



MARIA CLEIDIVANE ALVES BATISTA

**PREDIÇÃO DA DIVERSIDADE ARBÓREA NO SUDESTE DO
BRASIL UTILIZANDO SENSORIAMENTO REMOTO E
APRENDIZAGEM DE MÁQUINA**

LAVRAS - MG

2020

MARIA CLEIDIVANE ALVES BATISTA

**PREDIÇÃO DA DIVERSIDADE ARBÓREA NO SUDESTE DO BRASIL
UTILIZANDO SENSORIAMENTO REMOTO E APRENDIZAGEM DE MÁQUINA**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia Florestal, área de concentração em Ciências Florestais, para a obtenção do título de Mestre.

Prof. Dr. Fausto Weimar Acerbi Júnior

Orientador

Dra. Marcela de Castro Nunes Santos Terra

Coorientadora

Dra. Eduarda Martiniano de Oliveira Silveira

Coorientadora

LAVRAS - MG

2020

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Batista, Maria Cleidivane Alves.

Predição da diversidade arbórea no sudeste do Brasil utilizando
sensoriamento remoto e aprendizagem de máquina / Maria
Cleidivane Alves Batista. - 2020.

56 p. : il.

Orientador(a): Fausto Weimar Acerbi Júnior.

Coorientador(a): Marcela de Castro Nunes Santos Terra,
Eduarda Martiniano de Oliveira Silveira.

Dissertação (mestrado acadêmico) - Universidade Federal de
Lavras, 2020.

Bibliografia.

1. Diversidade arbórea. 2. Modelagem ambiental. 3. Random
Forest. I. Júnior, Fausto Weimar Acerbi. II. Terra, Marcela de
Castro Nunes Santos. III. Silveira, Eduarda Martiniano de Oliveira.
IV. Título.

MARIA CLEIDIVANE ALVES BATISTA

**PREDIÇÃO DA DIVERSIDADE ARBÓREA NO SUDESTE DO BRASIL
UTILIZANDO SENSORIAMENTO REMOTO E APRENDIZAGEM DE MÁQUINA
PREDICTION OF TREE DIVERSITY IN SOUTHEAST BRAZIL USING REMOTE
SENSING AND MACHINE LEARNING**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia Florestal, área de concentração em Ciências Florestais, para a obtenção do título de Mestre.

APROVADA em 17 de fevereiro de 2020.

Dr. Fausto Weimar Acerbi Júnior UFLA

Dra. Marcela de Castro Nunes Santos Terra UFLA

Dra. Mônica Canaan Carvalho IFSUDESTEMG

Prof. Dr. Fausto Weimar Acerbi Júnior

Orientador

Dra. Marcela de Castro Nunes Santos Terra

Coorientadora

Dra. Eduarda Martiniano de Oliveira Silveira

Coorientadora

LAVRAS – MG

2020

AGRADECIMENTO

A Deus, por me conceder vida, saúde e graças sempre;

À Universidade Federal de Lavras e ao programa de pós-graduação em Engenharia Florestal pela oportunidade;

À Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG) pelo auxílio financeiro concedido através das bolsas de mestrado;

A minha família - papai, mamãe, Cleidiane, Cleidimilson e Renata – por ser um berço de amor, carinho, confiança, incentivo e apoio;

Ao Eduardo, meu noivo, namorado, amigo, amor e companheiro, que com seu sincero e intenso amor me ajudou a superar as dificuldades encontradas;

Ao professor Fausto, orientador, e à Duda e Marcela, coorientadoras, pelo apoio, esforço e ajuda na realização desse trabalho;

Aos professores, amigos e colegas do LEMAF pelo conhecimento e companheirismo compartilhados;

E a todos que, de alguma forma, direta ou indiretamente, proporcionaram condições para a realização desse trabalho;

MUITO OBRIGADA!

RESUMO GERAL

Conhecer a biodiversidade de uma região consiste em importante ferramenta de gestão no que diz respeito ao desenvolvimento sustentável. Para isso, são utilizados índices que sintetizam informações de riqueza e uniformidade das espécies de uma área, com base em levantamentos de campo, que por sua vez são demorados e dispendiosos. A modelagem ambiental utilizando variáveis de fácil obtenção (como as provenientes de sensoriamento remoto) tem auxiliado muito nesse processo, diminuindo a necessidade de tais levantamentos. Por outro lado, tendo em vista a grande quantidade de dados disponíveis para o uso na modelagem, surge outro desafio que é a seleção das variáveis que mais contribuem para a predição da métrica em questão, sendo os algoritmos de aprendizagem de máquina bastante eficientes nessa questão. Diante disso, objetivou-se modelar e mapear a diversidade arbórea para o estado de Minas Gerais utilizando o algoritmo Random Forest (RF) e variáveis climáticas, de terreno e de sensoriamento remoto. Também procurou-se responder as seguintes questões: (1) Quais variáveis têm maior potencial preditor para cada métrica de diversidade? (2) Qual o padrão de distribuição espacial da diversidade arbórea para o estado de Minas Gerais? (3) Existe um consenso entre as métricas de diversidade? Foram utilizadas informações do inventário florestal de Minas Gerais de 2755 parcelas amostrais para obtenção das métricas de diversidade e 115 variáveis preditoras. A modelagem foi realizada no software Rstudio. Os modelos gerados pelo RF para todas as métricas tiveram em média R^2 de 0,46 e RMSE (%) de 20,73. De forma geral, as variáveis selecionadas mais representativas para as métricas foram precipitação (climática), produtividade primária bruta (sensoriamento remoto) e profundidade do vale (terreno). Os mapas gerados para o estado mostram que a diversidade arbórea aumenta de Norte a Sul e Noroeste a Sudeste, estando fortemente associada aos domínios morfoclimáticos. A maior diversidade arbórea se concentrou no domínio Atlântico, diminuindo um pouco no domínio do Cerrado e a menor diversidade foi para o domínio da Caatinga. O índice alfa de Fisher teve um maior consenso com a riqueza, e os índices de Shannon e Simpson acordaram mais entre si. Pielou, por representar uniformidade, apresentou padrão diferente. Conclui-se que as variáveis selecionadas pelo RF conseguiram captar as diferenças existentes entre os domínios e representar bem os padrões espaciais de diversidade arbórea de Minas Gerais.

Palavras-chave: Diversidade arbórea. Modelagem ambiental. Random Forest.

GENERAL ABSTRACT

Knowing the biodiversity of a region is an important management tool with regard to sustainable development. For that, indexes are used that synthesize information on the richness and uniformity of species in an area, based on field surveys, which in turn are time-consuming and expensive. Environmental modelling using easily obtainable variables (such as from remote sensing) has helped a lot in this process, reducing the need for such surveys. On the other hand, in view of the large amount of data available for use in modelling, another challenge arises, which is the selection of variables that most contribute to the prediction of the metric in question and the machine learning algorithms are quite efficient in this matter. Therefore, the objective was to model and map tree diversity for the state of Minas Gerais using the Random Forest (RF) algorithm and climatic, terrain and remote sensing variables. We also tried to answer the following questions: (1) Which variables have the greatest predictive potential for each diversity metric? (2) What is the pattern of spatial distribution of tree diversity for the state of Minas Gerais? (3) Is there a consensus among the diversity metrics? Information from the forest inventory of 2755 sample plots was used to obtain diversity metrics and 115 predictor variables. The modelling was performed using the Rstudio software. The models generated by the RF for all metrics had an average R^2 of 0.46 and RMSE (%) of 20,73. In general, the most representative selected variables for the metrics were precipitation (climatic), gross primary productivity (remote sensing) and valley depth (terrain). The maps generated for the state show that the tree diversity increases from North to South and Northwest to Southeast, being strongly associated with the morphoclimatic domains. The greatest tree diversity was concentrated in the Atlantic domain, decreasing slightly in the Cerrado domain and the least diversity was in the Caatinga domain. Fisher's alpha index had a greater consensus with richness, and the Shannon and Simpson indexes agreed more with each other. Pielou, for representing uniformity, presented a different pattern. It is concluded that the variables selected by the RF managed to capture the existing differences between the domains and represent well the spatial patterns of tree diversity in Minas Gerais.

Keywords: Tree diversity. Environmental modelling. Random Forest.

SUMÁRIO

PRIMEIRA PARTE	8
1. INTRODUÇÃO	8
2. REFERENCIAL TEÓRICO	10
2.1. A vegetação do estado de Minas Gerais	10
2.2. Medidas de diversidade biológica aplicadas à vegetação	11
2.3. Sensoriamento remoto como ferramenta para estimativa da diversidade biológica	13
3. CONSIDERAÇÕES GERAIS	16
REFERÊNCIAS	17
SEGUNDA PARTE - ARTIGO	20
Artigo 1 - Modelling and mapping tree diversity in Southeast Brazil using remote sensing and machine learning	20

PRIMEIRA PARTE

1. INTRODUÇÃO

Os serviços ecossistêmicos fornecidos pela biodiversidade proporcionam uma melhor qualidade de vida em todo o planeta, e conhecer tal biodiversidade é importante para o planejamento de um desenvolvimento sustentável, tendo em vista a preservação ambiental para as futuras gerações. Em termos de flora, o Brasil é o país mais rico do mundo. Possui atualmente cerca de 46.826 espécies reconhecidas (incluindo algas, fungos, gimnospermas, angiospermas, briófitas e samambaias e licófitas) com destaque para as 33.358 espécies de angiospermas (~57% endêmicas do país) (FLORA DO BRASIL, 2020; BFG, 2015). Essa diversidade está relacionada, além da ampla área do país, às variações de clima, solo e relevo, que afetam diretamente a distribuição geográfica da vegetação e proporcionam composições de diferentes paisagens. O estado brasileiro de Minas Gerais também possui ampla diversidade. Segundo Terra et al. (2017), o estado possui entre 3.592 e 3.743 espécies arbóreas distribuídas nos domínios Atlântico, Cerrado e Caatinga. Tais espécies configuram cerca de 24,5 bilhões de árvores, que correspondem a aproximadamente 6,29% do total de árvores da Bacia Amazônica (TER STEEGE et al., 2013), 1,23% do total de árvores da floresta tropical (SLIK et al., 2015) e 0,81% do total de árvores do mundo (CROWTHER et al., 2015). Estes valores encontrados são maiores do que os encontrados por Oliveira-Filho (2006) em uma estimativa de árvores realizada para o estado, que foi de 2500 espécies. Isso reforça a importância de estudos que utilizem variáveis relacionadas às variações da vegetação para a modelagem e quantificação da diversidade florestal.

Tendo em vista a dificuldade de obtenção de dados de campo (muitas vezes demorados e caros), é vantajoso utilizar variáveis de fácil obtenção para a estimativa da biodiversidade por meio da modelagem. As técnicas de sensoriamento remoto mostram-se altamente efetivas na análise da cobertura vegetal, uma vez que são de baixo custo e com capacidade de monitoramento em amplas áreas geográficas. De acordo com Rocchini, Hernández-Stefanoni e He (2015), a detecção remota é uma das abordagens mais poderosas para identificar hotspots de biodiversidade e prever mudanças na composição de espécies em tempo e custo reduzidos; isto porque permite obter repetições da cobertura espacial completa da Terra em um curto período de tempo. Atualmente, o sensoriamento remoto consiste numa ferramenta base para as chamadas 'Variáveis Essenciais da Biodiversidade' (EBV) que são utilizadas para a modelagem da diversidade global, desde composição genética a estrutura do ecossistema (GEO BON,

2020). Dessa forma, tem-se acesso a uma grande quantidade de dados e saber selecioná-los para a modelagem pode ser um desafio.

Os algoritmos de aprendizagem de máquina são uma ótima solução para esse problema, uma vez que ‘aprendem’ com exemplos e extrapolam esse aprendizado para áreas não amostradas. Em especial o algoritmo Random Forest tem se destacado no quesito de modelagem ambiental (CARVALHO et al., 2019), porém, até o momento, poucos estudos têm utilizado variáveis espacialmente explícitas, como de sensoriamento remoto, de terreno e ambientais, como variáveis preditoras para o mapeamento da diversidade de espécies arbóreas do estado de Minas Gerais (SILVEIRA et al., 2019). Vale ressaltar que existe uma certa dificuldade em realizar essa modelagem, pois além da necessidade de grande quantidade de dados de inventário florestal (representativos de todo o estado), as variáveis preditoras precisam explicar a alta variabilidade desses dados.

Diante disso, o objetivo desse estudo é modelar e mapear a diversidade arbórea do estado de Minas Gerais utilizando o algoritmo Random Forest e variáveis climáticas, topográficas e de sensoriamento remoto.

2. REFERENCIAL TEÓRICO

2.1. A vegetação do estado de Minas Gerais

O estado brasileiro de Minas Gerais (MG), localizado na região sudeste do país, integra a transição dos domínios morfoclimáticos brasileiros Atlântico, Cerrado e Caatinga (Floresta semiárida) que correspondem a 41%, 57% e 2% dos remanescentes de vegetação nativa do estado, respectivamente (SCOLFORO et al., 2015). Dentro de cada um desses domínios, devido a variações das características climáticas, do solo e relevo, ocorrem diferentes fitofisionomias, que vão desde formações florestais até campestres.

O domínio Atlântico compreende a segunda maior área de floresta tropical úmida da América do Sul, ficando atrás do domínio amazônico. Ambos os domínios são separados pela chamada diagonal de formações abertas, que incluem o Cerrado e as Caatingas. O domínio é composto principalmente pelas fitofisionomias Floresta Ombrófila e Floresta Estacional Semidecidual conforme o gradiente longitudinal costa-interior do país, que proporciona a ocorrência de formações mais abertas (OLIVEIRA-FILHO; FONTES, 2000). Possui elevada riqueza ameaçada de espécies vegetais e animais, com elevada taxa de endemismo e esse fato o coloca na lista mundial dos hotspots prioritários para a conservação da biodiversidade (MITTERMER et al., 2011). Em Minas Gerais o domínio Atlântico ocorre nas regiões centro-sul e leste do estado, predominando a Floresta Estacional Semidecidual uma vez que a Floresta Ombrófila se restringe ao extremo sul do estado (TERRA et al., 2018).

O domínio do Cerrado também é um hotspot mundial de biodiversidade. No Brasil, é o segundo maior domínio em área, ficando também atrás do domínio amazônico. Engloba formações florestais – fitofisionomias Mata Ciliar, Mata de Galeria, Mata Seca e Cerradão; formações savânicas – Cerrado sentido restrito (Cerrado Denso, Cerrado Típico, Cerrado Ralo e Cerrado Rupestre conforme o ambiente e a densidade arbóreo-arbustiva), Parque de Cerrado, Palmeiral e Vereda; e formações campestres – Campo Rupestre, Campo Sujo e Campo Limpo (RIBEIRO; WALTER, 2008). Conforme os mesmos autores, a distribuição da flora no Cerrado é influenciada pelo clima, solo, disponibilidade de água e nutrientes, geomorfologia e topografia, profundidade do lençol freático, latitude, pela frequência de queimadas e outras atividades antrópicas. No estado de Minas Gerais, o domínio do Cerrado abrange a maior parte do território, na região centro-oeste e norte.

O domínio das florestas semiáridas ou Caatingas é o único domínio exclusivo do Brasil. É quase totalmente restrito (98,8%) à região nordeste do país (SCOLFORO et al., 2015),

ocorrendo o restante em Minas Gerais. É representado pela fitofisionomia Floresta Estacional Decidual, adaptada à maior limitação meteorológica do país com elevadas temperaturas e baixa precipitação (SCOLFORO et al., 2015). Possui elevada taxa de endemismo (PENNINGTON; LAVIN; OLIVEIRA-FILHO, 2009) com ameaças antrópicas e mesmo assim o domínio não é considerado um hotspot mundial. Em Minas Gerais o domínio ocorre no norte do estado. Diante dessa grande biodiversidade existente no estado, percebe-se a importância de políticas e ações que visem a conservação de tal patrimônio.

2.2. Medidas de diversidade biológica aplicadas à vegetação

As medidas de diversidade biológica podem ser aplicadas a qualquer comunidade biológica. Tais medidas captam a riqueza (número de espécies dentro da unidade de estudo) e/ou a uniformidade (similaridade entre as espécies em termos de abundância e número de indivíduos) de uma comunidade.

O termo riqueza de espécies é a mais antiga e mais intuitiva medida de biodiversidade, proposto por McIntosh (1967). No entanto, sabemos que a abundância (número de indivíduos em cada espécie) varia e por isso o termo riqueza de espécies não pode ser considerado sinônimo de biodiversidade. Incluir a uniformidade nas medidas de diversidade ajuda a definir e discriminar comunidades ecológicas (MAGURRAN, 2013). Uma comunidade com maior uniformidade, ou seja, com espécies mais igualmente abundantes, é mais biodiversa do que uma comunidade onde haja dominância de algumas espécies.

