

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS DA NATUREZA
ENGENHARIA AMBIENTAL

Mateus Faria Santana

**Determinação do Efeito de Ilhas Pequenas a
partir dos Modelos Bent-Cable Probabilísticos**

Buri - São Paulo

2023

Mateus Faria Santana

Determinação do Efeito de Ilhas Pequenas a partir dos Modelos Bent-Cable Probabilísticos

Trabalho de conclusão de curso apresentado como exigência parcial para a obtenção do grau de Bacharel em Engenharia Ambiental na Universidade Federal de São Carlos.

Orientação: Prof. Dr. Iuri Emmanuel de Paula Ferreira

Buri - São Paulo

2023

Santana, Mateus Faria

Determinação do Efeito de Ilhas Pequenas a partir dos Modelos Bent-Cable Probabilísticos / Mateus Faria
Santana -- 2024.

41f.

TCC (Graduação) - Universidade Federal de São Carlos,
campus Lagoa do Sino, Buri

Orientador (a): Iuri Emmanuel de Paula Ferreira

Banca Examinadora: Iuri Emmanuel de Paula Ferreira,
Alexandre Camargo Martensen, Vinicius Ranier Boniolo

Bibliografia

1. Ecologia. 2. Relação Espécie área. 3. Estatística. I.
Santana, Mateus Faria. II. Título.

Ficha catalográfica desenvolvida pela Secretaria Geral de Informática
(SIn)

DADOS FORNECIDOS PELO AUTOR

Bibliotecário responsável: Lissandra Pinhatelli de Britto - CRB/8 7539

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS DA NATUREZA
CURSO DE GRADUAÇÃO EM ENGENHARIA AMBIENTAL

Folha de Aprovação

Assinatura dos membros da comissão examinadora que avaliou e aprovou a Defesa de Trabalho de Conclusão de Curso do candidato Mateus Faria Santana, intitulado “Determinação do Efeito de Ilhas Pequenas a partir dos Modelos Bent-Cable Probabilísticos”, realizada em 06/02/2024:

Documento assinado digitalmente
 IURI EMMANUEL DE PAULA FERREIRA
Data: 15/02/2024 13:05:44-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Iuri Emmanuel de Paula Ferreira – Orientador
Centro de Ciências da Natureza – UFSCar – Campus Lagoa do Sino.

Documento assinado digitalmente
 ALEXANDRE CAMARGO MARTENSEN
Data: 15/02/2024 16:17:30-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Alexandre de Camargo Martensen
Centro de Ciências da Natureza – UFSCar – Campus Lagoa do Sino.

Documento assinado digitalmente
 VINICIUS RAINER BONIOLO
Data: 15/02/2024 13:11:00-0300
Verifique em <https://validar.iti.gov.br>

Me. Vinicius Ranier Boniolo
Núcleo de Estudos em Ecologia Espacial e Desenvolvimento Sustentável – UFSCar –
Campus Lagoa do Sino.

Dedico este trabalho primeiramente a Deus, à minha
Família e à memória de um grande amigo Jarbert Ditzel
Azevedo.

Agradecimentos

Primeiramente, expresso minha gratidão a Deus por todas as oportunidades concedidas ao longo dos seis anos em uma das melhores universidades do país. Agradeço imensamente aos meus pais pelo suporte constante durante toda a minha graduação e por cada puxão de orelha, assim como pelos incentivos que marcaram cada etapa da minha vida. A presença deles foi fundamental para o meu crescimento acadêmico e pessoal, e sou profundamente grato por todo o apoio e orientação que me proporcionaram.

Manifesto meu sincero agradecimento ao meu orientador, Iuri Ferreira, por todos os ensinamentos valiosos proporcionados ao longo da graduação, assim como pelos conselhos que foram fundamentais para o meu desenvolvimento acadêmico. Quero expressar também minha gratidão ao grupo Cemeca, do qual faço parte, e aos seus integrantes, que contribuíram significativamente para o meu aprendizado. Destaco especialmente a influência de Denner Augusto, Kaik Fontana e do professor Cristiano Augusto de Souza, que, com suas experiências e conhecimentos, enriqueceram ainda mais minha jornada acadêmica.

Expresso minha sincera gratidão aos amigos da graduação, Pedro Chiari, Rafael Terra, Carlos Alexandre, e à memória de Jarbert Ditzel. Agradeço a todos por compartilharmos trabalhos, experiências e lembranças ao longo desse período. Também estendo meu especial agradecimento a Eduardo Ciríneo, Murilo Petrone, Erick Brito, Denner Augusto e Tiago Nóbrega pelos ensinamentos valiosos durante o curso. Quero agradecer a todos os amigos de sala e à turma de Engenharia Ambiental de 2018, especialmente a Julia Fraidemberge, Amanda Macambira, Taynara Soares e Raissa Tavares, por todos os puxões de orelha e por tornarem a jornada acadêmica ainda mais especial.

Quero expressar minha profunda gratidão à universidade pelas oportunidades de estágio oferecidas ao longo do curso. Um agradecimento especial à Omnia Consultoria, ao Rafael Mattos e aos colaboradores, que generosamente compartilharam seus ensinamentos.

Agradeço ao Consórcio Construtor pela oportunidade de estágio, em especial ao Lucas Areco, pela confiança e pela chance de encarar diversos desafios na empresa. Um reconhecimento especial aos técnicos Fábio William e João Bonfim, cujos conselhos foram fundamentais para o meu crescimento nessa fase. Estou sinceramente grato por ter evoluído ao lado de profissionais dedicados e experientes, uma experiência crucial para meu desenvolvimento pessoal e profissional. Estendo meu agradecimento aos auxiliares Rodrigo e Yasmin Vitória por seu companheirismo e ensinamentos essenciais no meu desenvolvimento. Agradeço também a toda a equipe de meio ambiente do Consórcio e à equipe de QSMR. Cada membro contribuiu significativamente para minha jornada, e sou grato por ter aprendido e crescido ao lado de profissionais tão dedicados.

*"Busquem, pois, em primeiro lugar o Reino de Deus e a sua justiça, e todas essas coisas
serão acrescentadas a vocês"*

Mateus 6:33

Resumo

Limiars ecológicos são pontos ou zonas de mudança entre dois ou mais estados de um sistema ecológico. A detecção de limiars ecológicos é fundamental para o desenvolvimento de políticas de conservação e sustentabilidade. O objetivo deste trabalho é a detecção de limiars aleatórios a partir do ajuste de modelos bent-cable probabilísticos, um tipo especial de modelos de regressão segmentada útil para descrever zonas de transição suaves. Neste trabalho, a abordagem foi usada para modelar o Efeito de Ilhas Pequenas, contribuindo para a detecção do Limiar de Área. A aplicação da técnica foi ilustrada nos dados de Matthews et al. sobre a riqueza de plantas em função da área das ilhas do Mar Egeu, na Grécia. A inferência estatística foi conduzida a partir de métodos Bayesianos, considerando a não-normalidade dos dados e a influência da área na dispersão e forma da distribuição da riqueza de espécies. Nossos resultados foram superiores quando comparados aos dos modelos *broken-line*, formulação frequentemente usada na representação de limiars ecológicos. O modelo bent-cable probabilístico possibilitou, ainda, recuperar a distribuição subjacente aos limiars aleatórios, possibilitando uma interpretação probabilística do Efeito de Ilhas Pequenas.

Palavras-chave: Regressão Segmentada, Transições Suaves, Relação Espécie-Área, Limiar de Área

Abstract

Ecological thresholds are change points or transitional zones within an ecological system signifying changes between different states. Detecting these thresholds is crucial for shaping conservation policies and sustainable practices. This research aims to spot random thresholds by employing probabilistic bent-cable models, a specialized type of segmented regression models adept at characterizing gradual transition zones. The study focused on modeling the Small Island Effect to identify the Area Threshold. To demonstrate this technique, data from Matthews et al. regarding the richness of plants in the archipelago of the Aegean Sea, Greece, were utilized. Statistical analysis utilized Bayesian methods, considering the non-normality of the data and how the area impacts the distribution and spread of species richness. Our results surpassed those of commonly used broken-line models for representing ecological thresholds. Additionally, the bent-cable model allowed for the retrieval of the underlying distribution of random thresholds, enabling a probabilistic interpretation of the Small Island Effect.

Keywords: Ecological thresholds, probabilistic models, bent-cable models, Small Island Effect, Bayesian methods, species richness, broken-line models.

