Metodologia

# 1 Fonte de Dados

## 1.1 TreeCo

Consultamos a base de dados TreeCo, que reuniu amplamente estudos fitossociológicos realizados na Floresta Atlântica (Lima et al. (2015)). Selecionamos nessa base de dados 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 altas confiabilidades taxonômica, no nível de morfoespécie, e na precisão da coordenada central da parcela. Detalhes do procedimento de avaliação dos inventários pode ser obtido em Lima et al. (2020). Em resumo, a base TreeCo classificou os inventários florestais quanto à confiabilidade da identificação das morfoespécies e das coordenadas geográficas centrais. A resolução taxonômica foi aprimorada a partir da verificação de espécimes de referência fornecidos por cerca de um terço dos inventários, com base em atualizações de identificação disponíveis na rede speciesLink. Além disso, os nomes das espécies foram padronizados conforme um catálogo taxonômico nacional de referência, corrigindo erros tipográficos, sinônimos e variantes ortográficas (Lima et al. (2020)). Morfoespécies marcadas como confer foram atribuídas à espécie sugerida, enquanto aquelas marcadas como affinis foram consideradas no nível de gênero. Quanto à localização, as coordenadas centrais de cada inventário foram verificadas para garantir correspondência com o fragmento florestal descrito. Inventários com localização inconsistente foram descartados. Quando necessário, as coordenadas foram ajustadas com base em mapas, descrições dos locais nos estudos originais ou outras fontes disponíveis online.

## 1.2 Mapbiomas

Para caracterizar a cobertura vegetal em torno da área de cada inventário, utilizamos os mapas da coleção 6 do MapBiomas com resolução de 30x30 m², que compilam dados anuais de 1985 a 2020 (Souza Jr et al. (2020)). 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 de inventário, com uma extensão máxima de 16x16 km². Nos ajustamos a resolução dos mapas para garantir correspondência entre a densidade de pixels do mapa e a densidade de indivíduos na parcela. Pixels foram classificados como unidades de habitat se o valor de cobertura fosse ≥ 0,7 (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 (posições para indivíduos), correspondendo ao número total de indivíduos na parcela. Quando não foi possível centralizar a parcela por conta da particular configuração espacial da cobertura florestal remanescente, ela foi ligeiramente deslocada para uma área adjacente que permitisse a demarcação de um quadrado adequado.

# 2 Paisagens Hipotéticas

As paisagens hipotéticas têm forma quadrada, foram criadas com a mesma densidade de pixels, área da parcela amostrada e dimensões espaciais.

Denominaremos paisagem fragmentada o mapa de cobertura florestal no ano mais próximo ao da amostragem do inventário florestal selecionado. Em alguns casos, os inventários podem informar o ano da amostragem ou apenas o ano de publicação do estudo (Lima et al. (2015)). Nesses casos, a data do mapa é estimada com base na informação disponível na base TreeCo.

Denominaremos paisagem prístina um mapa gerado convertendo todas as unidades de não-habitat da paisagem fragmentada em unidades de habitat. Assim, ela mantém as mesmas características da paisagem fragmentada, exceto pela perda de habitat e sua configuração espacial. Em outras palavras, as paisagens pristinas são mapas com 100% de cobertura florestal.

Chamaremos de paisagem aglomerada a obtida pela quantificação do número de unidades de habitat na paisagem fragmentada. Então é calculado a menor paisagem prístina que contém o número de unidades de habitat da paisagem fragmentada. Essa paisagem prístina é inserida no centro de uma paisagem sem cobertura florestal de mesma dimensão espacial que a paisagem fragmentada. Dessa forma, a paisagem aglomerada possui a mesma proporção de cobertura florestal da paisagem fragmentada e a mesma configuração espacial da paisagem prístina.

# 3 MNEE

## 3.1 Caracterização de MNEE

O modelo neutro espacialmente explícito (MNEE) é parte de uma família de modelos fundamentados pela Teoria Neutra da Biodiversidade (TNB, Hubbell (2011)). Na TNB, modelos demográficos buscam reproduzir padrões de biodiversidade pela representação das dinâmicas de morte, nascimento, dispersão e especiação de múltiplas espécies no espaço (Azaele et al. (2016)). 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 (Azaele et al. (2016)).