A série logarítmica de Fisher foi um dos primeiros trabalhos a descrever matematicamente a relação entre o número de espécies e o número de indivíduos destas espécies (FISHER; CORBET; WILLIAMS, 1943). O índice α e N resumem a série por meio da relação $S = \alpha \ln(1 + N/\alpha)$, onde S é o número total de espécies, N é o número total de indivíduos e α é o índice da série logarítmica de Fisher (alfa de Fisher) que é calculado baseado na relação N/S da amostra. Segundo Terra et al. (2017), os valores de alfa de Fisher podem ser usados para extrapolar a riqueza de espécies para uma região definida no caso de já se conhecer o número total de indivíduos, desde que satisfaça algumas condições, como por exemplo, a abundância das espécies apresentar uma distribuição de série logarítmica. Ter Steege et al. (2013) utilizaram essa relação para estimar a riqueza de espécies arbóreas na Amazônia e obtiveram resultados semelhantes à previsão modelada por Hubbell et al. (2008). Terra et al. (2017) também utilizaram esse índice para prever a riqueza arbórea para o estado de Minas Gerais.

O índice de Simpson (D) é um índice de diversidade que enfatiza a componente uniformidade/dominância em detrimento da componente riqueza (MAGURRAN, 2013). Ele calcula a probabilidade de dois indivíduos quaisquer, retirados aleatoriamente de uma comunidade infinitamente grande, pertencerem à mesma espécie (SIMPSON, 1949). É dado pela fórmula:

$$D = \sum_{i=1}^S \left(\frac{n_i[n_i - 1]}{N[N - 1]} \right)$$

Onde n_i é o número de indivíduos na i -ésima espécie, S é o número total de espécies e N é o número total de indivíduos.

Os valores do índice de Simpson variam de 0 a 1, sendo que quanto mais próximo de 1, menor a diversidade (menor a uniformidade e maior a dominância). O índice de Simpson, segundo Magurran (2013) é uma das medidas de diversidade mais significativas e robustas disponíveis, ficando na frente do mais popular índice de Shannon, que é sensível à variação do tamanho amostral.

O índice de Shannon (1948), como dito anteriormente, é o índice de diversidade mais popular. Ele assume que indivíduos são aleatoriamente amostrados de uma comunidade infinitamente grande e que todas as espécies são representadas na amostra (PIELOU, 1975). Dessa maneira, existe um erro associado ao índice, sendo que geralmente não se amostra todas as espécies de uma comunidade. É definido pela fórmula:

$$H' = - \sum_{i=1}^S \frac{n_i}{N} \ln \frac{n_i}{N}$$

Em que S , N e n_i foram definidos anteriormente.

O valor de H' permanece entre 1,5 e 3,5 e raramente ultrapassa 4 (MARGALEF, 1972). Quanto maior o valor, maior a diversidade da comunidade.

O índice de Shannon leva em consideração a uniformidade na abundância das espécies. Porém é possível calcular uma medida de uniformidade separadamente, a uniformidade de Pielou (J'), baseada na diversidade máxima (todas as espécies igualmente abundantes). O índice de uniformidade de Pielou (1975) é dado por:

$$J' = \frac{H'}{H_{\max.}} = \frac{H'}{\ln S}$$

Onde H_{\max} é a diversidade máxima possível, H' é o índice de diversidade de Shannon e S é o número total de espécies. Quanto maior o J' , maior a uniformidade e consequentemente, maior a diversidade.

2.3. Sensoriamento remoto como ferramenta para estimativa da diversidade biológica

Estudos que envolvem a identificação e monitoramento de espécies em grandes áreas geográficas encontram desafios e problemas continuamente. Além do trabalho ser demorado e dispendioso, existe a dificuldade relacionada ao sistema de amostragem a ser adotado, sem falar que em estudos de monitoramento esse trabalho deve se repetir ao longo do tempo. Apesar desse esforço, as pesquisas de campo às vezes apresentam baixa precisão (ROCCHINI et al., 2015). Scott e Hallam (2003) encontraram uma taxa média de identificação errada de 2,7% a 25,65% a depender da capacidade do pesquisador e da espécie envolvida. Sendo assim, as técnicas de sensoriamento remoto podem auxiliar nesses estudos, diminuindo a necessidade de levantamentos de campo, uma vez que permitem a cobertura de toda a superfície terrestre em curto período de tempo e fornecem continuamente informações sobre a distribuição da biodiversidade (HE et al., 2015).

Cada objeto na superfície terrestre se comporta de maneira diferente ao receber a energia eletromagnética vinda do Sol, absorvendo, transmitindo e refletindo essa energia. Dessa forma, as ferramentas de sensoriamento remoto mostram-se bastante efetivas para avaliar o comportamento desses objetos. No caso da vegetação, a manipulação dos valores de resposta espectral permite obter índices de vegetação que podem ser comparados com indicadores ecológicos da estrutura e funcionamento das florestas (VICENS; CRUZ; RIZZINI, 1998).

O Índice de Vegetação por Diferença Normalizada (NDVI – *Normalized Difference Vegetation Index*), relatado por Rouse et al. (1974) é o índice de vegetação mais comumente utilizado. Ele se baseia na relação entre a reflectância da vegetação nas bandas espectrais vermelho e infravermelho próximo captada pelos sensores de satélite. É dado pela fórmula:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

Em que NIR e RED correspondem à reflectância nos comprimentos de onda do infravermelho próximo e vermelho, respectivamente.

Segundo Vorovencii (2014), o NDVI realça a vegetação e ao mesmo tempo minimiza o efeito da sombra causado pela topografia e, por mais que seja sensível aos efeitos do solo e da

atmosfera, é um bom indicador da quantidade total de vegetação, estando também associado diretamente a outros parâmetros como porcentagem de cobertura do solo, atividade fotossintética da planta, índice de área foliar e quantidade de biomassa (ACERBI JÚNIOR et al., 2015). Isso explica o fato do NDVI ser amplamente utilizado em estudos que relacionem produtividade, biodiversidade e heterogeneidade de habitats (HE; ZHANG; ZHANG, 2009).

Em florestas pouco densas, a reflectância dos diferentes tipos de solo pode influenciar nos valores de NDVI. Para reduzir esse efeito, Huete (1988) desenvolveu o índice de vegetação ajustado ao solo (SAVI – *Soil-adjusted Vegetation Index*), aplicando a constante L cujos valores (determinados empiricamente) variam de 0 a 1 conforme a densidade da vegetação. O valor 1 é adotado para áreas com pouca cobertura vegetal; 0,5 para áreas com cobertura intermediária; 0,25 para áreas muito densas (SILVA et al., 2015). Quando o L é igual a 0, o SAVI é idêntico ao NDVI (MENESES; ALMEIDA, 2012). O SAVI é dado pela fórmula:

$$SAVI = \left(\frac{NIR - RED}{NIR + RED} \right) \cdot (1 + L)$$

O NDVI é pouco sensível à variação da estrutura do dossel, apresentando uma rápida saturação dos seus valores, não distinguindo as variações de biomassa por exemplo em uma floresta mais densa (ASRAR et al., 1984). Para resolver esse problema, o Índice de vegetação melhorado (EVI – *Enhanced Vegetation Index*) foi proposto por Justice et al. (1998). Além de reduzir as influências atmosféricas e do solo no dossel, o EVI é mais sensível à arquitetura do dossel e está mais relacionado com a dinâmica sazonal da produtividade das plantas do que o NDVI (XIAO et al., 2005). É dado por:

$$EVI = \left(\frac{G(NIR - RED)}{(L + NIR + C1.RED - C2.BLUE)} \right)$$

Em que L é o fator de ajuste para o solo; G=2,5 é o fator de ganho; NIR e RED são as reflectâncias nas bandas do infravermelho próximo e vermelho, respectivamente; C1= 6 é o coeficiente de correção dos efeitos atmosféricos para a banda do vermelho; C2= 7,5 é o coeficiente de correção dos efeitos atmosféricos para a banda do azul; e BLUE é a reflectância na banda do azul.

Além desses, existem diversos outros índices já calculados e disponíveis para o uso, como é o caso de vários produtos do satélite MODIS Terra (Moderate Resolution Imaging Spectroradiometer). Tal satélite registra informações muito importantes para estudos

ambientais, como atividade fotossintética e temperatura da superfície terrestre. Tais variáveis se relacionam à diversidade florestal, por exemplo, por estarem associadas aos fatores que explicam a distribuição das espécies como a radiação absorvida, produtividade e estrutura de uma floresta. Tais índices têm apresentado boa correlação com a diversidade arbórea (MEDEIROS et al., 2019; SILVEIRA et al., 2019) e por esse motivo foram implementados nesse estudo.

3. CONSIDERAÇÕES GERAIS

Destaca-se a importância desse estudo para a conservação dos remanescentes florestais de Minas Gerais: fornecerá informações (locais de baixa, média e alta diversidade arbórea) que poderão auxiliar no processo de licenciamento ambiental; será possível analisar se as políticas voltadas para a conservação do estoque de carbono satisfazem a conservação da diversidade de espécies arbóreas do estado e poderá servir de base para criação e gestão de unidades de conservação. Ou seja, o conhecimento da biodiversidade, sua distribuição e condicionantes é importante para o seu monitoramento e gestão, além do fato de atuar na mitigação e redução das mudanças climáticas e desastres ambientais.

REFERÊNCIAS

- ACERBI JÚNIOR, et al. Change detection in Brazilian Savannas using semivariograms derived from NDVI images. **Ciência e Agrotecnologia**, Lavras, v. 39, n. 2, p. 103-109, 2015.
- ASRAR, G. et al. Estimating absorbed photosynthetic radiation and leaf area index from spectral reflectance in wheat 1. **Agronomy Journal**, Madison, v. 76, n. 2, p. 300-306, 1984.
- BRAZIL FLORA GROUP. Growing knowledge: an overview of seed plant diversity in Brazil. **Rodriguésia**, Rio de Janeiro, v. 55, n. 4, 2015.
- CARVALHO M.C. et al. Algoritmos de aprendizagem de máquina na modelagem da distribuição potencial de habitats de espécies arbóreas. **Nativa**, Mato Grosso, v. 7, n. 5, p. 600-606.
- CROWTHER, T. W. et al. Mapping tree density at a global scale. **Nature**, v. 255, n. 7568, p. 201-205, 2015.
- FISHER, R. A.; CORBET, A. S.; WILLIAMS, C. B. The relation between the number of species and the number of individuals in a random sample of animal population. **The Journal of Animal Ecology**, London, v. 12, n. 1, p. 42-58, 1943.
- FLORA DO BRASIL 2020 EM CONSTRUÇÃO**. Jardim botânico do Rio de Janeiro. Disponível em: <<http://floradobrasil.jbrj.gov.br/>>. Acesso em 23 jan. 2020.
- Group on Earth Observations Biodiversity Observation Network - GEO BON. **Essential Biodiversity Variables**. GEO BON, 2020. Disponível em: <<https://geobon.org/ebvs/what-are-ebvs/>>. Acesso em: 23 jan. 2020.
- HE, K. S. et al. Will remote sensing shape the next generation of species distribution models? **Remote Sensing in Ecology and Conservation**, United Kingdom, v. 1, n. 1, p. 4-18, 2015.
- HE, K. S.; ZHANG, J.; ZHANG, Q. Linking variability in species composition and MODIS NDVI based on beta diversity measurements. **Acta Oecologica**, Esquel, v. 35, n.1, p. 14-21, 2009.
- HUBBELL S. P. et al. How many tree species are there in the Amazon and how many of them will go extinct? **Proceedings of the National Academy of Sciences of the United States of America**, EUA, v. 105, p. 11498-11504, 2008.
- HUETE, A. R. A soil-adjusted vegetation index (SAVI). **Remote Sensing of Environment**, Ontario, v. 25, n. 3, p. 295-309, 1988.
- JUSTICE, C. O. et al. The moderate resolution imaging spectroradiometer (MODIS): Land remote sensing for global change research. **IEEE Transactions on Geoscience and Remote Sensing**, Canada, v. 36, n. 4, p. 1228-1249, 1998.
- MAGURRAN, A. E. **Medindo a diversidade biológica**. Curitiba: Ed. da UFPR, 2013.

MARGALEF, R. Homage to Evelyn Hutchinson, or why is there an upper limit to diversity? **Transaction of the Connecticut Academy of Arts and Sciences**, Toronto, v. 44, p. 211-235, 1972.

McINTOSH, R. P. An index of diversity and the relation of certain concepts to diversity. **Ecology**, Washington, v. 48, n. 3, p. 392-404, 1967.

MEDEIROS, E. S. S. et al. Predicting plant species richness with satellite images in the largest dry forest nucleus in South America. **Journal of Arid Environments**, United States of America, v. 166, p. 43-50, 2019.

MENESES, P. R.; ALMEIDA, T. de. (Org.) **Introdução ao processamento de imagens de sensoriamento remoto**. Brasília: UnB, 2012.

MITTERMEIER, R. A. et al. Global biodiversity conservation: the critical role of hotspots. In: ZACHOS, F.; HABEL, J. (Eds.) **Biodiversity Hotspots**. Berlin: Springer, 2011. p. 3-22.

OLIVEIRA-FILHO, A. T.; FONTES, M. A. L. Patterns of floristic differentiation among Atlantic Forest in Southeastern Brazil and the influence of climate. **BIOTROPICA**, Gainesville, v. 32, n. 4, p. 793-810, 2000.

OLIVEIRA-FILHO, A. T. **Catálogo das árvores nativas de Minas Gerais – Mapeamento e inventário da flora nativa e dos reflorestamentos de Minas Gerais**. Lavras: UFLA, 2006.

PENNINGTON, R. T.; LAVIN, M.; OLIVEIRA-FILHO, A. Woody plant diversity, evolution, and ecology in the Tropics: perspectives from Seasonally Dry Tropical Forests. **Annual Review of Ecology, Evolution, and Systematics**, Palo Alto, v. 40, p. 437-457, 2009.

PIELOU, E. C. **Ecological diversity**. New York: Wiley, 1975.

RIBEIRO, J. F.; WALTER, B. M. T. As principais fitofisionomias do bioma Cerrado. In: SANO, S. M.; ALMEIDA, S. P. de; RIBEIRO, J. F. (Eds.). **Cerrado: ecologia e flora**. Planaltina: Embrapa, 2008. Cap. 6, p. 152- 212.

ROCCHINI, D. et al. Satellite remote sensing to monitor species diversity: potential and pitfalls. **Remote Sensing in Ecology and Conservation**, United Kingdom, v. 2, n. 1, p. 25-36, 2015.

ROCCHINI, D.; HERNÁNDEZ-STEFANONI, J. L.; HE, K. S. Advancing species diversity estimate by remotely sensed proxies: a conceptual review. **Ecological Informatics**, Bologna, v. 25, p. 22-28, 2015.

ROUSE, J. W. et al. **Monitoring the vernal advancement of retrogradation (greenwave effect) of natural vegetation**. NASA/GSFCT, Type III, Final Report, Greenbelt, MD, 371 p, 1974.

SCOLFORO, H. F. et al. Spatial distribution of aboveground carbon stock of the arboreal vegetation in Brazilian biomes of Savanna, Atlantic Forest and Semi- Arid woodland. **PLOS ONE**, California, v. 10, n. 6, p. e0128781, 2015.

