 3, width = 10)
```

```
hydra.coeff.s <- test.coef.Boot.density(B = 500, sim = hydra.s.sim.v, dis =
distgeo.s.v)
```

```
dev.off()

pdf(file = "lamin.s.pdf", height = 3, width = 10)
lamin.coeff.s <- test.coef.Boot.density(B = 500, sim = lamin.s.sim.v, dis =
  distgeo.s.v)
dev.off()

pdf(file = "leptu.s.pdf", height = 3, width = 10)
leptu.coeff.s <- test.coef.Boot.density(B = 500, sim = leptu.s.sim.v, dis =
  distgeo.s.v)
dev.off()

pdf(file = "limne.s.pdf", height = 3, width = 10)
limne.coeff.s <- test.coef.Boot.density(B = 500, sim = limne.s.sim.v, dis =
  distgeo.s.v)
dev.off()

pdf(file = "ochte.s.pdf", height = 3, width = 10)
ochte.coeff.s <- test.coef.Boot.density(B = 500, sim = ochte.s.sim.v, dis =
  distgeo.s.v)
dev.off()

pdf(file = "otior.s.pdf", height = 3, width = 10)
otior.coeff.s <- test.coef.Boot.density(B = 500, sim = otior.s.sim.v, dis =
  distgeo.s.v)
dev.off()

pdf(file = "phyll.s.pdf", height = 3, width = 10)
phyll.coeff.s <- test.coef.Boot.density(B = 500, sim = phyll.s.sim.v, dis =
  distgeo.s.v)
dev.off()

pdf(file = "polyd.s.pdf", height = 3, width = 10)
polyd.coeff.s <- test.coef.Boot.density(B = 500, sim = polyd.s.sim.v, dis =
  distgeo.s.v)
dev.off()

pdf(file = "psela.s.pdf", height = 3, width = 10)
psela.coeff.s <- test.coef.Boot.density(B = 500, sim = psela.s.sim.v, dis =
  distgeo.s.v)
dev.off()

pdf(file = "ptero.s.pdf", height = 3, width = 10)
ptero.coeff.s <- test.coef.Boot.density(B = 500, sim = ptero.s.sim.v, dis =
  distgeo.s.v)
dev.off()

pdf(file = "scara.s.pdf", height = 3, width = 10)
scara.coeff.s <- test.coef.Boot.density(B = 500, sim = scara.s.sim.v, dis =
  distgeo.s.v)
dev.off()

pdf(file = "scoly.s.pdf", height = 3, width = 10)
scoly.coeff.s <- test.coef.Boot.density(B = 500, sim = scoly.s.sim.v, dis =
  distgeo.s.v)
```

```

dev.off()

pdf(file = "silph.s.pdf", height = 3, width = 10)
silph.coeff.s <- test.coef.Boot.density(B = 500, sim = silph.s.sim.v, dis =
  distgeo.s.v)
dev.off()

pdf(file = "trech.s.pdf", height = 3, width = 10)
trech.coeff.s <- test.coef.Boot.density(B = 500, sim = trech.s.sim.v, dis =
  distgeo.s.v)
dev.off()

# Generación de tabla de datos con parámetros de los modelos

tabla.parametros.s<-round(rbind(apply(amara.coeff.s[[1]], 2,
  mean),apply(apion.coeff.s[[1]], 2, mean),apply(carab.coeff.s[[1]], 2,
  mean),apply(ceram.coeff.s[[1]], 2, mean),apply(chaet.coeff.s[[1]], 2,
  mean),apply(chrys.coeff.s[[1]], 2, mean),apply(crypt.coeff.s[[1]], 2,
  mean),apply(hydra.coeff.s[[1]], 2, mean),apply(lamin.coeff.s[[1]], 2,
  mean),apply(leptu.coeff.s[[1]], 2, mean),apply(limne.coeff.s[[1]], 2,
  mean),apply(ochte.coeff.s[[1]], 2, mean),apply(otior.coeff.s[[1]], 2,
  mean),apply(phyll.coeff.s[[1]], 2, mean),apply(polyd.coeff.s[[1]], 2,
  mean),apply(psela.coeff.s[[1]], 2, mean),apply(ptero.coeff.s[[1]], 2,
  mean),apply(scara.coeff.s[[1]], 2, mean),apply(scoly.coeff.s[[1]], 2,
  mean),apply(silph.coeff.s[[1]], 2, mean),apply(trech.coeff.s[[1]], 2,
  mean)), 5)

colnames(tabla.parametros.s)<-c("a.gauss", "b.gauss", "r2.gauss",
  "a.gamma", "b.gamma", "r2.gamma")

write.table(tabla.parametros.s, file="tabla-parametros-s.txt")

```

9.3 CÓDIGO DE R PARA EL ESTUDIO DE APLICACIÓN DE LOS MÉTODOS DE *BOOTSTRAP* DISPONIBLES EN LA LITERATURA

```

# - . # - . # - . # - . # - . # - . # - . # - . # - . # - . #
# ESTUDIO MÉTODOS BOOTSTRAP DE LA LITERATURA
# - . # - . # - . # - . # - . # - . # - . # - . # - . # - . #

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# 1. Librerías
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

library(sm)
library(vegan)
library(betapart)

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# 2. Codificación básica del Bootstrap de bloques móviles
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

# Dataframe de distancias espaciales (x) y similitudes bióticas (y), tras
#   la conversión de las matrices correspondientes en vectores

data <- data.frame(x,y)

# Condiciones del remuestreo

k <- ceiling(nrow(data)/bl) # Número de bloques en la remuestra
q <- nrow(data)-bl+1       # Número de posibles bloques
obs.blocks <- 1:nrow(data) #Posiciones a partir de las cuales construir
#   bloques

# Construcción de la matriz de bloques

for (i in 1:q){
  if (i == 1) blocks <- obs.blocks[i:(i + bl - 1)]
  if (i! = 1) blocks <- cbind(blocks,obs.blocks[i:(i + bl - 1)])
}
for (i in 1:B){
  i.resample <- sample(x = 1:ncol(blocks), size = k, replace = T)#
  # Vector de los índices de los k bloques seleccionados
  block.resample <- as.vector(blocks[,i.resample]) #Vector I cuyos
  # elementos son los índices de los valores muestrales
  resample <- data[block.resample,] resample <- resample[1:nrow(data), ]
  # Matriz cuyas columnas son X* e Y*
  x.resample <- resample[ ,1]
  y.resample <- resample[ ,2]
}

```

```

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# 3. Codificación básica del Bootstrap de bloques manuales
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

size <- 16 # Número de localizaciones de la muestra

x.m # Semimatriz de distancia geográfica.
y.m # Semimatriz de similitud biótica

# Creación de bloques para remuestreo
pre.b.blocks.x <- matrix(rep(0, size*2*size), ncol = size)
pre.b.blocks.y <- matrix(rep(0, size*2*size), ncol = size)

b.blocks.x <- matrix(rep(0, size*(size-1)), ncol = size)
b.blocks.y <- matrix(rep(0, size*(size-1)), ncol = size)

# Generación de las matrices que contendrán las observaciones de los
bloques (cada columna es un bloque)
for (i in 1:size){
  # Selección de las observaciones de la columna y fila i
  pre.b.blocks.x[ , i] <- c(x.m[i, ], x.m[ , i])
  pre.b.blocks.y[ , i] <- c(y.m[i, ], y.m[ , i])

  # Generación de las matrices con los bloques, de las que se han
  eliminado los valores NA correspondientes a la parte superior de una
  matriz completa (parte superior de semimatriz)
  b.blocks.x[ , i] <- pre.b.blocks.x[ , i][which(pre.b.blocks.x[,i ] !=
    "NA")]
  b.blocks.y[ , i] <- pre.b.blocks.y[ , i][which(pre.b.blocks.y[ , i] !=
    "NA")]
}
for (i in 1:B){
  b.i.resample <- sample(x = 1:size, size = size, replace = T) # s
  observaciones posibles
  b.x.resample <- b.blocks.x[ , b.i.resample] # Matriz remuestra de X
  con los valores remuestreados como columnas bloques
  b.y.resample <- b.blocks.y[ , b.i.resample] #Matriz remuestra de Y con
  los valores remuestreados como columnas bloques

  # Vectores remuestra de X e Y creados a partir de las matrices
  remuestra.
  v.b.x.resample <- as.vector(b.x.resample)
  v.b.y.resample <- as.vector(b.y.resample)
}

```

```

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# 4. Simulación de datos
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

# Identificación de valores de referencia de parámetros para simulación a
partir del patron-distance decay de Pterostichus

mod.base <- glm(ptero.s.sim ~ distgeo.s, family = gaussian(link="log"),
  start = c(0.005, -0.00001))
a <- exp(mod.base$coefficients[1])
b <- mod.base$coefficients[2]

# -----
# Función para obtener variables geográficas y realizar el cálculo de la
matriz de distancia entre ellas(X).
# position.x y position.y son coordenadas geográficas,
#size es el número de lugares muestrados (N).

x.dist.geo <- function(position.x, position.y,size){
  x_posx <- sample(seq(from = min(position.x), to = max(position.x)),
    size = size, replace = F)
  x_posy <- sample(seq(from = min(position.y), to = max(position.y)),
    size = size, replace = F)

  # Dataframe para cálculo de matriz de distancia entre coordenadas
simuladas
  x_pos <- data.frame(x_posx, x_posy)
  x_dist <- dist(x_pos, method = "euclidean") # Semimatriz de distancia
geográficas
  #Transformación de objeto tipo "dist" en uno de tipo "matrix" y
posteriormente en vector
  x <- as.matrix(x_dist)
  x[upper.tri(x)] <- 0
  x[diag(x)] <- 0
  x_v <- as.vector(x_dist)
  return(list(x, x_dist, x_v, x_pos))
}

# -----
# Función para obtener obtener valores de similitud biótica (Y), aplicando
los parámetros obtenidos con modelo de referencia
# X debe introducirse como un objeto de tipo "matrix".
# a.exp.N.log y b.exp.N.log son a y b.
simulated.data <- function(x, a.exp.N.log, b.exp.N.log){
  error_N_log_exp <- matrix(rnorm(n = (nrow(x)*ncol(x)), sd = 0.3*0.5),
    ncol = ncol(x), nrow = nrow(x))
  y.exp.N.log <- a.exp.N.log*exp(b.exp.N.log*x) + error_N_log_exp #
Construcción de variable respuesta (valores de similitud, Y)
  y.exp.N.log <- as.dist(y.exp.N.log) #Transformación en objeto dist

```

```

error_N_log_exp <- as.dist(error_N_log_exp)
return(list(y.exp.N.log, error_N_log_exp, coeff = c(a.exp.N.log,
          b.exp.N.log)))
}

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# 5. Definición de función para estimación de coeficientes de regresión y
#   desviación típica y representación gráfica
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

# position.x y position.y son coordenadas geográficas, size es el número de
# lugares muestrados (N).
# a.exp.N.log y b.exp.N.log son a y b.
# bl es el tamaño de bloque y B el numero de remuestras bootstrap
# Si plot = T se generan las gráficas de las distribuciones bootstrap y los
# ajustes

par.bias.sd <- function(position.x, position.y, size, a.exp.N.log,
  b.exp.N.log, bl, B, plot=F){
  # Simulación de X
  x.tot <- x.dist.geo(position.x, position.y, size)
  x.m <- x.tot[[1]]
  x.v <- x.tot[[3]]

  # Simulación de Y
  sim.data <- simulated.data(x.m, a.exp.N.log, b.exp.N.log)
  y.v <- as.vector(sim.data[[1]])
  y.m <- matrix(rep(0, size^2), ncol = size)
  y.m[lower.tri(y.m)] <- y.v

  # Creación de dataframe con datos de X y de Y en formato vectorial
  data <- data.frame(x.v, y.v)

#-----
##IRWLS#
# Construcción del modelo
mod <- glm(data[, 2] ~ data[,1],family = gaussian(link=c"log"),c
  start = c(0.005, -0.00001))
  a.N <- exp(mod$coefficients[1]). # Interceptos
  sd.a.N <- summary(mod)$coefficients[1, 2]# desviación típica

  b.N <- mod$coefficients[2] # Pendientes
  sd.b.N <- summary(mod)$coefficients[2, 2]# desviación típica

## Bootstrap##
## Bootstrap de bloques móviles
k <- ceiling(nrow(data)/bl)
q <- nrow(data) - bl + 1

```

```

obs.blocks <- 1:nrow(data)
for (i in 1:q){
  if (i == 1) blocks <- obs.blocks[i:(i+bl-1)]
  if (I != 1) blocks <- cbind(blocks, obs.blocks[i:(i+bl-1)])
}

