



ARTUR FERRO DE SOUZA

**MODELAGEM ESPACIAL DE INDICADORES ECOLÓGICOS
EM ÁREAS ATINGIDAS POR REJEITOS DE MINÉRIO DE
FERRO**

**LAVRAS-MG
2024**

ARTUR FERRO DE SOUZA

**MODELAGEM ESPACIAL DE INDICADORES ECOLÓGICOS EM ÁREAS
ATINGIDAS POR REJEITO DE MINÉRIO DE FERRO**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação do Mestrado Profissional em Tecnologias e Inovações Ambientais, para a obtenção do título de Mestre.

Prof^a. Dr^a. Soraya Alvarenga Botelho
Orientadora

**LAVRAS-MG
2024**

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

Souza, Artur Ferro de.

Modelagem espacial de indicadores ecológicos em áreas atingidas por rejeitos de minério de ferro / Artur Ferro de Souza. - 2024.

65 p. : il.

Orientador(a): Soraya Alvarenga Botelho.

Coorientador(a): Marcela Castro Nunes Santos Terra.

Dissertação (mestrado profissional) - Universidade Federal de Lavras, 2024.

Bibliografia.

1. Restauração ecológica. 2. Sensoriamento Remoto. 3. Aprendizado de máquina. I. Botelho, Soraya Alvarenga. II. Terra, Marcela Castro Nunes Santos. III. Título.

ARTUR FERRO DE SOUZA

**MODELAGEM ESPACIAL DE INDICADORES ECOLÓGICOS EM ÁREAS
ATINGIDAS POR REJEITO DE MINÉRIO DE FERRO
SPATIAL MODELING OF ECOLOGICAL INDICATORS IN AREAS AFFECTED
BY IRON ORE TAILINGS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação do Mestrado Profissional em Tecnologias e Inovações Ambientais, para a obtenção do título de Mestre.

APROVADA em 20 de fevereiro de 2024.

Prof^a. Dr^a. Soraya Alvarenga Botelho UFLA

Prof^a. Dr^a. Marcela Castro Nunes Santos Terra UFSJ

Prof. Dr. Fausto Weimar Acerbi Junior UFLA

Dr. Carlos Delano Cardoso de Oliveira UFLA

**LAVRAS-MG
2024**

*Aos meus pais, Onery e Débora;
e a minha namorada, Thayná
por toda atenção, amor e carinho.
Dedico.*

AGRADECIMENTOS

Agradeço à minha família, em especial aos meus pais por apoiarem a minha decisão de continuar os estudos após a graduação. Agraço também ao meu irmão, Raul, pelas discussões profundas e o incentivo ao pensamento crítico desde muito cedo na minha vida.

Agradeço à minha namorada, Thayná, pelos longos anos de companheirismo, dedicação e apoio às minhas decisões profissionais e pessoais.

Aos meus grandes amigos de república, agradeço todo tempo de qualidade compartilhado na República Vegas, responsáveis por tornarem essa trajetória mais leve.

Agradeço à minha querida orientadora, Soraya Alvarenga Botelho, por todas as oportunidades concedidas a mim desde o começo da minha trajetória acadêmica. Ela acreditou em meu trabalho antes mesmo de mim, me trazendo desafios e contribuindo imensuravelmente com meu crescimento.

Em especial, a professora Marcela Castro Nunes Santos Terra, gostaria de agradecê-la imensamente por toda a dedicação e orientação para a construção deste trabalho. Tenho ela como uma inspiração pela sua forma de fazer ciência com brilho nos olhos.

Agradeço ao Zé Pedro, grande técnico e colega de campo, responsável por me ensinar o pragmatismo de campo. Junto ao Zé, gostaria de agradecer a Letícia e a Tamires, colegas do laboratório de solos da UFLA, que nos ajudaram muito com as últimas coletas dos dados do Projeto Rio Doce.

Agradeço à agência de fomento FAPEMIG junto a Renova, responsáveis por financiar o projeto “Ciência e tecnologia para a recuperação ambiental da bacia do rio Doce: métodos, estratégias e indicadores de restauração de ecossistemas florestais”. Através dos financiadores pude concluir meu mestrado com bolsa e gerar este trabalho como um dos produtos exigidos pelo projeto, feito este que me orgulho grandemente, principalmente pelo desafio proposto e aprendizado gerado.

Por fim, agradeço a todos que de alguma forma contribuíram com este trabalho e estiveram presentes nesse processo.

RESUMO

A degradação ambiental decorrente de atividades humanas representa um dos desafios mais urgentes e complexos enfrentados pela sociedade contemporânea. Um exemplo desse cenário é o desastre ocorrido em Mariana, Minas Gerais, em novembro de 2015, quando o rompimento de uma barragem de rejeitos de minério de ferro resultou em consequências catastróficas para o meio ambiente local. Desde então, esforços têm sido direcionados para a restauração dessas áreas afetadas, que devem ser monitoradas para verificar a trajetória dos processos ecológicos e evolução no novo ecossistema, pois fornecem dados de erros e acertos, permitindo readequação dos métodos utilizados. Diante desse cenário, o principal objetivo desta pesquisa consiste em mapear e modelar indicadores ecológicos de diversidade (riqueza de espécies, modelo I) e estrutura (cobertura de dossel, modelo II) nas margens do rio Gualaxo do Norte, em Mariana, MG. Avaliou-se dois cenários (cenário A e cenário B), a fim de compreender a influência das variáveis de observação de campo nos modelos (cenário A, gerando os modelos IA e IIA), bem como o uso exclusivo de variáveis de sensoriamento remoto (cenário B, gerando os modelos IB e IIB). Os indicadores foram coletados em 24 transectos em seis áreas diferentes, junto aos indicadores, foram coletados dados de observação de campo (variáveis preditoras categóricas). Os dados de sensoriamento remoto (SR) foram gerados a partir de imagens Planet e utilizados na modelagem. Para realizar a modelagem e o mapeamento dos indicadores, foi utilizado o algoritmo de aprendizado de máquina Random Forest (RF), sendo gerados os modelos preliminares e, após a seleção das variáveis, gerados os modelos finais. A avaliação do desempenho dos modelos foi feita através do coeficiente de determinação (R^2) e métricas de erro (*Root Mean Square Error* – RMSE e *Mean Square Error* - MAE). Os resultados mostram que para estimar riqueza de espécies, a utilização de dados categóricos de campo são fundamentais para a mitigação do efeito de saturação dos dados, além da melhora do ajuste global e da precisão do modelo. Em contrapartida, o modelo de cobertura de dossel se ajustou melhor utilizando apenas as variáveis de sensoriamento remoto. As variáveis que mais contribuíram com a estimativa espacial de riqueza de espécies (modelo IB) e cobertura de dossel (modelo IIB) foram, respectivamente, média textural do infravermelho próximo e índice de vegetação NDVI.

Palavras-chave: Restauração ecológica. Sensoriamento Remoto. Aprendizado de máquina.

ABSTRACT

Environmental degradation resulting from human activities represents one of the most urgent and complex challenges faced by contemporary society. An example of this scenario is the disaster that occurred in Mariana, Minas Gerais, in November 2015, when the collapse of an iron ore tailings dam resulted in catastrophic consequences for the local environment. Since then, efforts have been directed towards the restoration of these affected areas, which must be monitored to verify the trajectory of ecological processes and evolution in the new ecosystem, as they provide data on errors and successes, allowing readjustment of the methods used. Given this scenario, the main objective of this research is to model and map ecological indicators of diversity (species richness, Model I) and structure (canopy cover, Model II) on the banks of the Gualaxo do Norte River, in Mariana, MG. Two scenarios were evaluated (Scenario A and Scenario B) to understand the influence of field observation variables on the models (Scenario A, generating Models IA and IIA), as well as the exclusive use of remote sensing variables (Scenario B, generating Models IB and IIB). The indicators were collected in 24 transects in six different areas. In addition to the indicators, field observation data (categorical predictor variables) were collected. Remote sensing (RS) data was generated from Planet imagery and used in the modeling. To carry out the modeling and mapping of indicators, the Random Forest (RF) machine learning algorithm was used, generating preliminary models and, after selecting the variables, generating the final models. The models' performance was assessed using the Coefficient of Determination (R^2) and error metrics (Root Mean Square Error – RMSE and Mean Absolute Error - MAE). The results show that to estimate species richness, the use of categorical field data is essential for mitigating the effect of data saturation, in addition to improving the global fit and precision of the model. In contrast, the canopy cover model was better fitted using only the remote sensing variables. The variables that contributed most to the spatial estimate of species richness (Model IB) and canopy cover (Model IIB) were, respectively, near-infrared textural mean and NDVI vegetation index.

Keywords: Ecological Restoration. Remote Sensing. Machine Learning.

INDICADORES DE IMPACTO

O colapso da barragem de rejeitos de minério de ferro em Mariana, Minas Gerais, em 2015, desencadeou um dos maiores desastres ambientais do Brasil, com impactos devastadores no ecossistema local e nas comunidades ribeirinhas. Este estudo tem como objetivo modelar e mapear indicadores ecológicos de diversidade e estrutura da vegetação nas margens do rio Gualaxo do Norte, uma das áreas mais afetadas pelo desastre. Utilizando dados de observação em campo e sensoriamento remoto, foram avaliados dois cenários para compreender a influência das variáveis preditoras na modelagem desses indicadores. O modelo de riqueza de espécies vegetais, ao incorporar dados de campo, alcançou um R^2 de 97,56%, evidenciando a importância de variáveis categóricas para superar as limitações do uso exclusivo de dados de sensoriamento remoto, que apresentaram um R^2 de apenas 22,74%. Já o modelo de cobertura de dossel mostrou bom ajuste com o uso de variáveis de sensoriamento remoto, atingindo um R^2 de 84,62%, destacando a relevância da elevação SRTM como variável preditora. Os resultados obtidos apontam para a eficácia do uso combinado de métodos de sensoriamento remoto e dados de campo no monitoramento de áreas degradadas, com potenciais implicações sociais, tecnológicas e ambientais, sendo uma metodologia com potencial de direcionar esforços para recuperação de áreas prioritárias incluindo os produtores rurais nas estratégias de recuperação dessas áreas. Ao oferecer um modelo robusto para o acompanhamento da recuperação ecológica, o trabalho contribui para a formulação de políticas públicas voltadas à restauração ambiental, beneficiando não apenas o meio ambiente, mas também as populações locais que dependem dos recursos naturais para sua subsistência. Em um contexto mais amplo, a pesquisa promove avanços tecnológicos no uso de algoritmos de aprendizado de máquina e dados de sensoriamento remoto para a modelagem ecológica, o que pode ser replicado em outras regiões impactadas por desastres ambientais. Assim, este estudo não apenas avança o conhecimento científico sobre os processos de recuperação ecológica pós-desastres, mas também oferece ferramentas práticas para a gestão e mitigação de futuros impactos ambientais, tendo em vista a sustentabilidade e o bem-estar das comunidades afetadas.

IMPACT INDICATORS

The collapse of the iron ore tailings dam in Mariana, Minas Gerais, in 2015 triggered one of Brazil's largest environmental disasters, with devastating impacts on the local ecosystem and riparian communities. This study aims to model and map ecological indicators of vegetation diversity and structure along the banks of the Gualaxo do Norte River, one of the areas most affected by the disaster. Using field observation and remote sensing data, two scenarios were evaluated to understand the influence of predictor variables on the modeling of these indicators. The plant species richness model, incorporating field data, achieved an R^2 of 97.56%, highlighting the importance of categorical variables in overcoming the limitations of using only remote sensing data, which presented an R^2 of just 22.74%. On the other hand, the canopy cover model showed a good fit with remote sensing variables, achieving an R^2 of 84.62%, underscoring the relevance of SRTM elevation as a predictor variable. The results obtained point to the effectiveness of combining remote sensing methods and field data in monitoring degraded areas, with potential social, technological, and environmental implications, offering a methodology that can guide efforts towards the recovery of priority areas, including rural producers in the strategies for restoring these areas. By providing a robust model for monitoring ecological recovery, the study contributes to the formulation of public policies aimed at environmental restoration, benefiting not only the environment but also the local populations who depend on natural resources for their livelihoods. In a broader context, the research promotes technological advances in the use of machine learning algorithms and remote sensing data for ecological modeling, which can be replicated in other regions impacted by environmental disasters. Thus, this study not only advances scientific knowledge on post-disaster ecological recovery processes but also offers practical tools for managing and mitigating future environmental impacts, with a focus on sustainability and the well-being of affected communities.

SUMÁRIO

1.	INTRODUÇÃO	11
2.	REFERENCIAL TEÓRICO	12
2.1.	Histórico de degradação da Mata Atlântica	12
2.2.	Impactos ambientais gerados pela atividade minerária	13
2.3.	Monitoramento de áreas em processo de restauração florestal	14
2.4.	Indicadores ecológicos.....	15
2.5.	Sensoriamento remoto aplicado à restauração florestal.....	17
2.6.	Algoritmo Random Forest.....	18
3.	CONSIDERAÇÕES FINAIS	20
	REFERÊNCIAS	21
	SEGUNDA PARTE – ARTIGO	26
1.	INTRODUÇÃO	27
2.	MATERIAL E MÉTODOS.....	28
2.1.	Áreas de estudo.....	28
2.2.	Amostragem e coleta de dados em campo.....	30
2.3.	Amostragem e obtenção de dados de sensoriamento remoto	32
2.4.	Base de dados e pré-processamento.....	33
2.5.	Modelagem	37
3.	RESULTADOS.....	41
3.1.	Estatística descritiva das variáveis resposta	41
3.2.	Filtragem dos dados	43
3.3.	Modelos preliminares.....	43
3.4.	Variáveis selecionadas.....	47
3.5.	Modelos finais	50
3.6.	Mapas estimativos de indicadores.....	54
4.	DISCUSSÃO	56
4.1.	Influência dos dados de campo e de sensoriamento remoto nos modelos	56
4.2.	Variáveis mais importantes para a estimativa dos indicadores	57
4.3.	Problema de saturação e tendências nas estimativas espaciais.....	59
5.	CONCLUSÃO	60
	REFERÊNCIAS	61

PRIMEIRA PARTE

1. INTRODUÇÃO

A degradação ambiental decorrente de atividades humanas representa um dos desafios mais urgentes e complexos enfrentados pela sociedade contemporânea. A mineração, uma atividade vital para a economia global, gera impactos ambientais significativos, comprometendo ecossistemas naturais e ameaçando a biodiversidade. Além de gerar impactos devido a forma como é realizada a extração de minérios, acidentes ambientais podem ocorrer. Um exemplo desse cenário é o desastre ocorrido em Mariana, Minas Gerais, em novembro de 2015, quando o rompimento de uma barragem de rejeitos de minério de ferro resultou em consequências catastróficas para o meio ambiente local. Os rejeitos de minério de ferro foram transportados por mais de 650 km, percorrendo os rios Gualaxo do Norte, Carmo e Doce até atingir o oceano atlântico (PASSOS et al., 2020)

O acidente em Mariana gerou um fluxo de rejeito que causou diversos impactos ambientais, sociais e econômicos, como a contaminação da água e do solo, perda de biodiversidade e supressão da vegetação nativa (CARMO et al., 2017). Desde então, esforços têm sido direcionados para a restauração dessas áreas afetadas, visando recuperar a integridade ecológica e promover a resiliência dos ecossistemas.

A avaliação do progresso e da eficácia de ações de restauração é um desafio complexo, pois depende da disponibilidade de recursos financeiros e tecnológicos que muitas vezes é limitada, dificultando a realização de monitoramentos frequentes em campo. Logo, o monitoramento se torna uma atividade com elevado custo, devido a necessidade de avaliações contínuas em campo por profissionais capacitados (ZAHAWI et al., 2015; REIF; THEEL, 2017). Porém, o sucesso dos programas de restauração ecológica depende da avaliação e monitoramento das áreas em processo de restauração, pois fornecem dados de erros e acertos, permitindo readequação dos métodos (BARBOSA et al., 2003).

Dentre os desafios que residem no monitoramento de áreas em processo de restauração florestal, se destacam a seleção adequada de indicadores ecológicos sensíveis e a forma de coleta desses indicadores, que devem ser capazes de capturar os efeitos das ações de restauração a longo prazo. Superar essas dificuldades requer a combinação de abordagens multidisciplinares, como o uso de tecnologias de sensoriamento remoto, e a colaboração entre pesquisadores, gestores ambientais e comunidades locais.

Estudos utilizando novas tecnologias aplicadas ao sensoriamento remoto estão sendo desenvolvidos, havendo novas oportunidades e iniciativas de estudos abordando modelagens de variáveis ambientais utilizando dados advindos de geotecnologias (AHMED et al., 2015; DE ALMEIDA et al., 2021; DUBE; MUTANGA, 2015; GALLAUN et al., 2010; SILVEIRA et al., 2019; ZAMANI JOHARESTANI et al., 2019; ZHAO et al., 2018).

2. REFERENCIAL TEÓRICO

2.1. Histórico de degradação da Mata Atlântica

A Mata Atlântica foi uma das maiores florestas tropicais das Américas, originalmente cobrindo cerca de 150 milhões de hectares, em condições ambientais altamente heterogêneas (RIBEIRO et al., 2009). Hoje a vegetação remanescente de Mata atlântica encontra-se com, aproximadamente, 12% (16.377.472 hectares) da sua vegetação original, sendo seus remanescentes compostos, em sua maioria, por fragmentos pequenos (inferiores a 50 hectares), isolados e perturbados (RIBEIRO et al., 2009). Apesar de ser o quarto hotspot de biodiversidade do mundo, contendo 2% das espécies endêmicas do mundo (MYERS et al., 2000), a Mata Atlântica apresentou uma taxa de desmatamento de 2,5 hectares por hora entre os anos de 2020 e 2021, gerando uma emissão de 10,3 milhões de toneladas de CO₂ na atmosfera nesse período (SOS MATA ATLÂNTICA, 2022).

A antropização da Mata atlântica ocorre desde antes do Brasil colônia, durante o ápice da dominação Tupi, grupo indígena que ocupou a costa atlântica brasileira por séculos antes da chegada dos colonizadores europeus (DEAN, 1997). Ainda que em menor grau, os Tupi geravam perturbação no bioma devido ao seu modelo de agricultura nômade de derrubada e queima (DEAN, 1997). Segundo Pinto et al. (2014), devido a natureza específica e esporádica da prática adotada pelos Tupi, não houve impacto significativo na Mata Atlântica, podendo ser observado pela regeneração da floresta com o fim da sociedade Tupi.

Com a chegada dos colonos europeus, a principal exploração foi a do pau-brasil (*Caesalpinia echinata*) como fonte de tinta vermelha para tecidos da realeza, causando o risco de extinção da espécie ainda no primeiro século de exploração da colônia (CARDOSO, 2016; DEAN, 1997). Nesse mesmo contexto, a coroa portuguesa fez uma série de concessões para incentivar a ocupação das terras brasileiras e expandir a monocultura de cana-de-açúcar e com o esgotamento dos solos haviam novas concessões de terras, gerando extensas áreas degradadas (PINTO et al., 2014).

Ao fim do século XVII, os portugueses deram início ao terceiro ciclo econômico do país: a mineração. Com o crescimento econômico e populacional, novas terras passaram a ser exploradas não apenas pela mineração, mas também pela agricultura, gerando mais desmatamento na Mata Atlântica, resultando em mais 3 milhões de hectares de floresta destruídos ao final do século XVIII (DEAN, 1997). Hoje, cerca de 70% da população brasileira (aproximadamente 140 milhões de pessoas) vivem no bioma Mata Atlântica (MMA, 2014)

2.2. Impactos ambientais gerados pela atividade minerária

A mineração é responsável por uma série de impactos indesejáveis, incluindo perda de cobertura vegetal, destruição de cursos d'água, perda de biodiversidade, alterações no uso da terra, insegurança alimentar, aumento de vícios e conflitos sociais, elevado custo de vida e poluição atmosférica (WORLANYO; JIANGFENG, 2021).

Em Minas Gerais, observou-se a vulnerabilidade socioambiental na cidade de Itabira, sede da mineradora Vale, onde, segundo Silva e Souza (2002), apesar do aumento da consciência ambiental da sociedade, e a maior atuação do ministério Público em relação às exigências e cobranças do cumprimento das leis ambientais, a situação de vulnerabilidade devido a expansão das minas em direção às áreas urbanas persiste.

