Metodologia

## 0.1 Overview

A estimativa dos efeitos causais propostos ocorre por meio da interação de duas partes principais. A primeira parte é o modelo neutro espacialmente explícito, que, após ser empiricamente calibrado, funciona como um dispositivo de aproximação do processo gerador dos dados observados ([1]). A segunda parte envolve programas de pesquisa que escolhem contrafactuais, cujos contrastes em pares oferecem interpretações sobre como a perda de cobertura florestal influencia a biodiversidade remanescente ([2]; [3]).

Cada contraste de pares de contrafactuais permite interpretar o específico efeito proposto da paisagem na conectividade e na congruência com a evidência empírica. O contraste entre a calibração empírica dos processos demográficos, ajustado para simular a riqueza observada, foi interpretado como uma métrica funcional de conectividade relacionada à manutenção da riqueza local (logU/U). Já o contraste na congruência da distribuição de abundâncias das espécies da parcela (SAD) simulado com a SAD observada foi interpretado como métrica do efeito da paisagem na descrição da evidência empírica. Este sistema foi aplicado aos sítios de amostragem da base TreeCo ([4]), conforme critérios de seleção pré-estabelecidos.

Para cada sítio, exploramos o efeito escalar da dispersão em MNEE, simulando 20 graus de de dispersão, variando a proporção de propágulos que permanecem nas proximidades imediatas da árvore progenitora (parâmetro k). Em cada grau de limitação de dispersão, avaliamos a escala espacial mínima necessária para que a simulação não seja influenciada pela escala de observação da paisagem ao redor da parcela ([5]; [6]). Dessa forma, cada grau de limitação de dispersão foi simulado em sua respectiva extensão espacial suficiente.

Os resultados foram sumarizados utilizando modelos de regressão hierárquicos ([7]). Esses modelos de regressão descrevem os efeitos propostos da paisagem por sítio de amostragem e e o efeito médio comum entre os sítios, em função das métricas de conectividade funcional (logU/U) e de dispersão (k).

Nas próximas subseções da metodologia, apresentaremos: (i) a base de dados e os critérios de inclusão utilizados (evidência empírica); (ii) as justificativas para o uso do modelo neutro espacialmente explícito como dispositivo de mimese (validação teórica); (iii) a relação entre calibração e validação empírica; e (iv) a análise dos dados, incluindo a racionalidade do contraste entre contrafactuais. Na tabela 2 há o resumo dos procedimentos metodológicos necessários para obter os resultados finais.

| **Procedimento** | **Descrição** |
| --- | --- |
| **1. Coleta de dados** |  |
| 1.1 Seleção na base TreeCo | Seleção de dados de inventários florestais registrados na base TreeCo, abrangendo sítios amostrados ao longo de gradientes de riqueza e número de indivíduos. |
| 1.2 Seleção de mapas MapBiomas | Escolha de mapas da Coleção 6 do MapBiomas para representar a cobertura florestal contemporânea em cada sítio amostrado. |
| **2. Preparação para simulação** |  |
| 2.1 Adequação dos mapas | Conversão dos mapas de cobertura florestal em formato compatível com o modelo MNEE, considerando os requisitos para simulação da SAD. |
| 2.2 Estimativa dos coeficientes de dispersão | Cálculo dos coeficientes da função de dispersão para ajustar a proporção de propágulos atingindo a vizinhança imediata do progenitor para cada valor de k. |
| **3. Efeito escalar da dispersão no mecanismo simulado** |  |
| 3.1 Sorteio de sítios | Sorteio aleatório de sítios ao longo do gradiente de riqueza e abundância para análise do efeito escalar. |
| 3.2 Estimativa da taxa U | Cálculo da taxa U nas paisagens prístinas para 20 graus de limitação de dispersão (k) em extensões espaciais variando de 0,5 a 16 km de lado. |
| 3.3 Descrição estatística | Avaliação da relação entre U, k, e a extensão da paisagem em cada sítio, utilizando estatísticas descritivas. |
| 3.4 Definição da extensão suficiente | Identificação da extensão em que a taxa U estabiliza para cada grau de limitação de dispersão (k). |
| **4. Construção de paisagens hipotéticas** |  |
| 4.1 Réplicas dos mapas | Construção de paisagens adequadas para cada nível de dispersão simulado (segundo passo 3.4). |
| 4.2 Paisagens aglomerada e prístina | Construção das paisagens hipotéticas com toda a cobertura florestal agregada ou preservada conforme as condições prístinas. |
| **5. Simulação das SADs contrafactuais** |  |
| 5.1 Estimativa da taxa U | Estimativa da taxa U média por sítio e grau de limitação de dispersão (k). |
| 5.2 Geração de SADs | Simulação de 100 réplicas de SADs contrafactuais para cada combinação de sítio, k, e U. |
| **6. Contraste de contrafactuais** |  |
| 6.1 log(U/U) | Cálculo do logaritmo da razão entre as taxas U estimadas nos contrafactuais para cada sítio e k. |
| 6.2 log(OR) | Comparação entre SADs simuladas e observadas usando o teste Kolmogorov-Smirnov bootstrap. |
| **7. Análise estatística** |  |
| 7.1 Ajuste de modelos HGAM | Ajuste de 7 HGAMs com distribuições normais e método REML, variando nas preditoras e estrutura de suavização. |
| 7.2 Comparação de modelos | Comparação dos modelos via AICc, deviance explicada, e teste de autocorrelação espacial (Moran’s I) dos resíduos. |
| 7.3 Inferência | Baseada nos modelos mais plausíveis (>95% de peso de evidência), condicionada ao diagnóstico estatístico. |
| 7.4 Predição a posteriori | Geração de 1000 amostras das distribuições multivariadas dos coeficientes para predição, com resumo via quantis (5, 50 e 95%). Caso exista mais de um modelo plausível, então será feito a média de modelos baseado no peso de evidência |

