# 1 Materiais e Método

## 1.1 Sumário

Selecionamos inventários florestais realizados na Mata Atlântica e mapas de cobertura florestal contemporâneos aos respectivos anos de amostragem. Os mapas são centrados nas parcelas amostradas e processados de modo que a densidade de pixels corresponda à densidade de indivíduos registrada nos inventários. Em seguida, foi estimada a extensão espacial da paisagem adequada para simular diferentes graus de limitação à dispersão. Na terceira etapa, o Modelo Neutro Espacialmente Explícito (MNEE) foi calibrado: para cada paisagem hipotética e nível de limitação de dispersão simulado, a taxa de entrada de novas espécies na paisagem (U) foi ajustada para reproduzir a riqueza local observada no inventário correspondente. A validação do MNEE foi feita por meio da comparação entre a distribuição de abundância de espécies (SAD) observada e a simulada em diferentes paisagens hipotéticas e graus de limitação à dispersão. Por fim, nos sítios em que a SAD simulada reproduziu de forma satisfatória a observada, interpretamos os efeitos da paisagem com base na diferença entre os valores de log U estimados em pares de paisagens hipotéticas.

## 1.2 Fonte de Dados

### 1.2.1 TreeCo

Consultamos a base de dados TreeCo, que reúne amplamente estudos fitossociológicos realizados na Floresta Atlântica (Lima et al. (2015)). A partir dessa base, selecionamos inventários florestais que: amostraram indivíduos arbóreos com diâmetro à altura do peito (DAP) ≥ 4,8 cm; utilizaram parcelas únicas com pelo menos 1 ha de floresta; apresentaram alta confiabilidade na identificação taxonômica (ao nível de morfoespécie) e na precisão das coordenadas geográficas centrais da parcela.

Detalhes da avaliação da qualidade dos inventários podem ser consultados em Lima et al. (2020). Em resumo, a base TreeCo classificou os inventários segundo critérios de confiabilidade taxonômica e precisão geográfica.

A resolução taxonômica foi aprimorada por meio 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, corrigindo erros tipográficos, sinônimos e variantes ortográficas.

Morfoespécies marcadas como confer foram atribuídas à espécie sugerida; aquelas marcadas como affinis foram consideradas apenas no nível de gênero. As coordenadas centrais dos inventários foram verificadas quanto à correspondência com os fragmentos florestais descritos nos estudos. Inventários com localização inconsistente foram descartados. Quando necessário, as coordenadas foram ajustadas com base em mapas, descrições dos locais originais e outras fontes disponíveis online.

### 1.2.2 Mapbiomas

Para caracterizar a cobertura vegetal no entorno de cada inventário, utilizamos os mapas da coleção 6 do MapBiomas, com resolução de 30 x 30 m, cobrindo o período de 1985 a 2020 (Souza Jr et al. (2020)). Para cada parcela, foi selecionado o mapa de cobertura florestal correspondente ao ano mais próximo da amostragem, de forma a garantir a contemporaneidade entre os dados de vegetação e o levantamento florístico. Cada mapa foi recortado em quadrados centrados na coordenada da parcela, com extensão máxima de 16 x 16 km². A resolução dos mapas foi ajustada de modo que a densidade de pixels correspondesse à densidade de indivíduos amostrados na parcela. Após o ajuste de resolução, os valores de cobertura passaram a variar continuamente entre 0 e 1 (diferente do binarismo original). Consideramos como unidades de habitat os pixels com valor de cobertura ≥ 0,7. As parcelas foram representadas como quadrados centralizados no mapa, contendo N unidades de habitat (posições disponíveis para indivíduos), onde N corresponde ao número total de indivíduos registrados no inventário. Nos casos em que a configuração espacial da cobertura florestal impossibilitava a centralização exata, a parcela foi ligeiramente deslocada para uma área adjacente que permitisse a demarcação de um quadrado adequado.

## 1.3 Modelo Neutro Espacialmente Explícito (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 (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, sorteadas de forma independente e com a mesma distribuição (Rosindell et al. 2008; 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 genealogia da comunidade com uma equação que relaciona a riqueza observada à taxa média de introdução de novas espécies, permitindo calibrar o MNEE com base na riqueza local observada (Rosindell et al. (2008), Apêndice A).

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 muito severa) a 5% (limitação muito branda).