A degradação visual da paisagem é um impacto característico da mineração, conforme destacado por Matta (2001), especialmente nas lavras a céu aberto e nos garimpos, que inevitavelmente alteram a topografia do terreno. A mineração a céu aberto geralmente acumula grandes quantidades de estéril (substâncias minerais sem valor econômico). Segundo Viana (2007), quando esse material não é corretamente destinado, durante períodos de chuva, pode ser removido e transportado para áreas mais baixas dos cursos d'água e reservatórios, causando gradual assoreamento. O armazenamento de resíduos de rocha e rejeitos de mineração nas proximidades das áreas de mineração persiste como um problema ambiental de longa duração (AGBOOLA et al., 2020).

O rompimento da barragem de Fundão, em 5 de novembro de 2015, é considerado o maior desastre mundial no setor de mineração em termos de quantidade e distância percorrida pelo rejeito (CORDEIRO et al., 2019). Segundo Morgenstern et al. (2016), o rompimento foi causado pela liquefação do rejeito, um processo em que o material perde sua resistência devido ao acúmulo de água, adquirindo uma consistência líquida e exercendo maior pressão sobre a barragem, levando ao colapso.

O rompimento da barragem de Fundão resultou em vários impactos socioambientais ao longo do percurso do rejeito. Esses impactos incluem a destruição de propriedades públicas e privadas, a supressão de vegetação nativa e plantada, a poluição de corpos d'água, a perda da biodiversidade aquática e terrestre, além de causar a morte de 18 pessoas e um desaparecido (CARMO et al., 2017). Além disso, o acidente afetou o abastecimento de água e energia, e provocou a suspensão das atividades de pesca e turismo, acarretando em enormes prejuízos socioeconômicos (IBAMA, 2015). Omachi et al. (2018) concluíram que o impacto maior foi nos primeiros 74 km a jusante da barragem de Fundão, cobrindo uma área ribeirinha de 1.176,6 ha com lama contaminada e suprimindo 457,6 ha de Mata Atlântica.

2.3. Monitoramento de áreas em processo de restauração florestal

Devido a expansão da restauração florestal no Brasil e, no mundo, a fim de se atender demandas de regularização ambiental, mitigação de impactos e até mesmo para a captura de carbono da atmosfera, os métodos de restauração utilizados têm sido revisados e aprimorados (BRANCALION et al., 2012). Essa evolução ocorre devido ao princípio de observar, registrar e monitorar os tratamentos e as respostas dos tratamentos realizados em projetos de restauração, a fim de informar sobre possíveis alterações e diferentes abordagens em projetos futuros (MCDONALD et al., 2016).

O monitoramento de áreas em processo de restauração florestal é fundamental para o sucesso de projetos e ou programas de restauração, pois no monitoramento se define os indicadores que irão avaliar a eficácia das ações tomadas nesse processo, permitindo a identificação de medidas corretivas diante de indicadores insatisfatórios (WWF, 2017). E apesar da grande importância do tema, o monitoramento de áreas restauradas no Brasil carece de maior atenção. Um dos gargalos está na crença de que a restauração é realizada apenas para atender requisitos de certificação e licenciamento ambiental, faltando um compromisso real com a sustentabilidade ecológica. Em muitos casos, a restauração florestal é comumente confundida basicamente com plantios de mudas, sem acompanhamento ao longo do tempo para a avaliação do desenvolvimento do ecossistema (BRANCALION et al., 2012). Com isso, ações de manejos adaptativos são pouco realizadas, justamente pela não identificação dos problemas devido a falta do monitoramento (BRANCALION; GANDOLFI; RODRIGUES, 2015).

A avaliação envolve a análise de indicadores, dados e aspectos ambientais para determinar o estado da área restaurada e se os objetivos foram alcançados. Em teoria todo ecossistema pode ser decomposto em função e estrutura, sendo um caminho na definição dos

indicadores (DAVIDE; BOTELHO, 2015). No entanto, é necessário estabelecer parâmetros de referência para comparação.

A Sociedade Internacional de Restauração Ecológica (SER, 2004) define que os ecossistemas de referência servem como modelos para o planejamento e avaliação de um projeto de restauração. Conseqüentemente, o papel fundamental do ecossistema referência está em fornecer informações ecológicas cruciais sobre o local de interesse, trazendo informações do possível estado de uma área degradada anterior a perturbação e degradação daquela área (PALMER; ZEDLER; FALK, 2016). Dessa forma, a avaliação pode ser realizada através dos atributos que caracterizam um ecossistema restaurado: (1) presença de espécies do ecossistema de referência, (2) predominância de espécies nativas, (3) inclusão de todos os grupos funcionais essenciais, (4) capacidade do ambiente físico de sustentar populações, (5) funcionamento normal sem sinais de disfunção, (6) integração à paisagem, (7) eliminação ou redução de ameaças oriundas da matriz, (8) resiliência e (9) autossustentabilidade (SER, 2004). Esses critérios possibilitam avaliar o sucesso da restauração ecossistêmica ao compará-los com ecossistemas referência.

Enquanto a avaliação julga o valor sobre o sucesso ou insucesso do processo de restauração, o monitoramento é a constatação, e representa a coleta dirigida de dados sobre indicadores previamente estabelecidos permitindo descrever o estado atual de uma área. Ou seja, o monitoramento é um processo contínuo de coleta de dados que descreve o estado atual da área em restauração, permitindo acompanhar o progresso em relação aos objetivos (e.g. ecossistemas referência) definidos ao longo do tempo (BRANCALION; GANDOLFI; RODRIGUES, 2015).

2.4. Indicadores ecológicos

Os indicadores são variáveis cuja finalidade é medir alterações no tempo em um fenômeno ou processo. Em restauração ecológica, a seleção apropriada de indicadores deve possuir seis características fundamentais: (1) fácil mensuração; (2) sensibilidade a tensões sobre o sistema; (3) que sejam previsíveis as respostas às tensões; (4) que suas alterações sejam capazes de serem previstas, a fim de evitá-las através do manejo; (5) sejam integrativos; e (6) tenham respostas conhecidas aos distúrbios e que essas respostas sejam pouco variáveis (DALE; BEYELER, 2001). Além da seleção apropriada dos indicadores, a eficácia do

monitoramento depende da interpretação precisa desses indicadores e da credibilidade na amostragem, conforme destacado por Viani et al. (2018).

Os indicadores utilizados na avaliação de áreas em restauração podem ser classificados de diversas formas, por exemplo: com base na forma de coleta ou medição (quantitativos ou qualitativos); no atributo do ecossistema avaliado (estrutura, composição, funcionamento, serviços ecossistêmicos); e na época de avaliação (fase de implantação, pós-implantação e vegetação formada) (BRANCALION; GANDOLFI; RODRIGUES, 2015).

Indicadores qualitativos, obtidos subjetivamente, são vantajosos devido à facilidade de monitoramento e menor custo operacional, embora variem entre avaliadores e gerem valores subjetivos. Já os indicadores quantitativos são diretamente mensuráveis, proporcionando uma avaliação segura da área, mas implicam em custos elevados devido ao tempo e necessidade de avaliações contínuas em campo (REIF; THEEL, 2017). Porém, a redução do número de indicadores utilizados no monitoramento pode facilitar e tornar o monitoramento mais acessível. Desta forma, Suganuma e Durigan (2015) sugerem indicadores como cobertura de copas, área basal, densidade e riqueza do sub-bosque, considerando que esses atributos são representações de atributos ou processos mais difíceis de monitorar, ou seja, os indicadores sugeridos são capazes de representar outras variáveis não diretamente mensuráveis ou difíceis de se observar.

Um desafio encontrado ao realizar o monitoramento e avaliação usando indicadores em áreas em restauração está relacionado à falta de parâmetros para a seleção dos indicadores. Devido à grande diversidade de indicadores e à particularidade de cada área, falta consenso na escolha dos indicadores mais adequados para avaliar o sucesso da trajetória de restauração de uma área (SIQUEIRA; MESQUITA, 2007). Contudo, estudos indicam uma hierarquia de previsibilidade nos indicadores de restauração, sendo esta, em ordem decrescente: estrutura física, diversidade taxonômica, diversidade funcional e composição taxonômica (BRUDVIG et al., 2017; LAUGHLIN et al., 2017). Essa classificação influencia a escolha e uso de indicadores em projetos de restauração. Entretanto, diversos pesquisadores alegam que não há um modelo para avaliação e monitoramento que possa ser considerado universal e aplicável em diversos ecossistemas (RODRIGUES; GANDOLFI, 2004).

O sensoriamento remoto é considerado uma ferramenta que, junto à avaliação das áreas, tem elevado potencial para o monitoramento de condições da vegetação (BRIGGS; FREUDENBERGER, 2006). Sendo uma alternativa para reduzir custos, tempo, recursos humanos e, conseqüentemente, facilitar a escolha de indicadores.

2.5. Sensoriamento remoto aplicado à restauração florestal

O sensoriamento remoto é a tecnologia que permite a obtenção de imagens e outros tipos de dados da superfície terrestre por meio da captação e do registro da energia eletromagnética refletida ou emitida pela superfície. A captação e o registro das imagens e dados obtidos ocorrem através de um sistema de sensoriamento remoto, os quais são constituídos por sensores que podem estar instalados em plataformas terrestres (tratores, implementos, etc.), aéreas (drones, aeronaves, etc.) e orbitais (satélites artificiais). Essa captação é feita de forma remota, ou seja, sem o contato físico entre o sensor e o objeto na superfície terrestre (FLORENZANO, 2007).

O sensoriamento remoto com o papel de apoiar o planejamento e monitoramento de projetos de restauração florestal aumentou consideravelmente nas últimas décadas, devido aos novos sensores de satélites com melhores resoluções espaciais e maior disponibilidade de dados, além da diminuição de custo de aquisição de dados (KLEMAS, 2013; GILLESPIE; WILLIS; OSTERMANN-KELM, 2015). As imagens de satélite fornecem uma visão panorâmica e contínua das áreas em restauração, permitindo a detecção de mudanças na cobertura vegetal. Sabendo disso, imagens de satélite foram usadas para rastrear mudanças na cobertura da terra e apoiar estudos de biodiversidade (GILLESPIE et al., 2015; WILLIS, 2015). Sensores como o landsat que estão a décadas em operação, proporcionam a oportunidade de estudar as mudanças associadas à perda e degradação do habitat durante um longo período de tempo (WILLIS, 2015).

Avanços no campo dos micro-satélites de baixo custo têm impactado a capacidade de coleta de dados globais com resoluções espaciais e temporais aprimorados. Nesse contexto, surge a Planet, empresa que adota uma abordagem aeroespacial ágil em todo o ciclo de vida de seus satélites, desde o projeto até o controle de missões e sistemas operacionais. Atualmente, a Planet opera três constelações de satélites de imagens: PlanetScope, RapidEye e Sky-Sat. Cada satélite da constelação PlanetScope é projetado como um CubeSat 3U, com dimensões compactas de 10 cm por 10 cm por 30 cm (PLANET TEAM, 2018). Com isso, novas portas para estudos de monitoramento remoto se abrem.

No entanto, existem alguns desafios associados aos satélites, que incluem a resolução espacial limitada em imagens de satélite de baixa a média resolução, que pode limitar a avaliação detalhada da estrutura da vegetação (REIF; THEEL, 2017). Além disso, a presença de nuvens, sombras e condições climáticas adversas pode afetar a qualidade das imagens, principalmente em regiões e ou épocas do ano muito úmidas (FLORENZANO, 2007; LEE,

2019). Quanto maior for a distância entre os sistemas de sensoriamento remoto e a superfície da terra, menor será a resolução espacial e maior será a interferência na qualidade das imagens (FLORENZANO, 2007). É importante considerar esses aspectos e combinar o uso de imagens de satélite com outras técnicas, ou até mesmo utilizar outras tecnologias de sensoriamento remoto.

Segundo Reif e Theel (2017), dada a necessidade de padrões de monitoramento para avaliar o sucesso da restauração florestal, uma estrutura de monitoramento universal deve incluir uma variedade de opções de sensoriamento remoto as quais possam medir métricas comuns de restauração. Para habitats de florestas e pastagens, isso inclui métricas para identificar a composição e estrutura da floresta, como densidade de cobertura de espécies vegetais, diâmetro da árvore, cobertura de dossel, riqueza de espécies, índice de diversidade de Shannon, entre outros.

Com as inúmeras possibilidades de obtenção de dados através do sensoriamento remoto, novas maneiras de processar e modelar dados têm sido utilizadas. Como por exemplo, a detecção de espécies de árvores e amplas categorias de vegetação usando imagens hiperespectrais utilizando algoritmos de aprendizado de máquina a partir de treinamento e validação dos dados (ZAMANI JOHARESTANI et al., 2019).

O aprendizado de máquina é um grande aliado para a criação de modelos preditivos. A fim de estimar a concentração de material particulado fino no ar, Petropoulos, Arvanitis e Sigrimis (2012) utilizaram três metodologias de aprendizado de máquinas: random forest, aumento de gradiente extremo e deep-learning, as quais apresentaram modelos com desempenhos semelhantes. Silveira et al. (2019) também utilizaram o algoritmo Random Forest, porém, com o objetivo de estimar biomassa acima do solo em diferentes fitofisionomias dos biomas Cerrado e Mata Atlântica, no estado de Minas Gerais, Brasil.

2.6. Algoritmo Random Forest

O algoritmo Random Forest gera modelos preditivos de classificação e ou regressão através de um conjunto de árvores de decisão, as quais são criadas independentemente por um subconjunto de amostras de treinamento. Os nós das árvores de decisão são divididos a partir de parâmetros definidos pelo usuário (BREIMAN, 2001).

As árvores de decisão são produzidas a partir da definição do número de árvores que serão geradas (parâmetro Ntree) e o número de variáveis a serem utilizadas em cada árvore (Mtry). Desta maneira, conforme aumenta-se o número de árvores de decisão, o algoritmo cria

árvores com alta variação e baixa tendência, diminuindo a possibilidade do modelo se ajustar excessivamente aos dados de treinamento (overfitting) e melhorando o modelo (BREIMAN, 2001).

Esse algoritmo é capaz de descrever relações entre múltiplas variáveis complexas, gerando soluções mais acuradas em relação aos métodos convencionais, como as regressões lineares simples e múltiplas (WU et al., 2016). Outra vantagem desse método, é o potencial da utilização de dados espectrais na criação de modelos (GÖRGENS et al. 2015, WU et al. 2016).

Modelagem de atributos ecológicos utilizando algoritmos de aprendizado de máquinas e dados de sensoriamento remoto têm sido cada vez mais estudados (DE ALMEIDA et al., 2021; SILVEIRA et al., 2019). Desta forma, a utilização de geotecnologias mais avançadas, precisas e de menor custo, somadas a abordagens envolvendo aprendizado de máquina para classificação e ou regressão, se destacam para que mais estudos sejam desenvolvidos.

3. CONSIDERAÇÕES FINAIS

A Mata Atlântica vem sofrendo com a antropização a séculos, sendo a maior causa da perda de biodiversidade do bioma. Nesta revisão, o foco se deu nas atividades que intensificaram a perda de biodiversidade ao longo do tempo e os impactos que atividade minerária gera no bioma, dando destaque ao rompimento da barragem de Fundão e a dimensão do impacto causado pela onda de rejeito que se espalhou ao longo da bacia do rio Doce. Além do mais, foi apontada a importância do monitoramento de áreas em processo de restauração, os diferentes tipos de indicadores ecológicos e a complexidade de defini-los e monitorá-los periodicamente, deixando claro que a restauração florestal depende do monitoramento para que exista o sucesso nos projetos e programas de restauração ecológica, não sendo diferente para a restauração das áreas atingidas por rejeito de minério de ferro.

Ficou claro que a tecnologia é um forte aliado ao monitoramento de áreas em processo de restauração e que, além do sensoriamento remoto, algoritmos de aprendizado de máquinas tem se mostrado uma excelente ferramenta em estimar variáveis que se encaixam como indicadores ecológicos (e.g., estimativa de biomassa acima do solo) para áreas maiores, o que pode diminuir os custos e esforços de operações de campo para coletas periódicas. Dessa forma, é importante que mais trabalhos de monitoramento de restauração ecológica, utilizando tecnologias de sensoriamento remoto associadas a algoritmos de aprendizado de máquinas, sejam desenvolvidos.

REFERÊNCIAS

- AASEN, Helge et al. Generating 3D hyperspectral information with lightweight UAV snapshot cameras for vegetation monitoring: From camera calibration to quality assurance. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 108, p. 245-259, 2015.
- AGBOOLA, Oluranti et al. A review on the impact of mining operation: Monitoring, assessment and management. **Results in Engineering**, v. 8, p. 100181, 2020.
- ATLÂNTICA, SOS Mata. INPE, 2022. **Atlas dos remanescentes florestais da Mata Atlântica–período de**, v. 2022, 2022.
- BARBOSA, L. M. et al. Recuperação florestal com espécies nativas no Estado de São Paulo: pesquisas apontam mudanças necessárias. **Florestar estatístico**, v. 6, n. 14, p. 28-34, 2003.
- BITAR, Omar Yazbek. **Avaliação da recuperação de áreas degradadas por mineração na região metropolitana de São Paulo**. 1997. Tese de Doutorado. Universidade de São Paulo.
- BRANCALION, Pedro Henrique Santin et al. Finding the money for tropical forest restoration. **Unasyva**, v. 63, n. 1, p. 25-34, 2012.
- RODRIGUES, Ricardo Ribeiro; GANDOLFI, Sergius; BRANCALION, Pedro Henrique Santin. **Restauração florestal**. Oficina de Textos, 2015.
- BRANCALION, Pedro Henrique Santin et al. Avaliação e monitoramento de áreas em processo de restauração. **Restauração ecológica de ecossistemas degradados**, v. 2, p. 262-293, 2012.
- BREIMAN, L. **Random Forests**. Machine Learning, v. 45, n. 1, p. 5–32, 2001.
- BRUDVIG, Lars A. et al. Interpreting variation to advance predictive restoration science. **Journal of Applied Ecology**, v. 54, n. 4, p. 1018-1027, 2017.
- CARDOSO, Josiane Teresinha. A Mata Atlântica e sua conservação. **Revista Encontros Teológicos**, v. 31, n. 3, 2016.
- DO CARMO, Flávio Fonseca et al. Fundão tailings dam failures: the environment tragedy of the largest technological disaster of Brazilian mining in global context. **Perspectives in ecology and conservation**, v. 15, n. 3, p. 145-151, 2017.
- CBH-DOCE_COMITÊ DA BACIA HIDROGRÁFICA DO RIO DOCE. Disponível em: <<http://www.cbhdoce.org.br/institucional/abacia>>. Acesso em: 03 jun 2023
- CORDEIRO, Marcelle C. et al. Insights on the freshwater microbiomes metabolic changes associated with the world's largest mining disaster. **Science of the Total Environment**, v. 654, p. 1209-1217, 2019.
- DALE, Virginia H.; BEYELER, Suzanne C. Challenges in the development and use of ecological indicators. **Ecological indicators**, v. 1, n. 1, p. 3-10, 2001.

DAVIDE, A. C. et al. Fundamentos e métodos de restauração de ecossistemas florestais: 25 anos de experiência em matas ciliares. **Lavras: UFLA**, 2015.

DE ALMEIDA, Danilo Roberti Alves et al. Monitoring restored tropical forest diversity and structure through UAV-borne hyperspectral and lidar fusion. **Remote Sensing of Environment**, v. 264, p. 112582, 2021.

DEAN, Warren. **With broadax and firebrand: the destruction of the Brazilian Atlantic Forest**. Univ of California Press, 1997.

FARR, Tom G. et al. The shuttle radar topography mission. **Reviews of geophysics**, v. 45, n. 2, 2007.

FLORENZANO, Teresa Gallotti. **Iniciação em sensoriamento remoto**. Oficina de textos, 2007.

GILLESPIE, Thomas W.; WILLIS, Katherine S.; OSTERMANN-KELM, Stacey. Spaceborne remote sensing of the world's protected areas. **Progress in Physical Geography**, v. 39, n. 3, p. 388-404, 2015.

GITELSON, Anatoly A.; KAUFMAN, Yoram J.; MERZLYAK, Mark N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. **Remote sensing of Environment**, v. 58, n. 3, p. 289-298, 1996.

GITELSON, Anatoly A. et al. Novel algorithms for remote estimation of vegetation fraction. **Remote sensing of Environment**, v. 80, n. 1, p. 76-87, 2002.

GÖRGENS, Eric Bastos; MONTAGHI, Alessandro; RODRIGUEZ, Luiz Carlos Estraviz. A performance comparison of machine learning methods to estimate the fast-growing forest plantation yield based on laser scanning metrics. **Computers and Electronics in Agriculture**, v. 116, p. 221-227, 2015.

HARALICK, Robert M., SHANMUGAM, Karthikeyan, et al. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, 1973, no 6, p. 610-621.

HUETE, Alfredo et al. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. **Remote sensing of environment**, v. 83, n. 1-2, p. 195-213, 2002.