A TNB se desenvolveu em contraste com teorias que enfocam o componente adaptativo da biodiversidade, as quais consideram diferenças funcionais entre as espécies (Leroi et al. (2020)). Na TNB, assume-se a equivalência funcional, onde todas as espécies compartilham taxas demográficas per capita idênticas (Azaele et al. (2016); Leroi et al. (2020)). 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 (ALONSO et al. (2006)).

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 (Chisholm and Pacala (2010); Hammal et al. (2015)). Essa simplicidade torna o MNEE uma ferramenta eficiente para explorar padrões de biodiversidade em cenários com dados limitados.

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 (McGill (2010); May et al. (2015), May et al. (2016)). 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 (Rosindell et al. (2008)). 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 (Campos et al. (2012), Campos et al. (2013)).

O MNEE opera sob o pressuposto de equilíbrio dinâmico (Rosindell et al. (2008)), 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 (Condit et al. (2012); Azaele et al. (2016); Munoz and Huneman (2016)).

Estudos de simulação testaram a validade do MNEE em cenários de paisagens fragmentadas, variando os níveis de perda de habitat e autocorrelação espacial (Campos et al. (2012), Campos et al. (2013), Claudino et al. (2015)). 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 (Campos et al. (2012), Campos et al. (2013), Claudino et al. (2015)).

## 3.2 Descrição da simulação de MNEE

No MNEE, a dinâmica das comunidades ecológicas é modelada por meio de uma abordagem coalescente, que rastreia a ancestralidade das espécies até um conjunto comum de indivíduos fundadores na paisagem. O MNEE utiliza uma abordagem coalescente para modelar a dinâmica das comunidades ecológicas (Rosindell et al. (2008); Thompson et al. (2020)). 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 (Rosindell et al. (2008); Thompson et al. (2020)). 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 (Bullock et al. (2017); Thompson et al. (2020)). Para determinar a origem do propágulo, o modelo centraliza a função de dispersão no habitat vago e sorteia posições na paisagem até encontrar um 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 (Rosindell et al. (2008)). 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, Rosindell et al. (2008)).

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 Rosindell et al. (2008) (Apêndice A). Essa abordagem combina a simulação da árvore genealógica da comunidade sob a hipótese da taxa U tendendo a zero e uma equação construída a partir dessa simulação para relacionar a riqueza de espécies observadas com a taxa U média. Dessa forma, é possível calibrar de MNEE com a riqueza de espécies observada (Rosindell et al. (2008), Apêndice A).

## 3.3 parametrização do modelo: paisagem hipotética, k e U

Para realizar as simulações da SAD, o MNEE requer informações sobre a paisagem hipotética (fragmentada, aglomerada ou prístina), o parâmetro k (proporção de propágulos que permanecem na vizinhança imediata conforme uma distribuição exponencial espelhada) e a taxa U (entrada de novas espécies no sistema).

A proporção de propágulos que permanecem na vizinhança imediata (k) foi estimada por simulação, considerando a distância entre vizinhos imediatos. O objetivo foi determinar o parâmetro de escala (b) da distribuição de Laplace em duas dimensões, de forma que k propágulos fossem alocados entre a origem e o centro do vizinho imediato. Como a densidade de indivíduos pode variar entre inventários e paisagens, o valor de b foi calculado individualmente para cada inventário florestal e cenário de limitação de dispersão. Foram simulados 20 cenários de limitação de dispersão, com k variando de 99% (limitação severa) a 5% (limitação branda). Para cada combinação de sítio de amostragem, tipo de paisagem hipotética e cenário de limitação de dispersão, foi estimada a taxa U necessária para simular a riqueza observada. A estimativa da taxa média U foi obtida a partir de 20 réplicas para cada bateria de simulação.