## 1.4 Estimativa da extensão espacial controlado pelo grau de limitação de dispersão simulado

Em nossa abordagem, levamos em consideração o efeito da escala espacial sobre a limitação de dispersão no mecanismo demográfico simulado. Buscamos determinar a extensão espacial mais adequada da paisagem para cada grau de limitação de dispersão, avaliando a sensibilidade da estimativa da taxa U em função da escala espacial simulada. Essa estimativa foi realizada em paisagens sem perda de cobertura, de modo que a única influência sobre U foi a capacidade de dispersão e a extensão da paisagem, para um mesmo sítio.

Quando o MNEE é simulado em paisagens infinitas e contínuas (i.e., sem perda de habitat), a taxa U necessária para manter a riqueza local apresenta um máximo global sob um grau intermediário de limitação de dispersão (May et al. (2012)). Isso reflete o balanço entre deriva ecológica e dispersão. Em cenários de dispersão muito limitada, a substituição tende a ocorrer entre coespecíficos, reduzindo a perda de espécies e, consequentemente, a necessidade de introdução de novas espécies (baixa U). À medida que a dispersão se torna menos restrita, aumenta a substituição entre indivíduos de espécies diferentes, elevando a perda de espécies locais e exigindo maior U. Quando a dispersão é ampla demais, há maior contribuição de subpopulações mais distantes da parcela, o que reduz novamente a estimativa da taxa U (May et al. (2012)), porém à custa de uma maior homogeneização na escala da paisagem (Mouquet and Loreau (2003)).

Para avaliar a sensibilidade da extensão espacial em relação à limitação de dispersão, sorteamos 36 sítios amostrais e simulamos a taxa U em paisagens prístinas com 6 diferentes tamanhos (de 0,5 km a 16 km de lado), sob 20 níveis de limitação de dispersão (valores de k). A expectativa era que o aumento da extensão espacial resultasse na redução da taxa U, até um ponto em que a inclusão de áreas adicionais não alterasse significativamente essa estimativa. Isso é esperado, pois, em paisagens pequenas há poucos progenitores fora da parcela capazes de introduzir novas espécies. Com o aumento da extensão, mais progenitores contribuem com a parcela, aumentando a chance de imigração por novas espécies. No entanto, dada a limitação de dispersão, a influência de progenitores muito distantes tende a se estabilizar.

Para quantificar a variação na taxa U média associada ao aumento da extensão espacial, ajustamos um modelo gaussiano onde o logito da taxa U foi a variável resposta. As covariáveis explicativas incluíram o número de espécies e indivíduos (ambos na escala logarítmica), além dos graus de limitação de dispersão e da extensão da paisagem (tratadas como fatores), com intercepto aleatório para cada sítio amostral. Para cada valor de k, identificamos a extensão espacial em que a média de U se estabilizava, acumulando 95% da variação observada entre 0,5 km e 16 km. Essa escala foi considerada como extensão adequada para aquele grau de limitação de dispersão. A abordagem assume uma relação monotônica entre extensão espacial e estabilização da taxa U.

## 1.5 Calibração de MNEE nas paisagens hipotéticas

As paisagens hipotéticas têm forma quadrada, e para um mesmo sítio de amostragem, simulado, foram criadas com a mesma densidade de pixels e área da parcela amostrada. Denominaremos paisagem fragmentada o mapa de cobertura florestal no ano mais próximo ao da amostragem do inventário florestal selecionado. 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. Denominaremos de paisagem aglomerada a paisagem em que todo o habitat remanescente na paisagem fragmentada é aglomerado ao redor da parcela em forma de quadrado. Dessa forma, a paisagem aglomerada possui a mesma proporção de cobertura florestal da paisagem fragmentada e a mesma configuração espacial da paisagem prístina.

A extensão espacial da construção das paisagens hipotéticas variou entre graus de limitação de dispersão conforme a extensão espacial suficiente estimada. Para cada sítio de amostragem, paisagem hipotética (land) e grau de limitação de dispersão (k) é estimado uma taxa U média a partir de 20 réplicas. Então, dado a média das estimativas da taxa U, são simuladas 100 SADs réplicas para cada k e land.

## 1.6 Validação do predito por MNEE

### 1.6.1 Comparação com a SAD observada

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.

### 1.6.2 Descrição estatística da congruência com a SAD observada

A descrição estatística da probabilidade de uma SAD simulada ter boa congruência com a SAD observada por paisagem hipotética foi obtida por um modelo aditivo generalizado hierárquico (Wood (2017), Pedersen et al. (2019)). 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.

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 ser 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)).