- SCOTT, W. A.; HALLAM, C. J. Assessing species misidentification rates through quality assurance of vegetation monitoring. **Plant Ecology**, Perth, v. 165, n. 1, p. 101-115, 2003.
- SHANNON, C. E. A mathematical theory of communication. **Bell System Technical Journal**, Canada, v. 27, n. 3, p. 379-423, 1948.
- SILVA, M. V. R. da et al. Aplicação do índice de vegetação ajustado ao solo – SAVI para a identificação de fragmentos de Caatinga em cultivos de *Agave sisalana* Perrine na região semiárida do Brasil. **Anais XVII Simpósio Brasileiro de Sensoriamento Remoto, INPE, Brasil**, p. 5850-5857, 2015.
- SILVEIRA et al. Carbon-diversity hotspots and their owners in Brazilian southeastern Savanna, Atlantic Forest and Semi-Arid Woodland domains. **Forest Ecology and Management**, United States of America, v. 452, p. 117575, 2019.
- SIMPSON, E. H. Measurement of diversity. **Nature**, v. 163, p. 688, 1949.
- SLIK, J. W. F. et al. An estimate of the number of tropical tree species. **Proceedings of the National Academy of Sciences**, Champaign, v. 112, n. 24, p. 7472-7477, 2015.
- Ter STEEGE, H. et al. Hyperdominance in the Amazonian tree flora. **Science**, Washington, v. 342, n. 6156, p. 1243092, 2013.
- TERRA, M. de C. N. S. et al. Tree dominance and diversity in Minas Gerais, Brazil. **Biodiversity and Conservation**, London, v. 26, n. 9, p. 2133-2153, 2017.
- TERRA, M. de C. N. S. et al. Water availability drives gradients of tree diversity, structure and functional traits in the Atlantic-Cerrado-Caatinga transition, Brazil. **Journal of Plant Ecology**, Oxford, p. 1-12, 2018.
- VICENS, R. S.; CRUZ, C. B.M.; RIZZINI, C. M. Utilização de técnicas de sensoriamento remoto na análise da cobertura vegetal da Reserva Florestal de Linhares, ES, Brasil. **Anais IX Simpósio Brasileiro de Sensoriamento Remoto, INPE, Brasil**, p. 1561-1572, 1998.
- VOROVENCI, I. Assessment of some remote sensing techniques used to detect land use/land cover changes in South-East Transilvania, Romania. **Environmental Monitoring and Assessment**, Maine, v. 186, n. 5, p. 2685-2699, 2014.
- XIAO, X. et al. Modeling gross primary production of an evergreen needleleaf forest using MODIS and climate data. **Ecological Applications**, California, v. 15, n. 3, p. 954-969, 2005.

SEGUNDA PARTE - ARTIGO

Artigo 1 – Prediction of tree diversity in Southeast Brazil using remote sensing and machine learning

Maria Cleidivane Alves Batista; Marcela de Castro Nunes Santos Terra; Eduarda Martiniano de Oliveira Silveira; Fausto Weimar Acerbi Júnior

Abstract

Knowledge about the biodiversity of a region is of utmost importance to support actions aimed at sustainable development. Traditional methods based on diversity indices require field campaigns, which are time consuming and expensive. An option to reduce time and costs is modelling using easily obtainable variables, such as remote sensor variables, which have low cost and the ability to monitor large geographical areas in a short time. Another important variables are the climatic and terrain variables since they are strongly associated with the vegetation spatial pattern of the region. Therefore, this study aimed to model and map tree diversity in the state of Minas Gerais, Brazil, using 115 predictor variables and the Random Forest algorithm. In general, the Atlantic Forest domain presented the greatest richness and species diversity. Concerned to modelling, the variables that were most representative for all diversity metrics were precipitation (climatic), gross primary productivity (remote sensing) and depth of the valley (terrain), which are directly associated with tree diversity. As a conclusion, the tree diversity in the state of Minas Gerais increases from North to South and from Northwest to Southeast, being strongly associated with the Morphoclimatic Domains (less diversity in the Caatinga and greater in the Atlantic Forest); the variables selected by the RF represented the spatial patterns of tree diversity in the state of Minas Gerais, which can assist in the decision-making process.

Keywords: Tree diversity. Environmental modelling. Random Forest.

Artigo em preparação para submissão no periódico Biodiversity and Conservation.

Introdução

Conhecer a biodiversidade de uma região compreende importante estratégia de apoio a ações que visam o desenvolvimento sustentável, uma vez que os serviços ecossistêmicos gerados por ela em conjunto com outros componentes do ecossistema proporcionam melhorias ambientais, sociais e econômicas. A biodiversidade garante o equilíbrio do planeta e atua indiretamente nos processos de mitigação das mudanças climáticas, aumentando a resiliência do ecossistema frente às perturbações (Prevedello et al. 2019). Em termos de biodiversidade florestal, sua relação com estoque de carbono e biomassa geralmente é fraca em grandes escalas (Morandi et al. 2018; Silveira et al. 2019) e tendo em vista que os esforços globais para a proteção do clima se baseiam em estoque de carbono, a conservação da diversidade florestal pode estar sendo prejudicada. Dessa forma, percebe-se a necessidade em se conhecer a diversidade florestal para tomar decisões mais acertadas em relação a conservação da biodiversidade e proteção do clima.

Em termos de diversidade arbórea o Brasil é um país rico, contando com 14,5% das espécies registradas no mundo, sendo que a metade dessas espécies são consideradas endêmicas (Beech et al. 2017). O estado de Minas Gerais (MG) conta com aproximadamente 42% da riqueza brasileira de árvores e 6% da riqueza mundial (Terra et al. 2017), sendo mais rico que países como o Equador, México e Madagascar. E há ainda muito a se conhecer, pois no país são registradas anualmente cerca de 250 novas espécies (Brazil Flora Group, 2015), evidenciando o risco de extinção de espécies ainda desconhecidas se a degradação florestal não cessar.

Medidas tradicionais utilizadas para quantificar a diversidade arbórea se baseiam em índices que captam a riqueza (número de espécies) e a uniformidade (similaridade entre as espécies em termos de abundância, número de indivíduos) de uma comunidade. São muitos os índices de diversidade apresentados na literatura e os mais utilizados em florestas são os de Shannon, Simpson, Margalef, equabilidade de Pielou e alfa de Fisher (Pelissari et al. 2018; Terra et al. 2018; Silveira et al. 2019). Os índices possuem a vantagem de sintetizar as informações sobre a diversidade, incluindo uniformidade e riqueza; além de permitirem a comparação entre áreas. Eles são calculados com base em levantamentos de campo, que na maioria das vezes são demorados e dispendiosos. Uma solução para esse problema é a modelagem dos índices de diversidade, conforme demonstrado por Silveira et al. (2019) para o índice alfa de Fisher, utilizando variáveis de fácil obtenção e que sejam correlacionadas com os mesmos, como variáveis de sensoriamento remoto (Rocchini, Hernández-Stefanoni and He 2015; Rocchini et al. 2016; Medeiros et al. 2019), topográficas (Revermann et al 2016; Song and Cao 2017) e climáticas (Kwon et al. 2019).

Tendo disponível grande quantidade de variáveis que podem estar correlacionadas aos índices de diversidade, faz-se necessário definir e selecionar as mais importantes, com maior potencial preditivo. E para isso tem-se os algoritmos de aprendizagem de máquina. De forma especial o algoritmo Random Forest tem se destacado no quesito de modelagem ambiental, uma vez que apresenta fácil ajuste e fornece mais informações sobre a relação entre as variáveis dentro do modelo se comparado a outros algoritmos (Carvalho et al. 2019).

Desta forma, utilizando um banco de dados constituído de variáveis climáticas, topográficas e de sensoriamento remoto, totalizando 115 variáveis preditoras, o objetivo deste estudo foi modelar e mapear a riqueza de espécies de árvores e quatro índices de diversidade: Fisher, Shannon, Simpson e Pielou. Procurou-se responder três questões científicas: (1) Quais variáveis têm maior potencial preditor da diversidade arbórea no estado? (2) Qual o padrão de distribuição espacial da diversidade arbórea por domínio no estado de Minas Gerais? (3) Existe um consenso entre as métricas de diversidade?

Materiais e métodos

Descrição da área de estudo

Este estudo foi conduzido no estado de Minas Gerais, localizado na região sudeste do Brasil (Fig. 1a). A altitude média do estado é cerca de 734 metros, variando de 49 a 2672 metros (Fig. 1c); a precipitação média anual varia de 690 a 2396 milímetros (Fig. 1d), e a temperatura média anual varia de 9,2 a 25,1 °C (Fig. 1e) conforme dados da base WorldClim 2 (Fick and Hijmans 2017), configurando um clima predominantemente tropical seguido de temperado quente conforme a classificação de Köppen (Aw e Cwa respectivamente), com invernos secos e verões chuvosos (Martins et al. 2018).

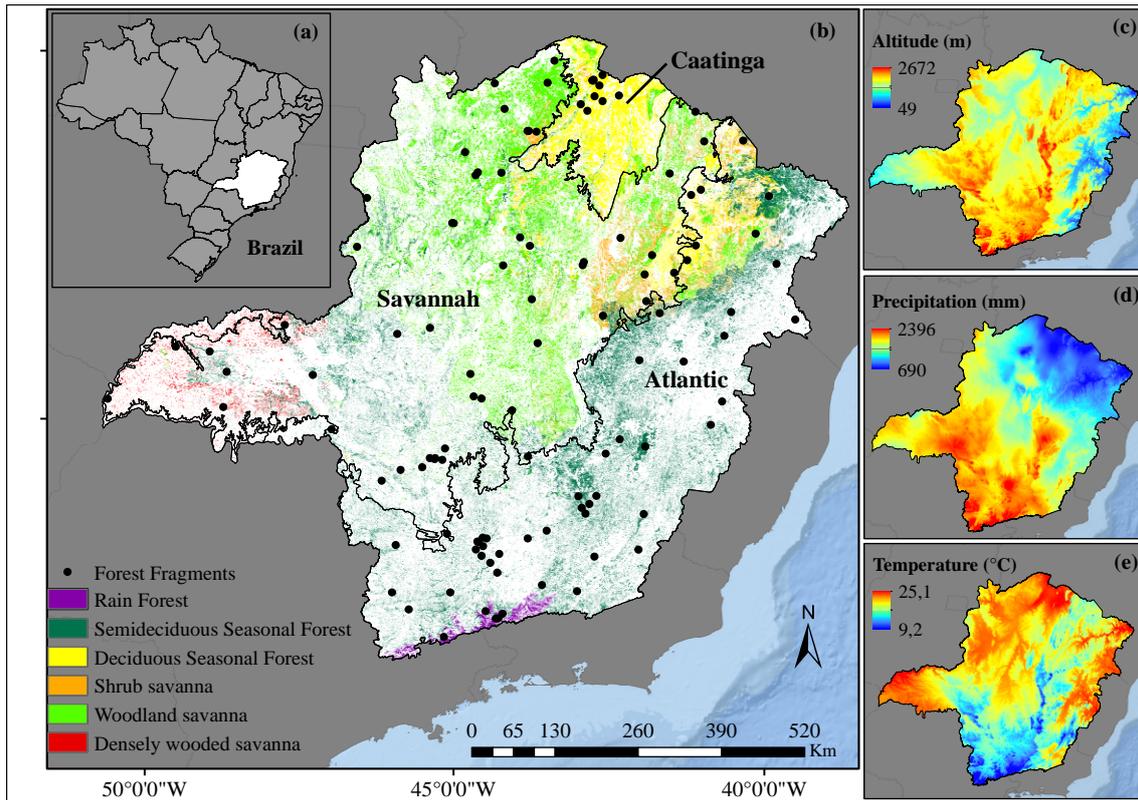


Fig. 1 Localização do estado de Minas Gerais no Brasil (a), seus domínios morfoclimáticos (Atlântico, Cerrado e Caatinga) e fitofisionomias e distribuição espacial dos 115 fragmentos florestais (b), a altitude em metros (c), a precipitação média anual em mm (d) e a temperatura média anual em °C (e)

O estado integra a transição dos domínios morfoclimáticos brasileiros Atlântico, Caatinga (Floresta semiárida) e Cerrado (Fig. 1b) que correspondem a 41%, 2% e 57% das áreas dos remanescentes de vegetação nativa, respectivamente (Scolforo et al. 2015). Dentro de cada um desses domínios, devido a variações das características climáticas, do solo e relevo, ocorrem as diferentes fitofisionomias, que vão de formações florestais a campestres.

O domínio Atlântico compreende a segunda maior área de floresta tropical úmida da América do Sul, ficando atrás do domínio Amazônico. É composto pelas fitofisionomias Floresta Ombrófila e Floresta Estacional Semidecidual conforme o gradiente longitudinal negativo de chuva à medida que aumenta a distância costa-interior do país, que proporciona a ocorrência de formações mais abertas (Oliveira-Filho and Fontes 2000). A precipitação pode variar de 800 mm a 4500 mm e alguns locais dependem exclusivamente da chuva orográfica (Ribeiro and Walter 2008). As Florestas Ombrófilas, também denominadas de sempre-verdes são caracterizadas por altas temperaturas e alta precipitação bem distribuída no ano, praticamente inexistindo período seco. Já as Florestas Estacionais Semidecíduais perdem entre 20 e 50% das suas folhas na estação seca (IGBE 2012). O domínio possui elevada riqueza de espécies vegetais e animais com elevada taxa de endemismo e ameaças e esse fato coloca o

domínio na lista mundial dos hotspots prioritários para a conservação da biodiversidade (Mittermeier et al. 2011). Em Minas Gerais o domínio Atlântico ocorre nas regiões centro-sul e leste do estado, predominando a Floresta Estacional Semidecidual uma vez que a Floresta Ombrófila se restringe ao extremo sul do estado (Terra et al. 2017).

O domínio do Cerrado também é um hotspot mundial de biodiversidade. No Brasil, é o segundo maior domínio, ficando também atrás do domínio amazônico. Engloba formações florestais (predominância de espécies arbóreas), savânicas (árvores e arbustos sobre um estrato herbáceo sem a formação de um dossel contínuo) e campestres (com predomínio de espécies herbáceas a algumas arbustivas, sem árvores) (Ribeiro and Walter 2008). Possui solos em sua maioria pobres, ácidos e arenosos e a temperatura do mês mais frio é superior a 18 °C. A precipitação média anual é de aproximadamente 1500 mm variando de 750 a 2000 mm, concentrados nos meses de outubro a março (Ribeiro and Walter 2008). Conforme os mesmos autores, a distribuição da flora no Cerrado é influenciada pelo clima, solo, disponibilidade de água e nutrientes, geomorfologia e topografia, profundidade do lençol freático, latitude, pela frequência de queimadas e outras atividades antrópicas.

No estado de Minas Gerais o domínio do Cerrado abrange a maior parte do território, na região centro-oeste e norte apresentando as fitofisionomias Cerradão – formação florestal com dossel contínuo (de 8 a 15 metros de altura) e cobertura arbórea variando de 50% a 90%, diminuindo na estação seca –, Cerrado Sentido Restrito – formação savânica com árvores baixas (máximo 8 metros), tortuosas, com ramificações irregulares, cascas com cortiça espessa e folhas coriáceas como estratégia a condições de seca; sua cobertura arbórea não ultrapassa 70% – e formações campestres. O domínio apresenta diversos níveis de caducifolia na época seca.

O domínio das florestas semiáridas ou Caatingas é o único domínio exclusivo do Brasil. É quase totalmente restrito (98,8%) à região nordeste do país, ocorrendo o restante em Minas Gerais. É representado pela fitofisionomia Floresta Estacional Decidual, adaptada à maior limitação meteorológica do país com elevadas temperaturas e baixa precipitação (de 250 mm a 800 mm, raramente atingindo 1000 mm), que é concentrada em 3 a 5 meses (Ribeiro and Walter 2008). É caracterizada por uma vegetação lenhosa com dossel descontínuo que perde mais de 50% das folhas na época seca; pela presença de plantas suculentas e espinhosas e espécies com folhas micrófilas (apenas uma nervura) como estratégia às condições adversas (Santos et al. 2012). Devido aos processos de adaptação a essas condições, a Caatinga possui elevada taxa de endemismo (Pennington, Lavin, Oliveira-Filho 2009), e é considerada por Gil (2002) como uma das áreas selvagens da Terra. Em Minas Gerais o domínio ocorre no norte do estado.

Conjunto de dados

Dados de diversidade arbórea

O conjunto de dados consiste em 2755 parcelas de 1000 m² (10x100m) referentes a 115 fragmentos florestais (Fig. 1b) provenientes do Inventário Florestal de Minas Gerais realizado no período de 2006 a 2008 pela Universidade Federal de Lavras em convênio com o Instituto Estadual de Florestas (Scolforo et al. 2006). O número de parcelas por fragmento foi em média 24, variando de 5 a 128. Dessas parcelas foram obtidas as medições do diâmetro a altura do peito (DAP) a 1,30 m de altura do solo, da altura total e identificação taxonômica (família, gênero e espécie) de todas as árvores com DAP maior ou igual a 5 cm, totalizando 1586 espécies identificadas. O número total de árvores amostradas foi de 300.484. A distribuição de espécies, a porcentagem de não identificadas e morfoespécies por fitofisionomia são descritas na Tabela 1. Vale ressaltar que o total de espécies apresentado na Tabela 1 inclui espécies compartilhadas entre as fitofisionomias, ou seja, existem 1840 espécies em comum nesses ambientes.

Tabela 1 Número de fragmentos, árvores e espécies por fitofisionomia amostrados no estado de MG. O total de espécies inclui as compartilhadas.