# 4 Estimativa da extensão espacial adequada

Em nossa metodologia consideramos o efeito escalar da dispersão no mecanismo demográfico simulado e então buscamos determinar a extensão espacial da análise de acordo. Como a capacidade de dispersão do mecanismo empírico que gera a SAD observada é desconhecida, simulamos diversos graus de limitação de dispersão para avaliar a sensibilidade dos resultados em relação à dispersão e sua consequente influência na escala espacial da paisagem. Para isso investigamos a estimativa da taxa U em paisagens prístinas, onde, para um particular sítio de amostragem, a única influência na estimativa da taxa U é a capacidade de dispersão e a extensão espacial da paisagem. Nós buscamos estimar a extensão espacial da paisagem que seja suficiente para estimar a taxa U sem que a simulação seja afetada pelas dimensões da paisagem. Uma expectativa é que a redução da limitação de dispersão aumente a extensão espacial mínima necessária para simular a dinâmica coalescente.

Quando o 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 (May et al. (2012)). Esse comportamento reflete a interação entre dispersão e deriva ecológica. Para graus severos de limitação à dispersão, a taxa U estimada é baixa, pois a perda de espécies é pequena: a maior parte das substituições ocorre entre coespecíficos (May et al. (2012)). 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 (May et al. (2012)). Com o aumento da capacidade de dispersão para 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 homogeneização na escala da paisagem (Mouquet and Loreau (2003)).

Em nosso estudo da sensibilidade da extensão espacial em relação à capacidade de dispersão simulada, foram sorteados 36 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. Esperamos 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 acumula 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. 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.

# 5 Estimativa da métrica funcional de conectividade

## 5.1 procedimento para obter logU/U

A métrica funcional de conectividade foi definida na introdução como o log da razão da taxa U de um par de paisagens hipotéticas. A taxa U média na bateria de simulação corresponde a uma particular configuração de sítio de amostragem, grau de limitação de dispersão simulado e paisagem hipotética. A razão entre taxas U (U/U) pode variar entre 0 e mais infinito, tornando a comparação assimétrica, Na escala log essa razão varia entre menos e mais infinito, com o valor zero correspondendo a equivalência de taxas U.

# 6 Congruência entre a SAD simulada e a SAD observada

## 6.1 procedimento para obter logOR

A congruência entre as SAD simuladas e a SAD observada é feita em duas etapas. Primeiro é feita a comparação de uma SAD simulada réplica com a SAD observada a partir do teste de hipótese Kolmogorov-Smirnov bootstrap (Dowd (2023)). Esse teste pressupõe a hipótese nula de que os dois vetores em comparação (SAD observada e SAD simulada) são amostras de uma mesma distribuição teórica; a estatística de interesse do teste é a maior diferença entre as curvas acumuladas das SAD (Dowd (2023)). Contabilizamos o número de SAD réplicas que são congruentes com a SAD observada para cada sítio, paisagem hipotética e grau de limitação de dispersão simulado.

Segundo, a partir do número de SADs réplicas com boa congruência é calculado o logito da proporção de SADs simuladas com boa congruência. Para cada par de paisagens hipotéticas propostas como contrastes para os efeitos explicativos descritos na introdução é calculado a diferença no logito pareado pelo sítio de amostragem e capacidade de dispersão simulado (k). A descrição do logito empírico quando o número de SADs congruentes é zero ou 100, o número total de réplicas simuladas, está na seção 7.3.2.1.

# 7 Análise dos Resultados

A seção dos resultados será composta de 3 seções (7.1, 7.2 e 7.3). Em 7.1 foi apresentada a análise exploratória dos dados empíricos selecionados, i.e., os dados de inventários florestais dentro do critério de inclusão e os respectivos mapas de cobertura florestal do ano mais próximo do inventário. Na subseção 7.2 foi mencionado os procedimentos para os principais resultados do estudo da análise do efeito escalar da extensão espacial da paisagem ao redor associado com o aumento da capacidade de dispersão per capita. Na subseção 7.3 a análise dos efeitos explicativos da paisagem na predição da biodiversidade local foi dividida em duas partes (7.3.1 e 7.3.2). Em 7.3.1 foi feita a análise exploratória da métrica de conectividade, definida como logU/U, para cada tipo de efeito explicativo. Em 7.3.2 foi desenvolvido a descrição da congruência da SAD simulada com a SAD observada, essa seção foi dividida em duas partes: 7.3.2.1 - descrição da congruência absoluta da SAD simulada por paisagem hipotética; e 7.3.2.2 - descrição da congruência relativa da SAD simulada entre as paisagens hipotéticas em contraste no efeito explicativo.