HUETE, Alfredo R. A soil-adjusted vegetation index (SAVI). **Remote sensing of environment**, v. 25, n. 3, p. 295-309, 1988.

IBAMA. **Laudo Técnico Preliminar: Impactos ambientais decorrentes do desastre envolvendo o rompimento da barragem de Fundão, em Mariana, Minas Gerais**. Minas Gerais: IBAMA, 2015. 74p

JORDAN, Carl F. Derivation of leaf-area index from quality of light on the forest floor. **Ecology**, v. 50, n. 4, p. 663-666, 1969.

KLEMAS, Victor. Using remote sensing to select and monitor wetland restoration sites: An overview. **Journal of Coastal Research**, v. 29, n. 4, p. 958-970, 2013.

LAUGHLIN, Daniel C. et al. The hierarchy of predictability in ecological restoration: are vegetation structure and functional diversity more predictable than community composition?. **Journal of Applied Ecology**, v. 54, n. 4, p. 1058-1069, 2017.

LEE, Kyu-Sung. Atmospheric correction issues of optical imagery in land remote sensing. **Korean Journal of Remote Sensing**, v. 35, n. 6_3, p. 1299-1312, 2019.

LOUHAICHI, Mounir; BORMAN, Michael M.; JOHNSON, Douglas E. Spatially located platform and aerial photography for documentation of grazing impacts on wheat. **Geocarto International**, v. 16, n. 1, p. 65-70, 2001.

MATTA, Paulo Magno da. Reflexo da mineração na qualidade ambiental das cidades. **Salvador, BA: Universidade Federal da Bahia**, 2001.

McDonald T, Gann GD, Jonson J, and Dixon KW (2016) International standards for the practice of ecological restoration – including principles and key concepts. **Society for Ecological Restoration**, Washington, D.C.

MINISTÉRIO DO MEIO AMBIENTE. Lei nº 9.795, de 27 de abril de 1999. Diário Oficial da União, Brasília, 2000. Disponível em: <<http://www.mma.gov.br>>. Acesso em: 06 fev 2024

MORGENSTERN, Norbert R. et al. Fundão tailings dam review panel report on the immediate causes of the failure of the Fundão dam. **Cleary Gottlieb Steen & Hamilton LLP, New York**, 2016.

MYERS, Norman et al. Biodiversity hotspots for conservation priorities. **Nature**, v. 403, n. 6772, p. 853-858, 2000.

OMACHI, Claudia Y. et al. Atlantic Forest loss caused by the world's largest tailing dam collapse (Fundão Dam, Mariana, Brazil). **Remote Sensing Applications: Society and Environment**, v. 12, p. 30-34, 2018.

PALMER, Margaret A.; ZEDLER, Joy B.; FALK, Donald A. Ecological theory and restoration ecology. **Foundations of restoration ecology**, p. 3-26, 2016.

PASSOS, Larissa Souza et al. Is the Doce River elutriate or its water toxic to *Astyanax lacustris* (Teleostei: Characidae) three years after the Samarco mining dam collapse?. **Science of the Total Environment**, v. 736, p. 139644, 2020.

PETROPOULOS, George P.; ARVANITIS, Kostas; SIGRIMIS, Nick. Hyperion hyperspectral imagery analysis combined with machine learning classifiers for land use/cover mapping. **Expert systems with Applications**, v. 39, n. 3, p. 3800-3809, 2012.

PLANET TEAM. **Planet Application Program Interface: In Space for Life on Earth**. San Francisco, CA. 2018

R Core Team, 2023. R: A Language and Environment for Statistical Computing. **R Foundation for Statistical Computing website**.

- REIF, Molly K.; THEEL, Heather J. Remote sensing for restoration ecology: Application for restoring degraded, damaged, transformed, or destroyed ecosystems. **Integrated environmental assessment and management**, v. 13, n. 4, p. 614-630, 2017.
- RIBEIRO, Milton Cezar et al. The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. **Biological conservation**, v. 142, n. 6, p. 1141-1153, 2009.
- RODRIGUES, Ricardo Ribeiro; GANDOLFI, Sergius. Conceitos, tendências e ações para a recuperação de florestas ciliares. **Matas ciliares: conservação e recuperação**, 2000.
- ROUSE, John Wilson et al. Monitoring vegetation systems in the Great Plains with ERTS. **NASA Spec. Publ**, v. 351, n. 1, p. 309, 1974.
- SAMARCO. **Relatório Bial 2015-2016**. Belo Horizonte: Samarco, 2017. 99p.
- SER. Princípios da SER International sobre a restauração ecológica, 2004
- SILVA, Maria Das Graças Souza; SOUZA, Maria do Rosário. Itabira-Vulnerabilidade Ambiental: impactos e riscos socioambientais advindos da mineração em área urbana. 2002. 2012.
- SILVEIRA, Eduarda MO et al. Pre-stratified modelling plus residuals kriging reduces the uncertainty of aboveground biomass estimation and spatial distribution in heterogeneous savannas and forest environments. **Forest ecology and management**, v. 445, p. 96-109, 2019.
- DE SIQUEIRA, Ludmila Pugliese; MESQUITA, Carlos Alberto Bernardo. **Meu pé de Mata Atlântica: experiências de recomposição florestal em propriedades particulares no Corredor Central**. Instituto BioAtlântica, 2007.
- SUGANUMA, Marcio S.; DURIGAN, Giselda. Indicators of restoration success in riparian tropical forests using multiple reference ecosystems. **Restoration Ecology**, v. 23, n. 3, p. 238-251, 2015.
- TAROLLI, Paolo. High-resolution topography for understanding Earth surface processes: Opportunities and challenges. **Geomorphology**, v. 216, p. 295-312, 2014.
- VIANA, Maurício Boratto. **Licenciamento ambiental de minerações em Minas Gerais: novas abordagens de gestão**. 2007.
- VIANI, Ricardo AG et al. Monitoring young tropical forest restoration sites: how much to measure?. **Tropical Conservation Science**, v. 11, p. 1940082918780916, 2018.
- WILLIS, Katherine S. Remote sensing change detection for ecological monitoring in United States protected areas. **Biological Conservation**, v. 182, p. 233-242, 2015.
- WORLANYO, Adator Stephanie; JIANGFENG, Li. Avaliando o impacto ambiental e económico da mineração para a restauração e uso da terra pós-mineração: Uma revisão. **Revista de Gestão Ambiental**, v. 279, pág. 111623, 2021.

WU, Chaofan et al. Comparison of machine-learning methods for above-ground biomass estimation based on Landsat imagery. **Journal of Applied Remote Sensing**, v. 10, n. 3, p. 035010-035010, 2016.

WULDER, Michael A. et al. The role of LiDAR in sustainable forest management. **The forestry chronicle**, v. 84, n. 6, p. 807-826, 2008.

WWF BRASIL. **Restauração ecológica no Brasil: desafios e oportunidades**. Brasília, 2017. Disponível em: <https://www.wwf.org.br/?60742/Restaurao-ecologica-no-Brasil-desafios-e-oportunidades>

ZAHAWI, Rakan A.; REID, John L.; HOLL, Karen D. Hidden costs of passive restoration. **Restoration Ecology**, v. 22, n. 3, p. 284-287, 2014.

ZAHAWI, Rakan A. et al. Using lightweight unmanned aerial vehicles to monitor tropical forest recovery. **Biological Conservation**, v. 186, p. 287-295, 2015.

ZAMANI JOHARESTANI, Mehdi et al. PM2.5 prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data. **Atmosphere**, v. 10, n. 7, p. 373, 2019.

SEGUNDA PARTE – ARTIGO

ATRIBUTOS TEXTURAIS SE DESTACAM COMO VARIÁVEIS PREDITORAS NA MODELAGEM DE INDICADORES ECOLÓGICOS

Resumo

Em novembro de 2015, ocorreu em Mariana, Minas Gerais, o rompimento de uma barragem de rejeitos de minério de ferro que resultou em consequências catastróficas ao meio ambiente. Desde então, esforços têm sido direcionados para a restauração dessas áreas afetadas, que devem ser monitoradas para verificar a trajetória dos processos ecológicos e evolução no novo ecossistema, pois fornecem dados de erros e acertos, permitindo readequação dos métodos utilizados. Diante desse cenário, o presente estudo tem como objetivo modelar e mapear indicadores ecológicos de diversidade (riqueza de espécies, modelo I) e estrutura (cobertura de dossel, modelo II) às margens do rio Gualaxo do Norte, utilizando dados de observação de campo e dados de sensoriamento remoto como variáveis preditoras. Avaliou-se dois cenários (cenário A e cenário B), a fim de compreender a influência das variáveis de observação de campo nos modelos (cenário A, gerando os modelos IA e IIA), bem como o uso exclusivo de variáveis de sensoriamento remoto (cenário B, gerando os modelos IB e IIB). Os indicadores foram coletados em 24 transectos em seis áreas diferentes, juntamente com os dados de observação de campo. Os dados de sensoriamento remoto foram gerados a partir de imagens PlanetScope, bem como dados do modelo digital de elevação da missão SRTM. Para realizar a modelagem e o mapeamento dos indicadores, foi utilizado o algoritmo de aprendizado de máquina Random Forest (RF). O modelo para riqueza de espécies vegetais alcançou R^2 de 97,56%, superando a limitação usando apenas dados de sensoriamento remoto ($R^2 = 22,74\%$). O modelo de cobertura de dossel obteve bom ajuste com dados de sensoriamento remoto ($R^2 = 84,62\%$) e demonstrou a importância da elevação SRTM como variável preditora.

Palavras-chave: Sensoriamento remoto. Restauração Ecológica. Aprendizagem de Máquina.

Abstract

In November 2015, an iron ore tailings dam collapsed in Mariana, Minas Gerais, resulting in catastrophic consequences for the environment. Since then, efforts have been directed towards the restoration of these affected areas, which must be monitored to verify the trajectory of ecological processes and evolution in the new ecosystem, as they provide data on errors and successes, allowing readjustment of the methods used. Given this scenario, the present study aims to model and map ecological indicators of diversity (species richness, Model I) and structure (canopy cover, Model II) on the banks of the Gualaxo do Norte River, using field observation data and Remote Sensing (RS) data as predictor variables. Two scenarios were evaluated (Scenario A and Scenario B) to understand the influence of field observation variables on the models (Scenario A, generating Models IA and IIA), as well as the exclusive use of remote sensing variables (Scenario B, generating Models IB and IIB). The indicators were collected on 24 transects in six different areas, together with field observation data. Remote sensing data were generated from PlanetScope imagery as well as data from the SRTM mission's digital elevation model. To carry out the modeling and mapping of indicators, the Random Forest (RF) machine learning algorithm was used. The model for plant species richness achieved an R^2 of 97.56%, overcoming the limitation using only remote sensing data ($R^2 = 22.74\%$). The canopy cover model obtained a good fit with remote sensing data ($R^2 = 84.62\%$) and demonstrated the importance of SRTM elevation as a predictor variable.

Keywords: Remote sensing. Ecological Restoration. Machine Learning.

1. INTRODUÇÃO

Em 2021 foi declarada pela Organização das Nações Unidas (ONU) a década da Restauração de Ecossistemas, programada até 2030, com o objetivo de impulsionar a recuperação de ecossistemas degradados em escala global. Essa iniciativa demanda uma abordagem transdisciplinar para assegurar seu êxito (FISCHER et al. 2020; EDRISI; ABHILASH, 2021). Assim, programas de restauração ecológica necessitam de um monitoramento eficaz, pois este é essencial para avaliar o êxito das iniciativas de restauração e assegurar a sustentabilidade dos ecossistemas recuperados (HERRICK et al., 2006). Essas iniciativas são fundamentais em ambientes fortemente degradados, como a Mata Atlântica (BRANCALION et al., 2019)

Os indicadores ecológicos desempenham um papel crucial no processo de monitoramento, fornecendo uma base para avaliar a eficácia das ações de restauração e das políticas ambientais (RAPPORT; HILDÉN, 2013). A seleção de indicadores apropriados é um processo complexo que requer objetivos e metas de restauração claros, bem como um modelo da trajetória esperada (PRACH et al., 2019). Indicadores ecológicos são fundamentais para avaliar a eficácia da restauração de ecossistemas florestais, porém, a mensuração de certos indicadores pode ser desafiadora, podendo ser facilitada através do uso de tecnologias de sensoriamento remoto (CAMARRETA et al., 2020).

A cobertura de dossel foi identificada como um dos principais indicadores para monitorar florestas em restauração e está diretamente ligado à recuperação de biomassa, proporcionando abrigo para a vida selvagem e contribuindo para serviços ecossistêmicos essenciais, como a restauração do microclima, a ciclagem de nutrientes, o controle da erosão e a regulação dos recursos hídricos (SUGANUMA; DURIGAN, 2015). Outro indicador importante é a riqueza de espécies vegetais, o qual fornece informações valiosas para o planejamento e gestão da conservação (ASHTON, 1992) e é comumente utilizado em avaliações de restauração florestal como um indicador de diversidade (GATICA-SAAVEDRA et al., 2017).

Diferentes trabalhos destacam a dificuldade em desenvolver ferramentas cientificamente confiáveis para medir o sucesso da restauração, e apontam a necessidade de protocolos de avaliação mais robustos, especialmente na restauração florestal, onde falta análise de indicadores ecológicos (DOREN, et al, 2009; GATICA-SAAVEDRA et al., 2017). Apesar de a coleta de indicadores ser dispendiosa e demorada, avanços tecnológicos como a detecção remota podem ajudar a mitigar esses desafios (REIF; THEEL, 2017).

O uso do sensoriamento remoto no planejamento e monitoramento da restauração tem se mostrado como uma abordagem eficiente, sendo capaz de medir atributos de recuperação ecológica, especialmente em projetos de restauração terrestre (MCKENNA et al., 2022). No entanto, embora o sensoriamento remoto seja eficiente no monitoramento da cobertura vegetal, a detecção remota pode ser menos eficaz na identificação de espécies individuais, especialmente em habitats complexos (SHUMAN; AMBROSE, 2003). Portanto, a combinação de métodos de detecção remota combinados com dados de campo pode ser uma abordagem mais eficaz para monitorar o sucesso da restauração.

Estudos utilizando índices de vegetação, medidas texturais de imagens de satélite, modelo digital de elevação e o algoritmo de aprendizagem de máquinas Random Forest, têm sido desenvolvidos para estimar variáveis ambientais, melhorando expressivamente as suas estimativas (AHMED et al., 2015; DE ALMEIDA et al., 2021; DUBE; MUTANGA, 2015; GALLAUN et al., 2010; SILVEIRA et al., 2019; ZAMANI JOHARESTANI et al., 2019; ZHAO et al., 2018).

Visando contribuir para a avaliação dos processos de restauração, o principal objetivo desta pesquisa consiste em modelar e mapear indicadores ecológicos de diversidade (riqueza de espécies) e estrutura (cobertura de dossel) em áreas afetadas por rejeito de mineração de ferro. Utilizando dados coletados em campo, informações provenientes de sensoriamento remoto e algoritmo de aprendizagem de máquinas, a pesquisa busca: (1) Analisar a influência de dados coletados em campo, atributos derivados de sensores ópticos (índices espectrais e atributos texturais) e dados da missão SRTM (elevação) na estimativa da riqueza de espécies e cobertura de dossel; (2) Classificar as variáveis mais relevantes com base em sua importância usando o algoritmo Random Forest para a estimativa de cada indicador; (3) Avaliar os modelos estimativos gerados pelo algoritmo Random Forest; (4) Gerar e avaliar modelos usando exclusivamente variáveis preditoras de sensoriamento remoto; e (5) Produzir mapas estimativos dos indicadores ao longo das margens do rio Gualaxo do Norte, no município de Mariana, MG.

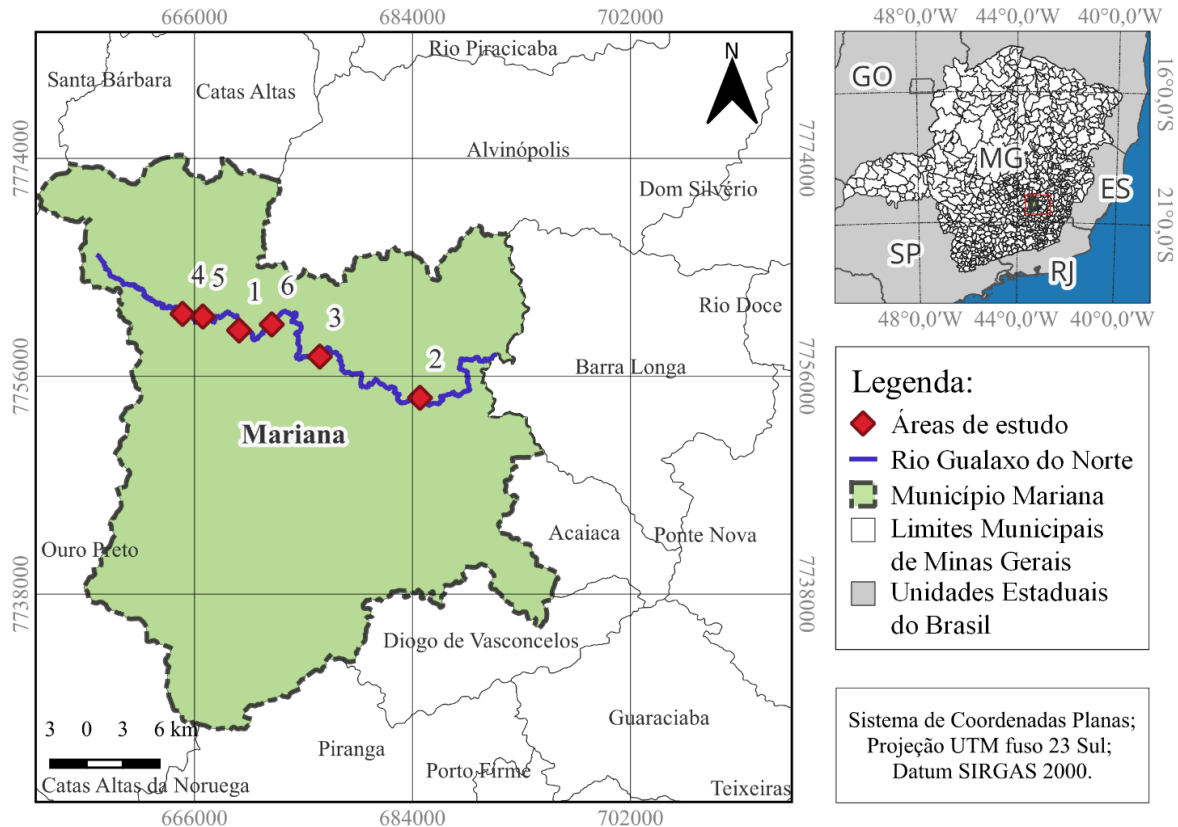
2. MATERIAL E MÉTODOS

2.1. Áreas de estudo

As áreas avaliadas estão inseridas na bacia hidrográfica do rio Doce, a qual apresenta uma área de drenagem de, aproximadamente, 87 mil km² e abrange dois estados, Minas Gerais e Espírito Santo (CBH RIO DOCE, 2018). O estudo foi conduzido ao longo das margens do rio

Gualaxo do Norte, nos primeiros 60 km a jusante da barragem de Fundão, em seis áreas pertencentes a produtores rurais, localizadas na região central do estado de Minas Gerais (IBGE, 2017), no município de Mariana (Figura 01).

Figura 01 – Localização das áreas de estudo.



A vegetação predominante na região é constituída por matas estacionais semidecíduas, pertencentes ao bioma Mata Atlântica (FONSECA, 1997) e, segundo a classificação de Köpen, apresenta clima temperado chuvoso (Cwa), com estação pluviosa de novembro a abril seguido de estação seca de maio a outubro. As taxas médias mínimas e máximas de temperatura e umidade variam, respectivamente, de 18° C a 25° C e 200 mm a 1050 mm (PIRES et al., 2017). Segundo dados fornecidos pela Agência Nacional de Águas (2013), o relevo é acidentado com predominância de Latossolos Vermelho - Amarelo e Argissolos Vermelho – Amarelo.

Dentre as seis áreas de estudo, cinco áreas apresentam deposição de rejeito da barragem de Fundão e estão em processo de restauração, enquanto que a sexta área foi avaliada em área sem deposição de rejeito, sendo esta considerada como ecossistema referência para a realização das avaliações (Tabela 01). A área 6 apresenta vegetação mais densa e está localizada em terrenos mais declivosos, com menor acesso para circulação de gado (Figura 02).

Tabela 01 – Descrição das áreas de estudo da bacia do rio Doce, no estado de Minas Gerais.