Lista de ilustrações

Figura 1 – Ajustes de modelos <i>bent-cable</i> aos dados de R.A. Cook sobre a espessura de camada de surfactante em função do fluxo laminar (SEBER; WILD, 2003). Dados experimentais (círculos), regimes lineares (linhas pontilhadas) e zonas de transição (linhas contínuas). Fonte: (FERREIRA; ZOCCHI, 2021).	26
Figura 2 – O ajuste do modelo broken-line com dois limiares proposto por Matthews e Rigal (2021) está apresentado na Figura (a). O gráfico quantil-quantil para os resíduos dessa análise está mostrado na Figura (b). Esses dados tratam da riqueza de plantas em 173 ilhas localizadas no Mar Egeu, Grécia, disponibilizados no pacote SARS (MATTHEWS et al., 2019). .	32
Figura 3 – O ajuste do modelo broken-line com dois limiares proposto por Matthews (2021), considerando a distribuição de Poisson para a riqueza, está representado na figura (a). O diagnóstico dos resíduos foi realizado utilizando o método DHARMa, apresentado na figura (b). Esses dados são provenientes do conjunto fornecido pelo pacote SARS.	33
Figura 4 – Ajuste do modelo broken-line com dois limiares sugerido por Matthews e Rigal (2021) considerando a distribuição Binomial Negativa para a riqueza (a); Diagnóstico dos resíduos pelo método DHARMa (b). Dados de Matthews et al. (2019).	33
Figura 5 – Ajuste dos Modelos 2 e 3 considerando a Binomial Negativa para a riqueza com influência do tamanho das ilhas (Log área) sobre o parâmetro de dispersão. Modelo 2: <i>broken-line</i> com dois limiares de área fixos; Modelo 3: <i>bent-cable</i> probabilístico Logístico (L-BC). Dados de Matthews et al. (2019).	35
Figura 6 – Ilustração das ilhas sendo constituídas por uma mistura de habitats, com limiares de área aleatórios. Arquipélago grego com ilha de Creta ao centro. Imagem adaptada. Fonte: Google Maps (2022)	37
Figura 7 – Distribuição dos limiares aleatórios recuperada pelo modelo <i>bent-cable</i> Logístico e intervalos com 95% credibilidade para a fração de unidades de habitats responsivas - que já ultrapassaram seu limiar intrínseco de área.	37

Lista de tabelas

Tabela 1 – Casos particulares da família <i>bent-cable</i> : Log e Exponential de Jimenez-Fernandez et al. (2016). Modelo de transição hiperbólico de Griffiths e Miller (1973) e Watts e Bacon (1974). <i>bent-cable</i> quadrático (BC) de Chiu, Lockhart e Routledge (2006). <i>bent-cable</i> generalizado (G-BC) de Khan e Kar (2018). Epanechnikov <i>bent-cable</i> (E-BC) adaptado de Zang (1980). <i>bent-cable</i> Normal (N-BC) e <i>bent-cable</i> Skew-Normal (SN-BC) de Ferreira e Zocchi (2021).	26
Tabela 2 – Critérios para a comparação dos modelos obtidos pelo método de validação cruzada LOO <i>Leave-One-Out</i>	36

Lista de Sigla

SAR - Species-Area Relationships

SIE - Small Island effect

TBI - Teoria da Biogeografia de Ilha

TEBI - Teoria do equilíbrio de biogeografia de ilhas

Sumário

1	INTRODUÇÃO	19
2	FUNDAMENTAÇÃO TEÓRICA	21
2.1	Diversidade Biológica	21
2.2	Biogeografia de Ilhas	21
2.3	Efeito de Pequenas Ilhas	22
2.4	Regressão Segmentada e Aproximações Suaves	22
2.5	Modelos Bent-Cable Probabilísticos	24
2.6	Relação espécie-Área	27
3	OBJETIVOS	27
3.1	Objetivos Específicos	27
4	METODOLOGIA	28
4.1	Base de Dados	28
4.2	Análise Estatística	28
5	RESULTADOS E DISCUSSÃO	32
6	CONCLUSÃO	36
	REFERÊNCIAS	38

1 Introdução

A definição de um limiar ecológico, segundo a *Resilience Alliance*, um grupo interdisciplinar de pesquisadores em ciências sociais e ecológicas, representa um ponto de bifurcação entre estados alternativos de um sistema. Quando ultrapassado, esse limiar provoca a transição do sistema, alterando suas características (WALKER; MEYERS, 2004). Apesar de o termo limiar ter origem na Física, especificamente nos estudos de bifurcação e transição de estados na termodinâmica, há tempos ele é aplicado em Biologia e Ecologia.

Na Ecologia, a identificação e a utilização de limiares e zonas de transição são fundamentais para a conservação e a sustentabilidade no manejo dos recursos naturais (HUGGETT, 2005). Estudos quantitativos são essenciais para compreender a perda de biodiversidade e os processos de extinção local, causados por fatores como a perda de habitat e sua conectividade, mudanças climáticas, degradação ambiental, poluição e exploração não sustentável dos recursos naturais (RIGUEIRA, 2015; HUGGETT, 2005; FICETOLA; DENOËL, 2009). Huggett (2005) identifica dois tipos de limiares: um pode ser um ponto claramente definido no espaço e/ou no tempo, enquanto o outro é intervalar, representado por uma zona onde a mudança entre os estados ocorre com certo grau de incerteza.

A análise de limiares ecológicos é crucial em várias circunstâncias, como nas relações espécie-área (*Species-Area Relationships - SAR*). Um aspecto interessante é o Efeito das Ilhas Pequenas (*Small Island Effect - SIE*), uma característica anômala da relação espécie-área (SAR), que demonstra a variação da riqueza de espécies abaixo de um determinado limite de área, independentemente do tamanho da ilha ou, no mínimo, com um aumento a uma taxa menor do que em ilhas maiores (GAO; WANG, 2022).

Apesar do consenso dos pesquisadores sobre a definição dos limiares ecológicos, ainda não existem metodologias amplamente aceitas para sua detecção (FICETOLA; DENOËL, 2009). A grande discrepância entre os métodos aplicados tem sido um desafio, e a falta de uma metodologia bem estabelecida é o principal obstáculo para a objetividade e promoção de estudos sobre limiares ecológicos. Ficetola e Denoël (2009) recomendam a utilização de regressão segmentada (*Piecewise Regression*), a qual permite modelar fenômenos não-lineares e detectar pontos de mudança. Os modelos segmentados são multifásicos, descrevendo o fenômeno estudado por meio da combinação de submodelos, representando as fases do processo. A transição entre as fases pode ser contínua ou descontínua, e as mudanças entre os estados podem ser abruptas ou suaves. A versão mais simples e amplamente utilizada são os modelos segmentados lineares com transições de fase abruptas (*broken-line models*), frequentemente empregados na determinação do limiar de área para

ilhas pequenas (MATTHEWS; RIGAL, 2021; GAO; WANG, 2022).

Muitas vezes, vários modelos segmentados são considerados como possibilidades e critérios da teoria da informação são usados para escolher o melhor modelo (AKAIKE, 2011), como mostrado no trabalho de (MATTHEWS; RIGAL, 2021). Alguns autores dizem que apenas um ponto de mudança de área não é suficiente para explicar a SAR e sugerem que outro ponto de mudança deveria existir para entender o Efeito das Ilhas Pequenas (MATTHEWS; RIGAL, 2021).

Recentemente, Matthews et al. (2019) criaram o pacote *sars* para ajustar e comparar vários modelos segmentados na descrição da SAR e do Efeito das Ilhas Pequenas. Esse pacote foi feito no programa R, é fácil de usar e compara diferentes modelos automaticamente. Mas esse pacote usa métodos matemáticos que não são adequados para dados de contagem, como o número de espécies, e não consegue lidar com transições suaves, que são comuns em Ecologia (TOMS; LESPERANCE, 2003). Em suma, o pacote podem resultar em ajustes que, da perspectiva estatística, são problemáticos.

Sabe-se que o ponto de transição, onde ocorre a mudança do Efeito de Ilhas Pequenas, pode responder a diversos fatores, tais como a riqueza de habitats e a estrutura da comunidade, além de características ambientais e alterações climáticas, como níveis das chuvas (GAO; WANG, 2022). Esses determinantes variam de região à região e, dessa forma, neste trabalho não vemos o limiar para o Efeito de Ilhas Pequenas como algo determinado, fixo. Neste trabalho, defendemos que os limiares são variados, no espaço e no tempo, sendo afetados por variáveis intrínsecas e extrínsecas, muitas vezes desconhecidas.

De acordo com Ferreira e Zocchi (2021), ao incorporar efeitos aleatórios latentes nos limiares da mudança, os modelos segmentados se transformam em modelos de mudanças mais suaves (SPLMs - Smooth Piecewise-Linear Models). Neste trabalho, os modelos Bent-Cable Probabilísticos (BCP) serão usados para descrever os limiares do Efeito das Pequenas Ilhas. Os modelos BCP são uma classe de curvas que permitem modelar transições suaves em processos multifásicos, além de recuperar a distribuição de limiares aleatórios (FERREIRA; ZOCCHI, 2021).

2 Fundamentação Teórica

2.1 Diversidade Biológica

A diversidade biológica abrange a variabilidade de organismos vivos de todas as origens, incluindo ecossistemas terrestres, marinhos e outros ambientes aquáticos, bem como os complexos ecológicos dos quais fazem parte. Isso compreende a diversidade dentro de espécies, entre espécies e de ecossistemas. Essa definição ressalta a interconexão vital entre os elementos vivos do planeta. Ao considerar as transformações antropogênicas nas paisagens, como a conversão de vegetação nativa em habitat fragmentado, percebemos que essas mudanças têm implicações diretas na diversidade biológica, afetando tanto a variedade dentro das espécies quanto a integridade dos ecossistemas. Portanto, é crucial abordar essas questões de maneira integrada, buscando soluções que preservem e promovam a riqueza da vida em todas as suas formas (PEIXOTO; LUZ; BRITO, 2016) .

O crescente interesse na preservação dos recursos naturais e sua utilização está intrinsecamente relacionado aos impactos das mudanças climáticas, especialmente devido às ações humanas. Este cenário ganha relevância diante da transformação das paisagens naturais em mosaicos antrópicos, nos quais áreas de vegetação nativa são convertidas em pequenas manchas de habitat, dispersas em uma matriz urbana. A fragmentação desses habitats compromete os processos ecológicos, culminando na significativa perda de biodiversidade (RIBEIRO; MELLO; VALENTE, 2020).

2.2 Biogeografia de Ilhas

A pesquisa sobre comunidades de ilhas tem contribuído para o desenvolvimento de princípios gerais relacionados à distribuição da diversidade biológica, conforme apresentado na Teoria da Biogeografia de Ilhas (TBI) de MacArthur e Wilson (PRIMACK., 2002). Um dos fundamentos essenciais dessa teoria postula que o número de espécies em uma determinada área aumenta proporcionalmente com o tamanho da ilha examinada (MORIN, 2005). Uma possível explicação para a relação área-espécie descrita pela teoria é que ela pressupõe que o número de espécies em uma ilha resulta de um equilíbrio dinâmico e da inter-relação entre as taxas de colonização e extinção (BORDIGNON et al., 2007). Contudo, conforme a Teoria, há uma previsão de que o aumento da área a uma certa distância do continente resulta no aumento do número de espécies, sendo essa área considerada como a fonte colonizadora. De acordo com essa perspectiva, não apenas o tamanho, mas também o grau de isolamento são fatores que podem influenciar a diversidade de espécies presentes. A riqueza de espécies em uma ilha é determinada pelo equilíbrio dinâmico entre

a taxa de imigração e a taxa de extinção, conforme proposto pela Teoria do Equilíbrio de Biogeografia de Ilhas (TEBI), formulada por MacArthur e Wilson em 1963 e 1967.