## 7.1 Dados Empíricos: inventários florestais e mapa de cobertura florestal contemporâneo

Nessa seção foi apresentado as características dos inventários florestais: coordenadas geográficas, área da parcela, número de indivíduos na amostra, riqueza de espécies, ano aproximado da amostragem. Os sítios serão contabilizados segundo a classe de perturbação usado na base de dados TreeCo, que considera o histórico e tipo de perturbações na parcela (mais detalhes na seção 6.3.2).

## 7.2 Efeito escalar da dispersão no mecanismo simulado

Nessa seção foi feita a análise exploratória e a descrição estatística dos resultados obtidos a partir da subseção 4.Estimativa da extensão espacial adequada. O objetivo foi determinar a extensão espacial da paisagem hipotética suficiente para simular cada cenário de limitação de dispersão.

## 7.3 Efeitos explicativos da paisagem na predição da biodiversidade local

### 7.3.1 Análise exploratória da métrica de conectividade funcional

A análise exploratória tem como objetivo observar a relação entre a métrica de conectividade, logU/U, com a variação da limitação de dispersão (k) e classe de perturbação. A métrica de conectividade deve variar apenas em função da limitação de dispersão e a diferença entre paisagens hipotéticas. Nessa primeira análise gráfica, os pontos foram coloridos pela proporção de cobertura florestal remanescente na paisagem (p) na escala suficiente para simular todos os k. Com isso foi possível avaliar a coerência entre as métricas de conectividade (logU/U) em função da p, métrica comum da paisagem usada no contexto do debate sobre fragmentação de habitat (Fahrig (2013); Villard and Metzger (2014); Püttker et al. (2020)).

Caso o logU/U apresente grande variação entre classes de perturbação, então isso pode indicar associação entre a configuração espacial da paisagem fragmentada e a classe de perturbação da parcela. Em estudo extenso sobre a base de dados TreeCo (Lima et al. (2020)), foi observado que dentro de áreas estritamente protegidas as perdas de biodiversidade local tendem a ser menores. Porém, foi estimado pouca influência do indice de impacto humano na paisagem ou da distância de grandes centros sobre biodiversidade local (Lima et al. (2020)). Assim, a hipótese de autocorrelação espacial dos resultados de congruência da SAD simulada com SAD local observada pode ser plausível dentro do conjunto de inventários selecionados. E uma primeira análise que pode indicar isso é a análise gráfica de logU/U.

### 7.3.2 Congruência da SAD simulada com a SAD observada

#### 7.3.2.1 Descrição da congruência absoluta das paisagens hipotéticas

A descrição estatística da probabilidade de uma SAD simulada ter boa congruência com a SAD observada por paisagem hipotética, complementa os efeitos explicativos que se baseiam na congruência relativa para um par de paisagens hipotéticas. Para essa descrição estatística foi ajustado um modelo aditivo generalizado hierárquico em que a probabilidade da SAD simulada ter boa congruência foi descrita em função do tipo de paisagem hipotética e de k, a proporção de propágulos na vizinhança imediata. Em todos os modelos estatísticos a variáveis k foi interpretada como variável contínua. Além de k e paisagem hipotética, outras possíveis variáveis fixas no modelo hierárquico foram a classe de perturbação e as coordenadas centrais (Wood (2017)).