Área	Dimensão (ha)	Rejeito	Intervenções de Restauração	Ano de plantio de mudas
1	2,0	P	P	2018
2	3,0	P	P	2018
3	1,0	P	P	2019
4	2,0	P	P	2019
5	2,0	P	P	2020
6	1,0	A	A	-

Em que: P = Presente; e A = Ausente.

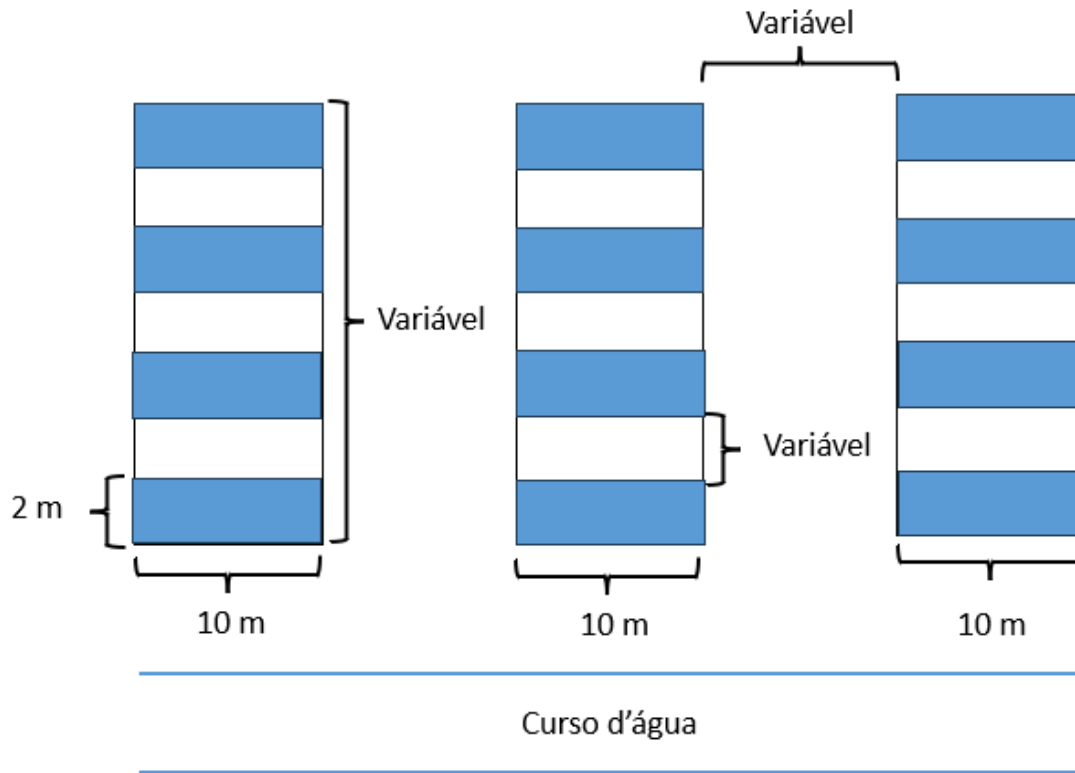
Fonte: Do autor (2024)

Houveram intervenções em cinco áreas que foram impactadas pelo rejeito mineral, utilizando diferentes técnicas de restauração ativa, tanto por meio da semeadura de um mix de sementes de espécies herbáceas/arbustivas de rápido crescimento, a partir de 2016, quanto o plantio de mudas de espécies florestais de ocorrência regional em diferentes épocas após o rompimento (tabela 01).

2.2. Amostragem e coleta de dados em campo

As coletas de dados foram realizadas entre os meses de janeiro e junho de 2021. Para a coleta dos dados em campo, foram demarcados transectos de 10 metros de largura e comprimento variável, de acordo com a distância da margem do rio e o final da área impactada pelo rejeito mineral em restauração, e dentro de cada transecto foram alocadas quatro parcelas de 2 m de comprimento por 10 m de largura (Figura 02). Na área 6, área referência, os transectos foram demarcados em área não impactada pelo rejeito. A intensidade de transectos demarcados variou de acordo com o tamanho de cada área. Dessa maneira, nas áreas 1, 4 e 5 foram avaliados cinco transectos em cada uma; na área 2, oito transectos; e nas áreas 3 e 6, três transectos. Como os transectos variam de dimensão, devido às características de cada área, os transectos variam de 177 a 1387m². Todos os transectos em que foram encontrados valores iguais a zero, foram excluídos da base de dados para a construção do modelo, restando 24 transectos para riqueza de espécies vegetais e 22 para cobertura de dossel. Cada transecto teve o ponto de seu vértice mais distante do rio georreferenciado com auxílio de um GPS.

Figura 02 – Representação esquemática dos transectos e parcelas para avaliação dos indicadores ecológicos nas áreas de estudo.



Fonte: Do autor (2024)

Com a finalidade de coletar indicadores que contemplem aspectos importantes do ecossistema florestal em estudo, os indicadores selecionados, foram: cobertura de dossel e riqueza de espécies.

Para a aferição da cobertura do dossel foi utilizado um densiômetro esférico convexo. As aferições foram realizadas no centro de cada parcela, em quatro direções (norte, sul, leste e oeste), a cerca de 1,30 m do solo. No espelho do densiômetro encontram-se 24 quadrantes. O total de quadrantes que refletiam a copa das árvores foram somados e multiplicados por 1,04. Através da média das quatro aferições, obteve-se a porcentagem de cobertura do dossel em cada parcela (LEMMON, 1957). A média das parcelas foram somadas e divididas pelo número de parcelas em cada transecto (quatro) e, por fim, obtivemos a porcentagem média de cobertura do dossel em cada transecto.

A coleta do indicador de diversidade foi realizada a partir de inventários do estrato arbóreo e a amostragem incluiu todos os indivíduos arbustivos e arbóreos com DAP (diâmetro a 1,30 m do solo) igual ou superior a 5 cm, presentes em cada transecto. A identificação das espécies foi realizada *in locu* e, quando necessário, foram preparadas exsicatas para a

identificação por especialistas ou através da literatura. A classificação se deu em famílias, nomes científicos e respectivos autores, de acordo com o banco de dados da Lista de Espécies da Flora Brasileira (FLORA DO BRASIL, 2020).

2.3. Amostragem e obtenção de dados de sensoriamento remoto

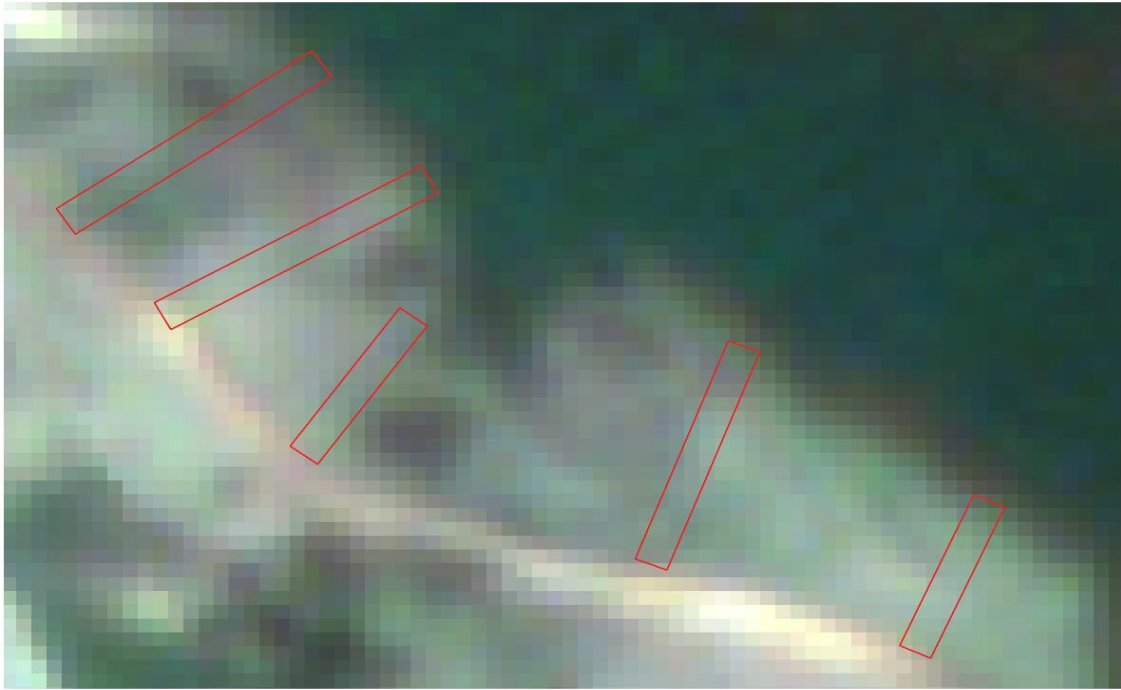
Os dados de sensoriamento remoto (SR) foram obtidos de sensores ópticos embarcados em satélite e as imagens obtidas são do período chuvoso, referentes ao primeiro semestre de 2021, assim como os dados de campo.

As imagens de satélite foram obtidas através da plataforma da empresa Planet, a qual disponibiliza imagens mensais, gratuitas, imageadas pela constelação de satélites PlanetScope. As imagens Planet possuem resolução espacial que varia, aproximadamente, de 3,0 a 4,5 metros, a depender da altitude do satélite (Ground Sample Distance – GSD observado foi de 4,5 m). As imagens Planet contêm quatro bandas espectrais, sendo elas: azul (*Blue*, 455-515 nm), verde (*Green*, 500-590 nm), vermelho (*Red*, 590-670 nm) e o infravermelho próximo (*NIR*, 780-860 nm). O tempo de revisita é diário na órbita síncrona ao Sol, com uma capacidade de 340 milhões de km² por dia (Planet Team, 2018).

O Modelo Digital de Elevação (MDE) com uma resolução espacial de 30 metros, obtido por meio da missão SRTM (FARR et al., 2007), foi empregado para extrair os dados referentes à elevação média.

Utilizou-se a camada vetorial do córrego Santarém e rio Gualaxo do Norte (caminho percorrido pelo rejeito) como máscara a fim de cortar as camadas raster. Para isso, foi feito um *buffer* de 900 m de raio nos vetores dos cursos d'água. Sendo assim, as imagens representam uma faixa de, aproximadamente, 900 m para cada margem do curso d'água. A partir do georreferenciamento dos transectos, foram extraídas as variáveis de sensoriamento remoto, tanto para as imagens planet, como para o MDE da missão SRTM.

Figura 03 - Relação entre o tamanho das parcelas e os pixels da imagem Planet com 4,5 m de resolução espacial.



Fonte: Do autor (2024)

2.4. Base de dados e pré-processamento

Os dados foram processados utilizando o Software QGIS 3.28, software de código aberto, gratuito para qualquer usuário, o que facilita a replicabilidade do trabalho.

2.4.1. Dados de campo

As variáveis de campo que constituem parte da base de dados, são: riqueza de espécies (nº de espécies), cobertura de dossel (%) e observações de campo (Tabela 02). Onde, riqueza de espécies e cobertura de dossel, são as variáveis respostas e não coexistem na base de dados. Ou seja, riqueza de espécies só foi mantida na base de dados quando o objetivo foi estimar riqueza de espécie, e o mesmo foi feito ao estimar cobertura de dossel.

Tabela 02 – Observações de campo feitas para cada área de estudo presentes na base de dados.

Área	Matriz	Rejeito	Técnicas de Restauração	Idade do plantio de mudas (anos)	Gado	Roçada	Uso anterior	Estágio
1	Con.	P	P	3	P	P	Con.	NR
2	Ant.	P	P	3	P	P	Deg.	NR
3	Con.	P	P	2	P	P	Con.	NR
4	Ant.	P	P	2	P	P	Deg.	NR
5	Con.	P	P	1	P	P	Con.	NR
6	Con.	A	A	> 30	A	A	Con.	Ref.

Em que: Con. = Conservada; Ant.= Antropizada; Deg.= Degradada; P = Presente; A = Ausente; NR = Não restaurada; e Ref.= Referência

Fonte: Do autor (2024)

Desta forma, 8 variáveis de observações de campo fizeram parte da base de dados e são, em sua maioria, dados categóricos, que serão utilizados para a modelagem dos indicadores. O pacote *randomForest* (BREIMAN, 2001) faz a conversão dos dados categóricos em numéricos de forma automatizada.

2.4.2. Índices de Vegetação

Os Índices de Vegetação (IV) foram gerados através da *calculadora raster* do software QGIS, sendo calculados 7 IV (Tabela 03), a fim de compor parte das variáveis preditoras do modelo.

Tabela 03 – Índices de vegetação (IV) que foram processados e incluídos na base de dados.

Índices de Vegetação (IV)	Fórmulas dos IV	Referência
Normalized Difference Vegetation Index (NDVI)	$NDVI = (\rho \text{ NIR} - \rho \text{ Vermelho}) / (\rho \text{ NIR} + \rho \text{ Vermelho})$	ROUSE et al., 1974
Enhanced Vegetation Index (EVI)	$EVI = 2.5 \times ((\rho \text{ NIR} - \rho \text{ Red}) / (\rho \text{ NIR} + 6 \times \rho \text{ Red} - 7.5 \times \rho \text{ Blue} + 1))$	HUETE et al., 2002
Simple Ratio (SR)	$SR = \rho \text{ NIR} / \rho \text{ Vermelho}$	JORDAN, 1969
Soil-Adjusted Vegetation Index (SAVI)	$SAVI = ((\rho \text{ NIR} - \rho \text{ Vermelho}) / (\rho \text{ NIR} + \rho \text{ Vermelho} + 0.5)) \times (1 + 0.5)$	HUETE, 1988
Atmospherically Resistant Index (VARI)	$VARI = (\rho \text{ Verde} - \rho \text{ Vermelha}) / (\rho \text{ Verde} + \rho \text{ Vermelha} - \rho \text{ Azul})$	GITELSON et al., 2002

Tabela 03 – Índices de vegetação (IV) que foram processados e incluídos na base de dados (continua).

Índices de Vegetação (IV)	Fórmulas dos IV	Referência
Green Normalized Difference Vegetation Index (GNDVI)	$GNDVI = \frac{(\rho \text{ NIR} - \rho \text{ Verde})}{(\rho \text{ NIR} + \rho \text{ Verde})}$	GITELSON et al., 1996
Green Leaf Index (GLI)	$GLI = \frac{(2x \rho \text{ Verde} - \rho \text{ Vermelha} - \rho \text{ Azu})}{(2x \rho \text{ Verde} + \rho \text{ Vermelha} + \rho \text{ Azu})}$	LOUHAICHI, BORMAN E JOHNSON, 2001

Nomenclatura das variáveis de IV na base de dados = media_IV_planet, onde **média** = valor médio da variável extraída de cada transecto; **IV** = índice de vegetação avaliado (NDVI, SAVI, etc); **planet** = fonte dos dados. Fonte: Do autor (2024)

2.4.3. Atributos texturais

A medida de textura é obtida a partir dos níveis de cinza dos pixels dentro de uma unidade de análise, conhecida como kernel. As medidas de textura podem ser classificadas como de primeira ordem, a qual resume estatisticamente os valores dos pixels, e como de segunda ordem, a qual se dá a partir da Matriz de Coocorrência de Níveis de Cinza (GLCM). A GLCM contém as frequências normalizadas das ocorrências de valores de pares de pixels adjacentes dentro da unidade de análise (kernel), ou seja, leva em consideração a disposição espacial dos pixels.

Neste estudo foram gerados 12 atributos texturais (Tabela 04), sendo a grande maioria atributos de segunda ordem, com exceção da Média (Mean), Variância (Variance) e Entropia (Entropy), considerados como atributos de primeira ordem. Todos os 12 atributos foram gerados para as 4 bandas (Red, Green, Blue e NIR). Portanto, foram gerados 48 atributos texturais (AT).

Tabela 04 – Atributos texturais (AT) que foram processados e incluídos na base de dados.

Atributos Texturais (AT)	Descrição	Referência
Energy	media_*_text_energy_planet	
Entropy	media_*_text_entropy_planet	
Correlation	media_*_text_corr_planet	
Inverse Difference Moment (IDM)	media_*_text_IDM_planet	
Inertia	media_*_text_iner_planet	HARALICK et al.,1973
Cluster Shade	media_*_text_shade_planet	e
Cluster Prominence	media_*_text_CProm_planet	HaralickTextureExtraction — Orfeo ToolBox 8.1.0 documentation, 2023
Mean	media_*_text_mean_planet	
Variance	media_*_text_var_planet	
Difference of Entropy	media_*_text_DiffEnt_planet	
Difference of Variance	media_*_text_DiffVar_planet	
IC1	media_*_text_IC1_planet	

Em que: **média** = valor médio da variável extraída de cada transecto; * = bandas da imagem; **text** = Medida de textura; **planet** = fonte dos dados. Fonte: Do autor (2024)

Assim como os índices de vegetação, os AT foram gerados utilizando o Software QGIS, porém, nesse caso, utilizou-se o plugin OrfeoToolBox, através da função *HaralickTextureExtraction*.

Os parâmetros da função foram mantidos padrão, com exceção dos seguintes parâmetros: unidade de análise (tamanho do kernel), definido em 3 x 3 pixels; do valor padrão máximo e mínimo de cada imagem, definidos com base no histograma de cada banda; e o valor do número do compartimento do histograma (Histogram number of bin), o qual foi adotado o valor de 64.

2.4.4. Elevação

Os dados de elevação foram extraídos da missão SRTM, como foi salientado anteriormente. A variável elevação (*SRTM_elev*) foi incluída na base de dados a partir da média da elevação em cada transecto.

2.4.5. Extração dos valores

A partir de cada transecto, utilizou-se a função *Estatísticas Zonais* do software QGIS com o objetivo de extrair os valores médios de cada IV, AT e elevação. O número de variáveis preditoras presentes na base de dados para a modelagem de cada um dos indicadores foi de 64, considerando os dados de observações de campo, IV, AT e elevação.

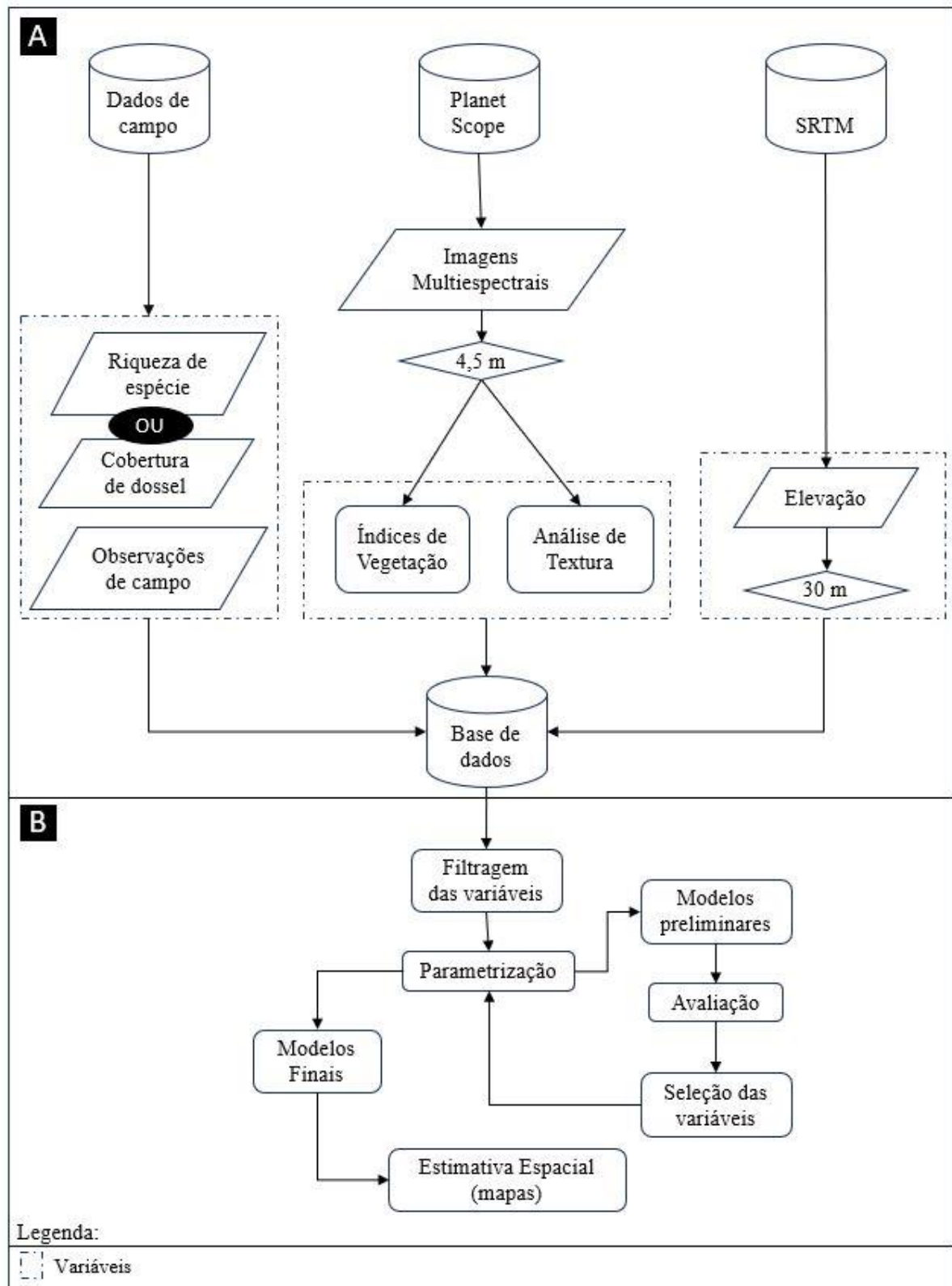
2.5. Modelagem

A fim de estimar riqueza de espécies (modelo I) e cobertura de dossel (modelo II) ao longo das margens do rio Gualaxo do Norte, foi utilizado o modelo *Random Forest* (BREIMAN, 2001). O fluxograma geral do trabalho traz de forma resumida a formação da base de dados e dos modelos (Figura 04). Na primeira etapa (Figura 04 – A), a base de dados obtidas de campo, das imagens Planet Scope e da missão SRTM foram concatenadas utilizando a linguagem de programação estatística R (R CORE TEAM, 2023). Na segunda etapa (Figura 04 – B), além do pacote *Random Forest*, utilizou-se o pacote *Caret* (KUHNS, 2008), no processo de filtragem das variáveis, parametrização, seleção das variáveis e otimização dos modelos.