De acordo com a Teoria de Biogeografia de Ilhas, o aumento da diversidade de espécies em uma ilha ocorre principalmente por meio da imigração de outras ilhas ou de uma massa continental, chamada de fonte. As espécies provenientes da área continental mais próxima representam um conjunto de colonizadores potenciais (RICKLEFS, 2003). À medida que mais colonizadores continentais potenciais se estabelecem na ilha, a proporção de indivíduos imigrantes pertencentes a novas espécies diminui. Importante notar que as "ilhas" não se limitam necessariamente a porções de terra cercadas por água do mar; exemplos incluem lagos em meio a uma "mar" de terra, cumes de montanhas, clareiras em florestas e áreas fragmentadas (TOWNSEND; BEGON; HARPER, 2006; BORDIGNON et al., 2007).

2.3 Efeito de Pequenas Ilhas

Fragmentos de terra maiores geralmente tendem a abrigar uma diversidade mais rica de espécies. No entanto, essa relação é mais evidente quando se considera uma ampla gama de tamanhos, como fragmentos que variam de 10 a 1.000.000 hectares (LOMOLINO; WEISER, 2001). Quando a variação de tamanho não é tão expressiva, a correlação positiva entre o número de espécies e o tamanho da área pode não ser aparente. Esse fenômeno é conhecido como "Efeito de Pequenas Ilhas" (Small Island Effect) e ocorre não apenas em mamíferos, mas na maioria dos grupos de organismos estudados. Além do tamanho, outros fatores, frequentemente chamados de "qualidade do fragmento", podem desempenhar um papel mais proeminente na determinação da riqueza de espécies em paisagens fragmentadas. (VIEIRA et al., 2003)

2.4 Regressão Segmentada e Aproximações Suaves

Considere um modelo de regressão simples que envolve apenas uma variável explanatória, usualmente expresso por

$$y_i = \eta(x_i) + \epsilon_i \quad (i = 1, 2, \dots, n),$$

onde y_i representa a i -ésima das n respostas observadas de uma variável aleatória Y , ϵ_i é o i -ésimo erro aleatório, e $\eta(x_i)$ é a resposta esperada condicionada ao i -ésimo valor da variável explanatória x_i . No contexto da análise de regressão, o valor de x_i é fixado pelo pesquisador ou mensurado de forma precisa, com erros desprezíveis. (RATKOOWSKY, 1990)

Na regressão segmentada, a parte fixa do modelo é desdobrada em

$$\eta(x) = \eta(x|\beta, \tau) = \begin{cases} g_1(x|\beta_1), & A \leq x \leq \tau_1, \\ g_2(x|\beta_2), & \tau_1 < x \leq \tau_2, \\ \vdots \\ g_D(x|\beta_D), & \tau_{D-1} < x \leq B, \end{cases}$$

onde as funções g_j 's, com $j = 1, 2, \dots, D$, denominadas também de regimes, geralmente são polinômios de 1º e 2º graus ou funções não lineares de x com poucos parâmetros que caracterizam a forma das D fases do processo estudado. Os β_j 's são os vetores de parâmetros associados a cada regime, de modo que $\beta^T = (\beta_1^T, \beta_2^T, \dots, \beta_D^T)$. A e B são os limites inferior e superior do domínio de regressão, e $\tau^T = (\tau_1, \tau_2, \dots, \tau_{D-1})$, com $A < \tau_1 < \tau_2 < \dots < \tau_{D-1} < B$. Os elementos do vetor τ são os valores das abscissas que delimitam a partição proposta do domínio de regressão e usualmente recebem nomes como pontos de junção, de transição, de mudança ou pontos finais. Por razões fenomenológicas, também os chamaremos de limiares. A modelagem de respostas contínuas exige, ainda, a imposição da seguinte restrição

$$g_l(x = \tau_l|\beta_l) = g_{l+1}(x = \tau_l|\beta_{l+1}), \quad \text{para } l = 1, 2, \dots, D - 1.$$

O pioneiro na regressão segmentada foi Quandt (1958), seguido por Sprent (1961), Hinkley (1969, 1971) e Feder (1975), que estudaram aspectos da estimação dos parâmetros e inferência para sistemas formados por duas retas que se interceptam. No trabalho de McGee e Carleton (1970), por sua vez, a ideia de Quandt (1958) foi estendida para um número indeterminado (e possivelmente desconhecido) de regimes polinomiais, cunhando-se o termo "piecewise regression", método para a busca da função contínua segmentada que melhor descreve o conjunto de dados. Os gráficos desses modelos caracterizam-se pela presença de quinas, que ocorrem nas interseções de regimes adjacentes, o que implica a inexistência da derivada de $\eta(x)$ sob os pontos de mudança. O problema é que sem derivadas contínuas da parte fixa do modelo, os métodos usuais de estimação e inferência em regressão não linear não são aplicáveis (SEBER; WILD, 1989). Além disso, mudanças repentinas são incomuns na natureza; as transições de fase geralmente são fenômenos graduais.

Gallant e Fuller (1973) propuseram, então, uma das primeiras formulações de modelos segmentados com junções suaves entre os regimes. A proposta generaliza a metodologia de McGee e Carleton (1970) ao impor restrições de continuidade para as derivadas de $\eta(x)$ em modelos segmentados com regimes polinomiais. Como caso particular dos modelos de Gallant e Fuller (1973), podem ser citados os q-"splines", curvas constituídas por segmentos polinomiais de grau q e com derivadas em relação a x contínuas até a ordem $q - 1$ (SEBER; WILD, 2005). A bibliografia especializada também apresenta outras

alternativas para suavizar as junções e obter modelos mais tratáveis do ponto de vista estatístico. Elas seguem praticamente o mesmo caminho: formulam modelos com o auxílio de operadores abruptos, como máximos, mínimos, módulos ou sinais, e, em seguida, os substituem por aproximações que conduzem a funções continuamente diferenciáveis (ou seja, funções da classe C^1). Na bibliografia especializada, destacam-se duas abordagens: as formulações sinal de Bacon e Watts (1971) e máx-mín de Tishler e Zang (1981a,b), as quais são descritas a seguir. (FERREIRA, 2015)

2.5 Modelos Bent-Cable Probabilísticos

Considere o modelo segmentado a seguir:

$$y_i = \eta(x_i) + \epsilon_i \quad (i = 1, 2, \dots, n),$$

no qual a parte fixa (η) é sub-dividida em D linhas retas (fases ou regimes). Os sub-modelos lineares são expressos por $g_j(x) = \alpha_j + \beta_j x$, para $j = 1, \dots, D$, $\tau_{j-1} \leq x \leq \tau_j$, e τ_j 's são os pontos de mudança (ou limiares). Os α 's, β 's e $\tau_1, \tau_2, \dots, \tau_{D-1}$ são parâmetros livres a serem estimados, enquanto $\tau_0 = x_{(1)}$ e $\tau_D = x_{(n)}$. A continuidade da curva de resposta é garantida por $g_l(\tau_l) = g_{l+1}(\tau_l)$ para $l = 1, \dots, D - 1$.

A fórmula geral dos modelos lineares segmentados contínuos (ou *broken-lines* é

$$\eta(x) = \alpha_1 + \beta_1 x + \sum_{l=1}^{D-1} \delta_l u(x - \tau_l),$$

onde $\delta_l = \beta_{l+1} - \beta_l$, e $u(z) = \max\{z, 0\}$ para qualquer z real. Estes modelos apresentam derivadas descontínuas em $x = \tau_l$, e a curva de resposta tem quinas sob estes pontos. Isto é consequência do emprego de operadores abruptos na formulação dos modelos. A ideia que embasa os SPLMs é substituir os operadores abruptos por aproximações suaves, com derivadas contínuas. Um exemplo disso é o modelo bent-cable de Chiu et al., no qual o operador de máximo $u(\cdot)$ é substituído por uma função polinomial quadrática nas vizinhanças do ponto de mudança (CHIU; LOCKHART; ROUTLEDGE, 2006). Dessa forma, o modelo de Chiu et al. é descrito por

$$\eta(x, \beta_1, \delta, \tau_1, \gamma) = \alpha_1 + \beta_1 x + \delta u_\gamma(x - \tau_1), \quad \gamma > 0, \quad (2.1)$$

em que $\delta = \beta_2 - \beta_1$ e

$$u_\gamma(x - \tau_1) = \frac{(x - \tau_1 + \gamma)^2}{4\gamma} I(|x - \tau_1| \leq \gamma) + (x - \tau_1) I(x - \tau_1 > \gamma). \quad (2.2)$$

Neste caso, a aproximação suave para u_γ é a função quadrática (o *bent-cable*), enquanto γ é o parâmetro de suavização, e a transição de fase ocorre no intervalo $G = [\tau_1 - \gamma, \tau_1 + \gamma]$.