# Vectores para guardar las estimaciones bootstrap de a y b
a.MBB.boots <- numeric(B)
b.MBB.boots <- numeric(B)

## Bootstrap de bloques manuales
pre.b.blocks.x <- matrix(rep(0, size*2*size), ncol = size)
pre.b.blocks.y <- matrix(rep(0, size*2*size), ncol = size)

b.blocks.x <- matrix(rep(0, size*(size - 1)), ncol = size)
b.blocks.y <- matrix(rep(0, size*(size - 1)), ncol = size)

for (i in 1:size){
  pre.b.blocks.x[ ,i] <- c(x.m[i, ], x.m[ , i])
  pre.b.blocks.y[ ,i] <- c(y.m[i, ], y.m[ , i])
  b.blocks.x[ ,i] <- pre.b.blocks.x[ ,i][~which(pre.b.blocks.x[ ,i]==0)]
  b.blocks.y[ ,i] <- pre.b.blocks.y[ ,i][~which(pre.b.blocks.y[ ,i]==0)]
}

# vector para guardar las estimaciones bootstrap de a y b
a.B.boots <- numeric(B)
b.B.boots <- numeric(B)

for (i in 1:B){
  # Bootstrap de bloques móviles
  i.resample <- sample(x = 1:ncol(blocks), size = k, replace = T)
  block.resample <- as.vector(blocks[, i.resample])
  resample <- data[block.resample,]
  resample <- resample[1:nrow(data), ]
  x.resample <- resample[ ,1]
  y.resample <- resample[ ,2]

  boot.mod <- glm(y.resample ~ x.resample, family = gaussian(link =
    "log"), start = c(0.005, -0.00001))

  a.MBB.boots[i] <- exp(boot.mod$coefficients[1])
  b.MBB.boots[i] <- boot.mod$coefficients[2]

  #Bootstrap de bloques manuales
  b.i.resample <- sample(x = 1:size, size = size, replace = T)
  b.x.resample <- b.blocks.x[, b.i.resample]
  b.y.resample <- b.blocks.y[, b.i.resample]

```

```

v.b.x.resample <- as.vector(b.x.resample)
v.b.y.resample <- as.vector(b.y.resample)

b.boot.mod <- glm(v.b.y.resample ~ v.b.x.resample, family =
  gaussian(link = "log"), start = c(0.005,-0.00001))

a.B.boots[i] <- exp(b.boot.mod$coefficients[1])
b.B.boots[i] <- b.boot.mod$coefficients[2] }

# Representación gráfica
if (plot == T){
  par(mfrow = c(1, 2))
  plot(density(a.MBB.boots), xlab = "Distribución de a (MBB)", xlim =
    c(min(a.MBB.boots, sim.data[[3]][1]), max(a.MBB.boots,
    sim.data[[3]][1])))
  abline(v = sim.data[[3]][1], col="black")
  abline(v = mean(a.MBB.boots), col = "blue")

  plot(density(a.B.boots), xlab = "Distribución de a (bloques
  manuales)", xlim = c(min(a.B.boots, sim.data[[3]][1]),
  max(a.B.boots, sim.data[[3]][1])))
  abline(v = sim.data[[3]][1], col = "black")
  abline(v = mean(a.B.boots), col = "blue")
}

mu.a.MBB.boots <- mean(a.MBB.boots)
bias.a.MBB.boots <- mu.a.MBB.boots - sim.data[[3]][1] #sesgo de a del
  Bootstrap de bloques móviles
sd.a.MBB.boots <- sd(a.MBB.boots) #desviación típica de a del
  bootstrap de bloques móviles

mu.a.B.boots <- mean(a.B.boots)
bias.a.B.boots <- mu.a.B.boots - sim.data[[3]][1] #sesgo de a del
  Bootstrap de bloques móviles
sd.a.B.boots <- sd(a.B.boots) #desviación típica de a del bootstrap de
  bloques móviles

bias.a.N <- a.N-sim.data[[3]][1] #sesgo de a de IRWLS

if (plot==T){
  par(mfrow = c(1, 2))
  plot(density(b.MBB.boots), xlab = "Distribución de b (MBB)", xlim =
    c(min(b.MBB.boots, sim.data[[3]][2]), max(b.MBB.boots,
    sim.data[[3]][2])))
  abline(v = sim.data[[3]][2], col = "black")
  abline(v = mean(b.MBB.boots), col = "blue")
  plot(density(b.B.boots), xlab = "Distribución de b (bloques
  manuales)", xlim = c(min(b.B.boots, sim.data[[3]][2]),
  max(b.B.boots, sim.data[[3]][2])))

```

```

    abline(v = sim.data[[3]][2], col = "black") abline(v =
    mean(b.B.boots), col = "blue")
  }

mu.b.MBB.boots <- mean(b.MBB.boots)
bias.b.MBB.boots <- mu.b.MBB.boots -s im.data[[3]][2] #sesgo de b del
  bootstrap de bloques móviles
sd.b.MBB.boots <- sd(b.MBB.boots) #desviación típica de b del
  bootstrap de bloques móviles

mu.b.B.boots <- mean(b.B.boots)
bias.b.B.boots <- mu.b.B.boots-sim.data[[3]][2] #sesgo de b del
  Bootstrap de bloques manuales
sd.b.B.boots <- sd(b.B.boots) #desviación típica de b del bootstrap de
  bloques manuales

bias.b.N <- b.N-sim.data[[3]][2] #sesgo de b de IRWLS

if (plot == T){
  par(mfrow = c(1, 1))
  plot(y.v ~ x.v)
  y.MBB <- mu.a.MBB.boots*exp(mu.b.MBB.boots*sort(x.v)) lines(y.MBB ~
  sort(x.v), col = "blue")
  y.B <- mu.a.B.boots*exp(mu.b.B.boots*sort(x.v)) lines(y.B ~
  sort(x.v), col= "green", lty = 2)
  y.norm <- sim.data[[3]][1]*exp(sim.data[[3]][2]*sort(x.v))
  lines(y.norm ~ sort(x.v), col = "black")
}

#-----
norm.bias.sd <- data.frame(a.N, bias.a.N, sd.a.N, b.N, bias.b.N, sd.b.N)
MBB.bias.sd <- data.frame(mu.a.MBB.boots, bias.a.MBB.boots, sd.a.MBB.boots,
  mu.b.MBB.boots, bias.b.MBB.boots, sd.b.MBB.boots)
B.bias.sd <- data.frame(mu.a.B.boots, bias.a.B.boots, sd.a.B.boots,
  mu.b.B.boots, bias.b.B.boots, sd.b.B.boots)
return(list(norm.bias.sd,MBB.bias.sd,B.bias.sd))
}

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# 6. Definición de función para contraste de significación del modelo
  basado en la deviance
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

niv.d <- function(nsim, position.x, position.y, size, a.exp.N.log,
  b.exp.N.log, bl, B, alpha){
  p.val.norm.D <- numeric(nsim)
  p.val.MBB.D <- numeric(nsim)
  p.val.B.D <- numeric(nsim)
  for (s in 1:nsim){
    x.tot <- x.dist.geo(position.x,position.y,size)

```

```

x.m <- x.tot[[1]]
x.v <- x.tot[[3]]

sim.data <- simulated.data(x.m,a.exp.N.log,b.exp.N.log)
y.v <- sample(sim.data[[1]]) #Aleatorización de Y
y.m <- matrix(rep(0,size^2),ncol = size)
y.m[lower.tri(y.m)] <- y.v

data <- data.frame(x.v,y.v

# ----
# Bootstrap de bloques móviles
k <- ceiling(nrow(data)/bl)
q <- nrow(data)-bl+1
obs.blocks <- 1:nrow(data)

for (i in 1:q){
  if (i==1) blocks <- obs.blocks[i:(i+bl-1)]
  if (i!=1) blocks <- cbind(blocks,obs.blocks[i:(i+bl-1)])
}

D.MBB.boots <- numeric(B) # Observaciones del estadístico D
                          (deviance modelo nulo - deviance modelo ajustado)
                          del bootstrap de bloques móviles

# ----
# Bootstrap de bloques manuales
pre.b.blocks.x <- matrix(rep(0, size*2*size), ncol = size)
pre.b.blocks.y <- matrix(rep(0,size*2*size),ncol = size)
b.blocks.x <- matrix(rep(0, size*(size - 1)),ncol = size)
b.blocks.y <- matrix(rep(0, size*(size - 1)),ncol = size)

for (i in 1:size){
  pre.b.blocks.x[ ,i] <- c(x.m[i, ], x.m[ ,i])
  pre.b.blocks.y[ ,i] <- c(y.m[i, ], y.m[ ,i])
  b.blocks.x[ ,i] <- pre.b.blocks.x[ ,i][-
    which(pre.b.blocks.x[ ,i] == 0)]
  b.blocks.y[ ,i] <- pre.b.blocks.y[ ,i][-
    which(pre.b.blocks.y[ ,i] == 0)]
}

D.B.boots <- numeric(B) # Observaciones del estadístico D
                      (deviance modelo nulo - deviance modelo ajustado)
                      del bootstrap de bloques

for (i in 1:B){
# ----
# Bootstrap de bloques móviles

```

```

i.resample <- sample(x = 1:ncol(blocks), size = k, replace = T)
block.resample.x <- as.vector(blocks[, i.resample])
r.block.resample.y <- as.vector(apply(blocks[, i.resample], 2,
  sample))

# Aleatorización de los bloques
x.resample <- data[block.resample.x,1]
x.resample <- x.resample[1:nrow(data)]
r.y.resample <- data[r.block.resample.y, 2]
r.y.resample <- r.y.resample[1:nrow(data)]
boot.mod.d <- glm(r.y.resample ~ x.resample, family gaussian(link
  = "log"), cstart = c(0.005, -0.00001))

D.MBB.boots[i] <- anova(boot.mod.d)$Deviance[2]

# ----
# Bootstrap de bloques manuales
b.i.resample <- sample(x = 1:size, size = size, replace = T)
b.x.resample <- b.blocks.x[, b.i.resample]
b.y.resample <- b.blocks.y[, b.i.resample]
r.b.y.resample <- apply(b.y.resample, 2, sample) #Aleatorización de
  los observaciones de los bloques (columnas)
v.b.x.resample <- as.vector(b.x.resample)
r.v.b.y.resample <- as.vector(r.b.y.resample)

b.boot.mod.d <- glm(r.v.b.y.resample ~ v.b.x.resample, family =
  gaussian(link = "log"), start = c(0.005, -0.00001))
D.B.boots[i] <- anova(b.boot.mod.d)$Deviance[2] }

mod <- glm(y.v ~ x.v, family = gaussian(link = "log"), start =
  c(0.005, -0.00001))

D <- anova(mod)$Deviance[2] #Valor muestral del estadístico

p.val.norm.D[s] <- anova(mod, test = "F")$`Pr(>F)`[2] #p-val del
  test, con distribucion asintotica
p.val.MBB.D[s] <- mean(D.MBB.boots > D) #p-val del test para
  bootstrap de bloques móviles
p.val.B.D[s] <- mean(D.B.boots > D) #p-val del test para bootstrap
  de bloques manuales
}
niv.norm.D <- mean(p.val.norm.D < alpha)
niv.MBB.D <- mean(p.val.MBB.D < alpha)
niv.B.D <- mean(p.val.B.D < alpha)
return(data.frame(niv.norm.D, niv.MBB.D, niv.B.D))
}

```

```

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# 7. Definición de función para selección de modelos mediante AIC
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
AIC.freq <- function(position.x, position.y, size, a.exp.N.log,
  b.exp.N.log, bl, B){
  x.tot <- x.dist.geo(position.x, position.y, size)
  x.m <- x.tot[[1]]
  x.v <- x.tot[[3]]

  sim.data <- simulated.data(x.m, a.exp.N.log, b.exp.N.log)
  y.v <- as.vector(sim.data[[1]])
  y.m <- matrix(rep(0, size^2), ncol = size)
  y.m[lower.tri(y.m)] <- y.v

  data<-data.frame(x.v, y.v)

  ## -----
  ## Bootstrap de bloques móviles
  K <- ceiling(nrow(data)/bl)
  q <- nrow(data)-bl+1
  obs.blocks <- 1:nrow(data)

  for (i in 1:q){
    if (i==1) blocks <- obs.blocks[i:(i + bl - 1)]
    if (i!=1) blocks <- cbind(blocks, obs.blocks[i:(i + bl - 1)])
  }

  # Vector para guardar cuál de los dos modelos es el mejor
  best.AIC.MBB.boots <- numeric(B)

  # Vector para guardar el valor del AIC del mejor modelo
  best.mod.MBB.boots <- numeric(B)

  ## -----
  # Bootstrap de bloques manuales
  pre.b.blocks.x <- matrix(rep(0, size*2*size), ncol = size)
  pre.b.blocks.y <- matrix(rep(0, size*2*size), ncol = size)
  b.blocks.x <- matrix(rep(0, size*(size-1)), ncol = size)
  b.blocks.y <- matrix(rep(0, size*(size-1)), ncol = size)

  for (i in 1:size){
    pre.b.blocks.x[ ,i] <- c(x.m[i, ], x.m[ ,i])
    pre.b.blocks.y[ ,i] <- c(y.m[i, ], y.m[ ,i])
    b.blocks.x[ ,i] <- pre.b.blocks.x[ ,i][~which(pre.b.blocks.x[
      ,i]==0)]
    b.blocks.y[ ,i] <- pre.b.blocks.y[ ,i][~which(pre.b.blocks.y[
      ,i]==0)]
  }

```