O algoritmo Random Forest (RF) é uma técnica de aprendizado de máquina não-paramétrica, portanto, adaptam-se de maneira mais flexível a padrões complexos e desconhecidos nos dados de treinamento. Neste método, uma parcela aleatória dos dados de treinamento é selecionada para construir uma árvore de decisão, enquanto a parte restante dos dados não utilizada é empregada para avaliar o erro associado a cada árvore. Em cada nó da árvore de decisão, um subconjunto aleatório de variáveis preditoras é escolhido para realizar a divisão dos dados. Esse processo envolve a construção de diversas árvores de decisão. Com isso, a predição ou estimativa de novos dados é feita através da agregação, por exemplo, da média das predições observadas pelas árvores construídas (LIAW; WIENER, 2002). Além disso, o RF possui a capacidade de identificar as variáveis cruciais para o modelo e gerar uma medida de erro de previsão independente (ADAM et al., 2014).

Para criar as árvores de decisão, é necessário definir dois parâmetros: o número de árvores a serem geradas (*Ntree*) e o número de variáveis a serem utilizadas em cada árvore (*Mtry*). Permitindo o aumento do número de árvores de decisão (*Ntree*), o algoritmo gera árvores caracterizadas por alta variação e baixa tendência (BREIMAN, 2001).

Figura 04 – Fluxograma geral para a formação da base de dados (A) e modelagem (B).



Fonte: Do autor (2024)

2.5.1. Filtragem das variáveis

Devido a quantidade de variáveis preditoras presentes na base de dados e a quantidade de amostras (24) relativamente baixa, foi utilizado o método de filtragem (*filter*) para uma primeira seleção das variáveis preditoras. Nesse caso, utilizou-se a função *cor* do pacote *stats* (R CORE TEAM, 2023), a fim de se obter a matriz de correlação de Spearman, eliminando assim variáveis com elevada correlação. Desse modo, todas as variáveis que apresentaram correlação superior a 0,8 foram excluídas da base de dados com o objetivo de se reduzir a multicolinearidade (TABACHNICK; FIDELL, 2013).

2.5.2. Parametrização

Com exceção do parâmetro *Mtry*, os parâmetros adotados para gerar os modelos foram os padrões do Random Forest para regressão. A fim de se definir o melhor *Mtry* para a base de dados, foi utilizado o método de Validação Cruzada Aninhada (VCA). Esse método é uma extensão da Validação Cruzada (VC) *k-fold*. Basicamente, a VC *k-fold* consiste em (1) particionar o conjunto de dados em subconjuntos, (2) analisar os subconjuntos (dados de treinamento) e (3) validar a análise em um dos subconjuntos (dados de teste). Os subconjuntos são particionados de maneira aleatória, onde esse processo de dividir, treinar e validar se repete *k* vezes.

A VCA é um conceito estatístico utilizado para estimar, de forma mais robusta, a capacidade de generalização de um modelo, dividindo os dados em conjuntos de treinamento e teste de forma estratificada. A VCA ocorre em dois níveis, um primeiro nível denominado de Validação Cruzada Externa (VCE) e, um segundo nível denominado de Validação Cruzada Interna (VCI). No primeiro nível, estima-se a capacidade preditiva do modelo, enquanto, no segundo nível, a partição dos dados é utilizada para otimizar a parametrização do modelo.

Neste estudo, definiu-se a partição dos dados em cinco ($k = 5$) para a VCE e em três para VCI ($i = 3$), dessa maneira os dados são particionados em cinco dobras (*folds*) durante a VCE. Em cada iteração externa, quatro *folds* são usados para treinamento, enquanto o quinto é reservado para teste (XIONG et al. 2020). Simultaneamente, para cada iteração externa, uma VCI com três *folds* é realizada nos dados de treinamento da iteração externa. A abordagem aninhada contribui para uma avaliação mais confiável do modelo, evitando viés nos resultados.

Nesse processo de VCA, visando o melhor *Mtry*, delimitou-se o *Mtry* mínimo e o máximo para que o modelo fosse coerente com a base de dados. Sendo assim, o valor mínimo

testado para o Mtry foi igual a duas variáveis preditoras (n), enquanto o valor máximo foi o número de variáveis preditoras presente na base de dados menos um ($n - 1$). A avaliação do parâmetro se deu através da métrica *Root Mean Square Error* (RMSE), sendo a escolha do Mtry relacionada ao menor valor do RMSE. Dessa maneira, as observações foram divididas aleatoriamente em 70% para dados de treinamento e 30% para dados de teste.

2.5.3. Seleção das variáveis

Existe uma melhora frequente da eficiência do algoritmo quando se reduz a dimensão dos dados (WITTEN; FRANK, 2005). Portanto, foram gerados modelos preliminares, onde foi feita a parametrização dos modelos e a avaliação das variáveis que mais contribuíram com os modelos. Feito isso, houve uma nova seleção de variáveis a partir da avaliação da importância relativa das variáveis preditoras, onde as variáveis que contribuíram menos com os modelos foram excluídas. Dessa maneira, o número de variáveis preditoras é reduzido e o modelo se torna mais simples e eficiente.

2.5.4. Avaliação dos modelos

A fim de se atender os objetivos específicos foram avaliados dois cenários (Tabela 05), um onde os modelos preliminares foram gerados utilizando dados de SR junto aos dados de observação de campo (cenário A) e, outro, onde os modelos foram gerados utilizando apenas os dados de SR (cenário B). Desta forma, pode-se observar a importância das variáveis de observação de campo para a modelagem de cada indicador.

Tabela 05 – Cenários para avaliação do desempenho dos modelos.

Cenários	Indicadores	Modelos	Variáveis preditoras
A	RE	IA	Variáveis SR + observações de campo;
B	RE	IB	Apenas as variáveis de SR;
A	CD	IIA	Variáveis SR + observações de campo;
B	CD	IIB	Apenas as variáveis de SR

RE = Riqueza de espécies; CD = Cobertura de dossel; SR = Sensoriamento remoto.

Fonte: Do autor (2024).

Os modelos IA e IIA somente seguiram para a etapa de modelagem final caso as variáveis de observação de campo apresentassem alta importância relativa no modelo, caso contrário, apenas os modelos IB e IIB seguiriam para a modelagem final.

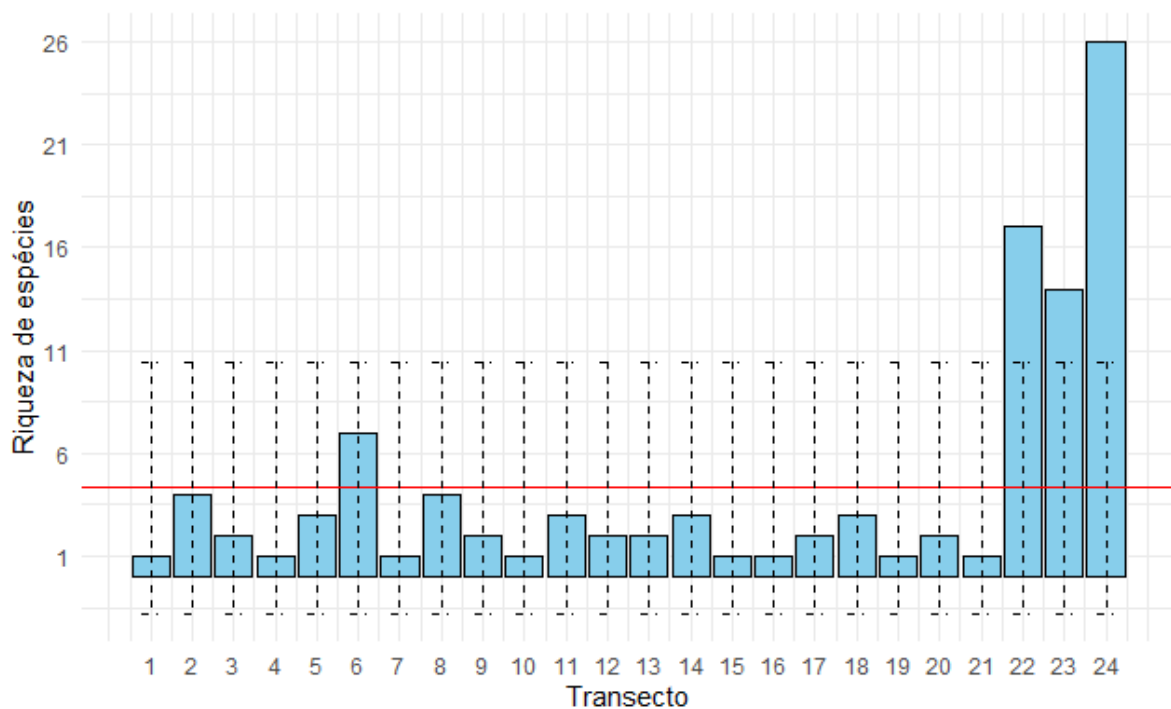
Os modelos foram avaliados por meio do coeficiente de determinação (R^2), raiz quadrada do erro médio (*Root Mean Square Error* – RMSE) e do erro médio absoluto (*Mean Absolut Error* – MAE). Dessa maneira, pode-se avaliar os melhores modelos e a capacidade estimativa de cada um deles através das medidas de erro. Os mapas estimativos dos indicadores riqueza de espécie e cobertura de dossel foram gerados pelos modelos IB e IIB, os quais utilizaram apenas as variáveis preditoras de sensoriamento remoto em suas modelagens.

3. RESULTADOS

3.1. Estatística descritiva das variáveis resposta

Na figura 05, pode-se observar os valores de riqueza de espécie para cada transecto avaliado. O valor do Coeficiente de Variação (CV) do indicador de riqueza de espécie foi de 141,05%. Os transectos 22, 23 e 24 representam a área referência, enquanto os demais transectos representam as áreas em processo de restauração.

Figura 05 – Riqueza de espécies (nº de espécies) por transectos amostrados.

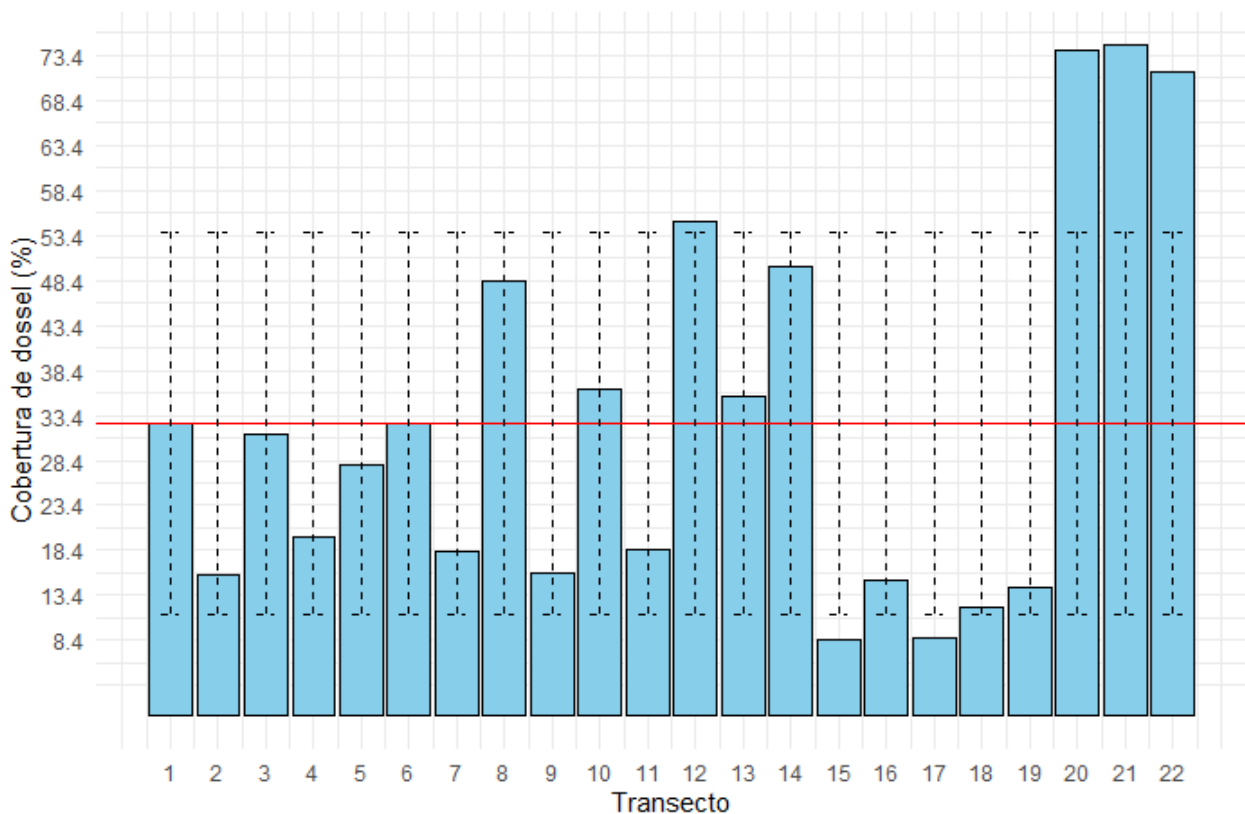


Média (linha vermelha) = 4 espécies; Desvio padrão (barra tracejada) = 6 espécies. Fonte: Do autor (2024)

Além do CV elevado, observa-se uma grande variabilidade na distribuição dos dados, com média igual a 4 (n° de espécies) e um desvio padrão de 6, evidenciando a variabilidade nos dados do indicador riqueza de espécies (figura 05).

Na Figura 06 encontram-se os valores de cobertura de dossel por transecto. A variável de cobertura de dossel apresentou uma considerável variabilidade, uma vez que o CV encontrado foi de 65,10%. Os transectos 20, 21 e 22 representam a área referência.

Figura 06 – Cobertura de dossel (%) por transectos amostrados.



Média (linha vermelha) = 32,5%; Desvio padrão (barra tracejada) = 21,2%. Fonte: Do autor (2024)

A média da cobertura de dossel foi de 32,5%, com um desvio padrão de 21,2%, refletindo a amplitude das observações em torno da média (Figura 06).

3.2. Filtragem dos dados

A tabela 06 apresenta o número de variáveis preditoras removidas da base de dados de cada modelo através do processo de filtragem pela correlação entre as variáveis. Esse processo foi realizado antes de se iniciar as modelagens. Após a geração dos modelos preliminares, as variáveis preditoras que contribuíram menos para os modelos foram excluídas. Os resultados dessa etapa serão abordados no tópico “4.4. Variáveis selecionadas”.

Tabela 06 – Número de variáveis removidas e mantidas nos modelos a partir do processo de filtragem.

Modelos	n° de variáveis preditoras	
	Removidas	Mantidas
IA e IB	33	32
IIA e IIB	34	31

Modelo IA = riqueza de espécies estimada com todas as variáveis; Modelo IB = riqueza de espécies estimada apenas com variáveis de SR; Modelo IIA = cobertura de dossel estimada com todas as variáveis; Modelo IIB = cobertura de dossel estimada apenas com variáveis de SR. Fonte: Do autor (2024)

Foram removidas as 8 variáveis de observação de campo dos modelos IB e IIB, após a filtragem. Feito isso, os modelos preliminares IB e IIB foram gerados contendo, respectivamente, 24 e 23 variáveis preditoras. Os modelos preliminares IA e IIA foram gerados com a quantidade de variáveis apresentada na Tabela 06.

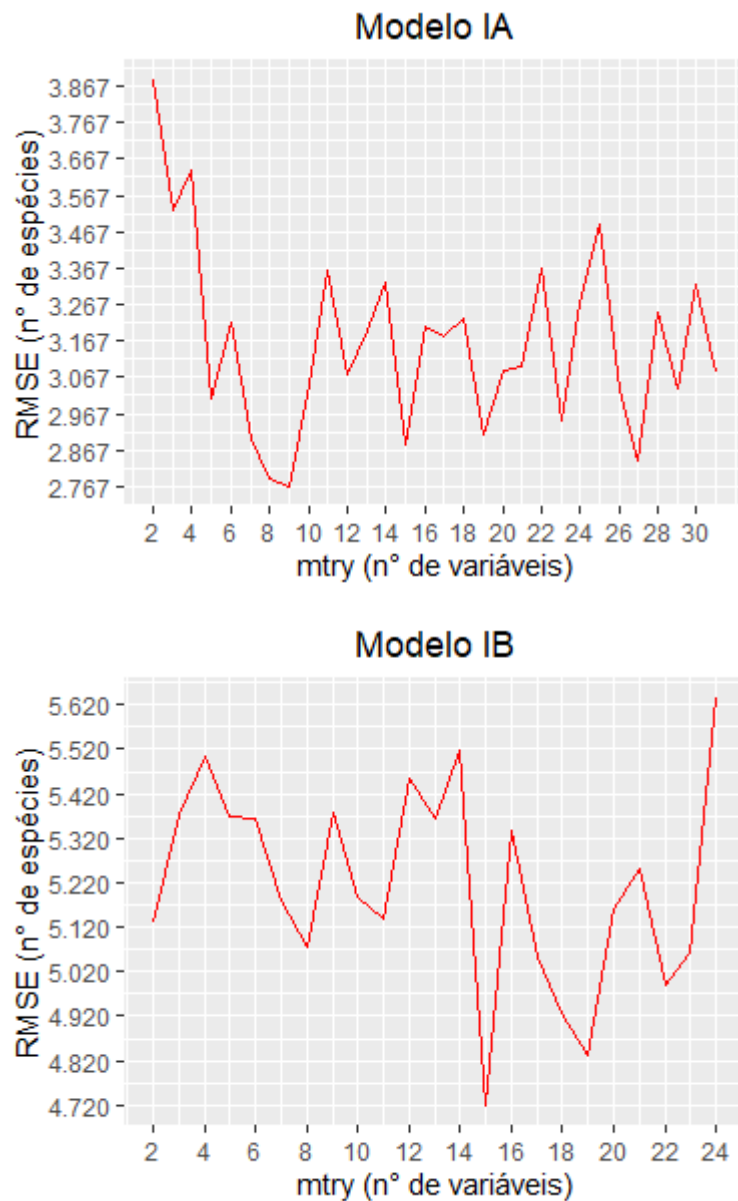
3.3. Modelos preliminares

3.3.1. Parametrização preliminar

Os modelos preliminares foram gerados utilizando um elevado número de variáveis preditoras, dessa forma, os modelos foram avaliados e otimizados para a geração dos modelos finais. Portanto, devido à alta correlação entre o número de variáveis preditoras e o parâmetro M_{try} , buscou-se determinar o valor ótimo de M_{try} para cada modelo, associando-o ao menor RMSE.

Na figura 07, observam-se os resultados da VCA na escolha do parâmetro M_{try} para os modelos IA e IB.

Figura 07 – Resultado da validação cruzada aninhada (VCA) para seleção do melhor valor de Mtry na construção dos modelos preliminares IA e IB.