Ferreira e Zocchi (2021) estenderam a ideia de *bent-cable* de Chiu et al. incorporando muitas aproximações suaves como alternativas à aproximação quadrática. A família de

modelos *bent-cable* por eles proposta surge da pressuposição de limiares aleatórios latentes. Nesta abordagem, cada unidade observacional possui sub-unidades operando em pelo menos dois estados possíveis (os regimes). As transições de fase são dependentes de x e mediadas por um limiar aleatório T , que opera ao nível das sub-unidades. Dessa forma, ao calcularmos o valor esperado de $u(\cdot)$, obtemos uma aproximação suave para o operador abrupto, dada por:

$$E(u(x - T)) = (x - \tau)F_T(x) + \int_x^{+\infty} (T - \tau) dF_T,$$

em que $F_T(x; \tau, \nu)$ é a função de distribuição para T , $E(T) = \tau$, e ν é o parâmetro de escala. E, por este caminho, Ferreira e Zocchi derivam a formulação geral dos modelos *bent-cable*:

$$\eta(x) = \alpha_1 + \beta_1 x + \delta \left[(x - \tau)F_T(x) + \int_x^{+\infty} (T - \tau) dF_T \right].$$

Este resultado vale apenas para T integrável, e existe um modelo *bent-cable* associado a cada distribuição contínua bem comportada ($E|T| < +\infty$) (FERREIRA; ZOCCHI, 2021). Nestes modelos, as mudanças estruturais ocorrem sob o ponto $x = \tau$ enquanto que a zona de transição de fase é estimada a partir de percentis da distribuição de T ; neste trabalho considera-se que a mudança ocorre ao longo da região $G = \{x : T_{20\%} < x < T_{80\%}\}$.

A maioria dos SPLMs encontrados na bibliografia encaixam-se na família *bent-cable*: por exemplo, o *bent-cable* quadrático de Chiu, Lockhart e Routledge (2006) ocorre quando T tem distribuição Uniforme; o modelo hiperbólico de Griffiths e Miller (1973) ocorre quando T tem distribuição t de Student não-padrão com dois graus de liberdade; a aproximação de Zang (1980) ocorre quando T tem distribuição de Epanechnikov; a aproximação log e exponencial de Jimenez-Fernandez et al. (2016) ocorre quando T tem distribuição Logística; e o *bent-cable* generalizado de Khan e Kar (2018) ocorre quando T tem distribuição de Uniforme Exponenciada conforme Ramires et al. (2019). Ferreira e Zocchi (2021) também introduzem novos SPLMs baseados nas distribuições Normal e Skew-Normal (Tabela 1, Figura 1).

Os modelos da família *bent-cable* possuem uma interpretação geométrica simples: enquanto nos modelos *broken-line* a derivada $d\eta/dx$ salta, abruptamente, de β_1 para β_2 sob o ponto de mudança, nos modelos *bent-cable* varia gradualmente de β_1 para β_2 através de uma combinação linear convexa com pesos iguais a $1 - F_T(x)$ e $F_T(x)$. Isto é verificado ao aplicarmos a regra de Leibniz para derivar a função *bent-cable*:

$$\frac{d}{dx}E(u(x - T)) = F_T(x),$$

e, assim,

$$\frac{d}{dx}\eta(x) = \beta_1 + \delta F_T(x) = \beta_1[1 - F_T(x)] + \beta_2 F_T(x).$$

Tabela 1 – Casos particulares da família *bent-cable*: Log e Exponential de Jimenez-Fernandez et al. (2016). Modelo de transição hiperbólico de Griffiths e Miller (1973) e Watts e Bacon (1974). *bent-cable* quadrático (BC) de Chiu, Lockhart e Routledge (2006). *bent-cable* generalizado (G-BC) de Khan e Kar (2018). Epanechnikov *bent-cable* (E-BC) adaptado de Zang (1980). *bent-cable* Normal (N-BC) e *bent-cable* Skew-Normal (SN-BC) de Ferreira e Zocchi (2021).

<i>Model</i> ^a	<i>T distribution</i>
Log and Exponential	Logistic
Hyperbolic transition	Non-standard t
Bent-Cable (BC)	Uniform
Generalised Bent-Cable (G-BC)	Exponentiated Uniform
Epanechnikov Bent-Cable (E-BC)	Epanechnikov
Normal Bent-Cable (N-BC)	Normal
Skew-Normal Bent-Cable (SN-BC)	Skew-Normal

^a Adaptado de Ferreira e Zocchi (2021)

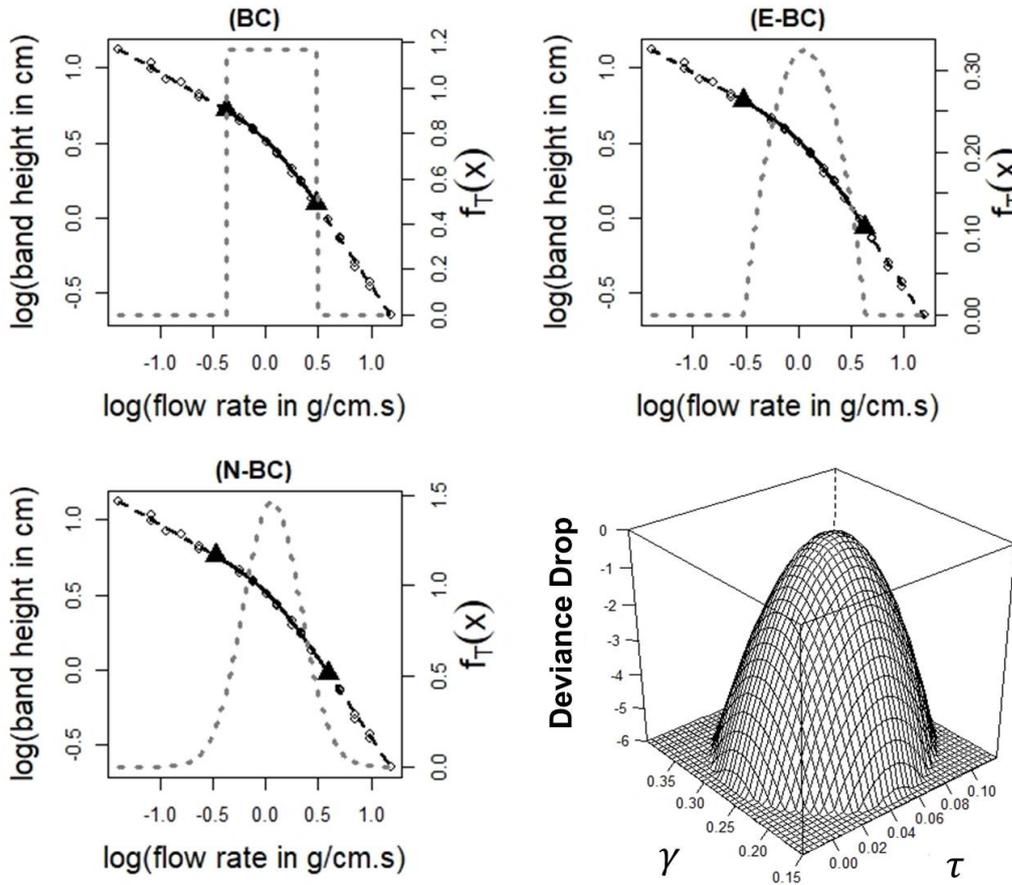


Figura 1 – Ajustes de modelos *bent-cable* aos dados de R.A. Cook sobre a espessura de camada de surfactante em função do fluxo laminar (SEBER; WILD, 2003). Dados experimentais (círculos), regimes lineares (linhas pontilhadas) e zonas de transição (linhas contínuas). Fonte: (FERREIRA; ZOCCHI, 2021).

2.6 Relação espécie-Área

Segundo Liu et al a relação espécie-área (SAR) é um padrão ecológico fundamental que descreve como a riqueza de espécies em uma determinada área varia conforme o tamanho da área amostral aumenta. Em outras palavras, a SAR mostra como o número de espécies observadas em uma região está relacionado ao tamanho da área estudada.

Essa relação é amplamente estudada em ecologia, biogeografia, ecologia de populações e biologia da conservação, sendo considerada uma das leis básicas que governam a distribuição e diversidade de espécies em diferentes habitats. A SAR é importante para entender como a biodiversidade é distribuída em diferentes escalas espaciais e pode ser usada para diversos fins, como identificar áreas de alta biodiversidade, planejar a conservação de habitats naturais e prever a extinção de espécies. (LIU; LIU; SHEN, 2022)

3 Objetivos

O objetivo geral deste trabalho é detectar limiares ecológicos aleatórios através do ajuste de modelos Bent-Cable Probabilísticos, visando descrever zonas de transição suaves em sistemas ecológicos. A técnica é aplicada na descrição do Efeito de Ilhas Pequenas.

3.1 Objetivos Específicos

- (a) Explorar e aprender sobre técnicas estatísticas que dividem dados ecológicos usando modelos de regressão segmentada.
- (b) Estudar uma família específica de modelos de partes lineares conhecida como bent-cable.
- (c) Identificar mudanças importantes na natureza e zonas onde essas mudanças acontecem ao analisar os modelos bent-cable.
- (d) Usar as técnicas aprendidas para investigar como o número de espécies varia com a área e como a destruição de habitats afeta essas variações.

4 Metodologia

4.1 Base de Dados

Neste trabalho, utilizaram-se dados sobre a riqueza de plantas em 173 ilhas do Mar Egeu, na Grécia. A tabela de dados apresenta apenas duas variáveis, a área das ilhas (em km^2) e o número de espécies de plantas. A variável área possui grande escala de variação, com valores indo de $0.0006 km^2$ para ilha da Tourleta até $8.270 km^2$ para a ilha de Creta. Dessa forma, a variável área foi analisada na escala logarítmica, para remover a influência da escala de variação e dos pontos discrepantes (ilhas com as maiores áreas). Os dados foram publicados e analisados previamente por Matthews et al. (2020) e Matthews e Rigal (2021), e disponibilizados publicamente como exemplo no pacote *sars* (MATTHEWS et al., 2019).