## 0.2 Evidência empírica

Selecionamos, na base de dados TreeCo, inventários florestais que amostraram indivíduos arbóreos com diâmetro à altura do peito (DBH) ≥ 4,8 cm, em parcelas únicas de pelo menos 1 ha de floresta. Selecionamos inventários com alta confiabilidade taxonômica tanto no nível de morfoespécie quanto na precisão da coordenada central da parcela. A base de dados TreeCo reuniu amplamente estudos fitossociológicos realizados na Floresta Atlântica ([4]).

Para caracterizar a cobertura vegetal, utilizamos os mapas da coleção 6 do MapBiomas com resolução de 30x30 m², que compilam dados anuais de 1985 a 2020 ([**mapbiomas?**]). Selecionamos o mapa de cobertura florestal mais próximo ao ano da amostragem de cada parcela para assegurar o pressuposto de contemporaneidade entre a amostragem da parcela e a cobertura florestal mapeada.

Cada mapa de cobertura foi recortado em quadrados centrados na coordenada central da parcela, com uma extensão máxima de 16x16 km². Para garantir uma correspondência entre a densidade de pixels do mapa e a densidade de indivíduos amostrados, ajustamos a resolução dos mapas. Pixels foram classificados como unidades de habitat se o valor de cobertura fosse ≥ 0,7 (observamos que, após o ajuste de resolução, surgiram valores intermediários entre 0 e 1, originalmente ausentes nos mapas binários).

As parcelas foram delineadas como quadrados centralizados no mapa, alocando um total de N unidades de habitat, correspondendo ao número total de indivíduos na parcela. Quando não foi possível centralizar a parcela, ela foi deslocada ligeiramente para uma área adjacente que permitisse a demarcação de um quadrado adequado.

## 0.3 Validação Teórica

### 0.3.1 Teoria Neutra da Biodiversidade

O modelo neutro espacialmente explícito (MNEE) é parte de uma família de modelos fundamentados pela Teoria Neutra da Biodiversidade (TNB, [8]). Na TNB, modelos demográficos buscam reproduzir padrões de biodiversidade por meio de simulações que incorporam dinâmicas de morte, nascimento, dispersão e especiação de múltiplas espécies no espaço ([9]). No MNEE, a única interação entre os indivíduos é da competição neutra por espaço, o que resulta em flutuações estocásticas nos tamanhos populacionais das espécies, um processo conhecido como deriva ecológica ([9]).

A TNB se desenvolveu em contraste com teorias que enfocam o componente adaptativo da biodiversidade, as quais consideram diferenças funcionais entre as espécies ([10]). Na TNB, assume-se a equivalência funcional, onde todas as espécies compartilham taxas demográficas per capita idênticas ([9]; [10]). A pesquisa dentro deste arcabouço teórico opera em dois eixos principais: (1) explorar pressupostos auxiliares que sustentam a equivalência funcional; e (2) investigar formas de relaxar esse pressuposto para acomodar variações propostas nos sistemas naturais ([11]).

Uma das vantagens dos modelos neutros é sua capacidade de reproduzir padrões estáticos de biodiversidade local de maneira equivalente a modelos que incorporam diferenças entre espécies, mas com um número menor de parâmetros demográficos ([12]; [13]). Isso torna o MNEE particularmente útil como uma primeira aproximação em contextos de escassez de dados detalhados.