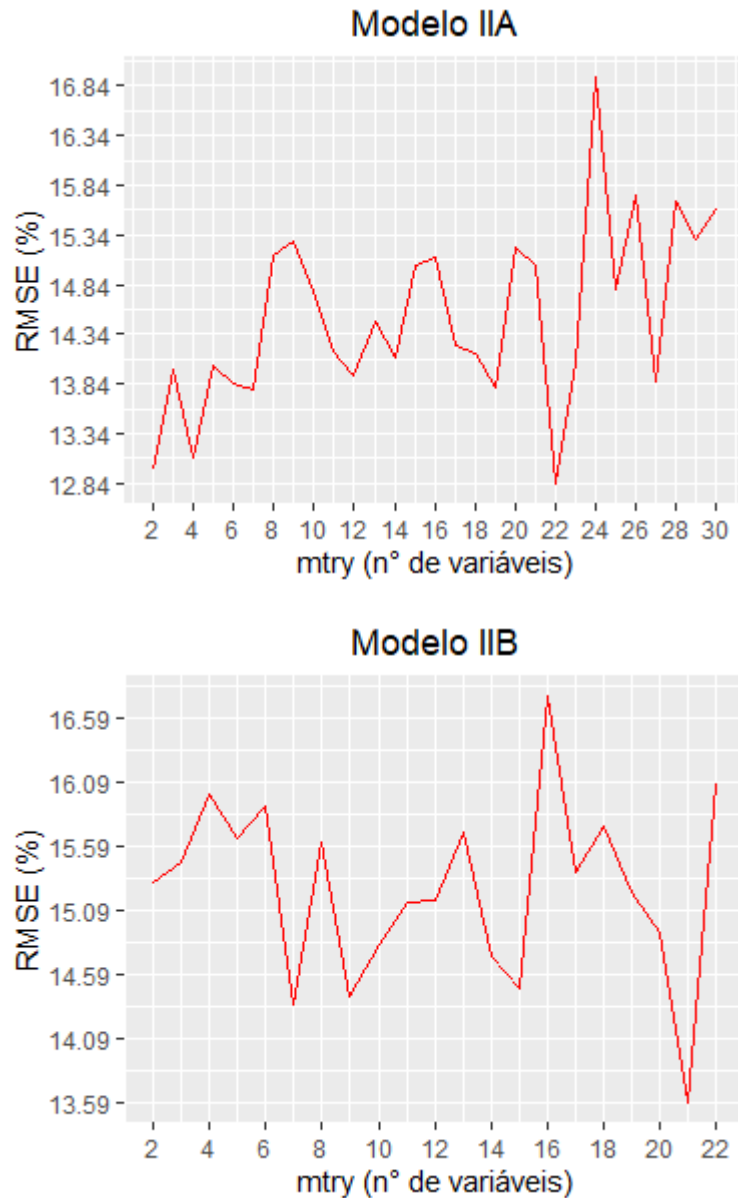


Fonte: Do autor (2024)

Ao avaliar o desempenho dos modelos IA e IB em relação ao parâmetro Mtry (Figura 07), observou-se que o menor valor de RMSE para o modelo IA foi de 2,767, obtido com Mtry igual a 9. Em contraste, o modelo IB apresentou um RMSE de 4,720 com Mtry igual a 15. Estes resultados indicam um ajuste superior do modelo IA (que inclui as variáveis de campo) em comparação com o Modelo IB (que exclui as variáveis de campo).

O mesmo foi feito para os modelos IIA e IIB (Figura 08), onde foi possível observar o melhor Mtry para cada modelo a partir do menor RMSE.

Figura 08 – Resultado da validação cruzada aninhada (VCA) para seleção do melhor valor de Mtry na construção dos modelos preliminares IIA e IIB.



Fonte: Do autor (2024)

Nos modelos IIA e IIB, os valores ótimos para o parâmetro Mtry foram 22 e 21, respectivamente. Embora haja uma pequena discrepância entre esses valores, é notável que o modelo IIA exibiu um RMSE inferior (12,84%), enquanto o modelo IIB demonstrou um RMSE ligeiramente mais elevado (13,59%)

3.3.2. Desempenho dos modelos preliminares

Buscando-se avaliar o impacto das variáveis de observação de campo, foram gerados os modelos preliminares utilizando os valores ótimos de Mtry. Na tabela 07, observam-se os coeficientes de determinação (R^2) e as métricas de erro geradas a partir de cada modelo preliminar. Vale ressaltar que valores menores para RMSE e MAE indicam um desempenho superior do modelo. No caso do coeficiente de determinação (R^2), um valor menor reflete um desempenho menos adequado do modelo.

Tabela 07 – Resultados do desempenho dos modelos preliminares através do coeficiente de determinação e das métricas de erro.

Modelos riqueza de espécies			
Modelos	R^2 (%)	RMSE (n° esp.)	MAE (n° esp.)
IA	87,53	2,69	2,00
IB	9,08	5,48	3,43
Modelos cobertura de dossel			
Modelos	R^2 (%)	RMSE (%)	MAE (%)
IIA	72,83	17,29	13,92
IIB	67,30	17,24	13,75

Fonte: Do autor (2024)

Nota-se que o modelo estimativo de riqueza de espécies IB apresenta desempenho inferior ao modelo IA, principalmente ao observar o coeficiente de determinação (R^2) que, após a exclusão das variáveis de campo, passa de 87,53% para 9,08%. O mesmo pode ser observado nas métricas RMSE e MAE, que apresentam valores maiores após a exclusão das variáveis de campo (Tabela 07).

Ao observar os modelos estimativos de cobertura de dossel, nota-se uma discreta diferença entre os coeficientes de determinação (R^2 , modelo IIA = 72,83%; modelo IIB = 67,30%). Entretanto, as métricas de erro, RMSE e MAE, não apresentaram alterações relevantes (Tabela 07).

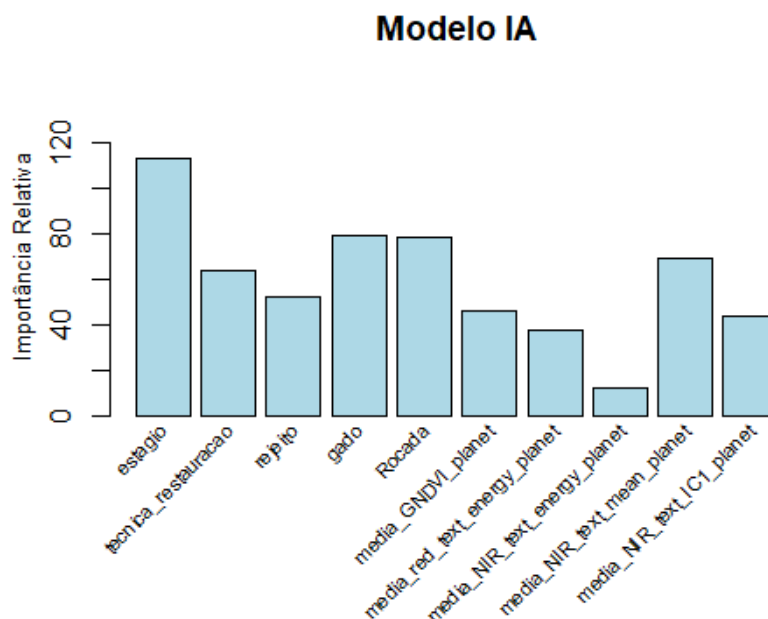
Os resultados obtidos sugerem que as variáveis de observação de campo exercem uma influência significativa na modelagem da riqueza de espécies. Em contrapartida, na modelagem da cobertura de dossel, essas variáveis não promovem alterações substanciais.

3.4. Variáveis selecionadas

Com os modelos preliminares gerados, além de identificar a importância dos dados de observações de campo para cada modelo, foi possível identificar quais variáveis preditoras contribuíram mais com os modelos preliminares. A fim de se otimizar os modelos preliminares e gerar os modelos finais, as variáveis preditoras de menor desempenho foram excluídas. Todas as variáveis removidas e mantidas para cada modelo final estão disponíveis para consulta no Apêndice A.

A exclusão das variáveis de menor importância nos modelos IA e IIA se deu a partir da permanência de 10 variáveis preditoras de maior contribuição (Apêndice A). Os valores de importância relativa associados às variáveis incluídas nos modelos têm uma tendência a se modificar após a remoção das variáveis preditoras de menor desempenho e a criação dos modelos finais. Essas alterações podem ser visualizadas nas Figuras 09, 10, 11 e 12 quando comparadas aos valores presentes no Apêndice A.

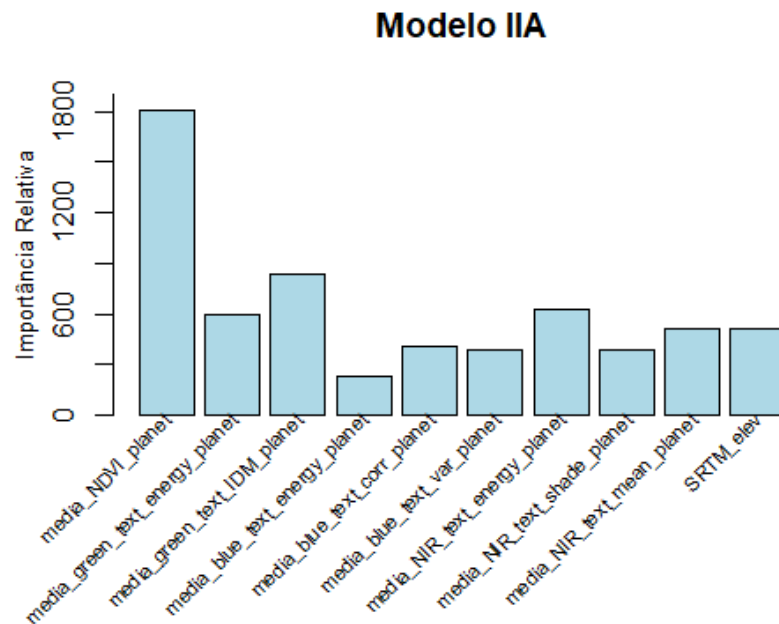
Figura 09 – Variáveis selecionadas para o modelo final de riqueza de espécies considerando dados de observação de campo (modelo IA).



Fonte: Do autor (2024)

No modelo final IA, as três variáveis que apresentaram maior importância foram as de observações de campo: *estagio*, *gado* e *roçada*, seguida das demais variáveis.

Figura 10 – Variáveis selecionadas para o modelo final de cobertura de dossel considerando dados de observação de campo (modelo IIA).



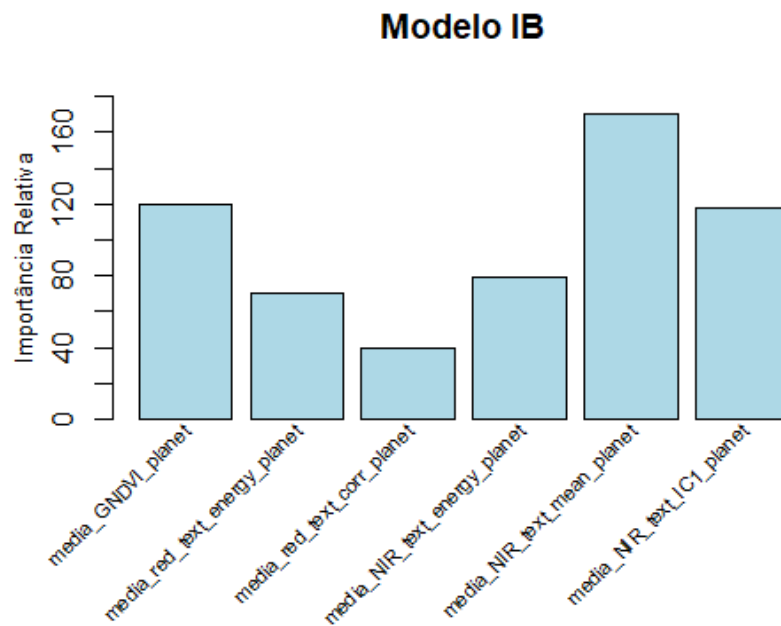
Fonte: Do autor (2024)

Nota-se que as variáveis de observação de campo tiveram uma contribuição limitada para o modelo final IIA, onde foram excluídas durante o processo de seleção das 10 variáveis de maior importância (Apêndice A), restando apenas variáveis de SR, conforme ilustrado na Figura 10. Dessa forma, a geração de um modelo IIA final não se justifica, uma vez que o objetivo do modelo IIA era a avaliação de um modelo contendo as variáveis observação de campo juntamente às variáveis de SR.

Quanto aos modelos IB e IIB, a intenção foi manter apenas 6 variáveis predictoras, resultando na exclusão de 18 e 15 variáveis, respectivamente, identificadas com menor importância relativa (Apêndice A). Observa-se nas Figuras 11 e 12, as 6 variáveis selecionadas e utilizadas para os modelos finais IB e IIB.

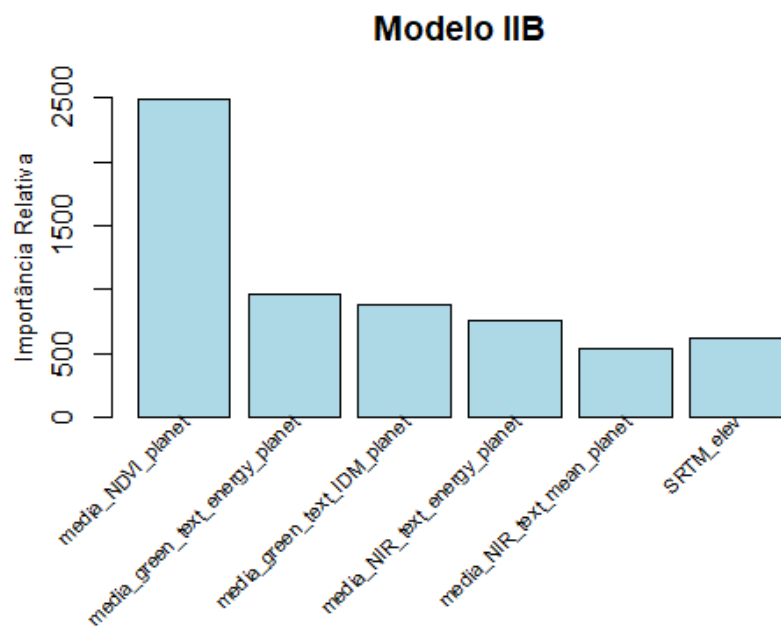
As variáveis mais importantes para o modelo final IB foram: a média textural do infravermelho próximo (*media_NIR_text_mean_planet*), o índice de vegetação GNDVI (*media_GNDVI_planet*) e o IC1 textural do infravermelho próximo (*media_NIR_text_IC1_planet*) (Figura 11). No modelo IIB, a variável de maior destaque foi o índice de vegetação NDVI (*media_NDVI_planet*), seguido pelas demais variáveis, que apresentaram importâncias relativas semelhantes (Figura 12). Nota-se que a variável de elevação (*SRTM_elev*) contribuiu apenas com o modelo IIB.

Figura 11 – Variáveis selecionadas para o modelo final de riqueza de espécies considerando apenas dados de sensoriamento remoto (modelo IB).



Fonte: Do autor (2024)

Figura 12 – Variáveis selecionadas para o modelo final de cobertura de dossel considerando apenas dados de sensoriamento remoto (modelo IIB).



Fonte: Do autor (2024).

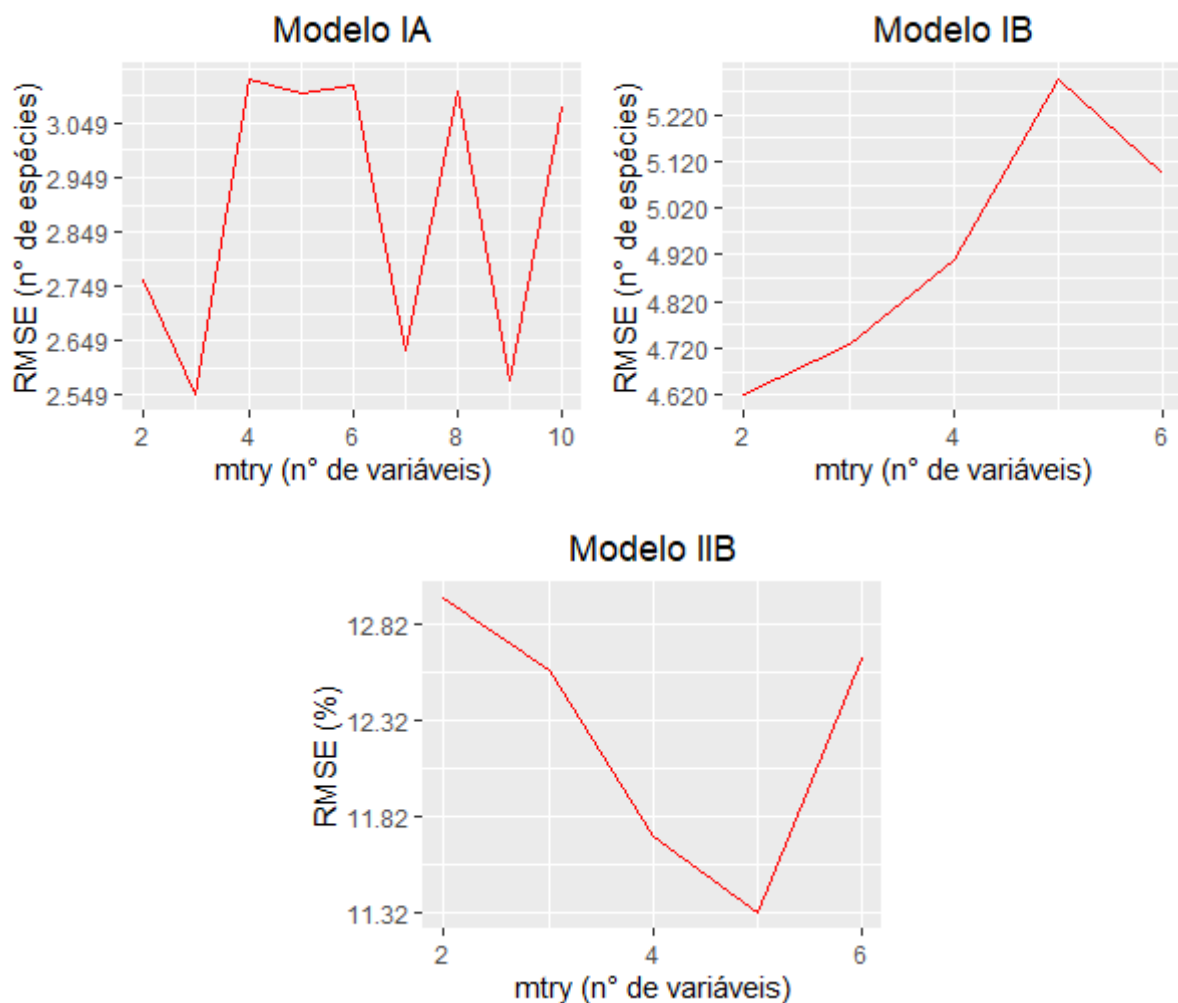
3.5. Modelos finais

O modelo final IA foi gerado e avaliado devido à alta influência das variáveis de campo no modelo estimativo para o indicador riqueza de espécies. Os modelos finais IB e IIB foram gerados e avaliados com o objetivo de se obter os mapas de estimativa dos indicadores ecológicos propostos neste trabalho.

3.5.1. Parametrização final

Abaixo é possível observar os melhores valores de Mtry para cada modelo final (Figura 13). Observa-se que os melhores valores de Mtry adotados para os Modelos IA, IB e IIB foram, respectivamente, 3, 2 e 5.

Figura 13 – Resultado da validação cruzada aninhada (VCA) para seleção do melhor valor de Mtry na construção dos modelos finais IA, IB e IIB.



Fonte: Do autor (2024)

3.5.2. Desempenho dos modelos finais

Com o objetivo de se avaliar a melhora nos três modelos após a otimização, foram comparados os resultados do coeficiente de determinação (R^2) e as medidas de erro (RMSE e MAE) entre os modelos preliminares (antes da otimização) e os modelos finais (após a otimização) (Tabela 08).

Tabela 08 – Resultados dos modelos prévios em comparação com os modelos finais.