```

    }

    # Vector para guardar cuál de los dos modelos es el mejor
    best.mod.B.boots <- numeric(B)

    # Vector para guardar el valor del AIC del mejor modelo
    best.AIC.B.boots <- numeric(B)

for (i in 1:B){
## -----
# Bootstrap de bloques móviles
i.resample <- sample(x = 1:ncol(blocks), size = k, replace = T)
block.resample <- as.vector(blocks[ ,i.resample])
resample <- data[block.resample, ]
resample <- resample[1:nrow(data), ]
x.resample <- resample[ ,1]
y.resample <- resample[ ,2]

boot.mod.1 <- glm(y.resample ~ x.resample, family = gaussian(link = "log"),
start = c(0.005, -0.00001))

boot.mod.2 <- glm(y.resample ~ log(x.resample), family = gaussian(link =
"log"), start = c(0.005, -0.00001))

MBB.mod.AIC <- c(boot.mod.1$aic, boot.mod.2$aic)
best.mod.MBB.boots[i] <- which.min(MBB.mod.AIC)
best.AIC.MBB.boots[i] <- MBB.mod.AIC[which.min(MBB.mod.AIC)]

## -----
# Bootstrap de bloques manuales
b.i.resample <- sample(x = 1:size, size = size, replace = T)
b.x.resample <- b.blocks.x[ ,b.i.resample]
b.y.resample <- b.blocks.y[ ,b.i.resample]
v.b.x.resample <- as.vector(b.x.resample)
v.b.y.resample <- as.vector(b.y.resample)

b.boot.mod.1 <- glm(v.b.y.resample ~ v.b.x.resample, family = gaussian(link
= "log"), start = c(0.005, -0.00001))

b.boot.mod.2 <- glm(v.b.y.resample ~ log(v.b.x.resample), family =
gaussian(link = "log"), start = c(0.005, -0.00001))
B.mod.AIC <- c(b.boot.mod.1$aic, b.boot.mod.2$aic)
best.mod.B.boots[i] <- which.min(B.mod.AIC)
best.AIC.B.boots[i] <- B.mod.AIC[which.min(B.mod.AIC)]
}

```

```
# Bootstrap de bloques móviles (resultados)
data.MBB <- data.frame(best.mod.MBB.boots, best.AIC.MBB.boots)
freq.MBB.1 <- length(which(best.mod.MBB.boots == 1))/B
freq.MBB.2 <- length(which(best.mod.MBB.boots == 2))/B
freq.MBB <- c(freq.MBB.1, freq.MBB.2)

#Bootstrap de bloques manuales (resultados)
data.B <- data.frame(best.mod.B.boots, best.AIC.B.boots)
freq.B.1 <- length(which(best.mod.B.boots == 1))/B
freq.B.2 <- length(which(best.mod.B.boots == 2))/B
freq.B <- c(freq.B.1, freq.B.2)

return(list(data.MBB, freq.MBB, data.B, freq.B))
}
```

9.4 CÓDIGO DE R PARA LA IMPLEMENTACIÓN DE PERMUTACIONES Y REMUESTREO POR BLOQUES DE LOCALIDAD

```

# - . # - . # - . # - . # - . # - . # - . # - . # - . # - . # - . #
# MÉTODOS DE PERMUTACIONES Y REMUESTREO POR LOCALIDAD PARA MODELOS DE
# DISTANCE-DECAY
# - . # - . # - . # - . # - . # - . # - . # - . # - . # - . #

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# 0. PREPARACIÓN DE LA SESIÓN
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

#Carga de librerías
library(sm)
library(vegan)
library(betapart)
library(geoR)
library(glm2)      # librería usada para proveer de mayor estabilidad a los
                    # modelos simulados, que podrían no converger usando glm().
library(ggplot2)
library(dplyr)

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# 1. SIMULACIÓN DE DATOS
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

# PARÁMETROS BÁSICOS DE SIMULACIÓN
# (Se corren simulaciones para 100 localidades y 150 especies)

set.seed(546454)

nsp = 150      # Número de especies
nsites = 100  # Número de sitios/localidades

#-----
## 1A. SIMULACIÓN DE LA RELACIÓN DE DISTANCE-DECAY BAJO LA HIPÓTESIS NULA
# (NO HAY DISTANCE-DECAY) ##

# SIMULACIÓN DE COORDENADAS ESPACIALES
# Coordenadas espaciales aleatorias para nsites en una cuadrícula de
# 9000x9000

coords.H0 <- cbind(runif(nsites, min = 0, max = 9000), # coordenada X
                  runif(nsites, min = 0, max = 9000)) # coordenada Y

```

```

# Distancia espacial (distancia euclídea) entre localidades
H0.Dx <- dist(coords.H0)

# SIMULACIÓN DE DATOS BIÓTICOS
# Simulación de datos de presencia/ausencia para n especies (nsp) en n
localidades (nsites).
# La probabilidad de ocupación de una especie sigue una distribución
lognormal (algunas especies comunes y la mayoría raras).

H0.sp <- matrix(ncol = nsp, nrow = nsites) # Matriz de presencia/ausencia
vacía
lnorm <- rlnorm(nsp)# Generación aleatoria de datos de ocupación para n
especies (nsp) siguiendo una distribución lognormal
threshold <- plnorm(lnorm, mean(lnorm)) # Transformación de la
probabilidad lognormal de ocupación en un valor umbral para convertir
los datos a presencia(1)/ausencia(0)

for (u in 1:nsp){
  H0.sp[,u] <- rbinom(n = nsites, size = 1, prob = threshold[u]) #
  Transformación en datos de presencia/ausencia usando el umbral lognormal
}

# Cálculo de la similitud en la composición específica entre localidades (1
- disimilitud de Simpson)
H0.Sy <- 1 - beta.pair(H0.sp)$beta.sim

# Representación gráfica del distance-decay
data.H0 <- data.frame(x = as.vector(H0.Dx), y = as.vector(H0.Sy))

H0.plot<-ggplot(data.H0, aes(x = x, y = y)) +
  geom_point(colour = "gray47", shape = 16, size = 2, alpha = 0.8) +
  xlab("Distancia geográfica (km)") +
  ylab("Similitud de Simpson") +
  ggtitle("Distance-decay bajo la hipótesis nula") +
  theme_bw()

H0.plot

#-----
## 1B. SIMULACIÓN DE LA RELACIÓN DE DISTANCE-DECAY BAJO LA HIPÓTESIS
ALTERNATIVA ##

## Generación de datos espacialmente autocorrelacionados de
presencia/ausencia y coordenadas espaciales usando campos aleatorios
Gaussianos (GRFs)

phi = 5000 # Se proporciona el ejemplo de una hipótesis alternativa con
fuerte autocorrelación espacial (phi = 5000). Modificar phi para otros
niveles de autocorrelación espacial en la composición específica de las
comunidades.

```

```

GRF.phi <- grf(n = nsites, cov.pars = c(sigma.sq = 1, phi), nsim = nsp, #
  Generación de variables normales autocorrelacionadas y coordenadas
  espaciales con GRF
    xlims = c(0, 9000),
    ylims = c(0, 9000)
  )

assign(paste("phi", phi, ".coords", sep=""), GRF.phi$coords) # Coordenadas
  espaciales
assign(paste("phi", phi, ".Dx", sep=""), dist(GRF.phi$coords)) # Distancias
  espaciales euclídeas entre localidades

M_phi = GRF.phi$data          # Variables normales autocorrelacionadas (n =
  nsp = 150)
M_norm_phi <- pnorm(M_phi) # Transformación de valores normales
  autocorrelacionados en densidades acumuladas (0-1) para aplicar un
  umbral de corte para su conversión a presencias/ausencias

# Generación de umbrales de corte para convertir valores
  autocorrelacionados en datos de presencia/ausencia siguiendo una
  distribución lognormal de datos de ocupación
lnorm <- rlnorm(nsp)          # Generación aleatoria de datos de ocupación para
  n especies (nsp) siguiendo una distribución lognormal
threshold <- plnorm(lnorm, mean(lnorm)) # Transformación de la probabilidad
  de ocupación lognormal en umbrales de corte

# Aplicación de los umbrales de corte a variables autocorrelacionadas para
  transformarlas en datos de presencia(1)/ausencia(0)
M_norm_phi <- (t(t(M_norm_phi) < threshold))*1

# Cálculo de la similitud en la composición específica entre localidades (1
  - disimilitud de Simpson)
assign(paste("phi", phi, ".Sy", sep=""), 1 - beta.pair(M_norm_phi)$beta.sim)
  # Semimatriz de similitud

# Representación visual del distance-decay
data.phi <- data.frame(x = as.vector(get(paste("phi", phi, ".Dx", sep =
  ""))), y = as.vector(get(paste("phi", phi, ".Sy", sep = ""))))

phi.plot <- ggplot(data.phi, aes(x = x, y = y)) +
  geom_point(colour = "gray47", shape = 16, size = 2, alpha = 0.8) +
  xlab("Distancia geográfica (km)") +
  ylab("Similitud de Simpson") +
  ggtitle(paste("Distance-decay bajo la hipótesis alternativa (phi =
  ", phi, ")", sep = "")) +
  theme_bw()

```



```

    perm.dev.R2.distr[i] <- (dev.null - dev.perm)/dev.null # Valores
    de pseudo-R2 (n = perm)
  }

mod <- glm2(ydis ~ xdis, family = gaussian(link = "log"), start =
  st.val) # Modelo de distance-decay de los datos observados

dev <- deviance(mod) # Deviance del modelo de distance-decay de los
  datos observados
dev.R2 <- (dev.null - dev)/dev.null # Valor pseudo-R2 del modelo de
  distance-decay en los datos observados
p.val <- mean(perm.dev.R2.distr > dev.R2) # P-valor del modelo de
  distance-decay en los datos observados
}

else if (model.type == "pow"){
  mod.null <- glm2(ydis ~ 1, start = st.val[1]) # Modelo nulo de distance-
  decay
  dev.null <- deviance(mod.null) # Deviance del modelo nulo

  for (i in 1:perm){
    ####Aleatorización por bloques
    perm.sites <- sample(1:nsites)
    ydismat <- as.matrix(ydis) # Conversión de la matriz triangular
    en cuadrada
    perm.ydis <- ydismat[perm.sites, perm.sites][lower.tri(ydismat)] #
    Permutación de la matriz de similitud (la permutación de filas y
    columnas en la matriz cuadrada debe ser simultánea para asegurar
    la preservación de la estructura a pares). En los análisis
    posteriores sólo se necesita la matriz triangular inferior

    mod.perm <- glm2(perm.ydis ~ log(xdis), family = gaussian(link =
    "log"), start = st.val) # Modelo de distance-decay ajustado con
    los datos permutados

    dev.perm <- deviance(mod.perm) # Deviance del modelo de
    distance-decay ajustado a los datos permutados
    perm.dev.R2.distr[i]<- (dev.null - dev.perm)/dev.null # Valores
    de pseudo-R2 (n = perm)
  }

  mod <- glm2(ydis ~ log(xdis), family = gaussian(link = "log"), start =
    st.val) # Modelo de distance-decay de los datos observados
  dev <- deviance(mod) # Deviance del modelo de distance-decay de los
  datos observados
  dev.R2 <- (dev.null - dev)/dev.null # Valor pseudo-R2 del modelo
  de distance-decay en los datos observados
  p.val <- mean(perm.dev.R2.distr > dev.R2) # P-valor del modelo de
  distance-decay en los datos observados
}

```



```

    blocks.ydis[,i] <- ydis.matrix[i,][-i]    # Almacenajee del bloque
    correspondiente al sitio i excluyendo el valor diagonal
  }

boots.dev.R2.distr <- numeric(resamples)

if (model.type == "exp"){
  for (i in 1:resamples){
    # Remuestreo de bloques con reemplazamiento: se remuestran las matrices
    con los bloques de localidad asegurando que los mismos casos se
    seleccionan en ambas matrices
    b.i.resample <- sample(x = 1:nsites, size = nsites, replace = T)    #
      Vector de índices para remuestrear con reemplazamiento

    b.xdis.resample <- blocks.xdis[, b.i.resample]    # Remuestreo con
      reemplazamiento de bloques de valores de distancia espacial
    b.ydis.resample <- blocks.ydis[, b.i.resample]    # Remuestreo con
      reemplazamiento de bloques de valores de similitud entre
      comunidades