4.2 Análise Estatística

Para a estimação dos parâmetros foi utilizada a abordagem Bayesiana. Neste estudo, considerou-se para riqueza de espécies

$$Y_i \sim D(\mu_i = \eta(x_i, \boldsymbol{\theta}), \boldsymbol{\phi}), \quad (4.1)$$

x_i é o Log da área na i -ésima ilha; $\boldsymbol{\theta}^T = [\alpha_1, \beta_1, \delta_1, \tau_1, \gamma_1, \dots]$ é o vetor de parâmetros associados à formulação do *bent-cable*, e $\mu_i = \eta(x_i)$ representa a média do processo condicionada ao valor de x ; e $\boldsymbol{\phi}$ é um vetor de parâmetros adicionais que determinam a forma, escala ou taxa na especificação da distribuição D . As especificações de modelos segmentados exploradas neste trabalho foram baseadas *bent-cable* Logístico (L-BC), uma reparametrização do modelo Log e Exponencial de Jimenez-Fernandez et al. (2016). O L-BC é expresso por:

$$\eta(x) = \alpha_1 + \beta_1 x + \sum_{l=1}^{D-1} \delta_l \gamma_l \log\{\exp[(x - \tau_l)/\gamma_l] + 1\}, \quad (4.2)$$

Modelos segmentados abruptos, do tipo *broken-line*, são obtidos da Eq. (4.2) quando os parâmetros γ 's são fixados em um valor arbitrariamente pequeno, próximo de zero.

A partir da Eq. (4.2) três situações foram consideradas:

- Modelo 1: *broken-line* com apenas um limiar fixo ($D = 1$ e $\gamma_1 = 0.01$)
- Modelo 2: *broken-line* com dois limiares fixos ($D = 2$ e $\gamma_1 = \gamma_2 = 0.01$)

- Modelo 3: *bent-cable* Logístico (L-BC) com um limiar aleatório ($D = 1$ e γ_1 desconhecido)

Os Modelos 1 e 2, do tipo *broken-line*, são comumente empregados na determinação do limiar de área para a SAR, como podemos ver nos trabalhos de Matthews e Rigal (2021) e Gao e Wang (2022). Geralmente são admitidos erros normais e homocedásticos, de maneira que os modelos empregados sejam casos particulares da Eq. (4.1) quando: (i) os γ tendem a zero e (ii) a distribuição $D(\boldsymbol{\mu}, \boldsymbol{\phi}) = N(\boldsymbol{\mu}, I\sigma^2)$. As suposições clássicas de normalidade e homogeneidade de variâncias devem ser questionadas, uma vez que dados de riqueza são contagens, onde a variância responde à média do processo. Além disso, a distribuição Normal é contínua e tem suporte ilimitado, podendo gerar valores negativos; enquanto isso, a riqueza é uma variável que assume valores naturais e, portanto, requer uma distribuição discreta com suporte \mathbb{R}^+ para a sua modelagem.

Em Estatística, a distribuição de Poisson é um modelo natural para lidar com contagens ilimitadas em uma região do espaço e/ou durante um período de tempo. Este modelo considera uma taxa de contagens constantes, e admite que cada item/unidade contada ocorra no espaço e/ou no tempo de forma independente. A distribuição de Poisson é expressa por

$$Pois(y | \mu) = \frac{\mu^y e^{-\mu}}{y!},$$

com $y \in \{0, 1, 2, \dots\}$ e $\mu \in \mathbb{R}^+$, para a qual $E(Y) = Var(Y) = \mu$.

Uma característica da Poisson, que decorre da independência dos itens/objetos contados, é que a média do processo e a variância são iguais. Porém, em problemas práticos, é comum o processo apresentar uma variância que exceda o valor da média de maneira significativa, situações nas quais diremos que existe super-dispersão nos dados. Isso pode ocorrer, por exemplo, quando os itens/objetos contados são interindependentes e aparecem em grupos na amostra. Super-dispersão é uma característica comum em dados de riqueza de espécies, por exemplo; isso porque há espécies com fortes interações ecológicas e que com frequência estão associadas ao mesmo habitat, aparecendo juntas quando detectadas.

O modelo Binomial Negativo pode ser compreendido como uma extensão do modelo de Poisson que, através de um parâmetro a mais, permite modelar a dispersão extra existente nos dados. O modelo Binomial Negativo utilizado neste trabalho é expresso por

$$NegBinomial(y | \mu, \phi) = \binom{y + \phi - 1}{y} \left(\frac{\mu}{\mu + \phi} \right)^y \left(\frac{\phi}{\mu + \phi} \right)^\phi,$$

com $\phi \in \mathbb{R}^+$, de maneira que

$$E(Y) = \mu \quad \text{e} \quad Var(Y) = \mu + \frac{\mu^2}{\phi}.$$

Desta forma, o termo $\frac{\mu^2}{\phi} > 0$ é a dispersão extra modelada pela Binomial Negativa em comparação com a Poisson (para mais detalhes, vide em *Stan's User Guide* <<https://mc-stan.org/docs/functions-reference/nbalt.html>>). Neste trabalho, ambas as distribuições de Poisson e Binomial Negativa foram usadas para modelar a riqueza de espécies Y , considerando as três curvas de regressão derivadas da equação (4.2), ou seja, os Modelos 1, 2 e 3.

Em biogeografia de ilhas, a dispersão da riqueza também pode ser afetada pelo tamanho das ilhas. Neste trabalho, o tamanho das ilhas x foi usado como co-variável para modelar a dispersão, sendo esta ligada ao parâmetro ϕ através de uma função de ligação logarítmica, ou seja:

$$\log(\phi_i) = a_0 + a_1 x_i.$$

Distribuições *a priori* pouco informativas foram fixadas para os parâmetros do modelo, com exceção dos pontos de mudança, τ_1 e τ_2 . Observa-se que é necessário incluir alguma informação inicial a respeito dos pontos de mudança, de maneira a obter valores iniciais que considerem ambos os regimes do modelo segmentado. Quando os valores iniciais dos τ 's estão muito afastados, um dos regimes pode ser negligenciado acarretando em problemas de mal-condicionamento do modelo.

No procedimento de estimação foram usadas as seguintes distribuições *a priori*:

$$\alpha_1, \beta_1, \delta_1, \delta_2, \gamma_1, \gamma_2 \sim G(0.01, 0.001) \quad \text{e} \quad \tau_1, \tau_2 \sim \text{Logística}(0.5, 3). \quad (4.3)$$

Enquanto isso, o parâmetro ϕ da distribuição Binomial Negativa teve distribuição *a priori* não-informativa automaticamente configurada pelo método *brms*.

As amostras das cadeias de Monte Carlo foram geradas a partir do algoritmo NUTS *No-U-Turn Sampler* um algoritmo do tipo Monte Carlo Hamiltoniano que automaticamente configura seus parâmetros de controle, evitando ajustes manuais. O amostrador NUTS implementado no *Stan* não requer *prioris* conjugadas (BÜRKNER, 2017), e a qualidade das amostras fornecidas por ele geralmente é superior a de amostras obtidas por outros métodos (HOFFMAN; GELMAN, 2014).

Os modelos foram implementados em linguagem R, com o auxílio do pacote *brms* (BÜRKNER, 2017). A performance dos modelos foi comparada através do método LOO *leave-one-out cross-validation*, a partir da *expected log-pointwise predictive density* 'elpd', do número efetivo de parâmetros 'p loo', do critério da informação pelo método *leave-one-out* 'looic' (VEHTARI; GELMAN; GABRY, 2017) e do coeficiente de determinação Bayesiano (GELMAN et al., 2019).

A convergência das cadeias foi explorada utilizando o critério *R-hat* de (GELMAN; RUBIN, 1992), e amostras efetivas com mais de 1.500 observações independentes foram

usadas para inferências. O diagnóstico do ajuste foi conduzido a partir de recursos gráficos dos pacotes *bayesplot* e *DHARMA*.

5 Resultados e discussão

Os dados foram inicialmente examinados por Matthews e Rigal (2021), que propuseram a aplicação de um modelo de segmentos denominado broken-line com dois pontos de transição (modelo 2). Eles ajustaram uma linha aos dados por meio do método de mínimos quadrados, presumindo a normalidade dos dados e variações homogêneas. Os resultados apresentados por Matthews e Rigal (2021) são ilustrados na figura 2. Embora a linha ajustada pareça se alinhar bem aos dados, nota-se que a variação não é uniforme; ela aumenta proporcionalmente à riqueza e ao tamanho das ilhas (Fig. 2-a). Ao examinarmos os desvios cometidos pelo modelo, observamos que eles são superiores às expectativas (Fig. 2-a), e os dados não exibem uma distribuição normal (Teste SW; $W = 0.83843$; valor $p < 0.0001$). Em relação à riqueza, os dados consistem em contagens, e para esse tipo de informação, a suposição de uma distribuição normal não é aconselhável.