O modelo mais plausível usado para descrever a probabilidade de congruência da SAD simulada nas paisagens hipotéticas foi usado em 3 situações. Primeiro, foi usado para calibrar a estatística I de Moran (detalhes na seção ‘Descrição Estatística’). Segundo, para estimar os logitos empíricos quando o número de SAD congruentes é zero ou 100 (detalhes na seção “procedimento para obter logOR”). Terceiro, para escolher as covariáveis do modelo estatístico usado para descrever os efeitos explicativos.

A probabilidade de uma distribuição de abundância de espécies (SAD) simulada apresentar boa congruência com a SAD observada foi modelada utilizando seis modelos estatísticos com diferentes níveis de complexidade.

O modelo mais completo (modelo cheio) inclui:

* Um intercepto por tipo de paisagem hipotética (comum aos sítios) e por sítio de amostragem;
* Um termo de suavização (spline) para o parâmetro k (grau de limitação de dispersão), também especificado por tipo de paisagem hipotética e por sítio de amostragem;
* Duas covariáveis adicionais comuns a todos os sítios: (i) a classe de perturbação do sítio de amostragem, que interage com o tipo de paisagem no spline para k, e (ii) um spline bidimensional para as coordenadas centrais do sítio.

Os cinco modelos restantes representam simplificações progressivas do modelo cheio, obtidas pela exclusão de uma ou mais covariáveis:

* Dois modelos mantêm apenas uma das duas covariáveis adicionais (classe de perturbação ou coordenadas);
* Um modelo possui apenas o spline de k por tipo de paisagem e por sítio, sem as covariáveis adicionais;
* Um modelo ainda mais simples inclui apenas o spline de k por tipo de paisagem (sem variação por sítio);
* O modelo mais simples inclui apenas interceptos por tipo de paisagem e por sítio, sem spline para k.

#### 7.3.2.2 Descrição dos efeitos explicativos da paisagem na biodiversidade local

A descrição do efeito explicativo apresentado na introdução requer que o modelo estatístico relacione a congruência relativa da SAD simulada com a SAD observada (expresso no logOR) e a métrica de conectividade (expresso em logU/U) por grau de limitação de dispersão (k) e sítio de amostragem. Assim, o objetivo da descrição estatística foi fazer um sumário quantitativo da variação de logOR, a métrica de congruência relativa, em função de logU/U, a métrica funcional de conectividade. Uma outra preditora básica no modelo estatístico é k, a proporção de propágulos na vizinhança imediata, pois os resultados de diferentes paisagens hipotéticas são comparados pareado por k. É importante que esse sumário quantitativo considere o padrão por sítio de amostragem e da população de sítios de amostragem, para poder expressar a natureza hierárquica dos resultados. Por isso foram usados modelos hierárquicos aditivos generalizados, que além de possibilitar expressar a estrutura dos dados por sítio de amostragem (Pedersen et al. (2019)), também possibilitam descrever efeitos explicativos não lineares por splines e tensores de suas preditoras contínuas (Wood (2017)). Os modelos foram ajustados pelo método da verossimilhança restrita (Wood (2017)), e os modelos foram comparados usando métricas derivadas do ‘Akaike Information Criteria’ para amostras pequenas (AICc), o delta AICc, o peso de evidência. Também incluímos o ‘deviance explained’ que pode foi interpretado de forma análoga ao coeficiente de determinação (R2), como forma de descrever a variabilidade explicada pelo modelo estatístico. Para essas tarefas foram utilizadas funções dos pacotes mgcv (Wood (2011)) e bblme (Bolker and R Development Core Team (2023)).