### 0.3.2 Modelo Neutro Espacialmente Explícito (MNEE)

Uma característica central que o MNEE reproduz é a geometria estocástica dos indivíduos, na qual a distribuição espacial é probabilística e limitada, resultando em forte agregação de coespecíficos e influenciada pela configuração espacial ([14]; [15], [16]). No MNEE, a dispersão é determinada tanto pela capacidade de dispersão dos propágulos quanto pela distância entre a fonte de propágulo e os habitats disponíveis ([17]). Ao replicar essa geometria estocástica de forma dinâmica, o MNEE assegura que a simulação expressa a conectividade entre a parcela amostrada e a paisagem circundante, refletindo a interação do espaço na competição entre indivíduos ([18], [19]).

O MNEE opera sob o pressuposto de equilíbrio dinâmico ([17]), alinhando-se ao pressuposto de contemporaneidade amplamente adotado na análise de dados observacionais pelos dois programas de pesquisa discutidos (ver Introdução). A situação de equilíbrio é atingida quando a taxa de extinção por deriva ecológica é compensada pela taxa de reposição de espécies (taxa U), que pode ocorrer via especiação, reposição do banco de propágulos, ou dispersão de longa distância de fora da paisagem ([20]; [9]; [21]).

A validade do MNEE foi testada em estudos de simulação que investigaram paisagens fragmentadas com diferentes níveis de perda de habitat e autocorrelação espacial ([18], [19], [22]). Esses estudos demonstraram que a configuração espacial da cobertura florestal remanescente exerce um papel significativo na conformação da distribuição de abundância de espécies locais (SAD) simulada pelo MNEE ([18], [19], [22]). Isso indica que o MNEE é sensível à estrutura espacial da paisagem, tornando-o uma ferramenta robusta para explorar a conectividade funcional em ecossistemas fragmentados.

### 0.3.3 Descrição da simulação de MNEE

O MNEE utiliza uma abordagem coalescente para modelar a dinâmica das comunidades ecológicas ([17]; [23]). A simulação inicia sem informação prévia sobre a identidade das espécies presentes na parcela amostral. Durante a simulação, a cada passo temporal, um indivíduo da parcela é sorteado aleatoriamente para ser substituído. A substituição pode ocorrer de duas maneiras: (i) por um singleton, ou seja, um indivíduo de uma nova espécie que surge na paisagem com probabilidade U; ou (ii) pela prole de um adulto residente na paisagem, com probabilidade 1−U ([17]; [23]). Nos casos em que a substituição ocorre por meio de um propágulo residente (1−U), o evento de dispersão é modelado através de um kernel de dispersão com distribuição exponencial espelhada, sendo i.i.d. (independentemente e identicamente distribuído) em duas dimensões ([24]; [23]). Para determinar a origem do propágulo, o modelo centraliza a função de dispersão no habitat vago e sorteia posições na paisagem até encontrar uma indivíduo adulto. O processo termina quando todos os indivíduos na parcela coalescem em um conjunto ancestral de indivíduos que são os primeiros de sua espécie na paisagem, criando uma árvore genealógica da comunidade local ([17]). Cada ramo dessa árvore é composto por indivíduos que compartilham um ancestral comum, representando a espécie que originalmente colonizou a paisagem. Isso permite a reconstrução da distribuição de abundância de espécies (SAD) a partir da árvore genealógica. A abordagem coalescente pressupõe que apenas um indivíduo pode ocupar uma unidade de habitat por vez, e que cada morte é imediatamente compensada pelo nascimento de um novo indivíduo adulto (pressuposto de soma zero, [17]).

A capacidade de dispersão é controlada variando-se a proporção de propágulos que permanecem próximos à planta progenitora (i.e., distância até o vizinho imediato). A taxa U, que representa a probabilidade de introdução de novas espécies por nascimento, é estimada usando um método analítico proposto por [17] (Apêndice A). Inicialmente, a simulação é executada com uma taxa U próxima de zero para simular uma árvore genealógica, e posteriormente utiliza-se uma equação para relacionar o número de espécies observadas com a taxa U média, permitindo uma calibração de MNEE com a riqueza de espécies observada ([17], Apêndice A).

## 0.4 Calibração e validação empírica

Nos estudos realizados pelos grupos de pesquisa que analisam dados observacionais, é comum utilizar a riqueza de espécies para ajustar e avaliar a congruência dos modelos estatísticos com os dados empíricos (e.g., [25]; [26]). Neste trabalho, utilizamos a riqueza de espécies para calibrar a taxa U do mecanismo demográfico representado no MNEE (ver ‘validação teórica’) e a Distribuição de Abundância de Espécies (Species Abundance Distribution, SAD) para validar empiricamente o resultado dessa simulação.