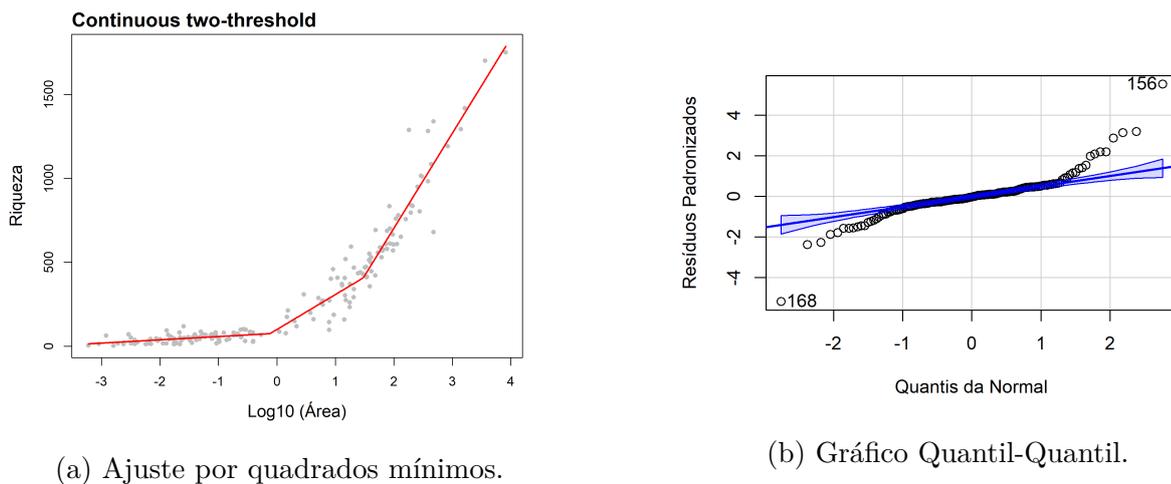
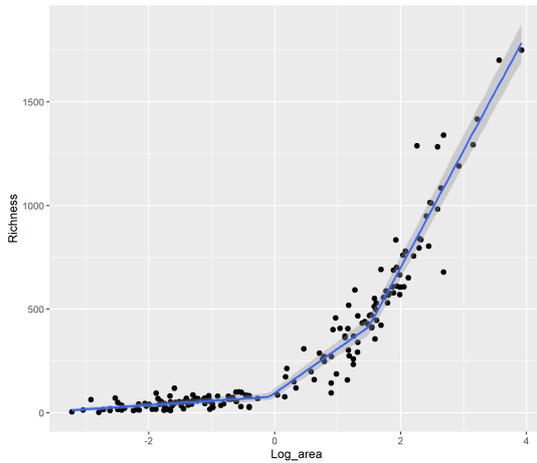


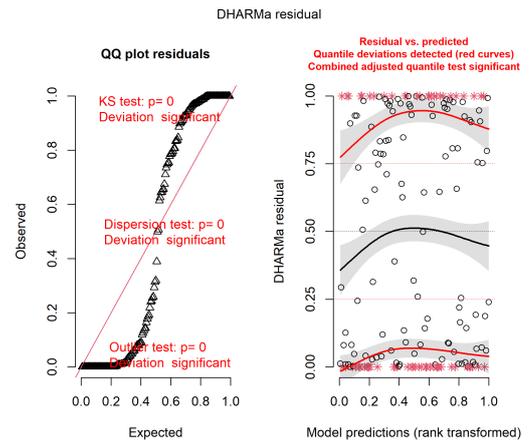
Figura 2 – O ajuste do modelo broken-line com dois limiares proposto por Matthews e Rigal (2021) está apresentado na Figura (a). O gráfico quantil-quantil para os resíduos dessa análise está mostrado na Figura (b). Esses dados tratam da riqueza de plantas em 173 ilhas localizadas no Mar Egeu, Grécia, disponibilizados no pacote SARS (MATTHEWS et al., 2019).

No presente estudo, investigou-se o desempenho do modelo de Poisson em relação à variável primária, que é a riqueza. Este modelo apresenta uma propriedade interessante em que a variabilidade dos dados é equivalente à média. Ao examinar a Figura 1a (referenciada como Fig. 3), observa-se que a variabilidade dos dados supera a previsão do modelo de Poisson, indicando uma dispersão excessiva nos dados. Tal dispersão é confirmada pela análise dos resíduos por meio do método DHARMA, conforme ilustrado na Figura 1b (referenciada como Fig. 3).

posteriormente, exploramos o modelo Binomial Negativo (BN) como uma alternativa

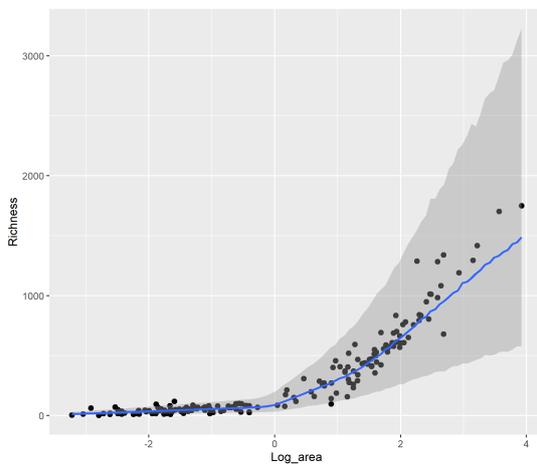


(a) Ajuste do modelo Poisson.

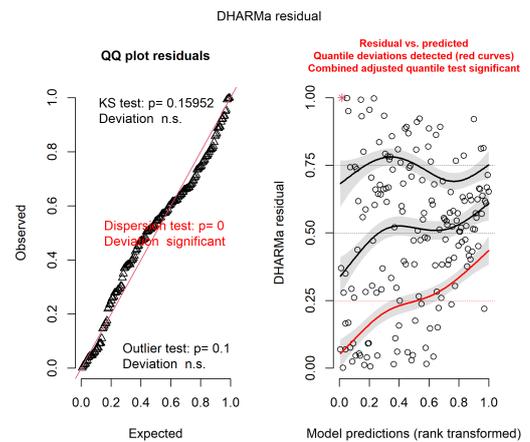


(b) Diagnóstico pelo método DHARMA.

Figura 3 – O ajuste do modelo broken-line com dois limiares proposto por Matthews (2021), considerando a distribuição de Poisson para a riqueza, está representado na figura (a). O diagnóstico dos resíduos foi realizado utilizando o método DHARMA, apresentado na figura (b). Esses dados são provenientes do conjunto fornecido pelo pacote SARS.



(a) Ajuste da Binomial Negativa



(b) Diagnóstico pelo método DHARMA

Figura 4 – Ajuste do modelo broken-line com dois limiares sugerido por Matthews e Rigal (2021) considerando a distribuição Binomial Negativa para a riqueza (a); Diagnóstico dos resíduos pelo método DHARMA (b). Dados de Matthews et al. (2019).

para abordar a dispersão adicional presente nos dados de contagem. A visualização do ajuste revelou que o modelo Binomial Negativo subestima a dispersão para ilhas pequenas, ao passo que a superestima para ilhas grandes (Fig. 4-a). A falta de adequação do ajuste é confirmada pela análise dos resíduos utilizando o método DHARMA (Fig. 4-b)

As conclusões anteriores oferecem evidências que sustentam a hipótese de que o tamanho da ilha pode influenciar a variabilidade do processo. Para explorar essa hipótese, incluí o tamanho das ilhas como uma covariável associada ao parâmetro ϕ do modelo

Binomial Negativo. Esse parâmetro regula a forma da curva e sua precisão (ou seja, ϕ é inversamente proporcional à dispersão).

Depois de incorporar a área como uma covariável em ϕ , os Modelos 1 e 2 passaram a apresentar resultados satisfatórios, sem evidências de falta de ajuste, dependência ou *outliers*, conforme demonstrado na figura 5. No entanto, o Modelo 1, que leva em consideração apenas um limiar de área fixo, não apresentou um bom ajuste aos dados em nenhuma das condições - distribuição Normal, de Poisson, Binomial Negativa ou Binomial Negativa com parâmetro ϕ heterogêneo. Portanto, as hipóteses de que (i) existem dois limiares de área fixos ou (ii) há um limiar de área aleatório ainda permanecem válidas.

Ao examinar a tabela de critérios de ajuste (Tabela 2), observa-se que a distribuição Binomial Negativa oferece os resultados mais favoráveis, apresentando o menor critério de informação *loaic* e o menor erro esperado para a log-densidade a posteriori, ponto a ponto *elpd*. É importante ressaltar que o ajuste é bem-sucedido para os Modelos 2 e 3, isto é, aqueles com dois limiares fixos e um limiar aleatório, respectivamente. No entanto, para o Modelo 1, que possui um limiar fixo, o método sequer convergiu. A análise gráfica revelou que o modelo não consegue descrever adequadamente as ilhas de grande porte, apresentando evidências claras de falta de ajuste.

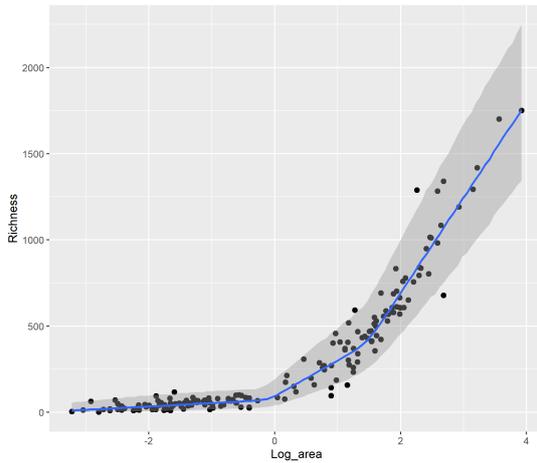
Com essa adaptação, os modelos SAR (1, 2 e 3) se ajustaram adequadamente aos dados, e não identificamos problemas nos gráficos de resíduos.

Comparado aos modelos do tipo linha interrompida, o modelo *curva angulada* possibilita ao pesquisador recuperar a distribuição de limiares de área dos habitats estudados. No entanto, como esses limiares podem ser interpretados? Considere que o arquipélago como um todo, ou até mesmo uma ilha isolada, pode ser percebido como uma combinação de habitats (Fig. 6). Na paisagem, existem locais que são mais ricos e diversos em recursos hídricos, nutrientes no solo e outras condições propícias para sustentar uma comunidade mais diversificada. Este trecho específico da paisagem responde de maneira consistente como se o limiar de área ocorresse anteriormente.