Fitofisionomia	Nº de fragmentos	Nº de árvores	Nº de espécies	Média de árvores não identificadas (%)	Média de morfoespécies (%)
Formações campestres	4	8642	253	3,07	0,55
Cerrado Sensu Stricto	24	76015	586	7,77	3,00
Cerradão	4	11422	322	10,77	5,62
Flor. Est. Decidua	14	49644	418	2,19	3,31
Flor. Est. Semidecidual	63	139099	1325	5,11	3,30
Flor. Ombrófila	6	15662	482	2,98	0,69
TOTAL	115	300484	3386	6,38	4,24

A distribuição das parcelas dentro dos domínios foi a seguinte: 1475 parcelas (69 fragmentos) no domínio Atlântico (1372 na fitofisionomia Floresta Estacional Semidecidual e 103 na Floresta Ombrófila); 846 parcelas (32 fragmentos) no domínio do Cerrado (67 na fitofisionomia Cerradão, 626 na fitofisionomia Cerrado Sensu

Stricto e 153 nas formações campestres); e 434 parcelas (14 fragmentos) no domínio da Caatinga referentes à fitofisionomia Floresta Estacional Decidual.

A partir da identificação taxonômica das árvores foi possível calcular índices de diversidade, que se baseiam na riqueza e uniformidade (relacionada à abundância da espécie), para cada parcela. Os índices analisados nesse estudo, além do número total de espécies (riqueza), foram o índice alfa de Fisher (Fisher), índice de diversidade de Shannon (H'), o índice de Simpson (1-D) e o índice de equabilidade de Pielou (J'). Tais índices foram calculados no programa RStudio (R Core Team 2019) utilizando o pacote vegan, exceto o J' , que foi calculado na planilha eletrônica Excel.

Por mais que a riqueza seja uma medida de diversidade bastante antiga e intuitiva, não pode ser considerada sinônimo de biodiversidade (Magurran 2013), pois a abundância de indivíduos em cada espécie varia e uma comunidade uniforme onde todas as espécies são igualmente abundantes é mais biodiversa do que uma comunidade onde haja dominância de algumas espécies; por isso a inclusão da uniformidade nas medidas de diversidade é importante. Um dos primeiros trabalhos a descrever matematicamente a relação entre riqueza e abundância foi Fisher et al. (1943), colocando a série logarítmica de Fisher dada pela fórmula $S = \alpha \ln(1 + N/\alpha)$, onde S é o número total de espécies, N é o número total de indivíduos e α é o índice da série logarítmica de Fisher (alfa de Fisher) que é calculado baseado na relação N/S da amostra. A vantagem desse índice é que ele não é afetado pelo tamanho da amostra, desde que N seja maior que 1000 (Magurran 2013), permitindo assim a comparação entre estudos com diferentes áreas de amostragem. Em Minas Geraia, Silveira et al. (2019) encontraram valores de alfa de Fisher variando de 2 (floresta estacional decidual) a 51,6 (floresta estacional Semidecidual).

O índice de Simpson (D) é um índice de diversidade que enfatiza a componente dominância/uniformidade em detrimento da componente riqueza (Magurran 2013). Ele calcula a probabilidade de dois indivíduos quaisquer, retirados aleatoriamente de uma comunidade infinitamente grande, pertencerem à mesma espécie (Simpson 1949). É dado pela fórmula:

$$D = \sum_{i=1}^S \left[\frac{n_i(n_i - 1)}{N(N - 1)} \right]$$

Onde n_i é o número de indivíduos na i -ésima espécie, S é o número total de espécies e N é o número total de indivíduos. Os valores do índice de Simpson variam de 0 a 1, sendo que quanto mais próximo de 1, menor a diversidade (maior a dominância e menor a uniformidade). Por isso o índice é normalmente representado por 1 - D (que é forma usada nesse estudo) e dessa forma quanto maior o valor, maior a diversidade. O índice de Simpson,

segundo Magurran (2013) é uma das medidas de diversidade mais significativas e robustas disponíveis, ficando na frente do mais popular índice de Shannon, que é sensível à variação do tamanho amostral.

O índice de Shannon (1948) assume que indivíduos são aleatoriamente amostrados de uma comunidade infinitamente grande e que todas as espécies são representadas na amostra. Dessa maneira, existe um erro associado ao índice, sendo que geralmente não se amostram todas as espécies de uma comunidade. É definido pela fórmula:

$$H' = - \sum_{i=1}^S \frac{n_i}{N} \ln \frac{n_i}{N}$$

Em que n_i , S e N foram definidos anteriormente. O valor de H' permanece entre 1,5 e 3,5 e raramente ultrapassa 4 (Margalef 1972). Quanto maior o valor, maior a diversidade da comunidade. O índice de Shannon também leva em consideração a uniformidade na diversidade das espécies. Porém é possível calcular uma medida de uniformidade separadamente, a uniformidade de Pielou (J'), baseada na diversidade máxima (todas as espécies igualmente abundantes). O índice de uniformidade de Pielou (1975) é dado por:

$$J' = \frac{H'}{H_{\max.}} = \frac{H'}{\ln S}$$

Onde $H_{\max.}$ é a diversidade máxima possível, H' é o índice de diversidade de Shannon e S é o número total de espécies. Quanto maior o J' , maior a uniformidade e consequentemente, maior a diversidade. Corsini et al. (2014) encontraram valores de Shannon variando de 2,24 a 4,52 e Pielou variando de 0,62 a 0,85 para florestas da região nordeste de Minas Gerais, com as florestas estacionais semidecíduais apresentando os maiores valores e florestas decíduais os menores valores.

Variáveis preditoras

O número de variáveis preditoras utilizadas para a modelagem dos índices de diversidade foi de 115, provenientes de dados climáticos, topográficos e espectrais. As variáveis climáticas (2) foram obtidas da base WorldClim 2 (Fick and Hijmans 2017) com resolução espacial de 1 km². Essas variáveis - temperatura média anual em °C (Bio 1) e precipitação anual em mm (Bio 12) - têm relação direta com o padrão de distribuição e composição das espécies (Oliveira-Filho and Fontes 2000; Terra et al. 2018), estando mais associadas a sobrevivência das mesmas. A escolha dessas duas variáveis climáticas para a modelagem é pelo fato de além de derivarem as demais variáveis disponíveis na base, elas são frequentemente selecionadas e utilizadas como boas preditoras na modelagem ambiental (Silveira et al. 2018, 2019). As variáveis topográficas (12) foram extraídas do modelo digital

de elevação (DEM) com resolução espacial de 30 m obtido da Missão Topográfica Radar Shuttle (SRTM), utilizando o software SAGA GIS (Conrad et al. 2015) e estão descritas na Tabela 2. As informações sobre o terreno podem ajudar a prever padrões da vegetação, influenciando sua diversidade, uma vez que representam dados relacionados à capacidade de retenção de água, altitude, gradiente de declividade, sombreamento e erosão que podem estar associados ao tipo de vegetação. As variáveis espectrais, por sua vez, foram procedentes dos satélites Landsat 5 TM (9) e de seis produtos do sensor MODIS (Tabela 3), obtidos por meio do United States Geological Survey for Earth Resources Observation and Science (USGS) para vários dias (dia juliano 17 a 353) do ano de 2008 (realização do inventário florestal), a fim de captar a variação por sazonalidade, quando existir, totalizando 92 variáveis. As variáveis de sensoriamento remoto, em especial os índices de vegetação, são correlacionadas com a diversidade florestal pois identificam características e padrões de reflexão e absorção da energia pela vegetação, que por sua vez podem auxiliar na identificação da distribuição espacial das espécies (Rocchini, Hernández-Stefanoni and He 2015). Todos os dados (raster) foram reamostrados para 100 m x 100 m (1 hectare) para apresentarem a mesma resolução espacial e projetados para ‘Albers Equal Area Conic’ pelo fato dessa projeção preservar a área e forma das imagens (Duro et al. 2012) e extraídos os valores para cada parcela para realizar a modelagem.

Tabela 2 Descrição das variáveis topográficas utilizadas na modelagem da diversidade arbórea de Minas Gerais

Variáveis topográficas	Descrição
Modelo digital de elevação (DEM)	Representa a altitude (m) de cada pixel da área de estudo
Aspecto (ASP)	Orientação de inclinação em direção ao Norte (WILSON; GALLANT, 2000)
Hillshade analítico (HILS)	Representa o ângulo entre a luz solar e a superfície do terreno
Nível base de rede de canal (CNBL)	Distância até um nível base de rede de canal
Fator LS (LSF)	Fator de comprimento de declive (DESMET; GOVERS, 1996)
Índice de multirresolução do nivelamento do topo do cume (MRRTF)	Identifica locais planos e em altitude elevada (GALLANT; DOWLING, 2003)
Índice de multirresolução do nivelamento do fundo do vale (MRVBF)	Identifica locais planos e baixos (GALLANT; DOWLING, 2003)
Posição de inclinação relativa (Slope)	Posição de cada célula em relação às áreas superiores e inferiores
Índice de posição topográfica (TPI)	Comparação de cada pixel com seus valores vizinhos (GUISAN; WEISS; WEISS, 1999)

Índice de umidade topográfica (TWI)	Mostra a viabilidade em acumular água de um local (BEVEN; KIRKBY, 1979)
Profundidade do vale (VD)	Distância vertical em relação ao vale
Distância vertical da rede de drenagem (VDCN)	Distância vertical de cada pixel em relação à rede de drenagem

Tabela 3 Descrição das variáveis espectrais utilizadas na modelagem da diversidade arbórea de Minas Gerais. Bandas individuais e índices do satélite Landsat 5 e de 6 produtos do sensor MODIS

Dado de sensoriamento remoto	Variável	Definição	Resolução espacial (m)
Landsat 5 TM	Banda 1	Azul	30
	Banda 2	Verde	
	Banda 3	Vermelho	
	Banda 4	Infravermelho próximo	
	Banda 5	Infravermelho de ondas curtas 1	
	Banda 7	Infravermelho de ondas curtas 2	
	EVI	Índice de vegetação melhorado	
	SAVI	Índice de vegetação ajustado ao solo	
	NDVI	Índice de vegetação por diferença normalizada	
MOD09Q1	Banda 1	Reflexão de superfície – 8 dias, vermelho	250
	Banda 2	Reflexão de superfície – 8 dias, infravermelho próximo	
MOD11A2	LSTd	Temperatura diurna diária da superfície – 8 dias	1000
	LSTn	Temperatura noturna diária da superfície – 8 dias	
	Banda 31	Emissividade	
	Banda 32	Emissividade	
MOD13Q1	Banda 1	Vermelho	250
	Banda 2	Infravermelho próximo	
	Banda 3	Azul	
	Banda 7	Infravermelho médio	
	NDVI	Índice de vegetação por diferença normalizada	
MOD15A2H	EVI	Índice de vegetação melhorado	500
	LAI	Índice de área foliar	
	FPAR	Fração da radiação fotossinteticamente ativa absorvida pela vegetação	
MOD17A2H	GPP	Produtividade primária bruta – 8 dias	500
	PSN	Fotossíntese líquida – 8 dias	
MOD44B	Tree cover (%)	Cobertura vegetal contínua	250

Modelagem da diversidade arbórea

A modelagem procurou determinar quais variáveis (climáticas e/ou topográficas e/ou de sensoriamento remoto) influenciam e determinam o padrão de cada métrica de diversidade (riqueza, índice alfa de Fisher, de Shannon, de Simpson e Pielou) no estado de Minas Gerais.

Para cada índice de diversidade haviam 115 variáveis preditoras como descrito anteriormente. Para reduzir esse número, foram removidas as variáveis altamente correlacionadas entre si ($>0,9$) e que possuíam menor correlação com o índice de diversidade em questão. Esse procedimento evita que o Random Forest selecione variáveis que possuam o mesmo potencial preditivo, além de diminuir o tempo de processamento. Dessa forma, das 115 variáveis restaram 71 variáveis para riqueza e índices alfa de Fisher e Shannon, 70 para o índice de Simpson e 72 para o índice de Pielou. A descrição de cada variável utilizada no processo de modelagem é apresentada no material suplementar 1.

Para a realização da modelagem foi utilizado o algoritmo Random Forest (Breiman 2001) disponível no pacote randomForest do software RStudio (R Core Team 2019). Ele tem se destacado em relação a outros algoritmos de aprendizagem de máquina no que diz respeito a modelagem ambiental (Carvalho et al. 2019, Reis et al. 2018). Os dados foram organizados em classes onde 70% dos dados de cada classe foram usados para treinar o algoritmo, enquanto os 30% restantes foram utilizados para validar o desempenho do modelo. Foram computados a raiz do erro quadrático médio percentual (RMSE%) e o coeficiente de determinação (R^2) para avaliar a capacidade de explicação do modelo gerado. Os gráficos de dispersão entre os valores observados dos índices de diversidade e os valores estimados podem ser observados no material suplementar 2.

Do conjunto de variáveis preditoras, o algoritmo precisa selecionar as mais importantes para cada métrica de diversidade. O método de seleção de variáveis utilizado se baseia na remoção recursiva das variáveis em ordem crescente de medida de importância (%IncMSE - aumento do erro médio quadrático, em %), que é fornecida pela função 'importance' do próprio pacote randomForest. Para o cálculo dessa medida, permuta-se os valores da variável de interesse entre exemplos do conjunto de teste e analisa-se o erro. A diferença do erro entre os dados originais e os permutados é a medida de importância (Carvalho et al. 2019, Breiman 2001). A cada iteração são eliminadas as variáveis menos importantes (20% das de menor %IncMSE) e construída uma Random Forest com as variáveis restantes. O parâmetro ntree (número de árvores) foi fixado em 1000, uma vez que pode ser arbitrariamente grande (Guan et al. 2013) a depender do limite de memória do computador. Já o parâmetro mtry

(número de variáveis amostradas aleatoriamente em cada divisão) foi obtido por meio da função ‘tuneRF’ que procura o valor ideal que apresente o menor erro OOB (Out-of-Bag) – erro médio de cada treinamento.

Predição dos índices de diversidade

Uma vez ajustados os modelos finais do Random Forest, as variáveis selecionadas foram aplicadas para todo o estado de Minas Gerais para prever as métricas de diversidade com resolução de 1 hectare, conforme esquema da Fig. 2. Depois, fez-se um ‘recorte’ dos mapas de cada métrica apenas para áreas com vegetação nativa conforme o mapa de cobertura do solo de MG fornecido por Carvalho et al. (2006).