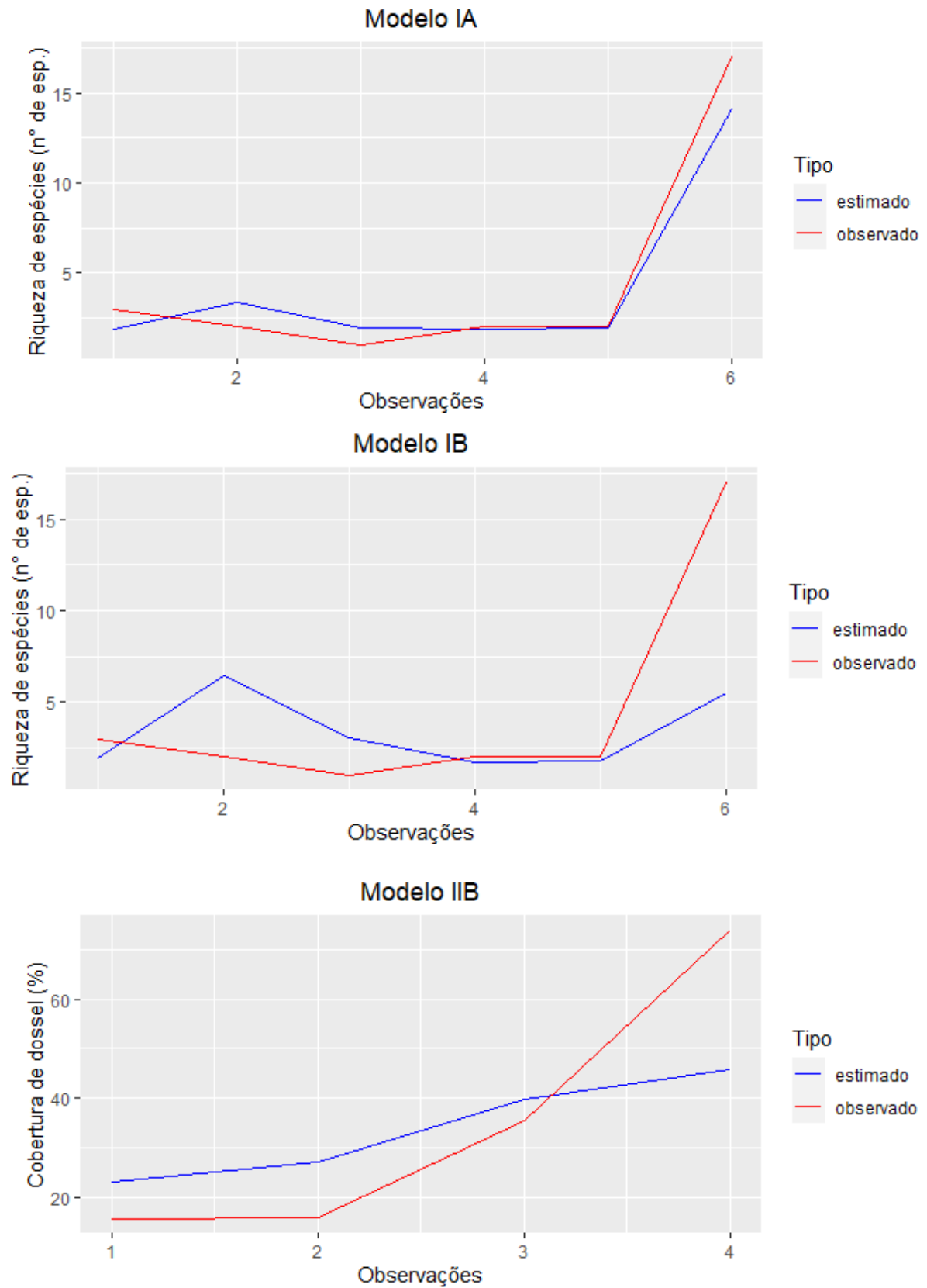
Modelos Preliminares	R^2 (%)	RMSE (u.v)	MAE (u.v)
IA	87,53	2,69	2,00
IB	9,08	5,48	3,43
IIB	67,30	17,24	13,75
Modelos Finais	R^2 (%)	RMSE (u.v)	MAE (u.v)
IA	97,56	1,42	1,07
IB	22,74	5,12	3,26
IIB	84,62	15,75	12,74

Em que: (u.v) = unidade da variável resposta, onde os modelos IA e IB são apresentados em nº de espécies e o modelo IIB é apresentado em %. Fonte: Do autor (2024)

Percebe-se que os três modelos demonstraram melhorias em suas capacidades estimativas. Notavelmente, o modelo IA registrou um aumento de 10,03% no coeficiente de determinação (R^2), acompanhado de reduções no RMSE e MAE, que diminuíram de 2,69 e 2,00 para 1,42 e 1,07, respectivamente. O modelo IB apesar de ter registrado um acréscimo elevado de 14,58% quando avaliado o R^2 , as medidas de erro RMSE e MAE quase não se alteraram, evidenciando pouca melhora após a otimização. Nota-se a melhora do modelo IIB, havendo um acréscimo de 17,32% entre o modelo preliminar e o modelo final, além das medidas RMSE e MAE demonstrarem uma queda de 1,52 e 1,01, respectivamente.

Buscando-se corroborar com os resultados apresentados na Tabela 08 e entender melhor, foram gerados gráficos de comparação entre os valores observados e estimados (Figura 14), onde o modelo de riqueza de espécies considerando as variáveis de observação de campo apresentou o melhor ajuste dos valores estimados em relação aos valores observados (modelo IA, $R^2= 97,56\%$), seguido do modelo de cobertura de dossel utilizando apenas as variáveis de SR (modelo IIB, $R= 84,62\%$). Já o modelo de riqueza de espécies considerando apenas as variáveis de SR, evidenciou desempenho inferior (modelo IB, $R^2= 22,74\%$).

Figura 14 – Gráficos de comparação entre os valores observados e os valores estimados de cada modelo final.

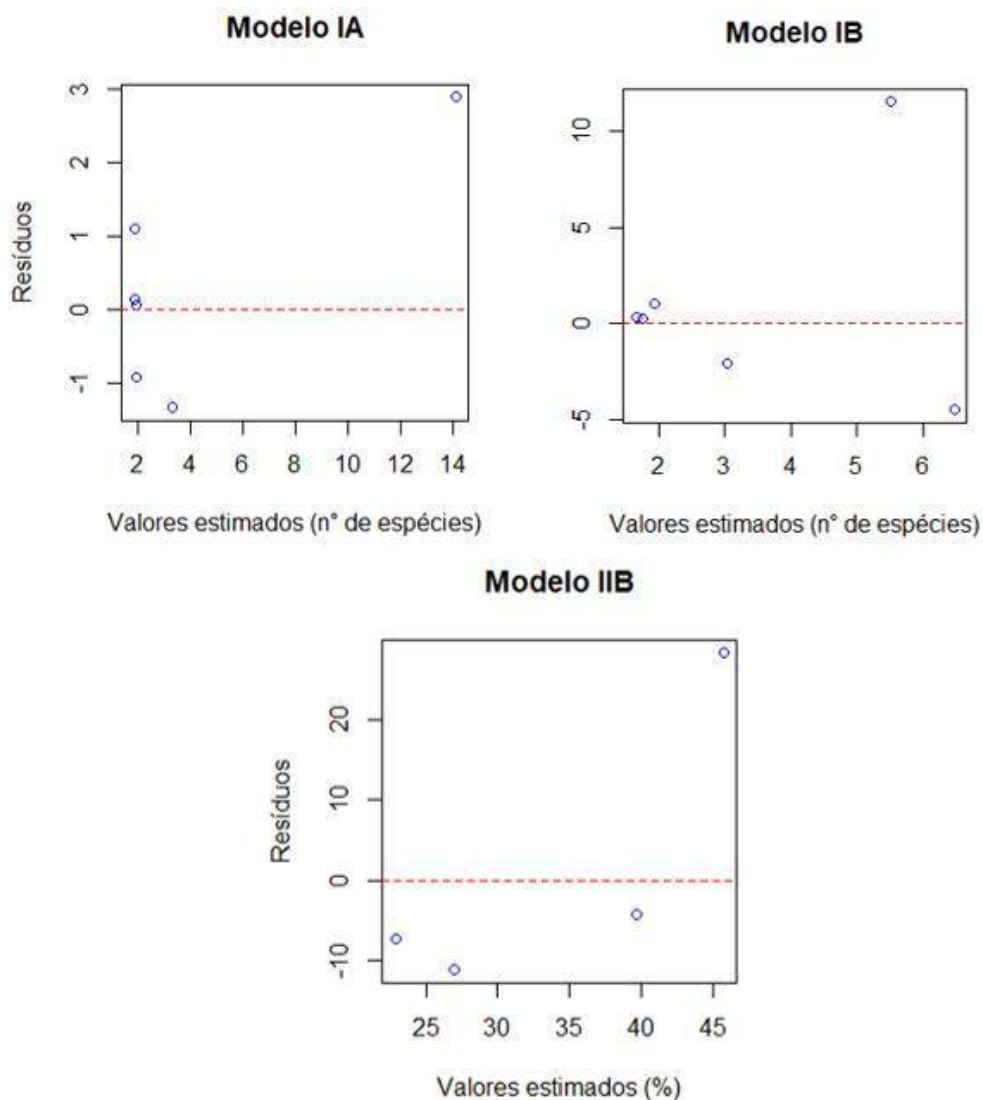


Fonte: Do autor (2024)

De maneira geral, os modelos gerados superestimam valores inferiores e subestimam valores elevados. Esse comportamento se torna mais evidente nos modelos que apresentaram o coeficiente de determinação (R^2) mais baixo e erros mais elevados, como pode ser observado no modelo IB, por exemplo (Figura 14).

Ao avaliar os gráficos de dispersão de resíduos de cada modelo, observa-se que quanto mais alto é o valor estimado, maior é o resíduo, ou seja, maior é a discrepância entre o valor observado e o estimado (Figura 15).

Figura 15 – Gráficos de dispersão residual de cada modelo final.



Fonte: Do autor (2024)

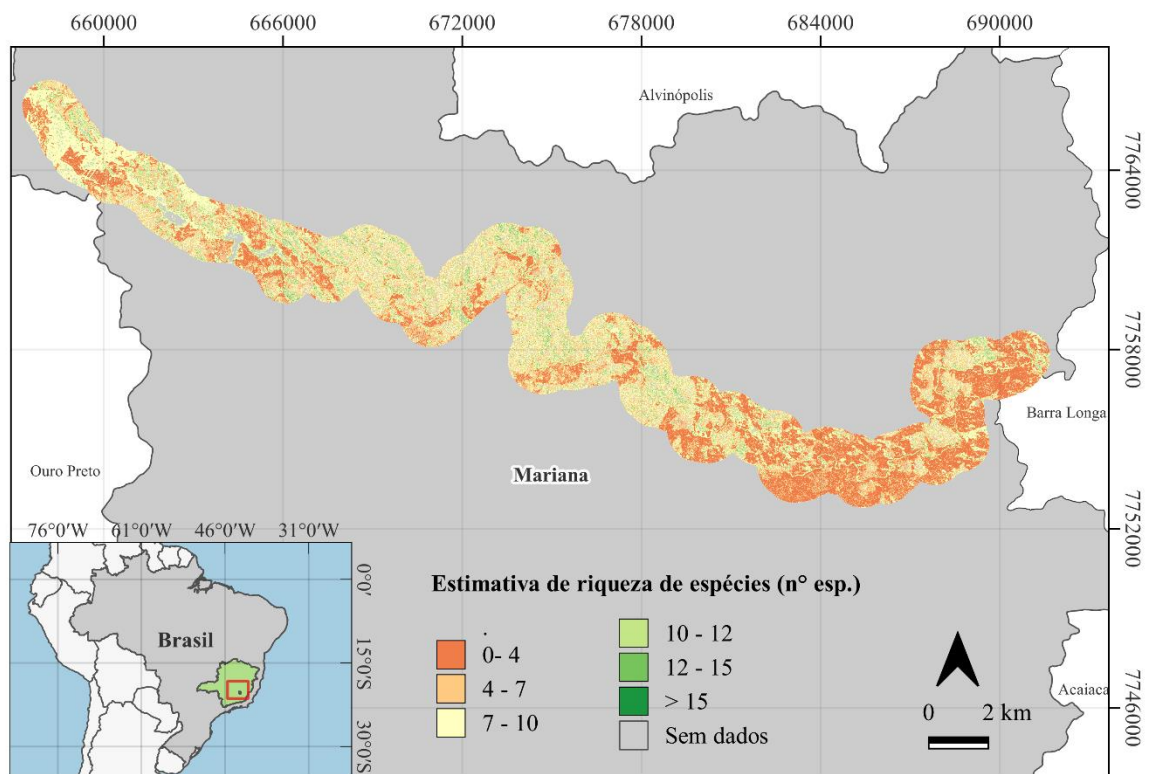
O modelo IA apresenta menor discrepância entre o valor observado e estimado, uma vez que o valor do resíduo máximo foi, aproximadamente, 3. Entretanto, o modelo IB apresenta um resíduo máximo próximo a 10, evidenciando o peso das variáveis de observação de campo para o modelo de riqueza de espécies. Já o comportamento residual do modelo IIB se mostrou

discrepante quando os valores estimados de cobertura de dossel ultrapassam 45%, onde o resíduo se aproxima ao valor de 30 (Figura 15).

3.6. Mapas estimativos de indicadores

Os mapas de estimativa de indicadores ecológicos foram gerados a partir dos modelos IB e IIB, os resultados se encontram nas Figura 16 e 17, respectivamente.

Figura 16 – Mapa estimativo de riqueza de espécies vegetais nas margens do rio Gualaxo do Norte.



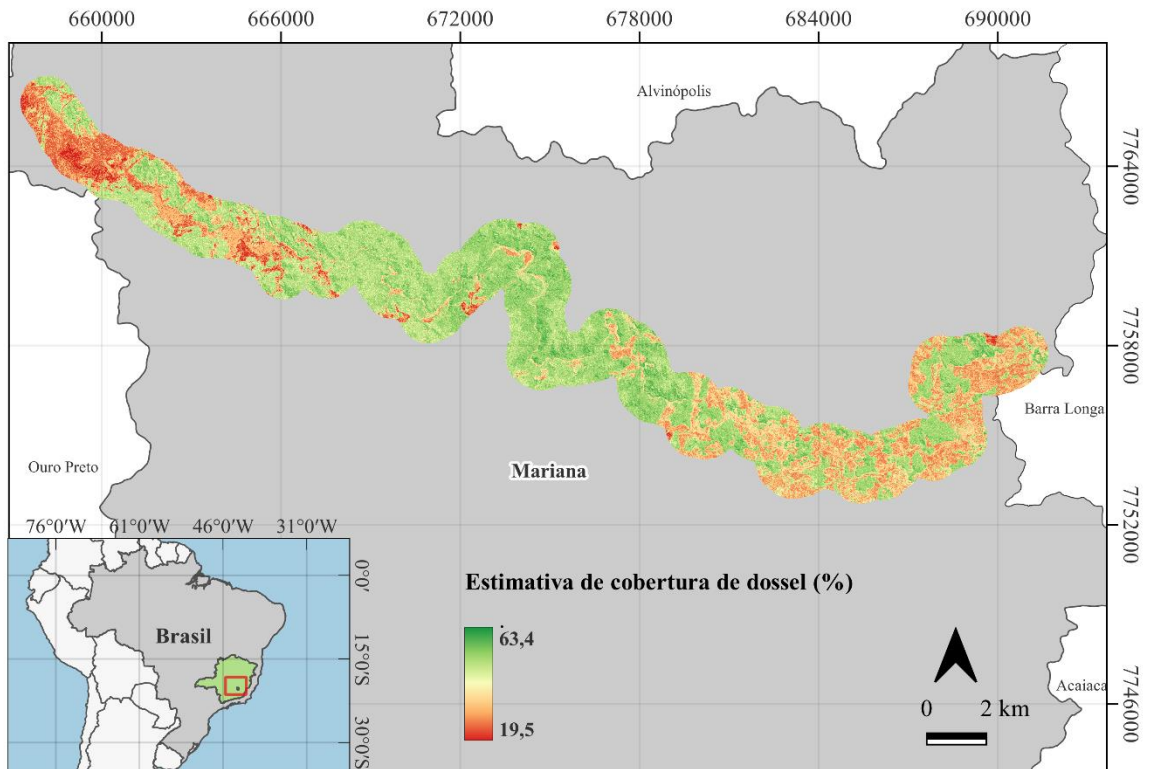
Fonte: Do autor (2024)

Conforme observado anteriormente, a estimativa espacial de riqueza de espécies (modelo IB, Figura 16) é menos precisa, portanto apresenta menor confiabilidade, uma vez que seu coeficiente de determinação (R^2) foi 22, 74%, juntamente à métrica de erro RMSE com valor de 5,12 espécies. Os valores estimados para riqueza de espécies variam de 4 a 15 espécies.

Quanto ao mapa estimativo de cobertura de dossel (modelos IIB, Figura 17), os valores observados no mapa apresentam alta confiabilidade, haja vista seu R^2 de 84,62%, lembrando

que seu erro RMSE de 15,75%, deve ser levado em consideração. Os valores estimados para cobertura de dossel variam de 19,5 a 63,4%.

Figura 17 – Mapa estimativo de cobertura de dossel nas margens do rio Gualaxo do Norte.



Fonte: Do autor (2024)

4. DISCUSSÃO

4.1. Influência dos dados de campo e de sensoriamento remoto nos modelos

Corroborando com o resultado encontrado no modelo final IA ($R^2 = 97,56\%$ e $RMSE = 1,42$ espécies), Guillera-Arroita et al. (2019) enfatizam a importância da seleção criteriosa de antecedentes ambientais na estimação da riqueza de espécies, indicando que dados categóricos de campo podem refinar o modelo. Apesar da hipótese da variação espectral, conforme proposta por Palmer et al. (2002), que sugere a mensuração remota e associação entre a variação nos padrões espectrais e a biodiversidade, o presente estudo evidencia que a utilização exclusiva de dados ópticos multiespectrais alcançou uma estimativa pouco precisa do indicador de riqueza de espécies (modelo final IB, $R^2 = 22,74\%$ e $RMSE = 5,14$ espécies). De acordo com Oldeland et al. (2010), a obtenção de alta resolução espectral, proporcionada por sensores hiperespectrais, pode desempenhar um papel crucial em estabelecer uma conexão bem-sucedida entre a diversidade ecológica e a variabilidade espectral. Em concordância, Rossi et al. (2021) ao abordarem a relação entre a diversidade espectral e a riqueza de espécies vegetais em pastagens alpinas, sugerem que futuras investigações devem se concentrar em compensar a resolução espacial por meio da resolução espectral ou mesmo temporal. Entretanto, os estudos evidenciam a importância do uso de diferentes sensores a fim de se estimar diversidade de espécies.

Os resultados obtidos sugerem que as variáveis de observação de campo exercem uma influência expressiva na modelagem da riqueza de espécies. Em contrapartida, essa expressiva diferença apresentada entre os modelos IA e IB, corroboram com a discussão de variáveis categóricas causarem distorções em modelos utilizando *Random Forest*. Variáveis categóricas podem impactar modelos gerados através de métodos baseados em árvores de decisão, introduzindo o problema dos "níveis ausentes", o qual surge quando uma divisão categórica no modelo encontra um nível de categoria durante a previsão que não estava presente durante a fase de treinamento (AU, 2017). Portanto, o problema dos "níveis ausentes" não deve ser ignorado, pois podem distorcer sistematicamente o modelo.

Diferente do modelo estimativo de riqueza de espécies, o modelo de cobertura de dossel utilizando apenas dados de SR (modelo IIB), não apresentou diferenças expressivas quando comparado com o modelo utilizando dados de observação de campo e dados de SR (modelo IIA). Entretanto, o modelo final IIB foi o único modelo gerado neste estudo em que os dados de elevação da missão SRTM contribuíram com a estimativa do indicador, apresentando um coeficiente de determinação (R^2) de 84,62% e um erro médio quadrático (RMSE) de 15,75%.

Esses resultados corroboram achados similares no estudo conduzido por Godinho et al. (2018), que empregou dados multiespectrais do Sentinel-2A e o algoritmo estocástico de aumento de gradiente para prever a cobertura de dossel em um sistema silvipastoril mediterrânico, alcançando um coeficiente de determinação de 82,80%. De maneira análoga, Narine et al. (2019) destacaram o potencial dos dados provenientes do ICESat-2 na estimativa da cobertura de dossel, atingindo valores de RMSE variando de 6,36% a 15,01%. Esses resultados demonstram a consistência e a eficácia de diferentes abordagens e fontes de dados de sensoriamento remoto na modelagem e estimativa da cobertura de dossel em diversos contextos ambientais.