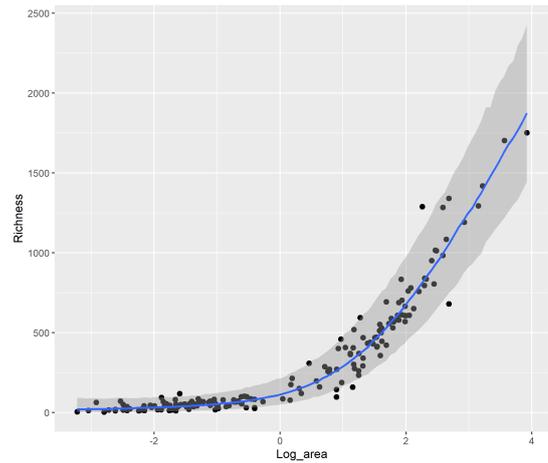
Por outro lado, existem regiões da paisagem com recursos menos diversos, ou até mesmo menos disponíveis. Nestas porções, é como se o ponto de transição da área ocorresse posteriormente. Dessa forma, cada pequena parcela da paisagem possui um ponto de transição de área subjacente, e a paisagem, como um todo, manifesta uma distribuição de pontos de transição. Assim, interpretamos o ponto de transição da área como uma variável latente e aleatória.

A distribuição dos limiares obtida pelo modelo Probabilístico *bent-cable* é apresentada abaixo:

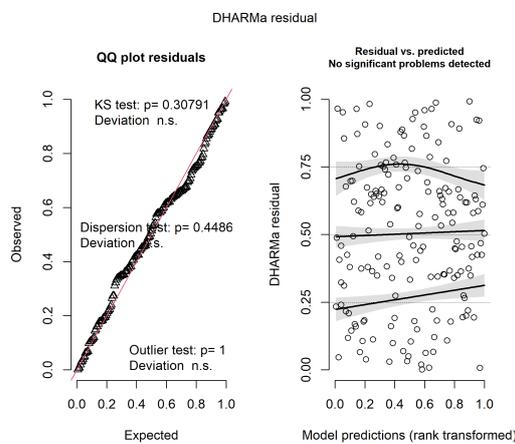
Com os valores amostrados, é possível inferir que o ponto de transição da área ocorre entre 0.94 e 2.05 (IC 95%) para a variável $\log_{10}(\text{área})$, resultando em 8.76 a 113.43, km²..



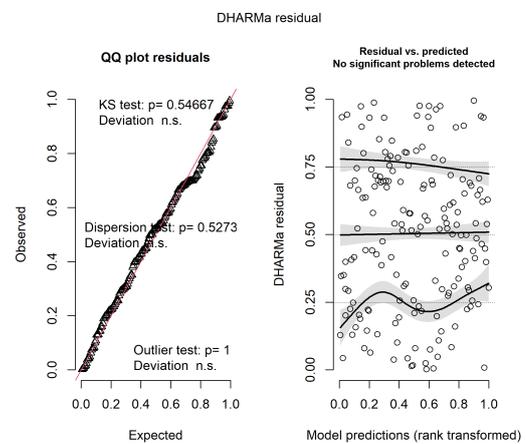
(a) Modelo 2: Broken-line com 2 limiares



(b) Modelo 3: L-BC (limiar aleatório)



(c) Diagnóstico pelo método DHARMA para o Modelo 2



(d) Diagnóstico pelo método DHARMA para o Modelo 3

Figura 5 – Ajuste dos Modelos 2 e 3 considerando a Binomial Negativa para a riqueza com influência do tamanho das ilhas (Log área) sobre o parâmetro de dispersão. Modelo 2: *broken-line* com dois limiares de área fixos; Modelo 3: *bent-cable* probabilístico Logístico (L-BC). Dados de Matthews et al. (2019).

Com os modelos *bent-cable* probabilísticos, torna-se possível calcular a fração dos habitats na paisagem (arquipélago, no caso) que responderiam a um aumento de área. A fração responsiva é definida como o percentual dos habitats que ultrapassaram seu limiar subjacente de área para o efeito das ilhas pequenas.

Por exemplo, para uma área de $1, \text{km}^2$, estimamos com 95% de credibilidade que a fração de habitats responsivos é de 8.8% a 19.0%; para $10, \text{km}^2$, estimamos de 22.1% a 55.3% de habitats responsivos; e para $100, \text{km}^2$, estimamos de 44.3% a 89.1% de habitats responsivos.

Tabela 2 – Critérios para a comparação dos modelos obtidos pelo método de validação cruzada LOO *Leave-One-Out*

Poisson			
Critério	Modelo 1	Modelo 2	Modelo 3
$elpd_{loo}$	-2510.55 (245.804)	-2097.045 (208.472)	-2118.578 (208.084)
p_{loo}	130.911 (22.32)	110.236 (16.842)	86.904 (14.058)
looic	5021.101 (491.607)	4194.089 (416.945)	4237.155 (416.169)
Bayes R^2	0.931 (0.001)	0.952 (0.001)	0.952 (0.001)

Binomial Negativa			
Critério	Modelo 1	Modelo 2	Modelo 3
$elpd_{loo}$	-959.416 (18.184)	-954.294 (17.903)	-955.815 (17.536)
p_{loo}	10.036 (3.589)	10.756 (3.569)	11.218 (3.825)
looic	1918.832 (36.367)	1908.588 (35.806)	1911.631 (35.071)
Bayes R^2	0.859 (0.029)	0.921 (0.035)	0.938 (0.017)

Binomial Negativa ¹ ($\log(\phi_i) = a_0 + a_1 x_i$)			
Critério	Modelo 1	Modelo 2	Modelo 3
$elpd_{loo}$	-4902.456 (343.06)	-919.275 (15.671)	-920.454 (15.431)
p_{loo}	3929.818 (350.395)	9.132 (1.538)	7.037 (1.121)
looic	9804.911 (686.12)	1838.55 (31.343)	1840.908 (30.862)
Bayes R^2	0.682 (0.394)	0.949 (0.004)	0.949 (0.004)

¹O procedimento de estimação não convergiu para o Modelo 1.

6 Conclusão

Este estudo ressaltou a importância de modelar a riqueza nas relações SAR com premissas adequadas, utilizando distribuições estatísticas discretas capazes de incorporar fenômenos como a super-dispersão, comum em dados de contagem. Observamos que, conforme sugerido por Matthews e Rigal (2021), o efeito de ilhas pequenas pode envolver dois limiares de área. No entanto, o modelo *bent-cable* probabilístico com um limiar de área aleatório também se mostrou consistente com os dados, além de ser menos complexo, sendo, portanto, a explicação mais simples. Além disso, a aplicação do modelo *bent-cable* probabilístico permitiu recuperar a distribuição subjacente ao limiar de área, possibilitando afirmações probabilísticas sobre a transição de estados, entre habitats não responsivos e habitats responsivos.

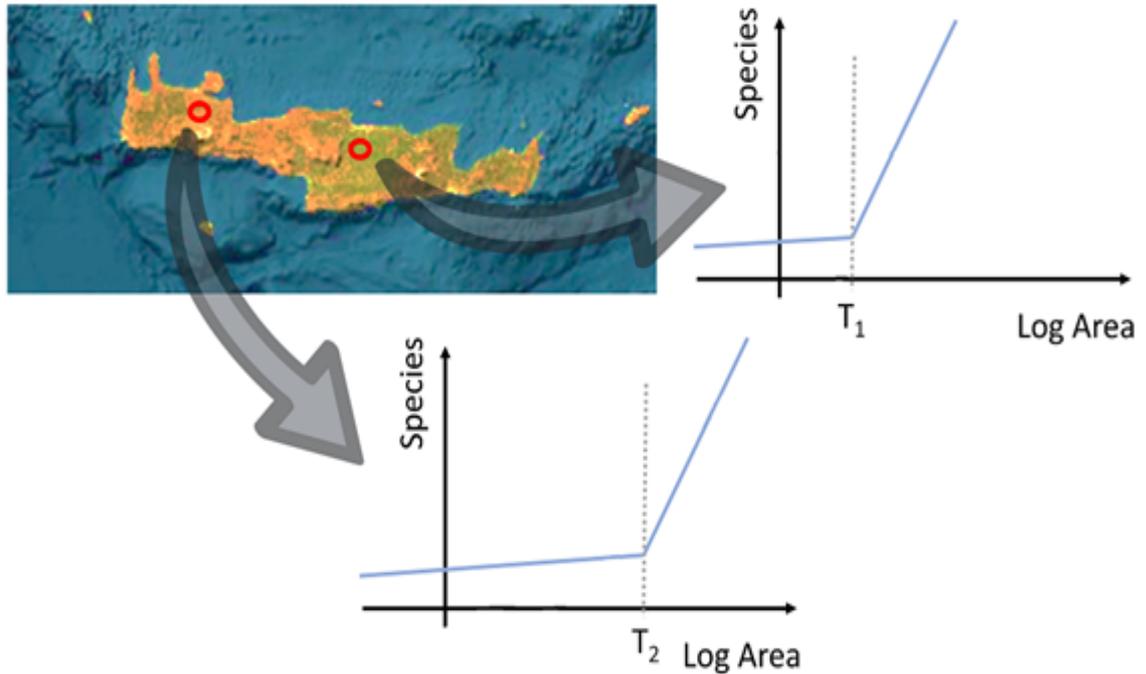
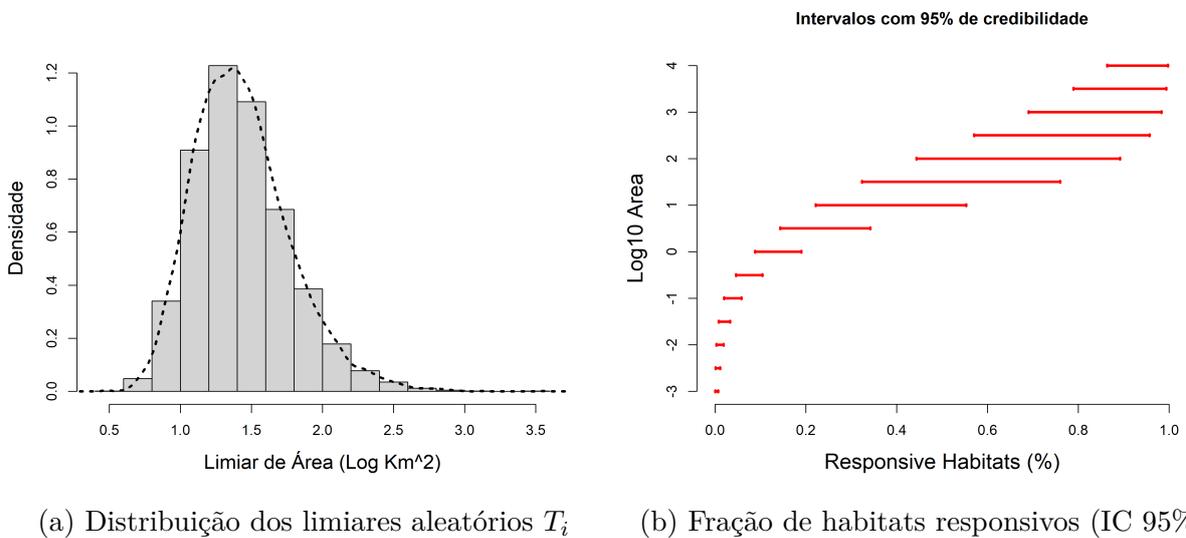