Uma vez estimada a taxa U necessária para simular a riqueza observada na paisagem contrafactual específica e grau de limitação de dispersão, a SAD é simulada sob o mesmo cenário contrafactual. A validação com a SAD observada é realizada por meio de um teste de hipótese, que classifica a forma da SAD simulada como congruente ou não congruente com a SAD empírica.

A taxa U pode ser ajustada diretamente pelo MNEE para refletir a riqueza de espécies observada na paisagem ([17]). Ao calibrar corretamente a riqueza nas paisagens contrafactuais, a SAD simulada é projetada para ter o mesmo número de indivíduos e uma quantidade aproximada de espécies em relação à SAD observada. Assim, quaisquer divergências na forma da SAD simulada podem ser atribuídas principalmente à variação na distribuição da abundância relativa das espécies, e não ao número total de espécies ou indivíduos. Isso é relevante porque, para um dado número de espécies e indivíduos, existem múltiplas formas possíveis para a SAD ([27]). Portanto, é esperado algum grau de variação natural entre as SAD simuladas e observadas devido à diversidade intrínseca de formas que a SAD pode assumir.

Esse design de simulação procura calibrar o mecanismo demográfico de baixo nível que interage diretamente com a configuração da paisagem, enquanto, a validação empírica é feita no nível do padrão emergente resultante desse mecanismo. Dessa forma, a informação usada na calibração e validação possuem graus de distinção, critério necessário para usar modelos baseados em indivíduos como representantes de mecanismos ([1]).

Para a validação empírica comparamos a congruência entre a SAD observada e simulada a partir do teste de hipótese Kolmogorov-Smirnov bootstrap ([28]). Esse teste tem pressupõe hipótese nula de que os dois vetores em comparação (SAD observada e SAD simulada) são amostras de uma mesma distribuição teórica ([28]). Contabilizamos o número de SAD simuladas que são congruentes com a SAD observada para cada sítio, paisagem contrafactual e nível de limitação de dispersão considerado.

## 0.5 Análise dos Dados

### 0.5.1 Lógica contrafactual para estimar os propostos efeitos da paisagem a partir do contraste de simulações de dinâmicas demográficas no espaço

O método proposto permite formular frases como: “A SAD simulada na paisagem contemporânea, em relação à SAD simulada na paisagem prístina, indica uma redução na conectividade local na ordem de **log⁡(U/U)** e apresenta uma melhora/piora na congruência com a SAD observada na ordem de **log⁡(OR)**, **de acordo com o pressuposto do MNEE e o grau de limitação de dispersão simulado**.” Esses termos em negrito facilitam a interpretação, atribuição e generalização dos efeitos estimados, seja para cada sítio de amostragem individual ou para um efeito médio comum. Essa estrutura de frase pode ser utilizada para descrever o efeito de fragmentação total, conforme pressuposto de interdependência (tabela 1 e box 1, introdução). Para os efeitos da paisagem sob o pressuposto de independência, temos uma frase por efeito, cada uma comparando pares de SADs contrafactuais distintos: fragmentação per se, comparando as paisagens contemporânea e aglomerada; e área per se, comparando as paisagens aglomerada e prístina (ver tabela 1, introdução).

Esses efeitos são estimados para cada sítio de amostragem da base TreeCo dentro do critério de inclusão. Por sítio de amostragem e paisagem contrafactual, foram simulados 20 graus de limitação de dispersão. A proporção de propágulos que permanecem na vizinhança imediata da planta progenitora (k) variou de 0,99 (limitação de dispersão severa) a 0,05 (limitação de dispersão branda). Para cada conjunto de simulações (combinando sítio de amostragem, paisagem contrafactual e k), estimamos 10 réplicas da taxa U necessária para manter a riqueza observada. A média dessas estimativas de U foi utilizada para parametrizar 100 simulações adicionais, as quais foram usadas para simular a Distribuição de Abundância de Espécies (SAD).

Nas próximas seções, detalhamos a interpretação do contraste na taxa U (log⁡(U/U)) como uma métrica funcional de conectividade local (0.5.2). Em seguida, descrevemos como atribuimos os efeitos propostos na congruência com a evidência empírica, ou seja, com a SAD observada (0.5.3). Por fim, argumentamos sobre a generalização dessa estimativa, ponderando o modelo mecanístico empregado e a escala do mecanismo, que é determinada pela capacidade de dispersão (0.5.4).