    # La hipótesis nula se genera por medio de la aleatorización dentro de
    los bloques de los valores de similitud para eliminar la relación entre
    xdis e ydis
    r.b.ydis.resample <- apply(b.ydis.resample,2,sample)

    # Reducción del tamaño de los bloques de localidad remuestreados: para
    mantener el tamaño muestral de las matrices observadas de
    distancia/similitud, tan sólo se mantiene el mismo número de casos que
    había en las matrices originales, eligiéndolos de forma aleatoria
    boot.index <- sample(1:length(r.b.ydis.resample), size = nsites*(nsites
      - 1)/2) # Vector de índices para la reducción de las remuestras de
      bloques de localidad
    boot.rb.ydis <- r.b.ydis.resample[boot.index]    # Reducción de
      la remuestra de similitud entre comunidades
    boot.b.xdis <- b.xdis.resample[boot.index]    # Reducción de
      la remuestra de Spatial distance

    # Modelos GLM models de las remuestras de bloques de localidad
    B.mod <- glm2(boot.rb.ydis ~ boot.b.xdis, family = gaussian(link =
      "log"), start = st.val) # Modelo de distance-decay de las
      remuestras de bloques de localidad
    B.mod.null <- glm2(boot.rb.ydis ~ 1, start = st.val[1])    #
      Modelo nulo de distance-decay

    dev.boots <- deviance(B.mod)    # Deviance del modelo de distance-
      decay de las remuestras de bloques de localidad
    dev.boots.null <- deviance(B.mod.null)    # Deviance del modelo nulo
    boots.dev.R2.distr[i] <- (dev.boots.null - dev.boots)/dev.boots.null #
      valores pseudo-R2
  }
}

```

```

mod <- glm2(ydis ~ xdis, family = gaussian(link = "log"), start = st.val)
# Modelo de distance-decay en datos observados
mod.null <- glm2(ydis ~ 1, start = st.val[1]) # Modelo nulo en datos
observados

dev <- deviance(mod) # Deviance del modelo de distance-
decay en datos observados
nulldev <- deviance(mod.null) # Deviance del modelo nulo en datos
observados
dev.R2 <- (nulldev - dev)/nulldev # Valor pseudo-R2 del modelo de
distance-decay en datos observados
p.val.boots.R2 <- mean(boots.dev.R2.distr > dev.R2) # P-valor del modelo
de distance-decay en los datos observados

return(list(Pseudo.R2 = dev.R2, p.value = p.val.boots.R2))

}

else if (model.type == "pow"){
for (i in 1:resamples){
# Remuestreo de bloques con reemplazamiento: se remuestran las matrices
con los bloques de localidad asegurando que los mismos casos se
seleccionan en ambas matrices
b.i.resample <- sample(x = 1:nsites, size = nsites, replace = T) #
Vector de índices para remuestrear con reemplazamiento

b.xdis.resample <- blocks.xdis[, b.i.resample] #
Remuestreo con reemplazamiento de bloques de valores de Spatial
distance
b.ydis.resample <- blocks.ydis[, b.i.resample] #
Remuestreo con reemplazamiento de bloques de valores de similitud
entre comunidades

# La hipótesis nula se genera por medio de la aleatorización dentro de
los bloques de los valores de similitud para eliminar la relación entre
xdis e ydis
r.b.ydis.resample <- apply(b.ydis.resample, 2, sample)

# Reducción del tamaño de los bloques de localidad remuestreados: para
mantener el tamaño muestral de las matrices observadas de
distancia/similitud, tan sólo se mantiene el mismo número de casos que
había en las matrices originales, eligiéndolos de forma aleatoria
boot.index <- sample(1:length(r.b.ydis.resample), size = nsites*(nsites -
1)/2) # Vector de índices para la reducción de las remuestras de
bloques de localidad
boot.rb.ydis <- r.b.ydis.resample[boot.index] # Reducción de la
remuestra de similitud entre comunidades
boot.b.xdis <- b.xdis.resample[boot.index] # Reducción de la
remuestra de Spatial distance

```

```

# Modelos GLM models de las remuestras de bloques de localidad
B.mod <- glm2(boot.rb.ydis ~ log(boot.b.xdis), family = gaussian(link =
  "log"), start = st.val) # Modelo de distance-decay de las remuestras
  de bloques de localidad
B.mod.null <- glm2(boot.rb.ydis ~ 1, start = st.val[1]) # Modelo nulo
  de distance-decay

dev.boots <- deviance(B.mod) # Deviance del modelo de distance-
  decay de las remuestras de bloques de localidad
dev.boots.null <- deviance(B.mod.null) # Deviance del modelo nulo
boots.dev.R2.distr[i] <- (dev.boots.null - dev.boots)/dev.boots.null #
  valores pseudo-R2
}

mod <- glm2(ydis ~ log(xdis), family = gaussian(link = "log"), start =
  st.val) # Modelo de distance-decay en datos observados
mod.null <- glm2(ydis ~ 1, start = st.val[1]) # Modelo nulo en datos
  observados
dev <- deviance(mod) # Deviance del modelo de distance-decay
  en datos observados
nulldev <- deviance(mod.null) # Deviance del modelo nulo en datos
  observados
dev.R2 <- (nulldev - dev)/nulldev # Valor pseudo-R2 del modelo de
  distance-decay en datos observados
p.val.boots.R2 <- mean(boots.dev.R2.distr > dev.R2) # P-valor del modelo de
  distance-decay en los datos observados

return(list(Pseudo.R2 = dev.R2, p.value = p.val.boots.R2))

}
}

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~
# 4. ESTIMACIÓN DEL ERROR DE TIPO I
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~

# Este ejemplo se da para modelos exponenciales con nsites = 50 y nsp
(número de especies) = 150

## PARÁMETROS ##
nsim = 1000 # número de simulaciones para la estimación del error tipo I
nsites = 50 # número de localidades de muestreo simuladas
nsp = 150 # número de especies simuladas
nresamples = 2500 # número de remuestras/permutaciones

# Vectores vacíos para almacenar los p-valores de los diferentes contrastes
permtest.pvalues.tI <- c()
bootstest.pvalues.tI <- c()

```

```

mantel.pvalues.tI <-c()
ftest.pvalues.tI <-c()

# Simulaciones de datos bajo la hipótesis nula y cálculo del p-valor con
distintos contrastes.

for (s in 1:nsim){

  ## SIMULACIÓN DE DATOS BAJO LA HIPÓTESIS NULA ##
  # Coordenadas espaciales aleatorias para nsites en una rejilla de 9000 x
  9000
  coords.H0 <- cbind(runif(nsites, min = 0, max = 9000), # Coordenada X
                    runif(nsites, min = 0, max = 9000)) # Coordenada Y

  # Distancias espaciales euclídeas entre localidades
  H0.Dx <- dist(coords.H0)

  # Simulación de datos de presencia/ausencia para n especies (nsp) en n
  localidades (nsites). La probabilidad de ocupación de las especies sigue
  una distribución lognormal (algunas especies comunes y la mayoría de
  ellas raras)
  H0.sp <- matrix(ncol = nsp, nrow = nsites) # Matriz vacía de
  presencia/ausencia de especies
  lnorm <- rlnorm(nsp) # Generación aleatoria de datos de ocupación para n
  especies (nsp) siguiendo una distribución lognormal
  threshold <- plnorm(lnorm, mean(lnorm)) # Transformación de la
  probabilidad lognormal de ocupación en valores umbral para convertir
  los datos a presencia(1)/ausencia(0)

  for (u in 1:nsp){
    H0.sp[,u] <- rbinom(n = nsites, size = 1, prob = threshold[u]) #
    Conversión en datos de presencia/ausencia usando umbrales lognormales
  }

  # Similitud en la composición específica entre localidades (1 -
  disimilitud de Simpson)
  H0.Sy <- 1 - beta.pair(H0.sp)$beta.sim

  # CONTRASTE DE PERMUTACIONES POR BLOQUES DE LOCALIDAD
  permtest.pvalues.tI[s] <- R2.permtest(H0.Dx, H0.Sy, nresamples,
  model.type = "exp")$p.value
  # CONTRASTE DE REMUESTREO POR BLOQUES DE LOCALIDAD
  bootstest.pvalues.tI[s] <- R2.bootstest(H0.Dx, H0.Sy, nresamples,
  model.type = "exp")$p.value
  # TEST DE MANTEL
  mantel.pvalues.tI[s] <- 1 - mantel(H0.Dx, H0.Sy, permutations =
  nresamples)$signif
  # TEST DE LA F
  mod.glm <- glm2(H0.Sy ~ H0.Dx, family = gaussian(link = "log"), start =
  c(0.005,-0.00001)) # modelo glm de distance-decay

```



```

GRF.phi <- grf(n = nsites, cov.pars = c(sigma.sq = 1, phi), nsim = nsp,
  # Generación de variables normales autocorrelacionadas y
  # coordenadas espaciales con campos aleatorios Gaussianos (GRFs)
  xlims = c(0, 9000),
  ylims = c(0, 9000)
)

phi.Dx <- dist(GRF.phi$coords) # Distancia espacial euclídea entre
  localidades

M_phi = GRF.phi$data          # Variables normales autocorrelacionadas
  (nsp = 150)
M_norm_phi <- pnorm(M_phi)    # Transformación de valores normales
  autocorrelacionados en densidades acumuladas (0-1) para aplicar un
  umbral de corte para su conversión a presencias/ausencias

# Generación de umbrales de corte para convertir valores
# autocorrelacionados en datos de presencia/ausencia siguiendo una
# distribución lognormal de datos de ocupación
lnorm <- rlnorm(nsp)         # Generación aleatoria de datos
  de ocupación para n especies (nsp) siguiendo una distribución
  lognormal
threshold <- plnorm(lnorm, mean(lnorm)) # Transformación de la
  probabilidad lognormal de ocupación en valores de umbral de corte

# Aplicación de los umbrales de corte a variables autocorrelacionadas
# para transformarlas en datos de presencia(1)/ausencia(0)
M_norm_phi <- (t(t(M_norm_phi) < threshold))*1

# Similitud en la composición específica entre localidades (1-
# disimilitud de Simpson)
phi.Sy <- (1 - beta.pair(M_norm_phi)$beta.sim) # semimatriz de similitud

# CONTRASTE DE PERMUTACIONES POR BLOQUES DE LOCALIDAD
permtest.pvalues.tII[s] <- R2.permtest(phi.Dx, phi.Sy, nresamples,
  model.type = "exp")$p.value
# CONTRASTE DE REMUESTREO POR BLOQUES DE LOCALIDAD
bootstest.pvalues.tII[s] <- R2.bootstest(phi.Dx, phi.Sy, nresamples,
  model.type = "exp")$p.value
# TEST DE MANTEL
mantel.pvalues.tII[s] <- 1 - mantel(phi.Dx, phi.Sy, permutations =
  nresamples)$signif
# TEST DE LA F
mod.glm <- glm2(phi.Sy ~ phi.Dx, family = gaussian(link = "log"), start
  = c(0.005,-0.00001)) # modelo glm de distance-decay
mod.null <- glm2(phi.Sy ~ 1, family = gaussian(link = "log"), start =
  c(0.005)) # modelo glm nulo

ftest.pvalues.tII[s] <- anova(mod.null, mod.glm, test = "F")[6][2,1]

```

```
message(paste("simulation#", s, sep=""))

}

# INFORME SOBRE LA POTENCIA ESTADÍSTICA DE LOS DISTINTOS MÉTODOS
paste("Potencia estadística (contraste de permutaciones por bloques de
localidad) =", mean(permtest.pvalues.tII < 0.05))# contraste de
permutaciones por bloques de localidad
paste("Potencia estadística (contraste de remuestreo por bloques de
localidad) =", mean(bootstest.pvalues.tII < 0.05)) # contraste de
remuestreo por bloques de localidad
paste("Potencia estadística (test de Mantel) =", mean(mantel.pvalues.tII <
0.05)) # test de Mantel
paste("Potencia estadística (test de la F) =", mean(ftest.pvalues.tII <
0.05)) # test de la F
```

9.5 APLICACIÓN DE LOS MÉTODOS *DISTANCE-DECAY* DESARROLLADOS EN ESTAS TESIS A DATOS REALES

Los métodos desarrollados en esta tesis han sido implementados en la función `decay.model()` de la librería "betapart".
 # En este apéndice se muestra cómo aplicar dicha función, así como otras relacionadas de esta librería, para el ajuste de modelos distance-decay no lineales y su representación gráfico.
 # Se usan los datos de distribución de coleópteros en territorios europeos

```
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# 1. Librerías
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
```

```
library(vegan)
library(betapart)
```

```
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# 2. Datos
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
```

Lectura de tablas de datos de presencia/ausencia de coleópteros