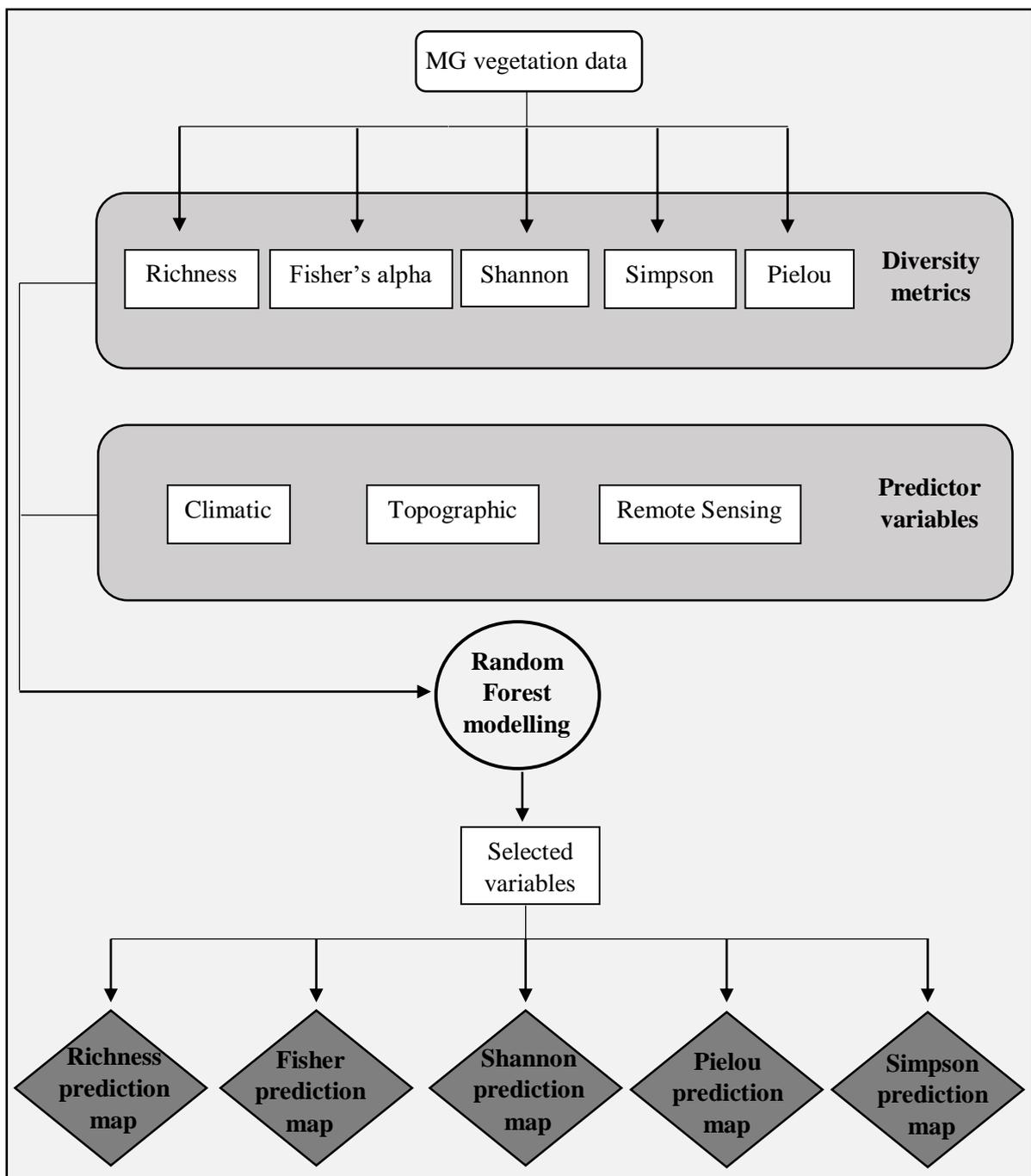


Fig. 2 Esquema mostrando os passos seguidos para elaboração dos mapas de predição das métricas de diversidade em Minas Gerais

Resultados

Diversidade arbórea

A Tabela 4 apresenta a média e coeficiente de variação (%) por parcela das medidas de diversidade observadas por fitofisionomia. Essas informações referentes às variáveis preditoras selecionadas podem ser observadas no material suplementar 3. Em relação a riqueza, a Floresta Ombrófila apresentou o maior número médio de espécies; já a Floresta Estacional Decidual apresentou a menor média de riqueza. Em relação ao índice alfa de Fisher, o maior e menor valor foram para Floresta Estacional Semidecidual e Floresta Estacional Decidual respectivamente. As fitofisionomias com maiores e menores valores do índice de Shannon foram Floresta Ombrófila e Floresta Estacional Decidual. Quanto ao índice de Simpson o maior valor foi para Floresta Ombrófila e o menor para Cerradão. E para a uniformidade de Pielou, a maior uniformidade foi para formações campestres e a menor foi para Cerradão. Quanto à variação de diversidade entre os fragmentos por fitofisionomia, Floresta Estacional Semidecidual exibiu a maior variação, possuindo fragmentos com os menores e com os maiores valores de todas as métricas de diversidade (Tabela 5).

Tabela 4 Média e coeficiente de variação (CV%) dos índices de diversidade das parcelas por fitofisionomia no estado de Minas Gerais

	Média (CV%)				
	Riqueza	Fisher	H'	D	J'
Formações campestres	17,37 (47,40)	9,93 (47,02)	2,40 (19,73)	0,87 (8,23)	0,88 (6,54)
Cerrado Sensu Stricto	23,93 (30,04)	10,09 (37,69)	2,58 (14,24)	0,87 (8,02)	0,83 (10,74)
Cerradão	33,43 (36,66)	13,55 (46,56)	2,64 (35,28)	0,82 (28,18)	0,75 (25,96)
Flor. Est. Decidual	17,00 (36,46)	7,37 (59,88)	2,21 (18,72)	0,83 (9,91)	0,80 (10,86)
Flor. Est. Semidecidual	28,07 (53,79)	17,04 (64,08)	2,71 (23,91)	0,88 (12,62)	0,85 (11,77)
Flor. Ombrófila	35,31 (32,07)	16,32 (40,97)	2,98 (12,42)	0,92 (4,75)	0,85 (6,35)
Geral	25,20 (51,30)	13,43 (68,77)	2,59 (22,46)	0,87 (11,70)	0,84 (11,82)

Tabela 5 Valores mínimos e máximos observados por fragmentos para cada fitofisionomia do estado de Minas Gerais

		Riqueza	Fisher	H'	1-D	J'
Formações campestres	Mínimo	51	7,52	2,15	0,84	0,87
	Máximo	131	13,72	2,8	0,91	0,88
	Mínimo	38	4,37	1,87	0,73	0,62

Cerrado Sensu Stricto	Máximo	187	14,9	2,82	0,91	0,91
Cerradão	Mínimo	86	10,84	2,09	0,64	0,60
	Máximo	169	16,4	3,18	0,94	0,85
Flor. Est. Decidual	Mínimo	30	3,97	1,59	0,73	0,72
	Máximo	123	15,74	2,56	0,88	0,94
Flor. Ombrófila	Mínimo	65	8,85	2,47	0,86	0,80
	Máximo	267	21,97	3,28	0,94	0,89
Flor. Est. Semidecidual	Mínimo	9	0,74	0,32	0,14	0,24
	Máximo	380	50,03	3,79	0,96	0,94

Seleção de variáveis e importância

As variáveis selecionadas ordenadas por ordem de importância (aumento percentual do erro quadrático médio) para cada métrica de diversidade estão apresentadas na Tabela 6. Em seguida, realizou-se a modelagem e as estatísticas coeficiente de determinação e raiz do erro quadrático médio estão apresentadas na Tabela 7.

A variável climática Bio12 (precipitação anual) foi selecionada para todas as métricas. Do conjunto de variáveis topográficas, foram selecionadas VD, CNBL, VDCN e Slope, sendo que a variável VD esteve ausente apenas no índice de equabilidade de Pielou. Em relação às variáveis de sensoriamento remoto, do satélite Landsat 5 foram selecionadas B1 e SAVI; já do satélite MODIS foram selecionadas GPP, LSTn, LSTd, LAI, EVI, banda azul, FPAR e ref2, sendo que a variável GPP foi a mais frequente não aparecendo apenas no índice alfa de Fisher. Observa-se também que as GPP's selecionadas foram de dias julianos referentes à períodos mais quentes e chuvosos do ano (final de agosto a novembro), época em que a produtividade aumenta (Souza et al. 2014).

Tabela 7 Coeficiente de determinação e raiz do erro quadrático médio obtidos da modelagem das métricas de diversidade para o estado de Minas Gerais

	Riqueza	Fisher	H'	1-D	J'
R ²	0,23	0,58	0,40	0,50	0,59
RMSE (%)	25,59	48,09	13,37	8,14	8,44

Mapeamento da diversidade arbórea em MG

Utilizando as variáveis selecionadas para cada índice conforme apresentadas na Tabela 5, foram elaborados os mapas da riqueza, índice de Fisher, Shannon, Simpson e Pielou (Figuras 3, 4, 5, 6 e 7 respectivamente).

Tabela 6 Variáveis selecionadas e respectivos %IncMSE para cada métrica de diversidade

Richness		Fisher's alpha index		Shannon index		Simpson index		Pielou index	
Variables	%IncMSE	Variables	%IncMSE	Variables	%IncMSE	Variables	%IncMSE	Variables	%IncMSE
17_241GPP	75,04	Bio12	91,41	17_241GPP	78,40	Bio12	53,46	17_289GPP	62,15
VD	74,00	VD	68,09	Bio12	71,89	17_321GPP	48,52	13_17blue	58,17
SAVI	64,60	15_209LAI	61,34	VD	71,69	VD	47,56	15_209FPAR	57,72
11_145LSTn	63,00	13_145EVI	59,57	B1	61,12	11_113LSTd	46,89	Bio12	56,30
CNBL	62,51			VDCN	60,87	11_145LSTn	46,73	9_257ref2	55,00
Bio12	62,21			13_17blue	60,86	SAVI	43,64	11_145LSTn	53,96
11_113LSTd	59,55			CNBL	58,95	Slope	39,76	CNBL	48,85

Where: XX_YYY is MODIS product XX_julian days YYY; GPP is Gross Primary Production; VD is Valley Depth; SAVI is Soil-adjusted Vegetation Index; LSTn is Nighttime Land Surface Temperature; CNBL is Channel Network Base Level; Bio12 is Annual precipitation (mm); LSTd is Daytime Land Surface Temperature; LAI is Leaf Area Index; EVI is Enhanced Vegetation Index; B1 is Band 1 from Landsat 5; VDCN is Vertical Distance to Channel Network; 13_17blue is band 3 from MODIS product 13, julian day 17; FPAR is Fraction of photosynthetically active radiation absorbed by vegetation; ref2 is Surface Reflectance Band 2 (NIR) and Slope is Position of each cell relative to upper and lower areas

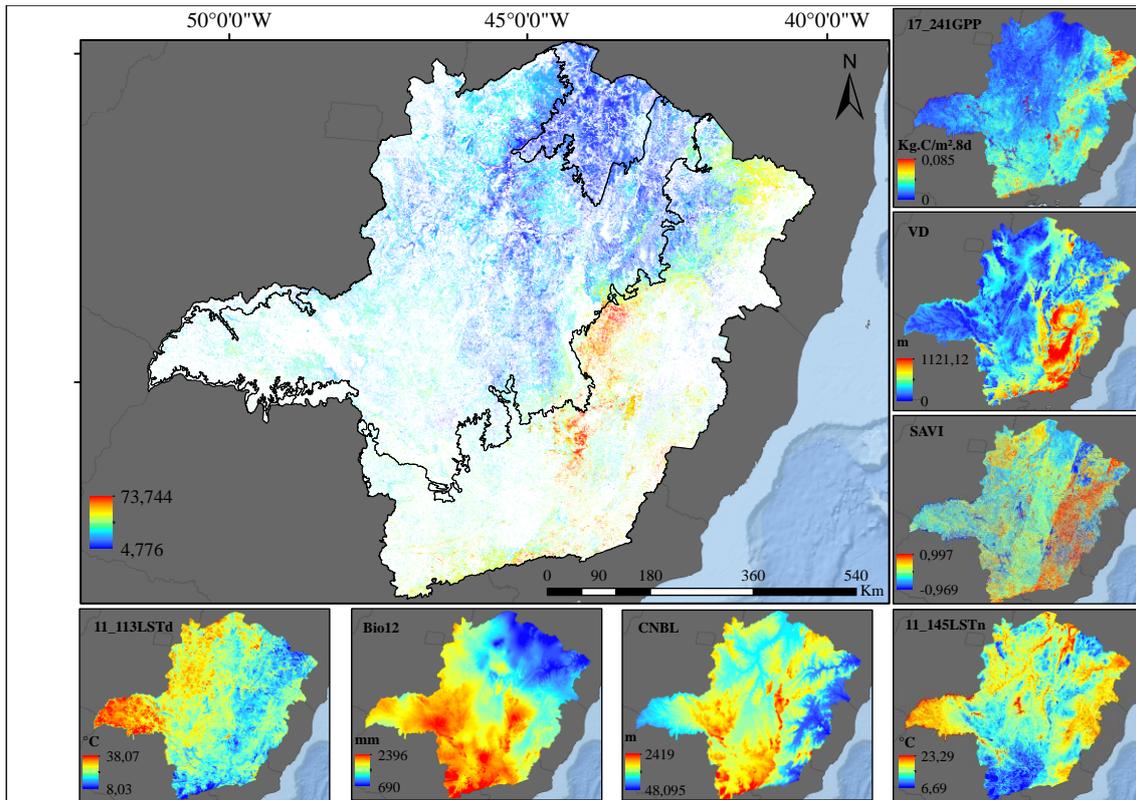


Fig. 3 Mapeamento da riqueza para o estado de MG, com suas respectivas variáveis selecionadas na modelagem e organizadas por ordem de importância: produtividade primária bruta (GPP), profundidade do vale (VD), índice de vegetação justado ao solo (SAVI), temperatura noturna da superfície terrestre (LSTn), nível base da rede de canais (CNBL), precipitação média anual (Bio12) e temperatura diurna da superfície terrestre (LSTd)

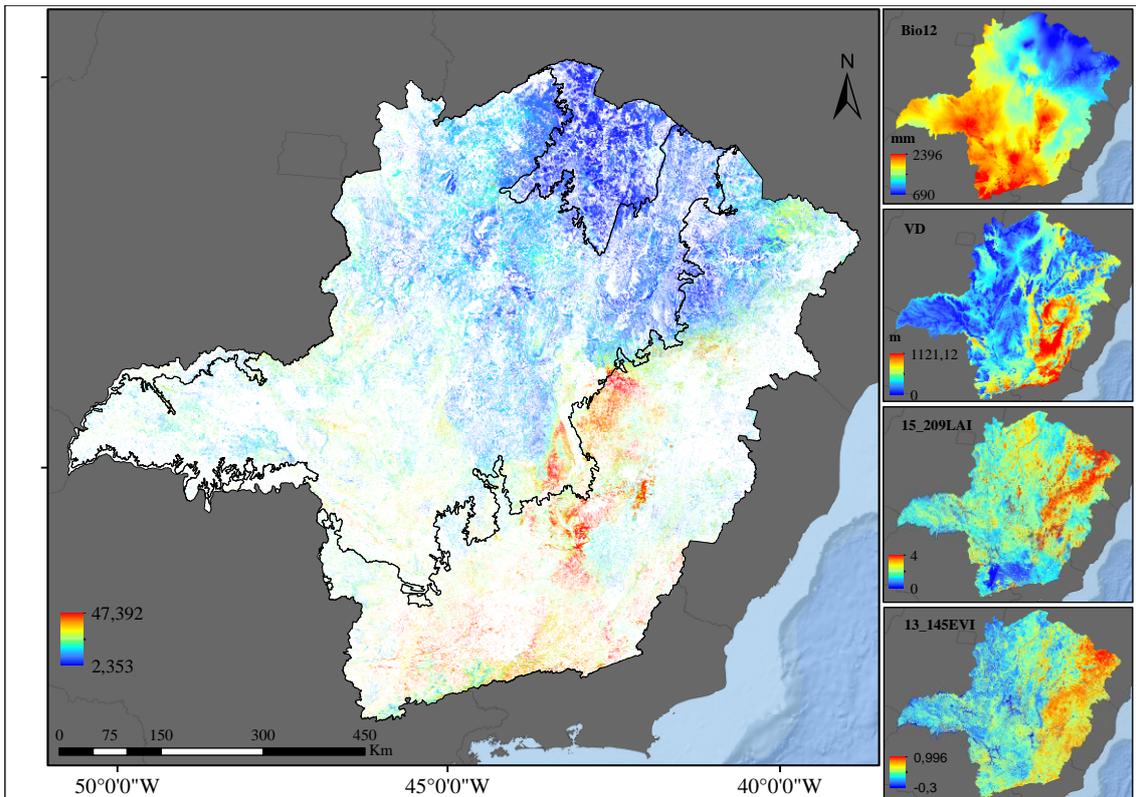


Fig. 4 Mapeamento do índice alfa de Fisher para o estado de MG com suas respectivas variáveis selecionadas na modelagem e organizadas por ordem de importância: precipitação média anual (Bio12), profundidade do vale (VD), índice de área foliar (LAI) e índice de vegetação melhorado (EVI)