Dado a possibilidade de autocorrelação espacial dos inventários florestais, incluímos no modelo estatístico cheio um spline para as coordenadas centrais das parcelas (Wood (2017)). Além disso, incluímos um teste de autocorrelação espacial dos resíduos dos modelos estatísticos. Uma vez que os splines são ideais para descrever relações contínuas, a autocorrelação espacial negativa pode não ser bem descrita pelo modelo estatístico. Assim, se houver autocorrelação espacial que não foi bem descrita pelo modelo estatístico, esperamos que a média, por sítio de amostragem, dos resíduos do modelo apresentem a estatística I de Moran diferente de zero. A estatística I de Moran varia entre -1 e 1, ou seja, entre a autocorrelação espacial negativa e positiva, respectivamente. E possui um teste de hipótese associado que avalia se o valor é significativamente diferente de zero, se o p valor for próximo de zero então a hipótese de ausência de autocorrelação espacial possui pouco suporte. A estatística I de Moran requer que um parâmetro livre seja escolhido. Para isso utilizamos os resíduos de um modelo estatístico de ranqueamento das paisagens hipotéticas mais plausível, tal como descrito no final da seção 6.3.1, análise exploratória. Para obter a estatística I de Moran e seu p valor associado utilizamos funções do pacote spdep (Pebesma and Bivand (2023)).

Outra preditora avaliada foi o nível de perturbação da parcela, tal como classificado na base TreeCo (Lima et al. (2020)). Essa classe é construída conforme a informação disponível sobre o tipo, intensidade e tempo da perturbação antrópica na área da parcela (Lima et al. (2020)). Foram consideradas 3 classes de perturbação: alta, com florestas severamente ou cronicamente perturbadas a pelo menos 50 anos antes do inventário; mediana, com florestas pouco ou esporadicamente perturbadas e/ou perturbada entre 50 e 80 anos; e baixa, com florestas não perturbadas por pelo menos 80 anos (Lima et al. (2020)). Existe grande variabilidade intra classe de perturbação, porém essa é a informação disponível nos trabalhos originais (Lima et al. (2020)).

A interpretação dos efeitos será feita a partir das estimativas de medianas e intervalos de confiança quantílicos (5% e 95%). Nós assumimos que a distribuição conjunta dos coeficientes do modelo mais plausível pode ser aproximada por uma normal multivariada, parametrizada pelas estimativas pontuais dos coeficientes (vetor de médias) e sua matriz de covariância estimada. Realizei 10.000 amostragens dessa distribuição multivariada, cada uma gerando um conjunto de coeficientes simulados. Para cada amostra, calculei as predições condicionais do modelo, obtendo assim uma distribuição empírica das predições. Dessa distribuição, extraí os quantis correspondentes aos percentis 5% (limite inferior), 50% (mediana) e 95% (limite superior). Essa operação de construção das medianas e intervalos de confiança quantílicos foi feito com funções do pacote mgcv (Wood (2011)) e do pacote MASS (Venables and Ripley (2002)).

Para descrever o log da razão da chance da SAD simulada em uma paisagem hipotética ter boa congruência com a SAD observada (logOR) foram usados modelos hierarquicos aditivos mistos (HGAM da sigla em inglês hierarchical generalized additive model Pedersen et al. 2019). Foi ajustado 3 HGAM variando em complexidade. o HGAM mais complexo possui um tensor entre o logU/U e k por sítio de amostragem e comum a todos os sítios (ver metodologia). Os outros HGAM são simplificações desse HGAM mais complexo. Assim, é possível que os HGAM descrevam o logOR em função da combinação de valores de logU/U e k por sítio ou até apenas em função de 1 intercepto por sítio. Esses HGAM foram comparados segundo 3 classes de métricas: AICc; proporção da deviance explicada; e pela estatística I de Moran aplicada à média dos resíduos por sítio.

# Referências

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

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

Bolker, B., and R Development Core Team. 2023. [Bbmle: Tools for general maximum likelihood estimation](https://CRAN.R-project.org/package=bbmle).

Bullock, J. M., L. Mallada González, R. Tamme, L. Götzenberger, S. M. White, M. Pärtel, and D. A. Hooftman. 2017. A synthesis of empirical plant dispersal kernels. Journal of Ecology 105:6–19.

Campos, P. R., E. D. Neto, V. M. de Oliveira, and M. Gomes. 2012. Neutral communities in fragmented landscapes. Oikos 121:1737–1748.

Campos, P. R., A. Rosas, V. M. de Oliveira, and M. A. Gomes. 2013. Effect of landscape structure on species diversity. PloS one 8:e66495.