```
amara <- read.table("./datos/amara.txt", header = T)
apion <- read.table("./datos/apionidae.txt", header = T)
carab <- read.table("./datos/carabus.txt", header = T)
ceram <- read.table("./datos/cerambycinae.txt", header = T)
chaet <- read.table("./datos/chaetocnema.txt", header = T)
chrys <- read.table("./datos/chrys.txt", header = T)
crypt <- read.table("./datos/crypt.txt", header = T)
hydra <- read.table("./datos/hydraenini.txt", header = T)
lamin <- read.table("./datos/lamiinae.txt", header = T)
leptu <- read.table("./datos/lepturinae.txt", header = T)
limne <- read.table("./datos/limnebiini.txt", header = T)
ochte <- read.table("./datos/ochtebiini.txt", header = T)
otior <- read.table("./datos/otiorhynchus.txt", header = T)
phyll <- read.table("./datos/phylllobini.txt", header = T)
polyd <- read.table("./datos/polydrosini.txt", header = T)
psela <- read.table("./datos/pselaphinae.txt", header = T)
ptero <- read.table("./datos/pterostichus.txt", header = T)
scara <- read.table("./datos/scarabaeidae.txt", header = T)
scoly <- read.table("./datos/scolytinae.txt", header = T)
silph <- read.table("./datos/silphidae.txt", header = T)
trech <- read.table("./datos/trechus.txt", header = T)
```

Lectura de tablas de datos con información sobre territorios europeos

```
vars <- read.table("./datos/col-eur-vars.txt", header = T, row.names = 1)
# Generación de tablas independientes para países del norte (1) y del sur
# (0) de Europa
```

```

amara.n <- subset(amara, vars$nortesur == 1)
apion.n <- subset(apion, vars$nortesur == 1)
carab.n <- subset(carab, vars$nortesur == 1)
ceram.n <- subset(ceram, vars$nortesur == 1)
chaet.n <- subset(chaet, vars$nortesur == 1)
chrys.n <- subset(chrys, vars$nortesur == 1)
crypt.n <- subset(crypt, vars$nortesur == 1)
hydra.n <- subset(hydra, vars$nortesur == 1)
lamin.n <- subset(lamin, vars$nortesur == 1)
leptu.n <- subset(leptu, vars$nortesur == 1)
limne.n <- subset(limne, vars$nortesur == 1)
ochte.n <- subset(ochte, vars$nortesur == 1)
otior.n <- subset(otior, vars$nortesur == 1)
phyll.n <- subset(phyll, vars$nortesur == 1)
polyd.n <- subset(polyd, vars$nortesur == 1)
psela.n <- subset(psela, vars$nortesur == 1)
ptero.n <- subset(ptero, vars$nortesur == 1)
scara.n <- subset(scara, vars$nortesur == 1)
scoly.n <- subset(scoly, vars$nortesur == 1)
silph.n <- subset(silph, vars$nortesur == 1)
trech.n <- subset(trech, vars$nortesur == 1)

amara.s <- subset(amara, vars$nortesur == 0)
apion.s <- subset(apion, vars$nortesur == 0)
carab.s <- subset(carab, vars$nortesur == 0)
ceram.s <- subset(ceram, vars$nortesur == 0)
chaet.s <- subset(chaet, vars$nortesur == 0)
chrys.s <- subset(chrys, vars$nortesur == 0)
crypt.s <- subset(crypt, vars$nortesur == 0)
hydra.s <- subset(hydra, vars$nortesur == 0)
lamin.s <- subset(lamin, vars$nortesur == 0)
leptu.s <- subset(leptu, vars$nortesur == 0)
limne.s <- subset(limne, vars$nortesur == 0)
ochte.s <- subset(ochte, vars$nortesur == 0)
otior.s <- subset(otior, vars$nortesur == 0)
phyll.s <- subset(phyll, vars$nortesur == 0)
polyd.s <- subset(polyd, vars$nortesur == 0)
psela.s <- subset(psela, vars$nortesur == 0)
ptero.s <- subset(ptero, vars$nortesur == 0)
scara.s <- subset(scara, vars$nortesur == 0)
scoly.s <- subset(scoly, vars$nortesur == 0)
silph.s <- subset(silph, vars$nortesur == 0)
trech.s <- subset(trech, vars$nortesur == 0)

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# 3. Matrices de similitud biótica y distancia espacial
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

# Cálculo de matrices de distancia espacial (distancia euclídea) entre los
centroides de los territorios europeos

distgeo <- dist(data.frame(vars$x.utm, vars$y.utm), method = "euclidean")
# Entre todos los países

```

```

distgeo.n <-dist(subset(data.frame(vars$x.utm, vars$y.utm), vars$nortesur
== 1), method = "euclidean") # Países del norte

distgeo.s <-dist(subset(data.frame(vars$x.utm,vars$y.utm),vars$nortesur ==
0), method = "euclidean") # Países del sur

# Cálculo de matrices de similitud biótica entre territorios europeos
(índice de similitud de Simpson)

amara.n.sim <- 1 - beta.pair(amara.n)$beta.sim
apion.n.sim <- 1 - beta.pair(apion.n)$beta.sim
carab.n.sim <- 1 - beta.pair(carab.n)$beta.sim
ceram.n.sim <- 1 - beta.pair(ceram.n)$beta.sim
chaet.n.sim <- 1 - beta.pair(chaet.n)$beta.sim
chrys.n.sim <- 1 - beta.pair(chrys.n)$beta.sim
crypt.n.sim <- 1 - beta.pair(crypt.n)$beta.sim
hydra.n.sim <- 1 - beta.pair(hydra.n)$beta.sim
lamin.n.sim <- 1 - beta.pair(lamin.n)$beta.sim
leptu.n.sim <- 1 - beta.pair(leptu.n)$beta.sim
limne.n.sim <- 1 - beta.pair(limne.n)$beta.sim
ochte.n.sim <- 1 - beta.pair(ochte.n)$beta.sim
otior.n.sim <- 1 - beta.pair(otior.n)$beta.sim
phyll.n.sim <- 1 - beta.pair(phyll.n)$beta.sim
polyd.n.sim <- 1 - beta.pair(polyd.n)$beta.sim
psela.n.sim <- 1 - beta.pair(psela.n)$beta.sim
ptero.n.sim <- 1 - beta.pair(ptero.n)$beta.sim
scara.n.sim <- 1 - beta.pair(scara.n)$beta.sim
scoly.n.sim <- 1 - beta.pair(scoly.n)$beta.sim
silph.n.sim <- 1 - beta.pair(silph.n)$beta.sim
trech.n.sim <- 1 - beta.pair(trech.n)$beta.sim

amara.s.sim <- 1 - beta.pair(amara.s)$beta.sim
apion.s.sim <- 1 - beta.pair(apion.s)$beta.sim
carab.s.sim <- 1 - beta.pair(carab.s)$beta.sim
ceram.s.sim <- 1 - beta.pair(ceram.s)$beta.sim
chaet.s.sim <- 1 - beta.pair(chaet.s)$beta.sim
chrys.s.sim <- 1 - beta.pair(chrys.s)$beta.sim
crypt.s.sim <- 1 - beta.pair(crypt.s)$beta.sim
hydra.s.sim <- 1 - beta.pair(hydra.s)$beta.sim
lamin.s.sim <- 1 - beta.pair(lamin.s)$beta.sim
leptu.s.sim <- 1 - beta.pair(leptu.s)$beta.sim
limne.s.sim <- 1 - beta.pair(limne.s)$beta.sim
ochte.s.sim <- 1 - beta.pair(ochte.s)$beta.sim
otior.s.sim <- 1 - beta.pair(otior.s)$beta.sim
phyll.s.sim <- 1 - beta.pair(phyll.s)$beta.sim
polyd.s.sim <- 1 - beta.pair(polyd.s)$beta.sim
psela.s.sim <- 1 - beta.pair(psela.s)$beta.sim
ptero.s.sim <- 1 - beta.pair(ptero.s)$beta.sim
scara.s.sim <- 1 - beta.pair(scara.s)$beta.sim
scoly.s.sim <- 1 - beta.pair(scoly.s)$beta.sim
silph.s.sim <- 1 - beta.pair(silph.s)$beta.sim
trech.s.sim <- 1 - beta.pair(trech.s)$beta.sim

```

```

# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .
# 4. Ajuste y representación gráfica de los modelos de distance-decay:
función exponencial negativa, power-law y gompertz
# ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ . ~ .

# y.type = "sim"      # La matriz biótica es una matriz de similitud
# model.type = "exp"  # Modelo con ajuste exponencial
# model.type = "pow"  # Modelo con ajuste power-law
# model.type = "gom"  # Modelo ajustado con función de Gompertz

#-----
# NORTE DE EUROPA

# - # Amara
amara.n.decay.exp <- decay.model(amara.n.sim, y.type = "sim", distgeo.n,
                                model.type = "exp", 100)
amara.n.decay.pow <- decay.model(amara.n.sim, y.type = "sim", distgeo.n,
                                model.type = "pow", 100)
amara.n.decay.gom <- decay.model(amara.n.sim, y.type = "sim", distgeo.n,
                                model.type = "gom", 100)

pdf(file = "amara.n.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(amara.n.decay.exp)
plot.decay(amara.n.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(amara.n.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
                      .(round(amara.n.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
                      .(round(amara.n.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
                      .(round(amara.n.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.n, amara.n.decay.exp$model$resid(), ylab = "Residuos (modelo
exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.n, amara.n.decay.pow$model$resid(), ylab = "Residuos (modelo
power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.n, amara.n.decay.gom$model$resid(), ylab = "Residuos (model
Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# - # Apionidae
apion.n.decay.exp <- decay.model(apion.n.sim, y.type = "sim", distgeo.n,
                                model.type = "exp", 100)
apion.n.decay.pow <- decay.model(apion.n.sim, y.type = "sim", distgeo.n,
                                model.type = "pow", 100)
apion.n.decay.gom <- decay.model(apion.n.sim, y.type = "sim", distgeo.n,
                                model.type = "gom", 100)

```

```

pdf(file = "apion.n.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(apion.n.decay.exp)
plot.decay(apion.n.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(apion.n.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(apion.n.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(apion.n.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(apion.n.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.n, apion.n.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.n, apion.n.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.n, apion.n.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# - # Carabus
carab.n.decay.exp <- decay.model(carab.n.sim, y.type = "sim", distgeo.n,
  model.type = "exp", 100)
carab.n.decay.pow <- decay.model(carab.n.sim, y.type = "sim", distgeo.n,
  model.type = "pow", 100)
carab.n.decay.gom <- decay.model(carab.n.sim, y.type = "sim", distgeo.n,
  model.type = "gom", 100)

pdf(file = "carab.n.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(carab.n.decay.exp)
plot.decay(carab.n.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(carab.n.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(carab.n.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(carab.n.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(carab.n.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.n, carab.n.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.n, carab.n.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.n, carab.n.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

```

```

# - # Cerambycinae
ceram.n.decay.exp <- decay.model(ceram.n.sim, y.type = "sim", distgeo.n,
  model.type = "exp", 100)
ceram.n.decay.pow <- decay.model(ceram.n.sim, y.type = "sim", distgeo.n,
  model.type = "pow", 100)
ceram.n.decay.gom <- decay.model(ceram.n.sim, y.type = "sim", distgeo.n,
  model.type = "gom", 100)

pdf(file = "ceram.n.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(ceram.n.decay.exp)
plot.decay(ceram.n.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(ceram.n.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(ceram.n.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(ceram.n.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(ceram.n.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.n, ceram.n.decay.exp$model$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.n, ceram.n.decay.pow$model$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.n, ceram.n.decay.gom$model$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# - # Chaetocnema
chaet.n.decay.exp <- decay.model(chaet.n.sim, y.type = "sim", distgeo.n,
  model.type = "exp", 100)
chaet.n.decay.pow <- decay.model(chaet.n.sim, y.type = "sim", distgeo.n,
  model.type = "pow", 100)
chaet.n.decay.gom <- decay.model(chaet.n.sim, y.type = "sim", distgeo.n,
  model.type = "gom", 100)

pdf(file = "chaet.n.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(chaet.n.decay.exp)
plot.decay(chaet.n.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(chaet.n.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(chaet.n.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(chaet.n.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(chaet.n.decay.gom$pseudo.r.squared, digits = 2))))

```

```

plot(distgeo.n, chaet.n.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.n, chaet.n.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.n, chaet.n.decay.gom$model$m$resid(), ylab = "Residuos (modelo
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# - # Chrysomelinae
chrys.n.decay.exp <- decay.model(chrys.n.sim, y.type = "sim", distgeo.n,
  model.type = "exp", 100)
chrys.n.decay.pow <- decay.model(chrys.n.sim, y.type = "sim", distgeo.n,
  model.type = "pow", 100)
chrys.n.decay.gom <- decay.model(chrys.n.sim, y.type = "sim", distgeo.n,
  model.type = "gom", 100)

pdf(file = "chrys.n.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(chrys.n.decay.exp)
plot.decay(chrys.n.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(chrys.n.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(chrys.n.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(chrys.n.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(chrys.n.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.n, chrys.n.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.n, chrys.n.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.n, chrys.n.decay.gom$model$m$resid(), ylab = "Residuos (modelo
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# - # Cryptocephalinae
crypt.n.decay.exp <- decay.model(crypt.n.sim, y.type = "sim", distgeo.n,
  model.type = "exp", 100)
crypt.n.decay.pow <- decay.model(crypt.n.sim, y.type = "sim", distgeo.n,
  model.type = "pow", 100)
crypt.n.decay.gom <- decay.model(crypt.n.sim, y.type = "sim", distgeo.n,
  model.type = "gom", 100)

pdf(file = "crypt.n.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(crypt.n.decay.exp)
plot.decay(crypt.n.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(crypt.n.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)

```