Uma vez que os splines são ideais para descrever relações contínuas, a autocorrelação espacial negativa pode não ser bem descrita pelo modelo estatístico, assim, incluímos um teste de autocorrelação espacial dos resíduos dos modelos estatísticos. 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. Determinamos esse valor por uma analise de sensibilidade dos resíduos dos modelos ajustados. Para obter a estatística I de Moran e seu p valor associado utilizamos funções do pacote spdep (Pebesma and Bivand (2023)).

### 1.6.3 Modelos concorrentes

Além de k e paisagem hipotética, outras possíveis variáveis fixas no modelo hierárquico foram a classe de perturbação e as coordenadas centrais. 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. 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. Por outro lado, 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, incluímos um spline para as coordenadas geográficas no modelo cheio (Wood (2017)).

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 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 (fixo) e por sítio de amostragem (aleatório);
* 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: 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 (e apenas um intercepto por sítio);
* O modelo mais simples inclui apenas interceptos por tipo de paisagem e por sítio, sem spline para k.

## 1.7 Interpretação do logU/U como métrica funcional dos efeitos da paisagem

A calibração e validação de MNEE ocorreu em extrema amplitude de limitações de dispersão, uma vez que para a maior parte das espécies, a maior parte dos propágulos permanece até a vizinhança imediata do progenitor (REF). Para interpretarmos os efeitos da paisagem selecionamos graus de limitação de dispersão entre acima de 0.50 (brando) até 0.99 (severo) dos propágulos na vizinhança imediata do progenitor.

A taxa U estimada reflete a perda local de espécies por deriva ecológica que não é compensada pela imigração da paisagem ao redor. Ao manter todos os outros parâmetros os mesmos e mudando apenas a paisagem hipotética, a diferença no log da taxa U informa o efeito da paisagem, codificado no contraste de paisagens hipotéticas, na interação entre competição neutra no espaço e entrada aleatória de novas espécies no sistema.

Para ter boa confiabilidade da estimativa do modelo, selecionamos apenas sítios em que todos os graus de limitação de dispersão tiveram boa congruência com a SAD observada. Consideramos sítios com boa congruência quando pelo menos 75% das SADs simuladas em todas as paisagens hipotéticas e graus de limitação de dispersão apresentaram boa congruência.

### 1.7.1 Descrição estatística da métrica funcional dos propostos efeitos da paisagem

Para descrever o logU/U dos sítios em que a SAD simulada apresenta boa congruência com a SAD observada, utilizamos 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)). Utilizamos os mesmo critérios e método de ajuste usados na descrição estatística da congruência, exceto a estatística I de Moran. Uma vez que a taxa U depende apenas dos parâmetros explícitos no modelo podemo desconsiderar a hipótese de autocorrelação espacial, relacionada com a congruência com a SAD.

Aqui, focamos em descrever a variabilidade do logU/U em função da proporção de cobertura florestal (%CF) na paisagem e do grau de limitação de dispersão (Villard and Metzger (2014)). A %CF não distingue entre a paisagem fragmentada e aglomerada, apenas a paisagem prístina. Contudo, existe uma relação não linear entre a variabilidade na configuração espacial e a %CF (Villard and Metzger (2014)). Existe maior variabilidade de configurações espaciais quando a %CF é intermediária (Villard and Metzger (2014)). Por isso fizemos uma descrição estatística de logU/U em função do grau de limitação de dispersão (k) e %CF onde há um spline para k por sítio que pode variar de forma independente. Com isso foi possível avaliar a coerência da nossa métrica funcional baseada na taxa U (logU/U) em função da %CF, métrica comum da paisagem usada no contexto do debate sobre fragmentação de habitat (Fahrig (2013); Villard and Metzger (2014); Püttker et al. (2020); Watling et al. 2020). A partir desse modelo em que a %CF é uma variável contínua, a predição foi agrupada em três faixas de %CF, abaixo de 30%, entre 30% e 60% e acima de 60% (Püttker et al. (2020)). Com os dados agrupados por faixa de %CF é possível visualizar a predição do modelo por sítio em função de k para o logU/U observado.

Para ter base de comparação da deviance explicada pelo modelo estatístico cheio, uma vez excluído a variabilidade por sítio, ajustamos um segundo modelo em que a variabilidade por sítio é modelada apenas por um intercepto. Os modelos foram ajustados usando funções do pacote mgcv (Wood (2011)). Na estrutura fixa do modelo foram usados um tensor entre a %CF e k, e na estrutura aleatória um spline por sítio ou apenas um intercepto por sítio (Pedersen et al. (2019)).

## Referências

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.

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.

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.

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

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.

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.