Figura 6 – Ilustração das ilhas sendo constituídas por uma mistura de habitats, com limiares de área aleatórios. Arquipélago grego com ilha de Creta ao centro. Imagem adaptada. Fonte: Google Maps (2022)



(a) Distribuição dos limiares aleatórios T_i (b) Fração de habitats responsivos (IC 95%)

Figura 7 – Distribuição dos limiares aleatórios recuperada pelo modelo *bent-cable* Logístico e intervalos com 95% credibilidade para a fração de unidades de habitats responsivas - que já ultrapassaram seu limiar intrínseco de área.

Referências

AKAIKE, H. Akaike's information criterion. In: _____. *International Encyclopedia of Statistical Science*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. p. 25–25. ISBN 978-3-642-04898-2. Disponível em: <https://doi.org/10.1007/978-3-642-04898-2_110>. Citado na página 20.

BORDIGNON, L. et al. Ilhas vegetacionais no pantanal matogrossense: um teste da teoria de biogeografia de ilhas. *Revista Brasileira de Biociências*, 2007. Citado 2 vezes nas páginas 21 e 22.

BÜRKNER, P.-C. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, v. 80, n. 1, p. 1–28, 2017. Citado na página 30.

CHIU, G.; LOCKHART, R.; ROUTLEDGE, R. Bent-cable regression theory and applications. *Journal of the American Statistical Association*, v. 101, n. n. 474, p. p. 542–553, 2006. Citado 4 vezes nas páginas 16, 24, 25 e 26.

FERREIRA, I.; ZOCCHI, S. A unified view of the smooth piecewise-linear models in the light of random thresholds. *ArXiv e-prints*, 2021. Citado 6 vezes nas páginas 15, 16, 20, 24, 25 e 26.

FERREIRA, I. E. de P. *Iuri Emmanuel de Paula Ferreira*. Tese (Doutorado) — Universidade de São Paulo, 2015. Citado na página 24.

FICETOLA, G. F.; DENOËL, M. Ecological thresholds: an assessment of methods to identify abrupt changes in species-habitat relationships. *Ecography*, v. 32, n. 6, p. 1075–1084, 2009. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0587.2009.05571.x>>. Citado na página 19.

GAO, D.; WANG, Y. A global synthesis of the small-island effect in amphibians and reptiles. *Ecography*, v. 2022, n. 1, 2022. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/ecog.05957>>. Citado 3 vezes nas páginas 19, 20 e 29.

GELMAN, A. et al. R-squared for bayesian regression models. *The American Statistician*, Taylor Francis, v. 73, n. 3, p. 307–309, 2019. Citado na página 30.

GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. *Statistical Science*, Institute of Mathematical Statistics, v. 7, n. 4, p. 457–472, 1992. ISSN 08834237. Disponível em: <<http://www.jstor.org/stable/2246093>>. Citado na página 30.

GRIFFITHS, D. A.; MILLER, A. J. Hyperbolic regression - a model based on two-phase piecewise linear regression with a smooth transition between regimens. *Communications in Statistics*, v. 2, n. n. 6, p. p. 561–569, 1973. Citado 3 vezes nas páginas 16, 25 e 26.

HOFFMAN, M. D.; GELMAN, A. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, v. 15, n. 47, p. 1593–1623, 2014. Disponível em: <<http://jmlr.org/papers/v15/hoffman14a.html>>. Citado na página 30.

- HUGGETT, A. J. The concept and utility of 'ecological thresholds' in biodiversity conservation. *Biological Conservation*, v. 124, n. 3, p. 301–310, 2005. ISSN 0006-3207. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0006320705000649>>. Citado na página 19.
- JIMENEZ-FERNANDEZ, V. M. et al. Transforming the canonical piecewise-linear model into a smooth-piecewise representation. *SpringerPlus*, Springer, n. 5, p. 1612, 2016. Citado 4 vezes nas páginas 16, 25, 26 e 28.
- KHAN, S. A.; KAR, S. C. Generalized bent-cable methodology for changepoint data: a bayesian approach. *Journal of Applied Statistics*, Taylor Francis, v. 45, n. 10, p. 1799–1812, 2018. Citado 3 vezes nas páginas 16, 25 e 26.
- LIU, Z.; LIU, X.; SHEN, T.-J. Applicability of common algorithms in species–area relationship model fitting. *Diversity*, v. 14, n. 3, 2022. ISSN 1424-2818. Disponível em: <<https://www.mdpi.com/1424-2818/14/3/212>>. Citado na página 27.
- LOMOLINO, M.; WEISER, M. D. Towards a more general species–area relationship: Diversity on all islands, great and small. *Journal of Biogeography*, 2001. Citado na página 22.
- MATTHEWS, T. et al. sars: an r package for fitting, evaluating and comparing species–area relationship models. *Ecography*, v. 42, p. 1446–1455, 2019. Citado 5 vezes nas páginas 15, 20, 28, 33 e 35.
- MATTHEWS, T. J.; RIGAL, F. Thresholds and the species-area relationship: a set of functions for fitting, evaluating and plotting a range of commonly used piecewise models in r. *Frontiers of Biogeography*, v. 13, n. 1, 2021. Disponível em: <<https://escholarship.org/uc/item/8x3151xr>>. Citado 6 vezes nas páginas 15, 20, 28, 29, 33 e 36.
- MATTHEWS, T. J. et al. Unravelling the small-island effect through phylogenetic community ecology. *Journal of Biogeography*, v. 47, n. 11, p. 2341–2352, 2020. Citado na página 28.
- MORIN, P. *Community Ecology*. [S.l.]: Blackwell Publishing, 2005. Citado na página 21.
- PEIXOTO, A.; LUZ, J. R.; BRITO, M. A. de. *Conhecendo Biodiversidade*. Ministério da Ciência, Tecnologia, Inovações e Comunicações - MCTIC, 2016. Disponível em: <https://ppbio.inpa.gov.br/sites/default/file/conhecendo_a_biodiversidade_livro.pdf>. Citado na página 21.
- PRIMACK, R. B. *Essentials of Conservation Biology*. [S.l.]: inauer Associates,, 2002. Citado na página 21.
- RAMIRES, T. G. et al. Exponentiated uniform distribution: An interesting alternative to truncated models. *Semina-ciências Agrárias*, v. 40, p. 107–114, 2019. Citado na página 25.
- RATKOOWSKY, D. *Handbook of nonlinear regression models*. [S.l.: s.n.], 1990. Citado na página 22.
- RIBEIRO, M. P.; MELLO, K. de; VALENTE, R. A. Avaliação da estrutura da paisagem visando à conservação da biodiversidade em paisagem urbanizada. *Ciência Florestal*, 2020. Citado na página 21.

- RICKLEFS, R. *A economia da natureza*. [S.l.]: Guanabara Koogan, 2003. Citado na página 22.
- RIGUEIRA, E. M.-N. E. D. M. G. Limiares. a perda abrupta de plantas na mata atlântica. *CIÊNCIAHOJE*, v. 54, n. 323, p. 16–19, 2015. ISSN 0006-3221. Citado na página 19.
- SEBER, G. A. F.; WILD, C. J. *Nonlinear Regression*. [S.l.]: John Wiley & Sons, Inc., 2003. 753 p. Citado 2 vezes nas páginas 15 e 26.
- SEBER, G. A. F.; WILD, C. J. *Nonlinear regression*. john wiley. *Wiley*, 2005. Citado na página 23.
- TOMS, J. D.; LESPERANCE, M. L. Piecewise regression: A tool for identifying ecological thresholds. *Ecology*, v. 84, n. 8, p. 2034–2041, 2003. Disponível em: <<https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/02-0472>>. Citado na página 20.
- TOWNSEND, C.; BEGON, M.; HARPER, J. *fundamentos em ecologia*. [S.l.]: Artmec, 2006. Citado na página 22.
- VEHTARI, A.; GELMAN, A.; GABRY, J. Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing*, v. 27, n. 5, p. 1413–1432, 2017. Citado na página 30.
- VIEIRA, M. et al. Mamíferos. In: _____. [S.l.: s.n.], 2003. p. 125–151. ISBN 87166484. Citado na página 22.
- WALKER, B.; MEYERS, J. A. Thresholds in ecological and social-ecological systems: a developing database. *Ecology and Society*, v. 9, n. 2, 2004. Citado na página 19.
- WATTS, D. G.; BACON, D. W. Using an hyperbola as a transition model to fit two-regime straight-line data. *Technometrics*, v. 16, n. n. 3, p. p. 369–373, 1974. Citado 2 vezes nas páginas 16 e 26.
- ZANG, I. A smoothing-out technique for min—max optimization. *Mathematical Programming*, v. 19, n. 1, p. 61–77, Dec 1980. ISSN 1436-4646. Citado 3 vezes nas páginas 16, 25 e 26.