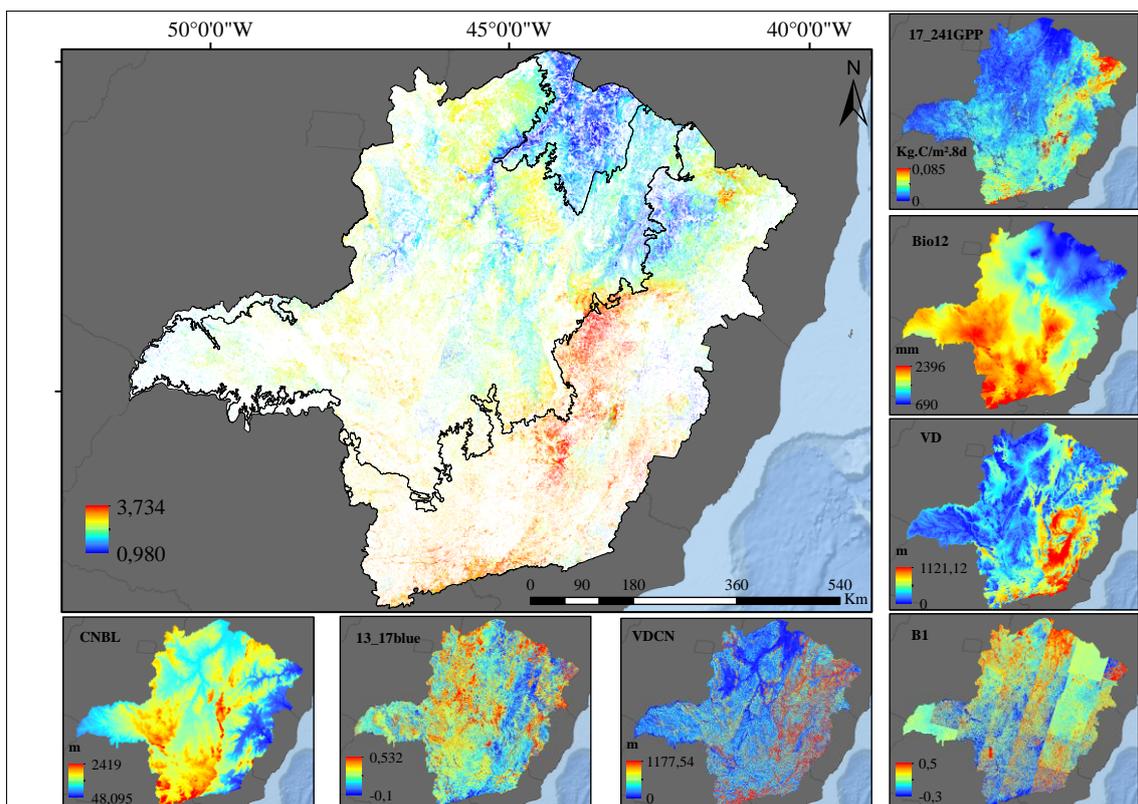


Fig. 5 Mapeamento do índice de Shannon para o estado de MG com suas respectivas variáveis selecionadas na modelagem e organizadas por ordem de importância: produtividade primária bruta (GPP), precipitação média anual (Bio12), profundidade do vale (VD), banda azul do satélite Landsat 5 (B1), distância vertical da rede de drenagem (VDCN), banda azul do produto 13 do sensor MODIS (blue) e nível base da rede de canais (CNBL)

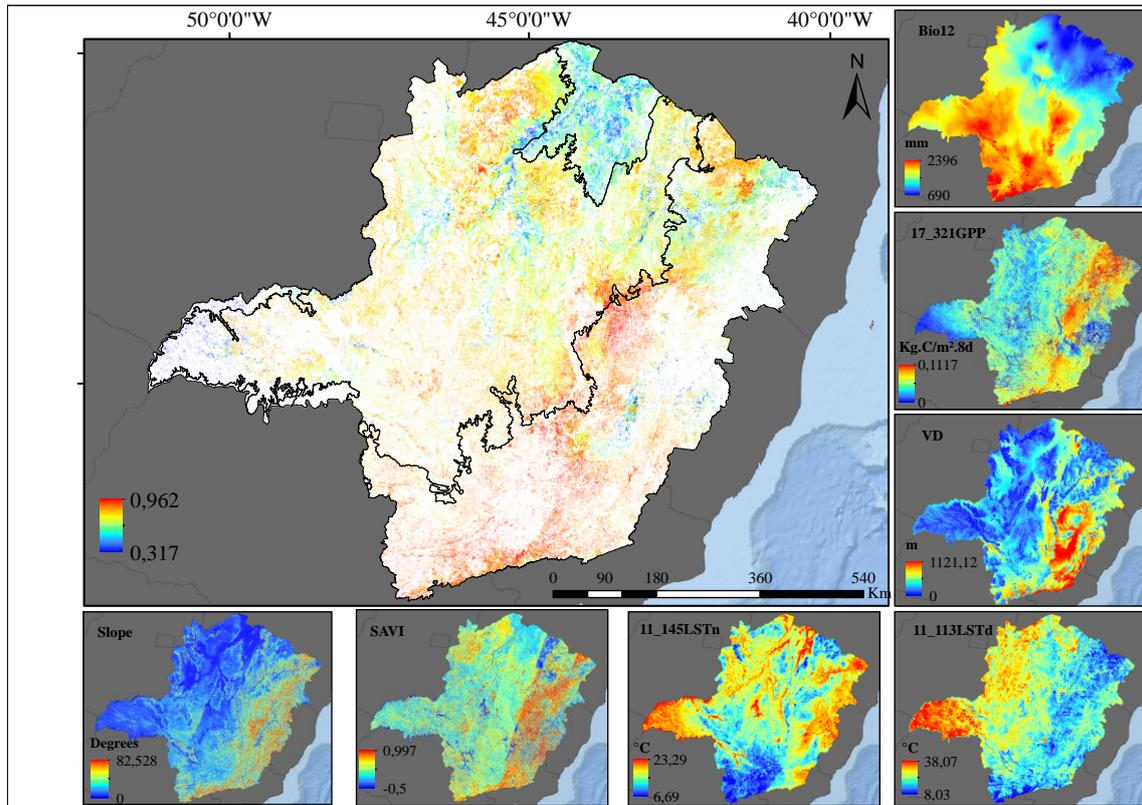


Fig 6 Mapeamento do índice de Simpson para o estado de MG com suas respectivas variáveis selecionadas na modelagem e organizadas por ordem de importância: precipitação média anual (Bio12), produtividade primária bruta (GPP), profundidade do vale (VD), temperatura diurna da superfície terrestre (LSTd), temperatura noturna da superfície terrestre (LSTn), índice de vegetação justado ao solo (SAVI) e inclinação do terreno (Slope)

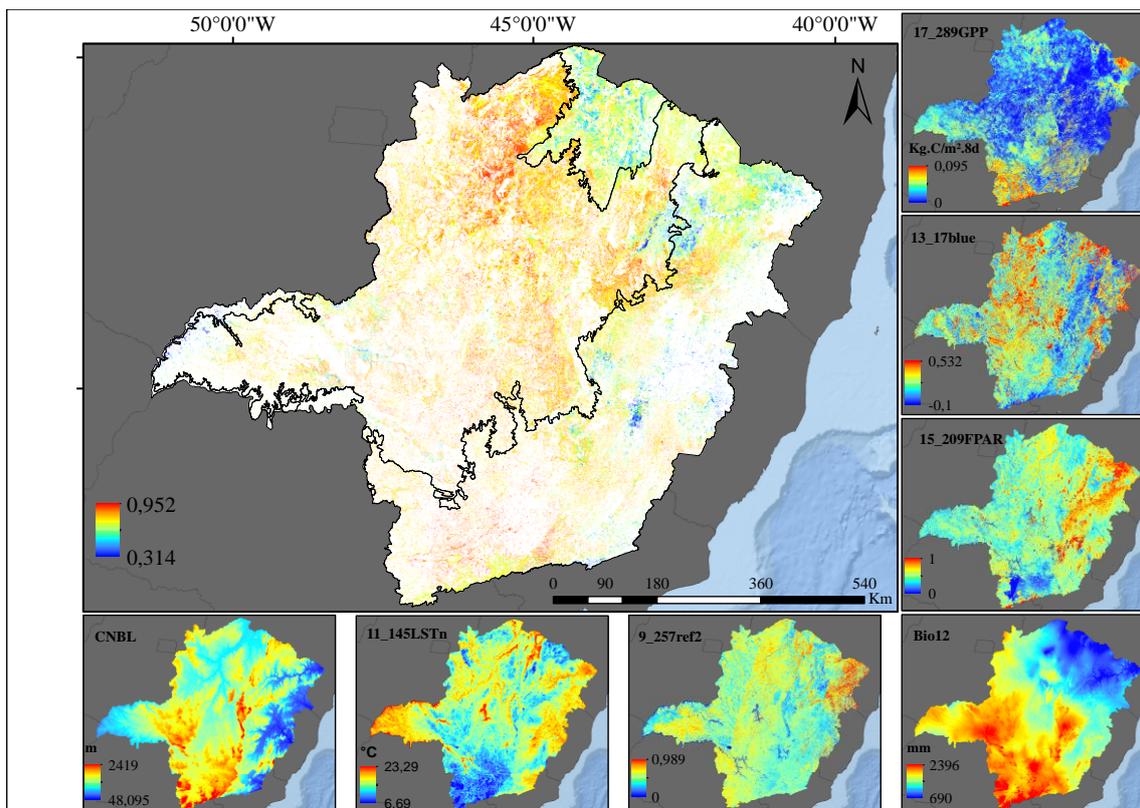


Fig 7 Mapeamento do índice de equabilidade de Pielou para o estado de MG com suas respectivas variáveis selecionadas na modelagem e organizadas por ordem de importância: produtividade primária bruta (GPP), banda azul do produto 13 do sensor MODIS (blue), fração da radiação fotossinteticamente ativa absorvida (FPAR), precipitação média anual (Bio12), banda do infravermelho próximo do produto 9 do sensor MODIS (ref2), temperatura noturna da superfície terrestre (LSTn) e nível base da rede de canais (CNBL)

Discussão

Diversidade arbórea

Com base nos resultados apresentados na Tabela 4, o domínio que apresentou maior biodiversidade foi o Atlântico, apresentando maiores valores médios de riqueza, índice de Shannon e Simpson para Floresta Ombrófila e maiores valores de alfa de Fisher para Floresta Estacional Semidecidual, seguido do domínio do Cerrado e por último o domínio das Caatingas (menores valores de riqueza e índices de Shannon e alfa de Fisher). Esse padrão também pode ser constatado nos mapas gerados dos índices, com maiores valores ao sul e sudeste (domínio Atlântico) e menores valores ao norte do estado (Caatinga).

Comportamento semelhante foi encontrado por Corsini et al. (2014) ao analisarem fragmentos de 4 fitofisionomias florestais de Minas Gerais (exceto Floresta Ombrófila e Cerradão analisados nesse estudo), onde os fragmentos de Floresta Estacional Decidual apresentaram as menores médias de Shannon (2,43) e Pielou (0,69)

e os fragmentos de Floresta Estacional Semidecidual apresentaram a maior média para Shannon (3,42) e Campo Cerrado a maior equabilidade de Pielou (0,78).

Essa variação de diversidade encontrada entre as fitofisionomias pode ser explicada pelos diferentes fatores que atuam na caracterização de cada fitofisionomia. As Florestas Ombrófilas e Semidecíduais apresentaram maiores valores de diversidade provavelmente devido à heterogeneidade ambiental existente no domínio Atlântico - como variações de relevo e altitude, por exemplo - além de ser um ambiente com maior quantidade de chuva e bem distribuída ao longo do ano (Ribeiro and Walter 2008). Já os valores intermediários e baixos dos índices de diversidade encontrados nos domínios do Cerrado e Caatinga respectivamente podem ser devido aos filtros ambientais que atuam nesses ambientes; como por exemplo o fogo, fatores climáticos e edáficos no Cerrado e principalmente fatores climáticos na Caatinga (Floresta Estacional Decidual), que possui os menores valores e má distribuição da precipitação anual (Ribeiro and Walter 2008). Tais fatores restringem a ocorrência de muitas espécies nessas condições, diminuindo a diversidade.

As parcelas pertencentes à fitofisionomia de Floresta Estacional Semidecidual apresentaram os menores e maiores valores para todas as métricas de diversidade. A menor riqueza encontrada foi de 3 espécies (também encontrada para Campo Cerrado e Floresta Estacional Decidual) e a maior foi de 90 espécies (sul do estado). O índice alfa de Fisher variou de 0,53 (leste) a 88,78 (sul). O menor valor de Shannon foi de 0,11 (leste) e o maior valor foi de 5,52 (sul). Os valores observados de Simpson (1-D) variaram de 0,04 (leste) a 0,98 (sul). A equabilidade de Pielou variou de 0,1 (leste) a 1,43 (sul). Vale ressaltar ainda que foi a mesma parcela que apresentou todos os menores valores. Essa grande variação encontrada se deve à ampla faixa latitudinal ocupada por essa fitofisionomia, que proporciona uma elevada heterogeneidade ambiental e conseqüentemente variações no padrão de diversidade.

Variáveis selecionadas

As variáveis selecionadas mais representativas para os índices foram: Bio12 (em todos as métricas), GPP (exceto para alfa de Fisher) e VD (exceto para Pielou). É claro a seleção da variável climática precipitação (Bio12) como sendo consideravelmente importante na predição da riqueza e diversidade arbórea. Percebe-se uma correlação direta entre precipitação e diversidade de árvores no estado, uma vez que a biodiversidade observada aumenta no sentido norte-sul da mesma forma que a precipitação, que variou de 690 a 2396 mm. A precipitação é indispensável, pois a água é um elemento essencial a qualquer ser vivo. Nas florestas, ela é demandada em

processos como germinação, crescimento e reprodução de cada indivíduo arbóreo, influenciando diretamente na capacidade de sobrevivência dos mesmos. Juntamente com a temperatura determina a distribuição das formações vegetais do planeta (Ledru et al. 1993) e é um dos principais gradientes de distribuição das fitofisionomias do estado (Terra et al. 2018). Por essa razão a precipitação é importante na predição da riqueza de espécies e de todas as métricas de diversidade no estado.

A variável de sensoriamento remoto produtividade primária bruta (GPP) está intrinsecamente ligada à biomassa arbórea, pois indica a capacidade das plantas em converter os recursos, durante a fotossíntese, em fixação bruta de carbono. Mesmo não sendo bem compreendida a relação carbono – diversidade em florestas tropicais (devido à dependência de vários fatores, como estágio de desenvolvimento, intervenção humana e escala), a resistência ou resiliência à perturbação ocasionadas pela maior diversidade arbórea possibilita uma maior estabilidade da produtividade a nível de paisagem ao longo do tempo (Oehri et al. 2017). Isso destaca o papel crucial da biodiversidade no funcionamento do ecossistema e em resposta às mudanças ambientais. Em Minas Gerais GPP variou de 0 a um máximo de 0,112 Kg.C/m².8 dias e os maiores valores se concentraram no domínio Atlântico, na região sul e nordeste do estado. Já a variável topográfica profundidade do vale (VD) se refere à distância vertical até o nível do topo, ou seja, valores altos de VD correspondem à locais de baixas altitudes. Deste modo VD pode estar diretamente relacionada com a disponibilidade de água no solo (devido à contribuição dos fluxos de água de áreas mais elevadas), que por sua vez é fator determinante para a composição e diversidade de árvores em Minas Gerais (Terra et al. 2018). VD variou de 0 a 1121,12 metros e os maiores valores de VD no estado se concentram na região Sudeste.

A variável de sensoriamento remoto SAVI foi selecionada para a riqueza de espécies e índice de Simpson. A variável SAVI corresponde ao NDVI ajustado à reflectância do solo que é influenciada pela cobertura de copa e consequentemente remete à biomassa florestal. No estado, SAVI variou de -0,969 a 0,997 e os maiores valores de SAVI se concentraram no domínio Atlântico, na região leste, onde também se concentram os maiores valores de biomassa no estado (Silveira et al. 2019). A variável LSTn (temperatura noturna da superfície) foi selecionada para riqueza, Pielou e Simpson e a variável LSTd (temperatura diurna da superfície) foi selecionada para riqueza e Simpson. Para tais métricas de diversidade, essas variáveis apresentaram correlação negativa. É notório que a cobertura florestal influencia diretamente o clima global, principalmente no controle da temperatura (Prevedello et al. 2019). E uma floresta mais biodiversa pode proporcionar mais chances de ocorrer maior cobertura vegetal (por possuir diferentes espécies com diferentes arquiteturas de copa). Então, a relação negativa encontrada entre a

temperatura da superfície terrestre e métricas de riqueza e diversidade provavelmente está associada à cobertura e regulação térmica associadas à maior diversidade. E essa relação pode ser explicada pelo fato de que a vegetação ameniza a temperatura superficial através da interceptação da radiação solar, redução do albedo e evapotranspiração (Prevedello et al. 2019), auxiliando na manutenção da umidade do solo que é um dos principais fatores de distribuição das fitofisionomias em Minas Gerais (Terra et al. 2018). Isso reforça a importância da biodiversidade e cobertura florestal para mitigar o aquecimento global. Em Minas Gerais LSTn variou de 6,69 a 23,29 °C, sendo que os valores altos foram bem distribuídos e os menores valores se concentraram no sul do estado. Já LSTd variou de 8,03 a 38,07 °C com os maiores valores se concentrando a oeste e norte e os menores valores a leste e sul, no domínio Atlântico.