### 0.5.2 Interpretação

Nós interpretamos log⁡(U/U) como uma métrica que reflete a proposta influência da paisagem na riqueza local de espécies. A influência da paisagem na abundância relativa das espécies locais está, em algum grau, associada à sua influência na riqueza local ([6], [5]), a qual é descrita pela taxa U. O contraste das taxas U estimadas entre diferentes pares de paisagens contrafactuais permite compreender como a diferença entre essas paisagens afeta a riqueza local. Para um dado sítio e grau de limitação de dispersão (k), quanto menor for a taxa U necessária para obter a riqueza observada, maior será a reposição simulada de espécies por meio da imigração de propágulos vindos da paisagem circundante para a parcela ([29]). Assim, o logaritmo da razão entre as taxas U ( log⁡(U/U) ) indica em qual paisagem contrafactual a dependência da imigração de espécies da paisagem circundante é maior.

Quando log⁡(U/U) é igual a zero, enão não há diferença entre os contrafactuais em relação à reposição de espécies locais pela paisagem. Se o valor for positivo, significa que a paisagem contrafactual no numerador tem uma taxa U mais elevada, indicando uma maior demanda de reposição de espécies locais pela paisagem. Por outro lado, se o valor for negativo, a paisagem no numerador apresenta uma menor demanda de reposição de espécies locais, sugerindo uma maior autossuficiência em termos de riqueza local.

### 0.5.3 Atribuição

A atribuição dos efeitos da paisagem é realizada comparativamente, analisando o contraste entre paisagens contrafactuais em termos de variação na congruência com a distribuição de abundância de espécies (SAD) observada. O processo de avaliação da congruência entre a SAD factual e as SADs simuladas contrafactuais foi realizado em etapas. Primeiro, foi calculada a proporção de SADs simuladas que apresentam boa congruência com a SAD observada para cada sítio de amostragem, grau de limitação de dispersão, e paisagem contrafactual. Em seguida, foi calculado o logaritmo da razão de chances (log odds ratio, log⁡OR) para os contrastes de paisagens contrafactuais propostos (ver Tabela 1 e box 1), agregados por sítio de amostragem e grau de limitação de dispersão. Quando a congruência da SAD simulada é igual para ambos os contrafactuais, o valor de log⁡OR é zero. Se a paisagem contrafactual no numerador apresenta maior proporção de SADs congruentes, então log⁡OR será maior que zero; caso contrário, será negativo.

O design de simulação empregado para estimar esses efeitos de contraste entre paisagens na congruência com a SAD observada introduz uma estrutura hierárquica aos dados, pois simula múltiplos cenários de dispersão para cada combinação de contrafactual e sítio de amostragem. Em situações onde os dados possuem uma estrutura hierárquica, é possível estimar efeitos causais não enviesados ([30]; [31]). A lógica subjacente é que, se o sítio de amostragem é mantido constante, então todas as possíveis covariáveis fixas são controladas implicitamente ([32]; [33]; [31]). Nesse contexto, é possível modelar a variabilidade em log⁡OR usando um modelo hierárquico de regressão, interpretando a predição deste modelo como uma síntese do efeito causal empírico representado por log⁡OR ( [30] ;[34] ). O objetivo é compreender essa variabilidade em termos de princípios organizadores e métricas interpretáveis.

#### 0.5.3.1 Descrição estatística dos efeitos da paisagem na congruência da simulação com o padrão empírico

Os efeitos causais médios da paisagem foram descritos com modelos hierárquicos aditivos generalizados (HGAM; [7]), com log⁡OR como variável resposta em função do logaritmo da razão das taxas U (log⁡(U/U) e da proporção de propágulos que permanecem na vizinhança imediata (k). Com até esse conjunto de preditores em interação, esperamos compreender a variabilidade dos efeitos causais estimados, separando a variabilidade associada ao sítio de amostragem daquela que é comum entre todos os sítios ([7]). Os HGAMs foram ajustados usando funções do pacote mgcv ([35]) com distribuição gaussiana (detalhes em Material Suplementar). Nós incluímos como preditora a capacidade de dispersão simulada, além da métrica funcional de influência da paisagem na riqueza, para representar qualquer efeito da paisagem na abundância relativa não contemplado pelo efeito da métrica funcional.