```

text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(crypt.n.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(crypt.n.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(crypt.n.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.n, crypt.n.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.n, crypt.n.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.n, crypt.n.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# - # Hydraenini
hydra.n.decay.exp <- decay.model(hydra.n.sim, y.type = "sim", distgeo.n,
  model.type = "exp", 100)
hydra.n.decay.pow <- decay.model(hydra.n.sim, y.type = "sim", distgeo.n,
  model.type = "pow", 100)
hydra.n.decay.gom <- decay.model(hydra.n.sim, y.type = "sim", distgeo.n,
  model.type = "gom", 100)

pdf(file = "hydra.n.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(hydra.n.decay.exp)
plot.decay(hydra.n.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(hydra.n.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(hydra.n.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(hydra.n.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(hydra.n.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.n, hydra.n.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.n, hydra.n.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.n, hydra.n.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# - # Lamiinae
lamin.n.decay.exp <- decay.model(lamin.n.sim, y.type = "sim", distgeo.n,
  model.type = "exp", 100)
lamin.n.decay.pow <- decay.model(lamin.n.sim, y.type = "sim", distgeo.n,
  model.type = "pow", 100)

```

```

lamin.n.decay.gom <- decay.model(lamin.n.sim, y.type = "sim", distgeo.n,
                                model.type = "gom", 100)

pdf(file = "lamin.n.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(lamin.n.decay.exp)
plot.decay(lamin.n.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(lamin.n.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
                      .(round(lamin.n.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
                      .(round(lamin.n.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
                      .(round(lamin.n.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.n, lamin.n.decay.exp$model$m$resid(), ylab = "Residuos (modelo
exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.n, lamin.n.decay.pow$model$m$resid(), ylab = "Residuos (modelo
power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.n, lamin.n.decay.gom$model$m$resid(), ylab = "Residuos (model
Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# - # Lepturinae
leptu.n.decay.exp <- decay.model(leptu.n.sim, y.type = "sim", distgeo.n,
                                model.type = "exp", 100)
leptu.n.decay.pow <- decay.model(leptu.n.sim, y.type = "sim", distgeo.n,
                                model.type = "pow", 100)
leptu.n.decay.gom <- decay.model(leptu.n.sim, y.type = "sim", distgeo.n,
                                model.type = "gom", 100)

pdf(file = "leptu.n.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(leptu.n.decay.exp)
plot.decay(leptu.n.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(leptu.n.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
                      .(round(leptu.n.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
                      .(round(leptu.n.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
                      .(round(leptu.n.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.n, leptu.n.decay.exp$model$m$resid(), ylab = "Residuos (modelo
exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.n, leptu.n.decay.pow$model$m$resid(), ylab = "Residuos (modelo
power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.n, leptu.n.decay.gom$model$m$resid(), ylab = "Residuos (model
Gompertz)", xlab = "Distancia espacial (km)", col = "blue")

```

```

dev.off()

# - # Limnebiini
limne.n.decay.exp <- decay.model(limne.n.sim, y.type = "sim", distgeo.n,
  model.type = "exp", 100)
limne.n.decay.pow <- decay.model(limne.n.sim, y.type = "sim", distgeo.n,
  model.type = "pow", 100)
limne.n.decay.gom <- decay.model(limne.n.sim, y.type = "sim", distgeo.n,
  model.type = "gom", 100)

pdf(file = "limne.n.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(limne.n.decay.exp)
plot.decay(limne.n.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(limne.n.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(limne.n.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(limne.n.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(limne.n.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.n, limne.n.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.n, limne.n.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.n, limne.n.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# - # Ochtebiini
ochte.n.decay.exp <- decay.model(ochte.n.sim, y.type = "sim", distgeo.n,
  model.type = "exp", 100)
ochte.n.decay.pow <- decay.model(ochte.n.sim, y.type = "sim", distgeo.n,
  model.type = "pow", 100)
ochte.n.decay.gom <- decay.model(ochte.n.sim, y.type = "sim", distgeo.n,
  model.type = "gom", 100)

pdf(file = "ochte.n.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(ochte.n.decay.exp)
plot.decay(ochte.n.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(ochte.n.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(ochte.n.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(ochte.n.decay.pow$pseudo.r.squared, digits = 2))))

```

```

text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(ochte.n.decay.gom$pseudo.r.squared, digits = 2)))

plot(distgeo.n, ochte.n.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.n, ochte.n.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.n, ochte.n.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# - # Otiorhynchus
otior.n.decay.exp <- decay.model(otior.n.sim, y.type = "sim", distgeo.n,
  model.type = "exp", 100)
otior.n.decay.pow <- decay.model(otior.n.sim, y.type = "sim", distgeo.n,
  model.type = "pow", 100)
otior.n.decay.gom <- decay.model(otior.n.sim, y.type = "sim", distgeo.n,
  model.type = "gom", 100)

pdf(file = "otior.n.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(otior.n.decay.exp)
plot.decay(otior.n.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(otior.n.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(otior.n.decay.exp$pseudo.r.squared, digits = 2)))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(otior.n.decay.pow$pseudo.r.squared, digits = 2)))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(otior.n.decay.gom$pseudo.r.squared, digits = 2)))

plot(distgeo.n, otior.n.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.n, otior.n.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.n, otior.n.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# - # Phyllobiini
phyll.n.decay.exp <- decay.model(phyll.n.sim, y.type = "sim", distgeo.n,
  model.type = "exp", 100)
phyll.n.decay.pow <- decay.model(phyll.n.sim, y.type = "sim", distgeo.n,
  model.type = "pow", 100)
phyll.n.decay.gom <- decay.model(phyll.n.sim, y.type = "sim", distgeo.n,
  model.type = "gom", 100)

```

```

pdf(file = "phyll.n.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(phyll.n.decay.exp)
plot.decay(phyll.n.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(phyll.n.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(phyll.n.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(phyll.n.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(phyll.n.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.n, phyll.n.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.n, phyll.n.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.n, phyll.n.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# - # Polydrusini
polyd.n.decay.exp <- decay.model(polyd.n.sim, y.type = "sim", distgeo.n,
  model.type = "exp", 100)
polyd.n.decay.pow <- decay.model(polyd.n.sim, y.type = "sim", distgeo.n,
  model.type = "pow", 100)
polyd.n.decay.gom <- decay.model(polyd.n.sim, y.type = "sim", distgeo.n,
  model.type = "gom", 100)

pdf(file = "polyd.n.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(polyd.n.decay.exp)
plot.decay(polyd.n.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(polyd.n.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(polyd.n.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(polyd.n.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(polyd.n.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.n, polyd.n.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.n, polyd.n.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.n, polyd.n.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

```

```

# - # Pselaphinae
psela.n.decay.exp <- decay.model(psela.n.sim, y.type = "sim", distgeo.n,
  model.type = "exp", 100)
psela.n.decay.pow <- decay.model(psela.n.sim, y.type = "sim", distgeo.n,
  model.type = "pow", 100)
psela.n.decay.gom <- decay.model(psela.n.sim, y.type = "sim", distgeo.n,
  model.type = "gom", 100)

pdf(file = "psela.n.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(psela.n.decay.exp)
plot.decay(psela.n.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(psela.n.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(psela.n.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(psela.n.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(psela.n.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.n, psela.n.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.n, psela.n.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.n, psela.n.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# - # Pterostichus
ptero.n.decay.exp <- decay.model(ptero.n.sim, y.type = "sim", distgeo.n,
  model.type = "exp", 100)
ptero.n.decay.pow <- decay.model(ptero.n.sim, y.type = "sim", distgeo.n,
  model.type = "pow", 100)
ptero.n.decay.gom <- decay.model(ptero.n.sim, y.type = "sim", distgeo.n,
  model.type = "gom", 100)

pdf(file = "ptero.n.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(ptero.n.decay.exp)
plot.decay(ptero.n.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(ptero.n.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(ptero.n.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(ptero.n.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(ptero.n.decay.gom$pseudo.r.squared, digits = 2))))

```

```

plot(distgeo.n, ptero.n.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.n, ptero.n.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.n, ptero.n.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

```

```
# - # Scarabaeidae
```

```
scara.n.decay.exp <- decay.model(scara.n.sim, y.type = "sim", distgeo.n,
  model.type = "exp", 100)
```

```
scara.n.decay.pow <- decay.model(scara.n.sim, y.type = "sim", distgeo.n,
  model.type = "pow", 100)
```

```
scara.n.decay.gom <- decay.model(scara.n.sim, y.type = "sim", distgeo.n,
  model.type = "gom", 100)
```

```
pdf(file = "scara.n.pdf", height = 3, width = 10)
```

```
par(mfrow = c(1, 4))
```

```
plot.decay(scara.n.decay.exp)
```

```
plot.decay(scara.n.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
```

```
plot.decay(scara.n.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
```

```
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(scara.n.decay.exp$pseudo.r.squared, digits = 2))))
```

```
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(scara.n.decay.pow$pseudo.r.squared, digits = 2))))
```

```
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(scara.n.decay.gom$pseudo.r.squared, digits = 2))))
```

```
plot(distgeo.n, scara.n.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
```

```
plot(distgeo.n, scara.n.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
```

```
plot(distgeo.n, scara.n.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
```

```
dev.off()
```

```
# - # Scolytinae
```

```
scoly.n.decay.exp <- decay.model(scoly.n.sim, y.type = "sim", distgeo.n,
  model.type = "exp", 100)
```

```
scoly.n.decay.pow <- decay.model(scoly.n.sim, y.type = "sim", distgeo.n,
  model.type = "pow", 100)
```

```
scoly.n.decay.gom <- decay.model(scoly.n.sim, y.type = "sim", distgeo.n,
  model.type = "gom", 100)
```

```
pdf(file = "scoly.n.pdf", height = 3, width = 10)
```

```
par(mfrow = c(1, 4))
```

```
plot.decay(scoly.n.decay.exp)
```

```
plot.decay(scoly.n.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
```

```

plot.decay(scoly.n.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(scoly.n.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(scoly.n.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(scoly.n.decay.gom$pseudo.r.squared, digits = 2))))

```

```

plot(distgeo.n, scoly.n.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.n, scoly.n.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.n, scoly.n.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

```

```
# - # Silphidae
```

```

silph.n.decay.exp <- decay.model(silph.n.sim, y.type = "sim", distgeo.n,
  model.type = "exp", 100)
silph.n.decay.pow <- decay.model(silph.n.sim, y.type = "sim", distgeo.n,
  model.type = "pow", 100)
silph.n.decay.gom <- decay.model(silph.n.sim, y.type = "sim", distgeo.n,
  model.type = "gom", 100)

```

```
pdf(file = "silph.n.pdf", height = 3, width = 10)
```

```
par(mfrow = c(1, 4))
```

```
plot.decay(silph.n.decay.exp)
```

```
plot.decay(silph.n.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
```

```
plot.decay(silph.n.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
```

```
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(silph.n.decay.exp$pseudo.r.squared, digits = 2))))
```

```
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(silph.n.decay.pow$pseudo.r.squared, digits = 2))))
```

```
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(silph.n.decay.gom$pseudo.r.squared, digits = 2))))
```

```
plot(distgeo.n, silph.n.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
```

```
plot(distgeo.n, silph.n.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
```

```
plot(distgeo.n, silph.n.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
```

```
dev.off()
```

```
# - # Trechus
```

```

trech.n.decay.exp <- decay.model(trech.n.sim, y.type = "sim", distgeo.n,
  model.type = "exp", 100)

```



```

trech.n.decay.pow <- decay.model(trech.n.sim, y.type = "sim", distgeo.n,
  model.type = "pow", 100)
trech.n.decay.gom <- decay.model(trech.n.sim, y.type = "sim", distgeo.n,
  model.type = "gom", 100)

pdf(file = "trech.n.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(trech.n.decay.exp)
plot.decay(trech.n.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(trech.n.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(trech.n.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(trech.n.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(trech.n.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.n, trech.n.decay.exp$model$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.n, trech.n.decay.pow$model$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.n, trech.n.decay.gom$model$resid(), ylab = "Residuos (modelo
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