A variável VDCN foi selecionada para o índice de Shannon. VDCN corresponde à distância vertical acima da rede de drenagem. Sendo assim, VDCN está ligada à altitude, mas nem sempre o ponto mais alto corresponde ao maior VDCN, pois a referência são os rios. Em MG, VDCN variou de 0 a 1177,54 concentrando os maiores valores no Domínio Atlântico, a Leste e Sul (relevo mais acidentado) e os menores valores na região Norte do estado (relevo mais plano). Já a variável CNBL foi selecionada para riqueza, índice de Shannon e Pielou. CNBL corresponde ao VDCN subtraído do DEM ($CNBL = DEM - VDCN$) e pode ser entendida como a interpolação das altitudes com base nas altitudes das redes de drenagem, estando sua distribuição espacial diretamente ligada à altitude (Figura 8). CNBL variou de 48,09 a 2419 metros.

O padrão de distribuição das espécies arbóreas tropicais em gradientes altitudinais ainda não é bem definido, podendo a riqueza diminuir com o aumento da altitude (Vázquez and Givnish 1998, Oliveira-filho and Fontes 2000, Caglioni et al. 2018) ou apresentar um padrão de ‘protuberância a médias altitudes’ (Zhao et al. 2005, Eisenlohr et al. 2013) em que ocorre um pico da riqueza arbórea em altitudes médias, diminuindo abaixo ou acima delas. Essa faixa de altitude apresenta maior riqueza provavelmente devido a uma combinação de temperatura, umidade e fatores ambientais locais (matéria orgânica, inclinação) favoráveis a uma maior riqueza (Sánchez - González and López - Mata 2005). Vale ressaltar que outros fatores, além da altitude, como distúrbios (antropogênicos ou naturais) e topografia também são fundamentais na distribuição das espécies (Eisenlohr et al. 2013). Como o presente estudo não objetivou uma análise em gradientes altitudinais, nossos resultados não nos permitem concluir se a relação da diversidade com a altitude é linear ou unimodal. A correlação positiva encontrada com VDCN e CNBL pode ser devido às condições climáticas mais favoráveis provocadas pelo aumento da altitude, como diminuição da temperatura e aumento da precipitação e nebulosidade (Oliveira-Filho and Fontes

2000); como também a dificuldade de acesso humano que diminui os distúrbios e possíveis características pedológicas que proporcionem maior fertilidade e disponibilidade hídrica.

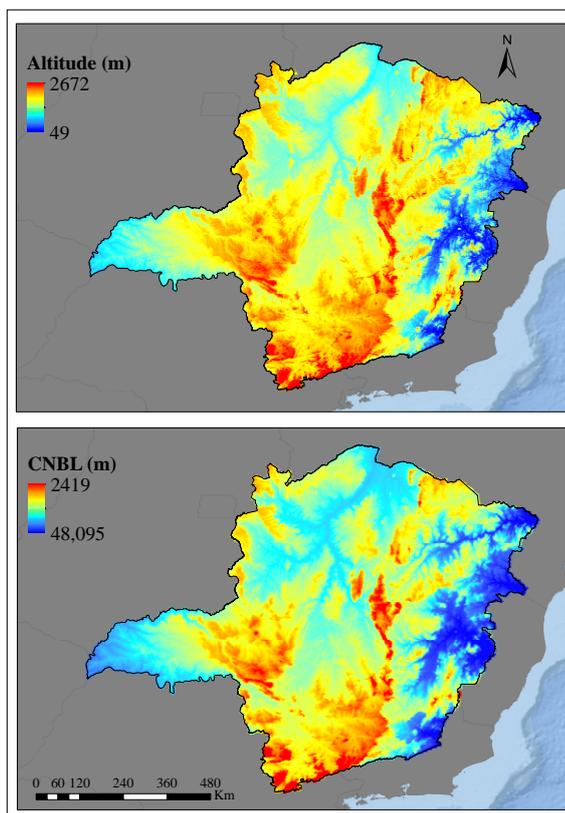


Fig. 8 Mapas de Altitude e CNBL do estado de Minas Gerais

As variáveis índice de área foliar (LAI) e índice de vegetação melhorado (EVI) foram selecionadas apenas para o índice de diversidade alfa de Fisher. O índice de área foliar se refere à área de copa por unidade de área da superfície, ou seja, à cobertura florestal. A relação dessa variável com a diversidade arbórea pode ser pelo fato de uma maior cobertura de dossel fornecer condições favoráveis ao desenvolvimento de outros grupos sucessionais, aumentando assim a riqueza e diversidade arbóreas. A maior cobertura do dossel também pode estar relacionada à diversidade devido à complementaridade de nicho como uma solução para a coexistência de diversas espécies em um mesmo ambiente, usando de forma diferente um mesmo recurso (Sapjanskas et al. 2014). Porém, nem sempre uma maior densidade de árvores corresponde a uma maior diversidade, pois tal cobertura pode pertencer a apenas algumas espécies dominantes, por exemplo. O EVI por sua vez, é um índice de vegetação mais sensível às variações na estrutura do dossel do que outros índices de vegetação (Huete et al. 2002), por ser menos sensível

aos efeitos do solo e atmosféricos. Uma floresta diversa apresenta diferentes texturas, formas, alturas e tonalidades de copa, dessa forma, podemos entender a importância da seleção dessa variável na predição de diversidade. No estado, o IAF variou de 0 a 4 e concentrou os maiores valores na região nordeste e menores valores na região sul, homogeneizando os valores nas demais áreas. Já EVI variou de -0,3 a 0,996, concentrando os maiores valores na margem leste do estado.

As variáveis referentes aos comprimentos de onda da faixa do azul foram selecionadas para Shannon (B1 do satélite Landsat 5 e 13_17blue do produto MODIS) e para Pielou (13_17blue do produto MODIS). O índice de Shannon apresentou correlação negativa entre ambas as variáveis (-0,15 para B1 e -0,02 para 13_17blue). Isso pode ser visto nos mapas: locais com alto valor de Shannon apresentam baixos valores de B1 e 13_17blue. Nagendra et al. (2010) e Medeiros et al. (2019) também encontraram correlação negativa (-0,297 e -0,41 respectivamente) entre banda azul e índice de Shannon em florestas tropicais secas na Índia e Brasil respectivamente. Já para o índice de Pielou (quando é considerada a máxima diversidade, em que todas as espécies são igualmente abundantes) a correlação é positiva (0,19).

A variável fração da radiação fotossinteticamente ativa (400 a 700 nm) absorvida pela vegetação (15_209FPAR) foi selecionada para o índice de Pielou. A FPAR é um parâmetro biofísico, medida em porcentagem (0 a 1), relacionado à atividade fotossintética das plantas, estando diretamente associada à produtividade primária da vegetação (Pilaš et al. 2014), pois é um fator limitante na taxa de assimilação de carbono. É sensível ao estresse da vegetação podendo indicar mudanças de sazonalidade (Gobron et al. 2005) e por este motivo é considerada uma das variáveis essenciais na caracterização do clima terrestre (GCOS 2016). Mahanand and Behera (2017) encontraram uma relação positiva entre FPAR e LAI e riqueza de espécies vegetais na Índia. Uma floresta diversa possui um dossel heterogêneo, com maior cobertura e densidade arbórea (Silveira et al. 2019), que otimiza a absorção da radiação a nível de paisagem. Dessa forma, entende-se a importância dessa variável na predição da diversidade florestal. Em Minas Gerais os maiores valores de FPAR se concentraram na região Nordeste e os menores valores ao Sul.

A variável 9_257ref2 corresponde à faixa do infravermelho próximo (NIR) do produto 9 do sensor MODIS e foi selecionada para o índice de Pielou. Infravermelho próximo corresponde à faixa espectral mais refletida pela vegetação, e além disso é a faixa mais sensível à variação de reflectância das folhas entre as espécies, correlacionando positivamente com a riqueza de espécies (Medeiros et al. 2019). Assim, compreende-se a

importância dessa variável na predição da diversidade florestal. Em Minas Gerais, os valores de H' variaram de 0 a 0,989, concentrando os maiores valores na região Nordeste.

A variável Slope corresponde à inclinação do terreno (em graus) e foi selecionada apenas para o índice de Simpson (1-D), com correlação positiva. Menores inclinações de terreno proporcionam uma menor erosão do solo e tendem a possuir solos mais profundos e ricos, com maior diversidade. Porém os resultados que encontramos mostram o contrário: maiores inclinações, maior diversidade. Isso pode ser devido às áreas planas serem utilizadas para cultivos agrícolas e atividades antrópicas em geral, restando remanescentes florestais quase que apenas em áreas íngremes ou de elevadas altitudes (Campanili and Prochnow 2006); daí a correlação positiva entre o índice de Simpson e Slope. Sánchez - González and López - Mata (2005) também encontraram correlação positiva entre riqueza e diversidade de espécies e grau de inclinação do terreno no México. Para a Amazônia, Laurance et al (2010) também encontraram correlação positiva entre inclinação e riqueza e diversidade de espécies. A explicação que encontraram foi que áreas mais íngremes, caracterizadas por solos mais pobres, possivelmente suportam poucas árvores grandes e competitivas, visto que a densidade de árvores (menores) foi maior nessas áreas. Em Minas Gerais, a inclinação variou de 0 a 82,53° concentrando os maiores valores no Domínio Atlântico, caracterizado por seus relevos acidentados que proporcionam maior substituição de espécies e possivelmente maior diversidade representada pelo índice de Simpson.

Mapeamento dos índices de diversidade em MG

De forma geral, observa-se que os mapas de distribuição de riqueza e índice alfa de Fisher tiveram uma maior semelhança entre si, nos fazendo perceber que alfa de Fisher é um índice que prioriza a componente riqueza em detrimento da uniformidade. Apresentaram maiores valores ao sul e centro-leste de Minas Gerais. Os índices de Shannon e Simpson por sua vez já carregam em sua fórmula a proporção de indivíduos em cada espécie. Os seus mapas também tiveram uma certa semelhança, apresentando maiores valores em todos os domínios. Já o mapa de uniformidade de Pielou apresentou um padrão um pouco diferente do encontrado nos demais mapas, mostrando uma maior uniformidade de indivíduos entre as espécies no norte do estado.

Riqueza

As variáveis selecionadas por ordem decrescente de importância para o número de espécies foram: 17_241GPP, VD, SAVI, 11_145LSTn, CNBL, Bio12 e 11_113LSTd. O número médio de espécies obtido da modelagem foi de 25, variando de 4,78 a 73,74. A distribuição de riqueza de árvores no estado seguiu um padrão

ascendente de norte a sul e de oeste a leste, semelhante com o que foi encontrado por Araújo et al. (2019). É nítida a diferença de riqueza encontrada entre os domínios morfoclimáticos, sendo a maior no domínio Atlântico e menor na Caatinga. Essa diferença se deve provavelmente às características relacionadas à disponibilidade de água, produtividade primária e temperatura (explicadas pelas variáveis selecionadas) que variam entre os domínios.

Alfa de Fisher

O modelo gerado pelo Random Forest selecionou as seguintes variáveis por ordem de importância para o índice alfa de Fisher: Bio12, VD, 15_209LAI e 13_145EVI. Silveira et al. (2019) também encontraram Bio12 e VD como sendo as variáveis mais importantes na modelagem de alfa de Fisher para Minas Gerais. Conforme os autores, ambas as variáveis juntas são indicadoras de umidade do solo, que por sua vez é determinante na distribuição da diversidade arbórea no estado (Terra et al. 2018).

A média do índice alfa de Fisher foi de 12,25 variando de 2,35 a 47,39, também parecido com os valores encontrados por Silveira et al. (2019) que variaram de 2 a 51,6. O índice alfa de Fisher apresentou uma distribuição bem parecida com riqueza. Isso pode ser devido a sua forma de cálculo ser baseada na relação número de espécies por número de indivíduos da área, e não número de indivíduos por espécie, como acontece nos demais índices.

Shannon

As variáveis selecionadas para o índice de Shannon foram: 17_241GPP, Bio12, VD, B1, VDCN, 13_17blue e CNBL. A média do índice de Shannon foi de 2,60 variando de 0,98 a 3,73. Observa-se que, em comparação à riqueza e Fisher, Shannon apresentou valores um pouco maiores no domínio do Cerrado. Isso nos faz notar que tal domínio aparentemente possui maior uniformidade entre as espécies (que será confirmada com o índice de Pielou), uma vez que Shannon já leva em consideração a proporção de indivíduos por espécie. Porém o padrão de diversidade continuou com maiores valores ao sul e leste do estado, coincidindo com o domínio Atlântico.

Simpson

As variáveis selecionadas para o índice de Simpson por ordem de importância foram: Bio12, 17_321GPP, VD, 11_113LSTd, 11_145LSTn, SAVI e Slope. Este índice, na sua forma original (D), é um índice que enfatiza a dominância; porém, na forma utilizada ($1 - D$) reflete a uniformidade e distribuição de indivíduos entre as espécies, mas ainda não é uma medida pura de uniformidade (Magurran 2013). Segundo a autora, tal índice é um

dos índices de diversidade mais informativos e robustos existentes. De forma geral, o mapa apresentou maiores valores em todo o estado se comparado com os mapas dos índices anteriores, exibindo valores ainda maiores para o domínio do Cerrado do que o índice de Shannon. Isso reflete ainda mais a ideia de maior dominância neste domínio, que poderá ser comprovada com o índice de uniformidade de Pielou. Mostra também que índices que envolvem apenas a riqueza de espécies podem não ser adequados quando o objetivo é estimar a diversidade.

A média do índice de Simpson para Minas Gerais obtida da modelagem foi de 0,86, variando de 0,32 a 0,96. Os maiores valores estão inseridos na região sul e centro-leste do estado, indicando uma maior diversidade nessa área (domínio Atlântico), com gradativa diminuição no Cerrado e menor diversidade nas florestas decíduais (Caatinga). Padrão aproximado foi encontrado por Araújo et al. (2019) ao utilizar a dominância de Simpson (forma original). Os menores valores encontrados por esses autores (maior diversidade) foram para a região sul e leste do estado, análogo ao encontrado nesse trabalho para os maiores valores de diversidade de Simpson. Porém, os maiores valores encontrados por eles (menor diversidade) diferiram um pouco do que encontramos, sendo que locais de menor diversidade na região norte encontrados por eles coincidem com áreas que apresentaram diversidade de média a alta nesse estudo; e no domínio da Caatinga onde encontraram valores de baixa a média diversidade, é o local de menor diversidade que encontramos, estando de acordo com o encontrado pelos demais índices. Essa pequena diferença pode ser devido às diferentes técnicas utilizadas entre os estudos, sendo que no presente trabalho o mapeamento foi a partir das variáveis que foram selecionadas na modelagem, como sendo importantes preditoras da diversidade arbórea no estado.

Pielou

As seguintes variáveis foram selecionadas em ordem de importância para prever o índice de uniformidade de Pielou: 17_289GPP, 13_17blue, 15_209FPAR, Bio12, 9_257ref2, 11_145LSTn e CNBL. Esse índice diz respeito a quão bem distribuídos estão os indivíduos entre as espécies, sendo maior a diversidade quanto maior for a uniformidade. A média do índice foi de 0,83, variando de 0,31 a 0,95. O padrão de distribuição desse índice diferiu dos demais. O domínio das Caatingas continuou sendo o de menor diversidade, porém a maior diversidade (que nos demais índices foi predominante no domínio Atlântico) agora foi predominante no domínio do Cerrado, em especial na fitofisionomia Cerrado Sensu Stricto e também nas florestas estacionais semidecíduais do sul do estado.

A menor uniformidade encontrada para as florestas estacionais decíduais (Caatinga) está de acordo com a menor riqueza e diversidade apresentada pelos demais índices. Isso significa que a região possui baixa diversidade arbórea devido à dominância de algumas poucas espécies, como *Combretum leprosum* Mart. e *Myracrodruon urundeuva* Allemão, que provocam o aumento de espécies raras (poucos indivíduos) nesse ambiente (Corsini et al. 2014). Além disso, a dominância de poucas espécies está relacionada à capacidade que elas possuem para suportar as condições ambientais adversas desse local, como elevadas temperaturas e baixa precipitação. Por sua vez, a maior uniformidade encontrada no domínio do Cerrado contrasta com os valores médios de riqueza. Isso mostra que por mais que o domínio não apresente a maior riqueza do estado, o número de indivíduos entre as espécies é semelhante devido à maior homogeneidade ambiental existente. Já em relação ao domínio Atlântico, valores altos de uniformidade se concentraram ao sul, dando a entender que a região Leste do estado possui uma maior riqueza, porém com maior dominância de algumas espécies e ocorrência de espécies raras. Percebe-se então que a distribuição dos indivíduos entre as espécies do Domínio do Cerrado é mais uniforme, mesmo não possuindo o maior número de espécies. Por outro lado, a maior riqueza do Domínio Atlântico pode possuir espécies dominantes e raras, uma vez que a uniformidade não foi tão alta nesse Domínio.