Para cada efeito causal médio da paisagem (área per se, fragmentação per se e fragmentação total), comparamos, via AICc ([36]), 7 modelos hierarquicos aditivos generalizados (tabela X): i) com um tensor entre k e logU/U comum e com um tensor dessas duas variáveis por sítio de amostragem (com o tensor por sítio compartilhando o parâmetro de penalizado com o tensor comum); ii) com um tensor entre k e logU/U comum e apenas um intercepto por sítio; iii) 1 spline comum para logU/U e com um spline por sítio (compartilhando a penalização); iv) 1 spline para logU/U e com apenas 1 intercepto por sítio; v) 1 spline comum para k e por sítio (compartilhando a penalização); vi) 1 spline comum para k e apenas 1 intercepto por sítio; vii) 1 intercepto comum e por sítio.

| HGAM | Efeito | Preditoras | Efeitos por Sítio | Descrição |
| --- | --- | --- | --- | --- |
| 1 | Tensor | k, logU/U | Tensor por sítio (mesma penalização) | Tensor compartilhado para k e logU/U entre os sítios, com um tensor adicional para cada sítio (mesma penalização). |
| 2 | Tensor | k, logU/U | Intercepto por sítio | Tensor compartilhado para k e logU/U entre os sítios, com um intercepto individual por sítio. |
| 3 | Spline | logU/U | Spline por sítio (mesma penalização) | Spline compartilhado para logU/U entre os sítios, com um spline adicional para cada sítio (mesma penalização). |
| 4 | Spline | logU/U | Intercepto por sítio | Spline compartilhado para logU/U entre os sítios, com um intercepto individual por sítio. |
| 5 | Spline | k | Spline por sítio (mesma penalização) | Spline compartilhado para k entre os sítios, com um spline adicional para cada sítio (mesma penalização). |
| 6 | Spline | k | Intercepto por sítio | Spline compartilhado para k entre os sítios, com um intercepto individual por sítio. |
| 7 | Intercepto | Nenhuma | Intercepto por sítio | Modelo simples de intercepto com intercepto comum e específico por sítio. |
| HGAM: sigla em inglês para Hierarchical Generalized Additive Models como em Pedersen et al. (2019). | | | | |

### 0.5.4 Generalização

Um ponto de discordância entre os programas de pesquisa é a escala espacial adequada para analisar a relação entre a perda de cobertura florestal e a biodiversidade remanescente (ver Introdução). Em nossa metodologia, consideramos o efeito escalar da dispersão no mecanismo demográfico simulado e então buscamos determinar a extensão espacial da análise de acordo. Na construção da paisagem contrafactual aglomerada, onde todo o habitat remanescente é aglomerado ao redor da parcela amostrada, a extensão espacial da paisagem determina a quantidade de habitat que será agregado ([3]). O processo gerador da SAD factual não é conhecido, e, portanto, não se sabe quais mecanismos estão envolvidos nem as escalas desses mecanismos ([5]; [6]). Para incluir essa incerteza, simulamos diversos graus de limitação de dispersão, uma vez que a capacidade de dispersão aumenta a área necessária para simular a árvore genealógica da comunidade ((0.3.3)). Nós buscamos ajustar a extensão espacial da paisagem para que a estimativa da taxa U não sofra interferência do recorte do mapa de cobertura florestal. Uma expectativa é que a redução da limitação de dispersão aumente a extensão espacial suficiente para simular a dinâmica coalescente.

Para avaliar a escala espacial adequada investigamos em paisagens prístinas o comportamento da taxa U sob o aumento da extensão espacial da paisagem ao redor. Quando o Modelo Neutro Espacialmente Explícito (MNEE) simula a dinâmica coalescente em paisagens infinitas e sem perda de habitat, a taxa U necessária para manter a riqueza na parcela apresenta um máximo global quando o grau de limitação de dispersão é moderado ([29]). Para graus severos de limitação à dispersão, a taxa U estimada em MNEE é baixa, pois a perda de espécies é pequena: a maior parte das substituições ocorre entre coespecíficos ([29]). Com o relaxamento da limitação de dispersão, a perda de espécies aumenta, pois a substituição de coespecíficos diminui, e a perda de espécies por deriva ecológica se torna mais expressiva ([29]). Com o aumento da capacidade de dispersão além de graus moderados, há maior contribuição de subpopulações mais distantes da parcela, o que reduz novamente a estimativa da taxa U, porém à custa de uma maior homogenização na escala da paisagem ([37]).

Investigamos como a variação na escala espacial da paisagem interfere na estimativa da taxa U para cada grau de limitação de dispersão simulado. Para isso, sorteamos 30 sítios entre os selecionados e simulamos a taxa U na paisagem prístina, variando a extensão espacial da paisagem entre 0,5 km e 16 km de lado para os 20 graus de limitação de dispersão (k). Descrevemos a variação na estimativa de U em função de k e da extensão espacial da paisagem. Espera-se que o aumento da extensão espacial da paisagem reduza a estimativa da taxa U cada vez menos até que incrementos na extensão espacial não resultem em redução na taxa U. Assim, a extensão espacial considerada “suficiente” foi definida como aquela que acumulou 95% de toda a variação observada entre as extensões de 0.5 km e 16 km (detalhes no Apêndice 1: Efeito Escalar).