#-----
# SUR DE EUROPA

# - # Amara
amara.s.decay.exp <- decay.model(amara.s.sim, y.type = "sim", distgeo.s,
  model.type = "exp", 100)
amara.s.decay.pow <- decay.model(amara.s.sim, y.type = "sim", distgeo.s,
  model.type = "pow", 100)
amara.s.decay.gom <- decay.model(amara.s.sim, y.type = "sim", distgeo.s,
  model.type = "gom", 100)

pdf(file = "amara.s.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(amara.s.decay.exp)
plot.decay(amara.s.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(amara.s.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(amara.s.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(amara.s.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(amara.s.decay.gom$pseudo.r.squared, digits = 2))))

```

```

plot(distgeo.s, amara.s.decay.exp$model$m$resid(), ylab = "Residuos (modelo
exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.s, amara.s.decay.pow$model$m$resid(), ylab = "Residuos (modelo
power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.s, amara.s.decay.gom$model$m$resid(), ylab = "Residuos (model
Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

```

```
# - # Apionidae
```

```

apion.s.decay.exp <- decay.model(apion.s.sim, y.type = "sim", distgeo.s,
model.type = "exp", 100)
apion.s.decay.pow <- decay.model(apion.s.sim, y.type = "sim", distgeo.s,
model.type = "pow", 100)
apion.s.decay.gom <- decay.model(apion.s.sim, y.type = "sim", distgeo.s,
model.type = "gom", 100)

```

```
pdf(file = "apion.s.pdf", height = 3, width = 10)
```

```
par(mfrow = c(1, 4))
```

```
plot.decay(apion.s.decay.exp)
```

```
plot.decay(apion.s.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
```

```
plot.decay(apion.s.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
```

```
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
.round(apion.s.decay.exp$pseudo.r.squared, digits = 2)))
```

```
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
.round(apion.s.decay.pow$pseudo.r.squared, digits = 2)))
```

```
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
.round(apion.s.decay.gom$pseudo.r.squared, digits = 2)))
```

```
plot(distgeo.s, apion.s.decay.exp$model$m$resid(), ylab = "Residuos (modelo
exponencial)", xlab = "Distancia espacial (km)")
```

```
plot(distgeo.s, apion.s.decay.pow$model$m$resid(), ylab = "Residuos (modelo
power-law)", xlab = "Distancia espacial (km)", col = "red")
```

```
plot(distgeo.s, apion.s.decay.gom$model$m$resid(), ylab = "Residuos (model
Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
```

```
dev.off()
```

```
# - # Carabus
```

```
carab.s.decay.exp <- decay.model(carab.s.sim, y.type = "sim", distgeo.s,
model.type = "exp", 100)
```

```
carab.s.decay.pow <- decay.model(carab.s.sim, y.type = "sim", distgeo.s,
model.type = "pow", 100)
```

```
carab.s.decay.gom <- decay.model(carab.s.sim, y.type = "sim", distgeo.s,
model.type = "gom", 100)
```

```
pdf(file = "carab.s.pdf", height = 3, width = 10)
```

```
par(mfrow = c(1, 4))
```

```
plot.decay(carab.s.decay.exp)
```

```

plot.decay(carab.s.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(carab.s.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(carab.s.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(carab.s.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(carab.s.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.s, carab.s.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.s, carab.s.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.s, carab.s.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# - # Cerambycinae
ceram.s.decay.exp <- decay.model(ceram.s.sim, y.type = "sim", distgeo.s,
  model.type = "exp", 100)
ceram.s.decay.pow <- decay.model(ceram.s.sim, y.type = "sim", distgeo.s,
  model.type = "pow", 100)
ceram.s.decay.gom <- decay.model(ceram.s.sim, y.type = "sim", distgeo.s,
  model.type = "gom", 100)

pdf(file = "ceram.s.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(ceram.s.decay.exp)
plot.decay(ceram.s.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(ceram.s.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(ceram.s.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(ceram.s.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(ceram.s.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.s, ceram.s.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.s, ceram.s.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.s, ceram.s.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# - # Chaetocnema
chaet.s.decay.exp <- decay.model(chaet.s.sim, y.type = "sim", distgeo.s,
  model.type = "exp", 100)

```

```

chaet.s.decay.pow <- decay.model(chaet.s.sim, y.type = "sim", distgeo.s,
  model.type = "pow", 100)
chaet.s.decay.gom <- decay.model(chaet.s.sim, y.type = "sim", distgeo.s,
  model.type = "gom", 100)

pdf(file = "chaet.s.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(chaet.s.decay.exp)
plot.decay(chaet.s.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(chaet.s.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(chaet.s.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(chaet.s.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(chaet.s.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.s, chaet.s.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.s, chaet.s.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.s, chaet.s.decay.gom$model$m$resid(), ylab = "Residuos (modelo
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# - # Chrysomelinae
chrys.s.decay.exp <- decay.model(chrys.s.sim, y.type = "sim", distgeo.s,
  model.type = "exp", 100)
chrys.s.decay.pow <- decay.model(chrys.s.sim, y.type = "sim", distgeo.s,
  model.type = "pow", 100)
chrys.s.decay.gom <- decay.model(chrys.s.sim, y.type = "sim", distgeo.s,
  model.type = "gom", 100)

pdf(file = "chrys.s.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(chrys.s.decay.exp)
plot.decay(chrys.s.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(chrys.s.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(chrys.s.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(chrys.s.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(chrys.s.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.s, chrys.s.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.s, chrys.s.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")

```

```

plot(distgeo.s, chrys.s.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# - # Cryptocephalinae
crypt.s.decay.exp <- decay.model(crypt.s.sim, y.type = "sim", distgeo.s,
  model.type = "exp", 100)
crypt.s.decay.pow <- decay.model(crypt.s.sim, y.type = "sim", distgeo.s,
  model.type = "pow", 100)
crypt.s.decay.gom <- decay.model(crypt.s.sim, y.type = "sim", distgeo.s,
  model.type = "gom", 100)

pdf(file = "crypt.s.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(crypt.s.decay.exp)
plot.decay(crypt.s.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(crypt.s.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(crypt.s.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(crypt.s.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(crypt.s.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.s, crypt.s.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.s, crypt.s.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.s, crypt.s.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# - # Hydraenini
hydra.s.decay.exp <- decay.model(hydra.s.sim, y.type = "sim", distgeo.s,
  model.type = "exp", 100)
hydra.s.decay.pow <- decay.model(hydra.s.sim, y.type = "sim", distgeo.s,
  model.type = "pow", 100)
hydra.s.decay.gom <- decay.model(hydra.s.sim, y.type = "sim", distgeo.s,
  model.type = "gom", 100)

pdf(file = "hydra.s.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(hydra.s.decay.exp)
plot.decay(hydra.s.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(hydra.s.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(hydra.s.decay.exp$pseudo.r.squared, digits = 2))))

```

```

text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(hydra.s.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(hydra.s.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.s, hydra.s.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.s, hydra.s.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.s, hydra.s.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# - # Lamiinae
lamin.s.decay.exp <- decay.model(lamin.s.sim, y.type = "sim", distgeo.s,
  model.type = "exp", 100)
lamin.s.decay.pow <- decay.model(lamin.s.sim, y.type = "sim", distgeo.s,
  model.type = "pow", 100)
lamin.s.decay.gom <- decay.model(lamin.s.sim, y.type = "sim", distgeo.s,
  model.type = "gom", 100)

pdf(file = "lamin.s.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(lamin.s.decay.exp)
plot.decay(lamin.s.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(lamin.s.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(lamin.s.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(lamin.s.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(lamin.s.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.s, lamin.s.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.s, lamin.s.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.s, lamin.s.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# - # Lepturinae
leptu.s.decay.exp <- decay.model(leptu.s.sim, y.type = "sim", distgeo.s,
  model.type = "exp", 100)
leptu.s.decay.pow <- decay.model(leptu.s.sim, y.type = "sim", distgeo.s,
  model.type = "pow", 100)

```

```

leptu.s.decay.gom <- decay.model(leptu.s.sim, y.type = "sim", distgeo.s,
                                model.type = "gom", 100)

pdf(file = "leptu.s.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(leptu.s.decay.exp)
plot.decay(leptu.s.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(leptu.s.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
                      .(round(leptu.s.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
                      .(round(leptu.s.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
                      .(round(leptu.s.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.s, leptu.s.decay.exp$model$m$resid(), ylab = "Residuos (modelo
exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.s, leptu.s.decay.pow$model$m$resid(), ylab = "Residuos (modelo
power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.s, leptu.s.decay.gom$model$m$resid(), ylab = "Residuos (modelo
Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# - # Limnebiini
limne.s.decay.exp <- decay.model(limne.s.sim, y.type = "sim", distgeo.s,
                                model.type = "exp", 100)
limne.s.decay.pow <- decay.model(limne.s.sim, y.type = "sim", distgeo.s,
                                model.type = "pow", 100)
limne.s.decay.gom <- decay.model(limne.s.sim, y.type = "sim", distgeo.s,
                                model.type = "gom", 100)

pdf(file = "limne.s.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(limne.s.decay.exp)
plot.decay(limne.s.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(limne.s.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
                      .(round(limne.s.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
                      .(round(limne.s.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
                      .(round(limne.s.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.s, limne.s.decay.exp$model$m$resid(), ylab = "Residuos (modelo
exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.s, limne.s.decay.pow$model$m$resid(), ylab = "Residuos (modelo
power-law)", xlab = "Distancia espacial (km)", col = "red")

```

```
plot(distgeo.s, limne.s.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()
```

```
# - # Ochtebiini
ochte.s.decay.exp <- decay.model(ochte.s.sim, y.type = "sim", distgeo.s,
  model.type = "exp", 100)
ochte.s.decay.pow <- decay.model(ochte.s.sim, y.type = "sim", distgeo.s,
  model.type = "pow", 100)
ochte.s.decay.gom <- decay.model(ochte.s.sim, y.type = "sim", distgeo.s,
  model.type = "gom", 100)

pdf(file = "ochte.s.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(ochte.s.decay.exp)
plot.decay(ochte.s.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(ochte.s.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(ochte.s.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(ochte.s.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(ochte.s.decay.gom$pseudo.r.squared, digits = 2))))
```

```
plot(distgeo.s, ochte.s.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.s, ochte.s.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.s, ochte.s.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()
```

```
# - # Otiorhynchus
otior.s.decay.exp <- decay.model(otior.s.sim, y.type = "sim", distgeo.s,
  model.type = "exp", 100)
otior.s.decay.pow <- decay.model(otior.s.sim, y.type = "sim", distgeo.s,
  model.type = "pow", 100)
otior.s.decay.gom <- decay.model(otior.s.sim, y.type = "sim", distgeo.s,
  model.type = "gom", 100)

pdf(file = "otior.s.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(otior.s.decay.exp)
plot.decay(otior.s.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(otior.s.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(otior.s.decay.exp$pseudo.r.squared, digits = 2))))
```

```

text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(otior.s.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(otior.s.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.s, otior.s.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.s, otior.s.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.s, otior.s.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# - # Phyllobiini
phyll.s.decay.exp <- decay.model(phyll.s.sim, y.type = "sim", distgeo.s,
  model.type = "exp", 100)
phyll.s.decay.pow <- decay.model(phyll.s.sim, y.type = "sim", distgeo.s,
  model.type = "pow", 100)
phyll.s.decay.gom <- decay.model(phyll.s.sim, y.type = "sim", distgeo.s,
  model.type = "gom", 100)

pdf(file = "phyll.s.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(phyll.s.decay.exp)
plot.decay(phyll.s.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(phyll.s.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(phyll.s.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(phyll.s.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(phyll.s.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.s, phyll.s.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.s, phyll.s.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.s, phyll.s.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# - # Polydrusini
polyd.s.decay.exp <- decay.model(polyd.s.sim, y.type = "sim", distgeo.s,
  model.type = "exp", 100)
polyd.s.decay.pow <- decay.model(polyd.s.sim, y.type = "sim", distgeo.s,
  model.type = "pow", 100)
polyd.s.decay.gom <- decay.model(polyd.s.sim, y.type = "sim", distgeo.s,
  model.type = "gom", 100)

```

```

pdf(file = "polyd.s.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(polyd.s.decay.exp)
plot.decay(polyd.s.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(polyd.s.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(polyd.s.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(polyd.s.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(polyd.s.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.s, polyd.s.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.s, polyd.s.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.s, polyd.s.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# - # Pselaphinae
psela.s.decay.exp <- decay.model(psela.s.sim, y.type = "sim", distgeo.s,
  model.type = "exp", 100)
psela.s.decay.pow <- decay.model(psela.s.sim, y.type = "sim", distgeo.s,
  model.type = "pow", 100)
psela.s.decay.gom <- decay.model(psela.s.sim, y.type = "sim", distgeo.s,
  model.type = "gom", 100)

pdf(file = "psela.s.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(psela.s.decay.exp)
plot.decay(psela.s.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(psela.s.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(psela.s.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(psela.s.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(psela.s.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.s, psela.s.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.s, psela.s.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.s, psela.s.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

```