A maior diversidade corresponde ao local de maior riqueza e maior uniformidade. Analisando os índices conjuntamente, observa-se que o índice de diversidade de Simpson conseguiu captar os padrões de riqueza e uniformidade em um único índice e dessa forma, conclui-se que a diversidade arbórea no estado de Minas Gerais aumenta de norte a sul, sendo maior no domínio Atlântico e menor no domínio das Caatingas.

Conclusões

Conclui-se que a diversidade florestal no estado de Minas Gerais tem um padrão ascendente de Norte a Sul e Leste a Oeste, com maiores valores no Domínio Atlântico e menores valores no Domínio das Caatingas. O algoritmo RandomForest foi adequado na seleção das variáveis, sendo a variável climática precipitação, a de sensoriamento remoto produtividade primária bruta e de terreno profundidade do vale as mais importantes na predição das métricas de diversidade no estado. Quanto aos mapas gerados a partir das variáveis selecionadas, percebe-se que o índice alfa de Fisher prioriza a componente riqueza e que o índice de Simpson foi o que melhor sintetizou as informações de riqueza e uniformidade, conseguindo representar bem a variação de diversidade entre os domínios. Por mais que haja diferenças na diversidade arbórea entre os domínios, associadas às diferentes condições ambientais e endemismos de cada um, é importante destacar a necessidade de preservação de todos os domínios como medida para proteção da diversidade arbórea em Minas Gerais.

Referências

- Araújo EJG, Morais VA, David HC, Scolforo JRS, Mello JM, Ebling AA (2019) Spatialization of tree species diversity in the state of Minas Gerais. *Floresta e Ambiente* 26: e20150206. <http://dx.doi.org/10.1590/2179-8087.020615>.
- Beech E, Rivers M, Oldfield S, Smith PP (2017) GlobalTreeSearch: The first complete global database of tree species and country distributions. *J Sustain Forest* 36: 454-489. <https://doi.org/10.1080/10549811.2017.1310049>.
- Beven KJ, Kirkby MJ (1979) A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrol Sci B* 24: 43-69. <https://doi.org/10.1080/02626667909491834>.
- Brazil Flora Group (2015) Growing knowledge: an overview of seed plant diversity in Brazil. *Rodriguésia* 66: 1085-1113. <http://rodriguesia.jbrj.gov.br>.
- Breiman L (2001) Random forests. *Mach Learn* 45: 5-32. <https://doi.org/10.1023/A:1010933404324>.
- Caglioni E et al. (2018) Altitude e solos determinam variações abruptas da vegetação em gradiente altitudinal de Mata Atlântica. *Rodriguésia* 69: 2055-2068. <http://dx.doi.org/10.1590/2175-7860201869436>.
- Campanili M, Prochnow P (2006) Mata Atlântica - Uma rede pela floresta. *Rede Mata Atlântica*, Brasília. 332 p.
- Carvalho LMT, Scolforo JR, Oliveira AD, Mello JM, Oliveira LT, Junior FWA, Cavalcanti HC, Filho RV (2006). Procedimentos de mapeamento. In: Scolforo JRS and Carvalho LMT (eds.) *Mapeamento e Inventário da Flora nativa e dos Reflorestamentos de Minas Gerais*. UFLA, Lavras, pp 37-57.
- Carvalho MC, França LCJ, Lopes IL, Araújo LA, Mello JM, Gomide LR (2019) Algoritmos de aprendizagem de máquina na modelagem da distribuição potencial de habitats de espécies arbóreas. *Nativa* 7: 600-606. <http://dx.doi.org/10.31413/nativa.v7i5.7214>.
- Conrad O et al. (2015) System for Automated Geoscientific Analyses (SAGA). *Geosci Model Dev* 8: 1991-2007. <https://doi.org/10.5194/gmd-8-1991-2015>.
- Corsini CR, Scolforo JRS, Oliveira AD, Mello JM, Machado ELM (2014) Diversidade e similaridade de fragmentos florestais nativos situados na região nordeste de Minas Gerais. *Cerne* 20: 1-10. <http://dx.doi.org/10.1590/S0104-77602014000100001>.

- Desmet PJJ, Govers G (1996) A GIS procedure for automatically calculating the USLE LS factor on topographically complex landscape units. *J soil water conserv* 51: 427-433.
- Duro DC, Franklin SE, Dubé MG (2012) A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sens Environ* 118: 259-272. <https://doi.org/10.1016/j.rse.2011.11.020>.
- Eisenlohr PV et al. (2013) Disturbances, elevation, topography and spatial proximity drive vegetation patterns along an altitudinal gradient of a top biodiversity hotspot. *Biodivers Conserv* 22:2767-2783. <https://doi.org/10.1007/s10531-013-0553-x>.
- Fick SE, Hijmans RJ (2017) Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int J Climatol* 37: 4302-4315. <https://doi.org/10.1002/joc.5086>.
- Fisher RA, Corbet AS, Willians CB (1943) The relation between the number of species and the number of individuals in a random sample of animal population. *The Journal of Animal Ecology* 12: 42-58.
- Gallant JC, Dowling TI (2003) A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resour Res* 39. <https://doi.org/10.1785/0120060267>.
- GCOS - Global Climate Observing System (2016) The global observing system for climate: implementation needs. GCOS-200 (GOOS-214).
- Gil PR (2002) *Wilderness: earth's last wild places*. CEMEX, México. 185 p.
- Gobron N, Pinty B, Mélin F, Tberner M, Verstraete MM, Belward A, Lavergne T, Widlowski JL (2005) The state of vegetation in Europe following the 2003 drought. *Int J Remote Sens* 26: 2013-2020. <https://doi.org/10.1080/01431160412331330293>.
- Guan H, Li J, Chapman M, Deng F, Ji Z, Yang X (2013) Integration of orthoimagery and lidar data for object-based urban thematic mapping using random forests. *Int J Remote Sens* 34: 5166-5186. <https://doi.org/10.1080/01431161.2013.788261>.
- Guisan A, Weiss SB, Weiss AD (1999) GLM versus CCA spatial modeling of plant species distribution. *Plant Ecol* 143: 107-122. <https://doi.org/10.1023/A:1009841519580>.

Huete A, Didan K, Miura T, Rodriguez EP, Gao X, Ferreira LG (2002) Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens Environ* 83: 195-213. [https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2).

Kwon Y, Lee T, Lang A, Burnette D (2019) Assessment on latitudinal tree species richness using environmental factors in the southeastern United States. *PeerJ* 7: e6781. <https://doi.org/10.7717/peerj.6781>.

Laurance SGW et al. (2010) Influence of soils and topography on Amazonian tree diversity: a landscape-scale study. *J Veg Sci* 21: 96-106. <https://doi-org.ez26.periodicos.capes.gov.br/10.1111/j.1654-1103.2009.01122.x>.

Ledru MP (1993) Late Quaternary Environmental and Climatic Changes in Central Brazil. *Quaternary Research* 39: 90-98. <https://doi.org/10.1006/qres.1993.1011>.

Magurran AE (2013) *Medindo a diversidade biológica*. UFPR, Curitiba.

Mahanand S, Behera MD (2017) Relationship Between Field-Based Plant Species Richness and Satellite-Derived Biophysical Proxies in the Western Ghats, India. *P Natl A Sci India A* 87: 927-939. <https://doi.org/10.1007/s40010-017-0460-8>.

Margalef R (1972) Homage to Hevelyn Hutchinson or why there is an upper limit to diversity? *Transp. Connect. Acad. Arts Sci* 44:214-235.

Martins FB, Gonzaga G, Santos DF, Reboita MS (2018) Classificação climática de Köppen e de Thornthwaite para Minas Gerais: cenário atual e projeções futuras. *Revista Brasileira de Climatologia – Dossiê climatologia de Minas Gerais*: 129-156. <http://dx.doi.org/10.5380/abclima.v1i0.60896>.

Medeiros ESS, Machado CCC, Galvêncio JD, Moura MSB, Araujo HFP (2019) Predicting plant species richness with satellite images in the largest dry forest nucleus in South America. *Journal of Arid Environments* 166: 43-50. <https://doi.org/10.1016/j.jaridenv.2019.03.001>.

Mittermeier RA, Turner WR, Larsen FW, Brooks TM, Gascon C (2011) Global biodiversity conservation: the critical role of Hotspots. *Biodiversity Hotspots*. In: Zachos F, Habel J (eds) *Biodiversity Hotspots*. Springer, Berlin, pp 3-22.

Morandi PS et al. (2018) Tree diversity and above-ground biomass in the South America Cerrado biome and their conservation implications. *Biodivers Conserv*. <https://doi.org/10.1007/s10531-018-1589-8>.

- Nagendra H, Rocchini D, Ghate R, Sharma B, Pareeth S (2010) Assessing Plant Diversity in a Dry Tropical Forest: Comparing the Utility of Landsat and Ikonos Satellite Images. *Remote Sens* 2: 478-496. <https://doi.org/10.3390/rs2020478>.
- Oehri J, Schmid B, Schaepman-Strub G, Niklaus PA (2017) Biodiversity promotes primary productivity and growing season lengthening at the landscape scale. *Proceedings of the National Academy of Sciences* 114: 10160-10165. <https://doi.org/10.1073/pnas.1703928114>.
- Oliveira-Filho AT, Fontes MAL (2000) Patterns of floristic differentiation among Atlantic Forests in Southeastern Brazil and the influence of climate. *Biotropica* 32: 793-810. <https://doi.org/10.1111/j.1744-7429.2000.tb00619.x>.
- Pelissari AL, Filho AF, Ebling AA, Sanquetta CR, Cysneiros VC, Corte APD (2018) Spatial variability of tree species diversity in a mixed tropical forest in Southern Brazil. *An Acad Bras Cienc* 90: 2491-2500. <http://dx.doi.org/10.1590/0001-3765201820170826>.
- Pennington RT, Lavin M, Oliveira-Filho A (2009) Woody plant diversity, evolution, and ecology in the tropics: perspectives from seasonally dry tropical forests. *Annual Reviews of Ecology, Evolution, and Systematics* 40: 437-457. <https://doi.org/10.1146/annurev.ecolsys.110308.120327>.
- Pielou EC (1975) *Ecological diversity*. Wiley, New York.
- Pilaš I, Medved I, Medak J, Medak D (2014) Response strategies of the main forest types to climatic anomalies across Croatian biogeographic regions inferred from FAPAR remote sensing data. *Forest Ecol Manag* 326: 58-78. <https://doi.org/10.1016/j.foreco.2014.04.012>.
- Prevedello JA, Winck GR, Weber MM, Nichols E, Sinervo B (2019) Impacts of forestation and deforestation on local temperature across the globe. *PloS one* 14: e0213368. <https://doi.org/10.1371/journal.pone.0213368>.
- R Core Team (2019) *R: A language and environment for statistical computing*. R foundation for statistical computing, Vienna, Austria. <https://www.R-project.org/>.
- Reis AA, Carvalho MC, Mello JM, Gomide LR, Ferraz Filho AC, Acerbi Junior FW (2018) Spatial prediction of basal area and volume in *Eucalyptus* stands using Landsat TM data: an assessment of prediction methods. *New Zealand Journal of Forestry Science* 48: 1. <https://doi.org/10.1186/s40490-017-0108-0>.

Revermann R, Finckh M, Stellmes M, Strohbach BJ, Frantz D, Oldeland J (2016) Linking land surface phenology and vegetation-plot databases to model terrestrial plant α -diversity of the Okavango Basin. *Remote Sensing* 8: 370. <https://doi.org/10.3390/rs8050370>.

Ribeiro JF, Walter BMT (2008) As principais fitofisionomias do Bioma Cerrado. In: Sano SM, Almeida SP, Ribeiro JF (eds) *Cerrado: ecologia e flora*. Embrapa Cerrados, Planaltina, pp.152 -212.

Rocchini D, Hernández-Stefanoni JL, He KS (2015) Advancing species diversity estimate by remotely sensed proxies: A conceptual review. *Ecological Informatics* 25: 22-28. <https://doi.org/10.1016/j.ecoinf.2014.10.006>.

Rocchini D, Boyd DS, Féret JB, Foody GM, He KS, Lausch A, Nagendra H, Wegmann M, Pettorelli N (2016) Satellite remote sensing to monitor species diversity: potential and pitfalls. *Remote Sensing in Ecology and Conservation* 2: 25-36. <https://doi.org/10.1002/rse2.9>.

Sánchez - González A, López - Mata L (2005) Plant species richness and diversity along an altitudinal gradient in the Sierra Nevada, Mexico. *Diversity and Distributions* 11: 567-575. <https://doi.org/10.1111/j.1366-9516.2005.00186.x>.

Santos RM, Oliveira-Filho AT, Eisenlohr PV, Queiroz LP, Cardoso DBOS, Rodal MJN (2012) Identity and relationships of the Arboreal Caatinga among other floristic units of seasonally dry tropical forests (SDTFs) of north-eastern and Central Brazil. *Ecol Evol* 2: 409-428. <https://doi.org/10.1002/ece3.91>.

Sapijanskas J, Paquette A, Potvin C, Kunert N, Loreau M (2014) Tropical tree diversity enhances light capture through crown plasticity and spatial and temporal niche differences. *Ecology* 95: 2479-2492. <https://doi.org/10.1890/13-1366.1>.

Scolforo JR et al. (2006) Procedimento do inventário florestal – flora nativa. In: Scolforo JR, Carvalho LMT (eds.) *Mapeamento e inventário da flora e dos reflorestamentos de Minas Gerais*. UFLA, Lavras, pp. 63-73.

Scolforo HF, Scolforo JRS, Mello CR, Mello JM, Filho ACF (2015) Spatial distribution of aboveground carbon stock of the arboreal vegetation in Brazilian Biomes of Savanna, Atlantic Forest and Semi-Arid Woodland. *PloS one* 10: e0128781. <https://doi.org/10.1371/journal.pone.0128781>.

Shannon CE (1948) A mathematical theory of communication. *Bell System Technical Journal* 27: 379-423.

- Silveira EMO, Bueno IT, Junior FWA, Mello JM, Scolforo JRS, Wulder MA (2018) Using Spatial Features to Reduce the Impact of Seasonality for Detecting Tropical Forest Changes from Landsat Time Series. *Remote Sens* 10: 808. <https://doi.org/10.3390/rs10060808>.
- Silveira EMO, Terra MCNS, Steege H ter, Maeda EE, Júnior FWA, Scolforo JRS (2019) Carbon-diversity hotspots and their owners in Brazilian southeastern Savanna, Atlantic Forest and Semi-Arid Woodland domains. *Forest Ecol Manag* 452: 117575. <https://doi.org/10.1016/j.foreco.2019.117575>.
- Simpson EH (1949) Measurement of diversity. *Nature* 163: 688. <https://doi.org/10.1038/163688a0>.
- Song C, Cao M (2017) Relationships between Plant Species Richness and Terrain in Middle Sub-Tropical Eastern China. *Forests* 8: 344. <https://doi.org/10.3390/f8090344>.
- Souza MC, Biudes MS, Danelichen VHM, Machado NG, Musis CR, Vourlitis GL, Nogueira JS (2014) Estimation of gross primary production of the Amazon-Cerrado transitional forest by remote sensing techniques. *Rev Bras Meteorol* 29:1-12. <http://dx.doi.org/10.1590/S0102-77862014000100001>.
- Terra MCNS, Santos RM, Fontes MAL, Mello JM, Scolforo JRS, Gomide LR, Júnior JAP, Schiavini I, ter Steege H (2017) Tree dominance and diversity in Minas Gerais, Brazil. *Biodivers Conserv* 26: 2133-2153. <https://doi.org/10.1007/s10531-017-1349-1>.
- Terra MCNS et al. (2018) Water availability drives gradients of tree diversity, structure and functional traits in the Atlantic-Cerrado-Caatinga transition, Brazil. *J Plant Ecol* 11: 1-12. <https://doi.org/10.1093/jpe/rty017>.
- Vázquez GJA, Givnish TJ (1998) Altitudinal gradients in tropical forest composition, structure, and diversity in the Sierra de Manantlán. *J Ecol* 86: 999-1020.
- Wilson JP, Gallant JC (2000) Digital terrain analyses. In: Wilson JP, Gallant JC (eds.) *Terrain analysis: Principles and applications*. Wiley, New York, pp 1-27.
- Zhao CM, Chen WL, Tian ZQ, Xie ZQ (2005) Altitudinal Pattern of Plant Species Diversity in Shennongjia Mountains, Central China. *J Integr Plant Biol* 47: 1431-1449. <https://doi.org/10.1111/j.1744-7909.2005.00164.x>.