Para descrever a variação na taxa U associada à variação na extensão espacial para um determinado grau de limitação de dispersão, exploramos um modelo gaussiano em que a taxa U média pode ser descrita em função do número de espécies e indivíduos (na escala log) e do grau de limitação de dispersão e da extensão espacial (como variáveis categóricas) com um intercepto aleatório por sítio de amostragem (detalhes no apêndice ‘Efeito de Escalar’). Avaliamos, por grau de limitação de dispersão, como a média variou em função do aumento da extensão espacial da paisagem. A escala que acumulou a maior parte da variação total (de aumentar o lado da paisagem local de 0.5 km até 16 km) foi considerada como escala adequada para aquele grau de limitação de dispersão. Essa abordagem pressupõe que o efeito de escalar é monotônico.

# Referências

1. Agent-based models and causal inference. Agent‐based models and causal inference. John Wiley & Sons, Ltd; 2022. pp. 49–68. doi:<https://doi.org/10.1002/9781119704492.ch4>

2. Bull JW, Strange N, Smith RJ, Gordon A. Reconciling multiple counterfactuals when evaluating biodiversity conservation impact in social-ecological systems. Conservation Biology. 2021;35: 510–521. doi:<https://doi.org/10.1111/cobi.13570>

3. Valente JJ, Gannon DG, Hightower J, Kim H, Leimberger KG, Macedo R, et al. Toward conciliation in the habitat fragmentation and biodiversity debate. Landscape Ecology. 2023;38: 2717–2730.

4. Lima RA de, Mori DP, Pitta G, Melito MO, Bello C, Magnago LF, et al. How much do we know about the endangered Atlantic Forest? Reviewing nearly 70 years of information on tree community surveys. Biodiversity and Conservation. 2015;24: 2135–2148.

5. Fletcher Jr RJ, Betts MG, Damschen EI, Hefley TJ, Hightower J, Smith TA, et al. Addressing the problem of scale that emerges with habitat fragmentation. Global Ecology and Biogeography. 2023;32: 828–841.

6. Jackson HB, Fahrig L. Are ecologists conducting research at the optimal scale? Global Ecology and Biogeography. 2015;24: 52–63.

7. Pedersen EJ, Miller DL, Simpson GL, Ross N. Hierarchical generalized additive models in ecology: An introduction with mgcv. PeerJ. 2019;7: e6876. doi:[10.7717/peerj.6876](https://doi.org/10.7717/peerj.6876)

8. Hubbell SP. The unified neutral theory of biodiversity and biogeography (MPB-32). Princeton University Press; 2011.

9. Azaele S, Suweis S, Grilli J, Volkov I, Banavar JR, Maritan A. Statistical mechanics of ecological systems: Neutral theory and beyond. Reviews of Modern Physics. 2016;88: 035003.

10. Leroi AM, Lambert B, Rosindell J, Zhang X, Kokkoris GD. Neutral syndrome. Nature human behaviour. 2020;4: 780–790.

11. ALONSO D, ETIENNE R, MCKANE A. The merits of neutral theory. Trends in Ecology &amp; Evolution. 2006;21: 451–457. doi:[10.1016/j.tree.2006.03.019](https://doi.org/10.1016/j.tree.2006.03.019)

12. Chisholm RA, Pacala SW. Niche and neutral models predict asymptotically equivalent species abundance distributions in high-diversity ecological communities. Proceedings of the National Academy of Sciences. 2010;107: 15821–15825. doi:[10.1073/pnas.1009387107](https://doi.org/10.1073/pnas.1009387107)

13. Hammal OA, Alonso D, Etienne RS, Cornell SJ. When Can Species Abundance Data Reveal Non-neutrality? PLOS Computational Biology. 2015;11: e1004134. doi:[10.1371/journal.pcbi.1004134](https://doi.org/10.1371/journal.pcbi.1004134)

14. McGill BJ. Towards a unification of unified theories of biodiversity. Ecology Letters. 2010;13: 627–642. doi:[10.1111/j.1461-0248.2010.01449.x](https://doi.org/10.1111/j.1461-0248.2010.01449.x)

15. May F, Huth A, Wiegand T. Moving beyond abundance distributions: Neutral theory and spatial patterns in a tropical forest. Proceedings of the Royal Society B: Biological Sciences. 2015;282: 20141657. doi:[10.1098/rspb.2014.1657](https://doi.org/10.1098/rspb.2014.1657)

16. May F, Wiegand T, Lehmann S, Huth A. Do abundance distributions and species aggregation correctly predict macroecological biodiversity patterns in tropical forests? Global Ecology and Biogeography. 2016;25: 575–585. doi:[10.1111/geb.12438](https://doi.org/10.1111/geb.12438)

17. Rosindell J, Wong Y, Etienne RS. A coalescence approach to spatial neutral ecology. Ecological Informatics. 2008;3: 259–271.