```

# - # Pterostichus
ptero.s.decay.exp <- decay.model(ptero.s.sim, y.type = "sim", distgeo.s,
  model.type = "exp", 100)
ptero.s.decay.pow <- decay.model(ptero.s.sim, y.type = "sim", distgeo.s,
  model.type = "pow", 100)
ptero.s.decay.gom <- decay.model(ptero.s.sim, y.type = "sim", distgeo.s,
  model.type = "gom", 100)

pdf(file = "ptero.s.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(ptero.s.decay.exp)
plot.decay(ptero.s.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(ptero.s.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(ptero.s.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(ptero.s.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(ptero.s.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.s, ptero.s.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.s, ptero.s.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.s, ptero.s.decay.gom$model$m$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# - # Scarabaeidae
scara.s.decay.exp <- decay.model(scara.s.sim, y.type = "sim", distgeo.s,
  model.type = "exp", 100)
scara.s.decay.pow <- decay.model(scara.s.sim, y.type = "sim", distgeo.s,
  model.type = "pow", 100)
scara.s.decay.gom <- decay.model(scara.s.sim, y.type = "sim", distgeo.s,
  model.type = "gom", 100)

pdf(file = "scara.s.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(scara.s.decay.exp)
plot.decay(scara.s.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(scara.s.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(scara.s.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(scara.s.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(scara.s.decay.gom$pseudo.r.squared, digits = 2))))

```

```

plot(distgeo.s, scara.s.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.s, scara.s.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.s, scara.s.decay.gom$model$m$resid(), ylab = "Residuos (modelo
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

```

```
# - # Scolytinae
```

```

scoly.s.decay.exp <- decay.model(scoly.s.sim, y.type = "sim", distgeo.s,
  model.type = "exp", 100)
scoly.s.decay.pow <- decay.model(scoly.s.sim, y.type = "sim", distgeo.s,
  model.type = "pow", 100)
scoly.s.decay.gom <- decay.model(scoly.s.sim, y.type = "sim", distgeo.s,
  model.type = "gom", 100)

```

```
pdf(file = "scoly.s.pdf", height = 3, width = 10)
```

```
par(mfrow = c(1, 4))
```

```
plot.decay(scoly.s.decay.exp)
```

```
plot.decay(scoly.s.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
```

```
plot.decay(scoly.s.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
```

```
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(scoly.s.decay.exp$pseudo.r.squared, digits = 2))))
```

```
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(scoly.s.decay.pow$pseudo.r.squared, digits = 2))))
```

```
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(scoly.s.decay.gom$pseudo.r.squared, digits = 2))))
```

```
plot(distgeo.s, scoly.s.decay.exp$model$m$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
```

```
plot(distgeo.s, scoly.s.decay.pow$model$m$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
```

```
plot(distgeo.s, scoly.s.decay.gom$model$m$resid(), ylab = "Residuos (modelo
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
```

```
dev.off()
```

```
# - # Silphidae
```

```

silph.s.decay.exp <- decay.model(silph.s.sim, y.type = "sim", distgeo.s,
  model.type = "exp", 100)

```

```

silph.s.decay.pow <- decay.model(silph.s.sim, y.type = "sim", distgeo.s,
  model.type = "pow", 100)

```

```

silph.s.decay.gom <- decay.model(silph.s.sim, y.type = "sim", distgeo.s,
  model.type = "gom", 100)

```

```
pdf(file = "silph.s.pdf", height = 3, width = 10)
```

```
par(mfrow = c(1, 4))
```

```
plot.decay(silph.s.decay.exp)
```

```
plot.decay(silph.s.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
```

```

plot.decay(silph.s.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(silph.s.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(silph.s.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(silph.s.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.s, silph.s.decay.exp$model$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.s, silph.s.decay.pow$model$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.s, silph.s.decay.gom$model$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# - # Trechus
trech.s.decay.exp <- decay.model(trech.s.sim, y.type = "sim", distgeo.s,
  model.type = "exp", 100)
trech.s.decay.pow <- decay.model(trech.s.sim, y.type = "sim", distgeo.s,
  model.type = "pow", 100)
trech.s.decay.gom <- decay.model(trech.s.sim, y.type = "sim", distgeo.s,
  model.type = "gom", 100)

pdf(file = "trech.s.pdf", height = 3, width = 10)
par(mfrow = c(1, 4))
plot.decay(trech.s.decay.exp)
plot.decay(trech.s.decay.pow, col = "red", add = TRUE, remove.dots = TRUE)
plot.decay(trech.s.decay.gom, col = "blue", add = TRUE, remove.dots = TRUE)
text(600,0.15, bquote("Exp.:" ~ italic(r)^2 ==
  .(round(trech.s.decay.exp$pseudo.r.squared, digits = 2))))
text(600,0.09, col = "red", bquote("Pow.:" ~ italic(r)^2 ==
  .(round(trech.s.decay.pow$pseudo.r.squared, digits = 2))))
text(600,0.03, col = "blue", bquote("Gom.:" ~ italic(r)^2 ==
  .(round(trech.s.decay.gom$pseudo.r.squared, digits = 2))))

plot(distgeo.s, trech.s.decay.exp$model$resid(), ylab = "Residuos (modelo
  exponencial)", xlab = "Distancia espacial (km)")
plot(distgeo.s, trech.s.decay.pow$model$resid(), ylab = "Residuos (modelo
  power-law)", xlab = "Distancia espacial (km)", col = "red")
plot(distgeo.s, trech.s.decay.gom$model$resid(), ylab = "Residuos (model
  Gompertz)", xlab = "Distancia espacial (km)", col = "blue")
dev.off()

# Tabla con los valores de r2 y p
vector.taxa <- c("amara", "apion", "carab", "ceram", "chaet", "chrys",
  "crypt", "hydra", "lamin", "leptu", "limne", "ochte", "otior",
  "phyll", "polyd", "psela", "ptero", "scara", "scoly", "silph",
  "trech")

```

```

-----
tabla.norte <- matrix(ncol = 9, nrow = 21)
rownames(tabla.norte) <- vector.taxa
colnames(tabla.norte) <- c("r2.exp", "p.val.exp", "p.val.BY.exp", "r2.pow",
  "p.val.pow", "p.val.BY.pow", "r2.gom", "p.val.gom" "p.val.BY.gom")

for (i in 1:21){
tabla.norte[i, 1] <- round(mget(paste(vector.taxa[i],".n.decay.exp", sep =
  ""))[[1]]$pseudo.r.squared, digits = 2)
tabla.norte[i, 2] <- round(mget(paste(vector.taxa[i],".n.decay.exp", sep =
  ""))[[1]]$p.value, digits = 3)

tabla.norte[i, 4] <- round(mget(paste(vector.taxa[i],".n.decay.pow", sep =
  ""))[[1]]$pseudo.r.squared, digits = 2)
tabla.norte[i, 5] <- round(mget(paste(vector.taxa[i],".n.decay.pow", sep =
  ""))[[1]]$p.value, digits = 3)

tabla.norte[i, 7] <- round(mget(paste(vector.taxa[i],".n.decay.gom", sep =
  ""))[[1]]$pseudo.r.squared, digits = 2)
tabla.norte[i, 8] <- round(mget(paste(vector.taxa[i],".n.decay.gom", sep =
  ""))[[1]]$p.value, digits = 3)
}

# Corrección de Benjamini-Yukiteli
tabla.norte[,3] <- round(p.adjust(tabla.norte[,2], method = "BY"), digits
  = 3)
tabla.norte[,6] <- round(p.adjust(tabla.norte[,5], method = "BY"), digits
  = 3)
tabla.norte[,9] <- round(p.adjust(tabla.norte[,8], method = "BY"), digits
  = 3)

write.table(tabla.norte, file = "tabla-norte.txt")

```

```

-----
tabla.sur <- matrix(ncol = 9, nrow = 21)
rownames(tabla.sur) <- vector.taxa
colnames(tabla.sur) <- c("r2.exp", "p.val.exp", "p.val.BY.exp", "r2.pow",
  "p.val.pow", "p.val.BY.pow", "r2.gom", "p.val.gom" "p.val.BY.gom")

for (i in 1:21){
tabla.sur[i, 1] <- round(mget(paste(vector.taxa[i],".s.decay.exp", sep =
  ""))[[1]]$pseudo.r.squared, digits = 2)
tabla.sur[i, 2] <- round(mget(paste(vector.taxa[i],".s.decay.exp", sep =
  ""))[[1]]$p.value, digits = 3)

tabla.sur[i, 4] <- round(mget(paste(vector.taxa[i],".s.decay.pow", sep =
  ""))[[1]]$pseudo.r.squared, digits = 2)
tabla.sur[i, 5] <- round(mget(paste(vector.taxa[i],".s.decay.pow", sep =
  ""))[[1]]$p.value, digits = 3)

```

```
tabla.sur[i, 7] <- round(mget(paste(vector.taxa[i], ".s.decay.gom", sep =
  ""))[[1]]$pseudo.r.squared, digits = 2)
tabla.sur[i, 8] <- round(mget(paste(vector.taxa[i], ".s.decay.gom", sep =
  ""))[[1]]$p.value, digits = 3)
}

# Corrección de Benjamini-Yukiteli
tabla.sur[, 3] <- round(p.adjust(tabla.sur[, 2], method = "BY"), digits =
  3)
tabla.sur[, 6] <- round(p.adjust(tabla.sur[, 5], method = "BY"), digits =
  3)
tabla.sur[, 9] <- round(p.adjust(tabla.sur[, 8], method = "BY"), digits =
  3)

write.table(tabla.sur, file = "tabla-sur.txt")
```

9.6. RELACIÓN DE PUBLICACIONES Y AUTORIZACIÓN DE LA REVISTA PARA INCLUSIÓN EN ESTA TESIS

ARTÍCULO

1. Martínez-Santalla S., R. Martín-Devasa, C. Gómez-Rodríguez, R.M. Crujeiras y A. Baselga (2022). Assessing the non-linear decay of community similarity: permutation and site-block resampling significance tests. *Journal of Biogeography* 49: 968-978. <https://doi.org/10.1111/jbi.14351>
Incluido en Metodología (apartado 4.5) y Resultados (apartado 5.4).

ISSN.: 1365-2699. Publicado por WILEY

Factor de impacto: 3.9, Cuartil: Q2. **Posición:** 13/49 en categoría: Geography, Physical y 46/171 en Ecology

Contribución de la doctoranda: Formulación del test, diseño y programación del estudio de simulación, análisis de datos reales, elaboración de figuras, contribución a la escritura.

Afiliación autores: Universidade de Santiago de Compostela (SMS, RMD, CGR, RMC y AB); University of Helsinki (RMD)

Autorización de la revista: No es necesaria la autorización de la revista ya que el artículo está publicado bajo la licencia Creative Commons Attribution-NonCommercial-NoDerivs, tal y como está especificado al pie de la primera página del mismo. El texto en inglés que especifica la licencia es el siguiente: *This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.*

APORTACIONES AL PAQUETE *BETAPART* (R) VERSIÓN 1.6.

Baselga, A., Orme, D., Villeger, S., De Bortoli, J., Leprieur, F., Logez, M., **Martínez-Santalla, S.**, Martín-Devasa, R., Gómez-Rodríguez, C., & Crujeiras, R.M. (2023). *betapart: Partitioning beta diversity into turnover and nestedness* (1.6). <http://CRAN.R-project.org/package=betapart>

Contribución de la doctoranda: Desarrollo del código que implementa el test de permutación en la función *decay.model()* y del test de remuestreo por bloques incluido en la función *zdep()*.

PERMISOS PARA LA REPRODUCCIÓN DE FIGURAS QUE NO SON DE ELABORACIÓN PROPIA

Todas las imágenes que no son de elaboración propia cuentan con licencia Creative Commons. La fuente original de la figura o la silueta está especificada en los pies de figura correspondientes. En el caso de las siluetas “freepik”, se puede consultar los términos de uso en http://www.freepik.com/terms_of_use. En el caso de las siluetas “phylopic”, servicio web que proporciona siluetas de libre uso de organismos, la información está disponible en <https://www.phylopic.org>

El descenso de la similitud biótica con la distancia (distance-decay) es un patrón universal de diversidad beta. Se modeliza analizando la relación entre matrices de similitud biótica y matrices de distancia. En estas matrices, cada observación no es independiente porque se origina de comparar dos comunidades, y cada comunidad está implicada en varias observaciones. Esta dependencia sesga el cálculo de varianzas y compromete la aplicación de contrastes de hipótesis. Esta tesis desarrolla métodos para modelos no lineales de distance-decay que tienen en cuenta esa dependencia, culminando en un contraste de significación y un método bootstrap por bloques de localidad para estimar la varianza de los parámetros de los modelos. Los nuevos métodos se aplican al estudio de los patrones de distance-decay en coleópteros europeos.