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

Claudino, E. S., M. A. F. Gomes, and P. R. A. Campos. 2015. [Extinction debt and the role of static and dynamical fragmentation on biodiversity](https://doi.org/10.1016/j.ecocom.2014.11.011). Ecological Complexity 21:150–155.

Condit, R., R. A. Chisholm, and S. P. Hubbell. 2012. Thirty years of forest census at Barro Colorado and the importance of immigration in maintaining diversity. PloS one 7:e49826.

Dowd, C. 2023. [Twosamples: Fast Permutation Based Two Sample Tests](https://CRAN.R-project.org/package=twosamples).

Fahrig, L. 2013. Rethinking patch size and isolation effects: The habitat amount hypothesis. Journal of Biogeography 40:1649–1663.

Hammal, O. A., D. Alonso, R. S. Etienne, and S. J. Cornell. 2015. [When Can Species Abundance Data Reveal Non-neutrality?](https://doi.org/10.1371/journal.pcbi.1004134) PLOS Computational Biology 11:e1004134.

Hubbell, S. P. 2011. The unified neutral theory of biodiversity and biogeography (MPB-32). Princeton University Press.

Leroi, A. M., B. Lambert, J. Rosindell, X. Zhang, and G. D. Kokkoris. 2020. Neutral syndrome. Nature human behaviour 4:780–790.

Lima, R. A. de, D. P. Mori, G. Pitta, M. O. Melito, C. Bello, L. F. Magnago, V. P. Zwiener, D. D. Saraiva, M. C. Marques, A. A. de Oliveira, and others. 2015. How much do we know about the endangered Atlantic Forest? Reviewing nearly 70 years of information on tree community surveys. Biodiversity and Conservation 24:2135–2148.

Lima, R. A. de, A. A. Oliveira, G. R. Pitta, A. L. de Gasper, A. C. Vibrans, J. Chave, H. Ter Steege, and P. I. Prado. 2020. The erosion of biodiversity and biomass in the atlantic forest biodiversity hotspot. Nature communications 11:6347.

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

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

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

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

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

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

Pebesma, E., and R. S. Bivand. 2023. [Spatial data science with applications in R](https://r-spatial.org/book/). Chapman & Hall.

Pedersen, E. J., D. L. Miller, G. L. Simpson, and N. Ross. 2019. [Hierarchical generalized additive models in ecology: An introduction with mgcv](https://doi.org/10.7717/peerj.6876). PeerJ 7:e6876.

Püttker, T., R. Crouzeilles, M. Almeida-Gomes, M. Schmoeller, D. Maurenza, H. Alves-Pinto, R. Pardini, M. V. Vieira, C. Banks-Leite, C. R. Fonseca, and others. 2020. Indirect effects of habitat loss via habitat fragmentation: A cross-taxa analysis of forest-dependent species. Biological Conservation 241:108368.

Rosindell, J., Y. Wong, and R. S. Etienne. 2008. A coalescence approach to spatial neutral ecology. Ecological Informatics 3:259–271.

Souza Jr, C. M., J. Z. Shimbo, M. R. Rosa, L. L. Parente, A. A. Alencar, B. F. Rudorff, H. Hasenack, M. Matsumoto, L. G. Ferreira, P. W. Souza-Filho, and others. 2020. Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth engine. Remote Sensing 12:2735.

Thompson, S. E., R. A. Chisholm, and J. Rosindell. 2020. Pycoalescence and rcoalescence: Packages for simulating spatially explicit neutral models of biodiversity. Methods in Ecology and Evolution 11:1237–1246.

Venables, W. N., and B. D. Ripley. 2002. [Modern applied statistics with s](https://www.stats.ox.ac.uk/pub/MASS4/). Fourth. Springer, New York.

Villard, M.-A., and J. P. Metzger. 2014. Beyond the fragmentation debate: A conceptual model to predict when habitat configuration really matters. Journal of Applied Ecology 51:309–318.

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

Wood, S. N. 2017. Generalized additive models: An introduction with r. Second edition. Chapman; Hall/CRC.