18. Campos PR, Neto ED, Oliveira VM de, Gomes M. Neutral communities in fragmented landscapes. Oikos. 2012;121: 1737–1748.

19. Campos PR, Rosas A, Oliveira VM de, Gomes MA. Effect of landscape structure on species diversity. PloS one. 2013;8: e66495.

20. Condit R, Chisholm RA, Hubbell SP. Thirty years of forest census at Barro Colorado and the importance of immigration in maintaining diversity. PloS one. 2012;7: e49826.

21. Munoz F, Huneman P. From the neutral theory to a comprehensive and multiscale theory of ecological equivalence. The Quarterly Review of Biology. 2016;91: 321–342.

22. Claudino ES, Gomes MAF, Campos PRA. Extinction debt and the role of static and dynamical fragmentation on biodiversity. Ecological Complexity. 2015;21: 150–155. doi:[10.1016/j.ecocom.2014.11.011](https://doi.org/10.1016/j.ecocom.2014.11.011)

23. Thompson SE, Chisholm RA, Rosindell J. Pycoalescence and rcoalescence: Packages for simulating spatially explicit neutral models of biodiversity. Methods in Ecology and Evolution. 2020;11: 1237–1246.

24. Bullock JM, Mallada González L, Tamme R, Götzenberger L, White SM, Pärtel M, et al. A synthesis of empirical plant dispersal kernels. Journal of Ecology. 2017;105: 6–19.

25. Watling JI, Arroyo-Rodríguez V, Pfeifer M, Baeten L, Banks-Leite C, Cisneros LM, et al. Support for the habitat amount hypothesis from a global synthesis of species density studies. Ecology letters. 2020;23: 674–681.

26. Püttker T, Crouzeilles R, Almeida-Gomes M, Schmoeller M, Maurenza D, Alves-Pinto H, et al. Indirect effects of habitat loss via habitat fragmentation: A cross-taxa analysis of forest-dependent species. Biological Conservation. 2020;241: 108368.

27. Locey KJ, White EP. How species richness and total abundance constrain the distribution of abundance. Ecology Letters. 2013;16: 1177–1185. doi:[10.1111/ele.12154](https://doi.org/10.1111/ele.12154)

28. Dowd C. Twosamples: Fast Permutation Based Two Sample Tests. 2023. Available: <https://CRAN.R-project.org/package=twosamples>

29. May F, Giladi I, Ziv Y, Jeltsch F. Dispersal and diversity–unifying scale-dependent relationships within the neutral theory. Oikos. 2012;121: 942–951.

30. Feller A, Gelman A. Hierarchical models for causal effects. Emerging Trends in the Social and Behavioral Sciences. 2015; 1–16. doi:[10.1002/9781118900772.etrds0160](https://doi.org/10.1002/9781118900772.etrds0160)

31. Weinstein EN, Blei DM. Hierarchical causal models. arXiv; 2024. doi:[10.48550/ARXIV.2401.05330](https://doi.org/10.48550/ARXIV.2401.05330)

32. Marshall BDL, Galea S. Formalizing the role of agent-based modeling in causal inference and epidemiology. American Journal of Epidemiology. 2014;181: 92–99. doi:[10.1093/aje/kwu274](https://doi.org/10.1093/aje/kwu274)

33. Arnold KF, Harrison WJ, Heppenstall AJ, Gilthorpe MS. DAG-informed regression modelling, agent-based modelling and microsimulation modelling: A critical comparison of methods for causal inference. International Journal of Epidemiology. 2018;48: 243–253. doi:[10.1093/ije/dyy260](https://doi.org/10.1093/ije/dyy260)

34. Grace JB. An integrative paradigm for building causal knowledge. Ecological Monographs. 2024;94: e1628. doi:<https://doi.org/10.1002/ecm.1628>

35. Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society (B). 2011;73: 3–36.

36. Burnham KP, Anderson DR, Burnham KP, Anderson DR. Practical use of the information-theoretic approach. Springer; 1998.

37. Mouquet N, Loreau M. Community patterns in source‐sink metacommunities. The American Naturalist. 2003;162: 544–557. doi:[10.1086/378857](https://doi.org/10.1086/378857)