Um melhor entendimento das combinações de variáveis categóricas de campo e dados de Sensoriamento Remoto (SR) pode aprimorar os modelos de estimativa de riqueza de espécies, aliado à capacidade dos dados de SR em produzir modelos confiáveis de cobertura de dossel, desempenha um papel crucial no processo decisório de técnicos, gestores e pesquisadores que atuam na área ambiental. Estudos como de Fitzgerald et al. (2006) e Lesch et al. (2005) revelaram a eficácia de modelos na previsão de atributos das culturas e do solo, demonstrando a capacidade de exigir menos pontos de dados e, conseqüentemente, reduzir custos. Resultados como estes demonstram que o conhecimento de valores estimados de indicadores ecológicos em áreas extensas, pode reduzir a frequência de coleta de dados em campo e até mesmo a intensidade amostral, podendo impactar em custo e operacionalidade no monitoramento de certos indicadores.

4.2. Variáveis mais importantes para a estimativa dos indicadores

O estudo ressalta importância do uso de diferentes fontes de dados a fim de se obter as melhores variáveis preditoras para modelagem de indicadores ecológicos. A estimativa de riqueza de espécies utilizando dados de observação de campo (modelo IA) tiveram as variáveis categóricas de campo como as variáveis mais importantes na modelagem, com destaque para a variável estágio. Esta variável distingue entre áreas não restauradas (NR) e áreas de referência (Ref.). Além da variável estágio, a variável gado (presença ou ausência) e a variável roçada (presença ou ausência) também foram destaque na contribuição do modelo. Este resultado reitera a conclusão de Guillera-Arroita et al. (2019), mencionada anteriormente, pois enfatiza a importância da seleção criteriosa de antecedentes ambientais. Portanto, o presente estudo sugere que os dados categóricos de campo refinam o modelo de riqueza de espécies.

As variáveis mais importantes utilizadas no modelo de riqueza de espécies sem os dados de campo (modelo IB) foram obtidas, exclusivamente, de sistemas sensores multiespectrais (dados Planet), demonstrando a utilidade desses sensores na avaliação de riqueza de espécies vegetais. A utilidade de sensores multiespectrais para avaliação desse indicador também foi identificada em outros trabalhos (ROCCHINI et al. 2007; ROCCHINI et al. 2011). Contudo, a variável que mais se destacou na estimativa espacial de riqueza de espécies vegetais foi o atributo textural *mean* no espectro do infravermelho próximo, contribuindo com Rocchini et al. (2007), que identificaram a banda do infravermelho próximo como um preditor chave da riqueza de espécies. Importante ressaltar que o presente estudo utilizou um limite de inclusão das espécies vegetais, avaliando apenas indivíduos arbustivos e arbóreos com diâmetro à altura do (DAP) igual ou superior a 5 cm.

O modelo estimativo de cobertura de dossel (modelo IIB) teve o índice de vegetação NDVI como a variável de maior contribuição para o modelo. Essa observação é respaldada pelos resultados de La et al. (2013), os quais destacaram uma correlação mais forte entre o NDVI e a cobertura de dossel quando comparado a outros índices de vegetação. Apesar da importância relativa da variável *elevação SRTM* não ter sido tão alta quando comparada ao NDVI, vale ressaltá-la devido a seleção desta variável para a geração do modelo, sendo o único modelo em que essa variável se mostrou importante. Diversos estudos apontam o uso de dados de elevação da missão SRTM como bom estimador de atributos estruturais de dossel (BROWN et al., 2010; GHOSH et al., 2020; KELLNDORFER et al., 2004). Entretanto, os estudos encontrados estimam altura de dossel utilizando dados de elevação da missão SRTM, enquanto o presente trabalho estima cobertura de dossel. Dessa forma, o estudo revela que a utilização de variáveis de sensores multiespectrais, com destaque ao NDVI, juntamente à variável de elevação SRTM, é capaz de estimar cobertura de dossel através do algoritmo Random Forest com boa precisão.

Os três modelos finais gerados tiveram atributos texturais como variáveis importantes em suas modelagens. Essa tendência é observada em diversos estudos que têm utilizado medidas de textura como variáveis preditoras em algoritmo Random Forest em aplicações diversas na área de conhecimento do sensoriamento remoto. Rodríguez-Galiano et al. (2011) evidenciaram que a inclusão de medidas de variabilidade espacial, como textura, em um classificador Random Forest resultou em uma melhoria na precisão da classificação da cobertura do solo. De maneira semelhante, Beguet et al. (2014) alcançaram uma maior precisão na previsão de variáveis de estrutura florestal ao empregar uma análise de textura multiescala. Outros estudos utilizaram com sucesso atributos texturais a partir de dados de SR para a

modelagem de Biomassa Acima do Solo (BAS) (BASTIN et al., 2014; CUTLER et al., 2012; DUBE; MUTANGA, 2015; WU et al., 2016). Apesar da importância atribuída às medidas de textura em diversos estudos, a literatura destaca que a escolha e seleção dessas medidas exigem atenção aos parâmetros do método GLCM, como o tamanho de janela mais adequado, direção e distância de deslocamento. No caso deste estudo, a configuração utilizada foi descrita na seção 2.4.3.

4.3. Problema de saturação e tendências nas estimativas espaciais

De maneira geral, os valores estimados mostraram certa limitação em todos os modelos gerados, onde os menores valores são superestimados e os maiores valores são subestimados. Conforme apontado por Rodríguez-Veiga et al. (2019), a explicação para esse padrão de comportamento pode estar relacionada à escassez de amostras de referência nos extremos, uma constatação que coincide com nosso estudo. Esse comportamento pode ser observado nas Figuras 05 e 06, principalmente onde os valores são mais altos, sendo representados por apenas três transectos.

A baixa intensidade amostral para a formação das amostras de referência (dados de inventário florestal e do densiômetro) para a geração dos modelos pode ter influenciado nos problemas de saturação e tendências na estimativa dos indicadores. Luan et al. (2020) notaram que o desempenho preditivo do modelo Random Forest foi substancialmente melhorado quando o tamanho da amostra aumentou ao avaliarem o desempenho de modelos Random Forest na distribuição de espécies de peixes marinhos.

Nos mapas gerados através dos modelos finais dos dois indicadores, foram observados valores estimados mais elevados em regiões de sombra nas imagens. Em respaldo ao que foi observado no presente estudo, Alavipanah et al. (2022) e Zhang et al. (2015) destacam o impacto significativo das sombras na estimativa de várias propriedades, incluindo vegetação, superfícies impermeáveis e água. Lopatin et al. (2019) enfatizam ainda o efeito negativo das sombras na precisão da classificação de espécies de plantas invasoras.

Em conjunto, essas constatações destacam a necessidade de considerar cuidadosamente as limitações associadas à amostragem e à presença de sombras para interpretações mais precisas e confiáveis dos resultados. Além do mais, outros estudos que empregaram o algoritmo Random Forest (RF) encontraram desafios semelhantes, atribuindo o problema de saturação a um efeito de agregação inerente aos algoritmos de aprendizado de máquina baseados em árvores de decisão, conhecido como *bagging* (BREIMAN, 2001).

5. CONCLUSÃO

Este estudo destaca os atributos texturais como variáveis preditoras importantes em estimar indicadores ecológicos de diversidade (riqueza de espécies vegetais) e estrutura (cobertura de dossel), com a finalidade de monitoramento de áreas em processo de restauração florestal impactadas por rejeito mineral. De maneira geral, todos os modelos superestimam valores inferiores e subestimam valores superiores, dessa forma os valores intermediários apresentam estimativas mais precisas.

O modelo final de riqueza de espécie vegetais apresentou desempenho limitado quando utilizado apenas os dados de sensoriamento remoto ($R^2 = 22,74\%$), porém, ao utilizar dados categóricos de campo juntamente aos dados de sensoriamento remoto na modelagem, apresenta bom desempenho ($R^2 = 97,56\%$), mitigando o efeito de saturação dos dados ao estimar riqueza de espécies vegetais. Portanto, a pesquisa propõe uma abordagem eficaz para mitigar o efeito de saturação de dados, especialmente em estimativas de valores mais elevados.

Já o modelo de cobertura de dossel apresentou bom ajuste utilizando exclusivamente dados de SR ($R^2 = 84,62\%$), não sofrendo influência significativa dos dados de campo. Os resultados reforçam o uso de medidas de textura para a geração dos modelos de estimativa espacial para os dois indicadores, além do uso de índices de vegetação, GNDVI para riqueza de espécies e, NDVI para cobertura de dossel. A variável elevação SRTM mostrou-se como variável preditora importante para a modelagem de cobertura de dossel. Dessa forma, esta pesquisa sugere uma abordagem eficiente e menos dependente de dados de campo para monitoramento da cobertura de dossel.

Portanto, essa pesquisa não apenas avança nosso entendimento sobre a aplicabilidade de modelos estimativos em ambientes impactados, mas também fornece ferramentas práticas para aprimorar o monitoramento e a gestão de áreas em processo de restauração florestal após a exposição a rejeitos minerais. Estudos futuros devem avaliar a utilização de outras fontes de dados, como sensores hiperespectrais e sensores ativos, bem como aumentar a quantidade de amostras referências e avaliar diferentes configurações do método GLCM, a fim de melhorar a capacidade estimativa dos indicadores ecológicos estudados.

REFERÊNCIAS

- AASEN, Helge et al. Generating 3D hyperspectral information with lightweight UAV snapshot cameras for vegetation monitoring: From camera calibration to quality assurance. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 108, p. 245-259, 2015.
- ADAM, Elhadi et al. Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers. **International Journal of Remote Sensing**, v. 35, n. 10, p. 3440-3458, 2014.
- AHMED, Oumer S. et al. Characterizing stand-level forest canopy cover and height using Landsat time series, samples of airborne LiDAR, and the Random Forest algorithm. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 101, p. 89-101, 2015.
- ALAVIPANAH, Seyed Kazem et al. The Shadow Effect on Surface Biophysical Variables Derived from Remote Sensing: A Review. **Land**, v. 11, n. 11, p. 2025, 2022.
- ASHTON, Peter S. Species richness in plant communities. In: **Conservation Biology: The Theory and Practice of Nature Conservation Preservation and Management**. Boston, MA: Springer US, 1992. p. 3-22.
- AU, Timothy C. Random forests, decision trees, and categorical predictors: the "absent levels" problem. **Journal of Machine Learning Research**, v. 19, n. 45, p. 1-30, 2018.
- BASTIN, Jean-François et al. Aboveground biomass mapping of African forest mosaics using canopy texture analysis: toward a regional approach. **Ecological Applications**, v. 24, n. 8, p. 1984-2001, 2014.
- BEGUET, Benoit et al. Automated retrieval of forest structure variables based on multi-scale texture analysis of VHR satellite imagery. **ISPRS journal of photogrammetry and remote sensing**, v. 96, p. 164-178, 2014.
- BREIMAN, Leo. Random forests. **Machine learning**, v. 45, p. 5-32, 2001.
- BROWN, Charles G.; SARABANDI, Kamal; PIERCE, Leland E. Model-based estimation of forest canopy height in red and Austrian pine stands using shuttle radar topography mission and ancillary data: A proof-of-concept study. **IEEE transactions on geoscience and remote sensing**, v. 48, n. 3, p. 1105-1118, 2009.
- CAMARRETTA, Nicolò et al. Monitoring forest structure to guide adaptive management of forest restoration: a review of remote sensing approaches. **New Forests**, v. 51, n. 4, p. 573-596, 2020.
- DO CARMO, Flávio Fonseca et al. Fundão tailings dam failures: the environment tragedy of the largest technological disaster of Brazilian mining in global context. **Perspectives in ecology and conservation**, v. 15, n. 3, p. 145-151, 2017.
- CBH-DOCE_COMITÊ DA BACIA HIDROGRÁFICA DO RIO DOCE. Disponível em: <<http://www.cbhdoce.org.br/institucional/abacia>>. Acesso em: 03 jun 2023

CUTLER, M. E. J. et al. Estimating tropical forest biomass with a combination of SAR image texture and Landsat TM data: An assessment of predictions between regions. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 70, p. 66-77, 2012.

DE ALMEIDA, Danilo Roberti Alves et al. Monitoring restored tropical forest diversity and structure through UAV-borne hyperspectral and lidar fusion. **Remote Sensing of Environment**, v. 264, p. 112582, 2021.

DOREN, Robert F. et al. Ecological indicators for system-wide assessment of the greater everglades ecosystem restoration program. **Ecological indicators**, v. 9, n. 6, p. S2-S16, 2009.

DUBE, Timothy; MUTANGA, Onesimo. Investigating the robustness of the new Landsat-8 Operational Land Imager derived texture metrics in estimating plantation forest aboveground biomass in resource constrained areas. **ISPRS Journal of Photogrammetry and Remote sensing**, v. 108, p. 12-32, 2015.

EDRISI, Sheikh Adil; ABHILASH, Purushothaman Chirakkuzhyil. Need of transdisciplinary research for accelerating land restoration during the UN Decade on Ecosystem Restoration. **Restoration Ecology**, v. 29, n. 8, p. e13531, 2021.

FARR, Tom G. et al. The shuttle radar topography mission. **Reviews of geophysics**, v. 45, n. 2, 2007.

FISCHER, Joern et al. Making the UN decade on ecosystem restoration a social-ecological endeavour. **Trends in ecology & evolution**, v. 36, n. 1, p. 20-28, 2021.

FITZGERALD, Glenn J. et al. Directed sampling using remote sensing with a response surface sampling design for site-specific agriculture. **Computers and electronics in agriculture**, v. 53, n. 2, p. 98-112, 2006.

GALLAUN, Heinz et al. EU-wide maps of growing stock and above-ground biomass in forests based on remote sensing and field measurements. **Forest Ecology and Management**, v. 260, n. 3, p. 252-261, 2010.

GHOSH, Sujit Madhab; BEHERA, Mukunda Dev; PARAMANIK, Somnath. Canopy height estimation using sentinel series images through machine learning models in a mangrove forest. **Remote Sensing**, v. 12, n. 9, p. 1519, 2020.

GITELSON, Anatoly A.; KAUFMAN, Yoram J.; MERZLYAK, Mark N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. **Remote sensing of Environment**, v. 58, n. 3, p. 289-298, 1996.

GITELSON, Anatoly A. et al. Novel algorithms for remote estimation of vegetation fraction. **Remote sensing of Environment**, v. 80, n. 1, p. 76-87, 2002.

GODINHO, Sérgio; GUIOMAR, Nuno; GIL, Artur. Estimating tree canopy cover percentage in a mediterranean silvopastoral systems using Sentinel-2A imagery and the stochastic gradient boosting algorithm. **International Journal of Remote Sensing**, v. 39, n. 14, p. 4640-4662, 2018.

GUILLERA-ARROITA, Gurutzeta; KÉRY, Marc; LAHOZ-MONFORT, José J. Inferring species richness using multispecies occupancy modeling: Estimation performance and interpretation. **Ecology and Evolution**, v. 9, n. 2, p. 780-792, 2019.

HARALICK, Robert M.; SHANMUGAM, Karthikeyan; DINSTEN, Its' Hak. Textural features for image classification. **IEEE Transactions on systems, man, and cybernetics**, n. 6, p. 610-621, 1973.

HERRICK, Jeffrey E.; SCHUMAN, Gerald E.; RANGO, Albert. Monitoring ecological processes for restoration projects. **Journal for Nature Conservation**, v. 14, n. 3-4, p. 161-171, 2006.

HUETE, Alfredo et al. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. **Remote sensing of environment**, v. 83, n. 1-2, p. 195-213, 2002.

HUETE, Alfredo R. A soil-adjusted vegetation index (SAVI). **Remote sensing of environment**, v. 25, n. 3, p. 295-309, 1988.

KELLNDORFER, Josef et al. Vegetation height estimation from shuttle radar topography mission and national elevation datasets. **Remote sensing of Environment**, v. 93, n. 3, p. 339-358, 2004.

KUHN, Max. Building predictive models in R using the caret package. **Journal of statistical software**, v. 28, p. 1-26, 2008.

LA, Hien Phu et al. Analysis of correlation between canopy cover and vegetation indices. **International Journal of Digital Content Technology and Its Applications**, v. 7, n. 11, p. 10, 2013.

LESCH, S. M. Sensor-directed response surface sampling designs for characterizing spatial variation in soil properties. **Computers and Electronics in Agriculture**, v. 46, n. 1-3, p. 153-179, 2005.

LIAW, Andy et al. Classification and regression by randomForest. **R news**, v. 2, n. 3, p. 18-22, 2002.

LOPATIN, Javier et al. How canopy shadow affects invasive plant species classification in high spatial resolution remote sensing. **Remote Sensing in Ecology and Conservation**, v. 5, n. 4, p. 302-317, 2019.

LOUHAICHI, Mounir; BORMAN, Michael M.; JOHNSON, Douglas E. Spatially located platform and aerial photography for documentation of grazing impacts on wheat. **Geocarto International**, v. 16, n. 1, p. 65-70, 2001.

LUAN, Jing et al. The predictive performances of random forest models with limited sample size and different species traits. **Fisheries Research**, v. 227, p. 105534, 2020.

OLDELAND, Jens et al. Does using species abundance data improve estimates of species diversity from remotely sensed spectral heterogeneity?. **Ecological Indicators**, v. 10, n. 2, p. 390-396, 2010

PALMER, Michael W. et al. Quantitative tools for perfecting species lists. *Environmetrics: The official journal of the International Environmetrics Society*, v. 13, n. 2, p. 121-137, 2002.

PLANET TEAM. **Planet Application Program Interface: In Space for Life on Earth**. San Francisco, CA. 2018

R Core Team, 2023. R: A Language and Environment for Statistical Computing. **R Foundation for Statistical Computing website**.

RAPPORT, David J.; HILDÉN, Mikael. An evolving role for ecological indicators: from documenting ecological conditions to monitoring drivers and policy responses. **Ecological indicators**, v. 28, p. 10-15, 2013.

REIF, Molly K.; THEEL, Heather J. Remote sensing for restoration ecology: Application for restoring degraded, damaged, transformed, or destroyed ecosystems. **Integrated environmental assessment and management**, v. 13, n. 4, p. 614-630, 2017.

ROCCHINI, Duccio et al. Landscape complexity and spatial scale influence the relationship between remotely sensed spectral diversity and survey-based plant species richness. **Journal of Vegetation Science**, v. 22, n. 4, p. 688-698, 2011.

ROCCHINI, Duccio et al. Landscape complexity and spatial scale influence the relationship between remotely sensed spectral diversity and survey-based plant species richness. **Journal of Vegetation Science**, v. 22, n. 4, p. 688-698, 2011.

ROCCHINI, Duccio; RICOTTA, C.; CHIARUCCI, Alessandro. Using satellite imagery to assess plant species richness: The role of multispectral systems. **Applied Vegetation Science**, v. 10, n. 3, p. 325-331, 2007.

RODRÍGUEZ-GALIANO, Víctor Francisco et al. Incorporating spatial variability measures in land-cover classification using Random Forest. **Procedia Environmental Sciences**, v. 3, p. 44-49, 2011.

RODRÍGUEZ-VEIGA, Pedro et al. Forest biomass retrieval approaches from earth observation in different biomes. **International Journal of Applied Earth Observation and Geoinformation**, v. 77, p. 53-68, 2019.

ROSSI, Christian et al. Spatial resolution, spectral metrics and biomass are key aspects in estimating plant species richness from spectral diversity in species-rich grasslands. **Remote Sensing in Ecology and Conservation**, v. 8, n. 3, p. 297-314, 2022.

ROUSE, John Wilson et al. Monitoring vegetation systems in the Great Plains with ERTS. **NASA Spec. Publ**, v. 351, n. 1, p. 309, 1974.

SAMARCO. **Relatório Bienal 2015-2016**. Belo Horizonte: Samarco, 2017. 99p

SILVEIRA, Eduarda MO et al. Pre-stratified modelling plus residuals kriging reduces the uncertainty of aboveground biomass estimation and spatial distribution in heterogeneous savannas and forest environments. **Forest ecology and management**, v. 445, p. 96-109, 2019.

SPANHOVE, Toon et al. Can remote sensing estimate fine-scale quality indicators of natural habitats?. **Ecological Indicators**, v. 18, p. 403-412, 2012.

TABACHNICK, Barbara G.; FIDELL, Linda S.; ULLMAN, Jodie B. **Using multivariate statistics**. Boston, MA: pearson, 2013.

TAROLLI, Paolo. High-resolution topography for understanding Earth surface processes: Opportunities and challenges. **Geomorphology**, v. 216, p. 295-312, 2014.

VIANI, Ricardo AG et al. Monitoring young tropical forest restoration sites: how much to measure?. **Tropical Conservation Science**, v. 11, p. 1940082918780916, 2018.

WITTEN, Ian H. et al. Practical machine learning tools and techniques. In: **Data mining**. Amsterdam, The Netherlands: Elsevier, 2005. p. 403-413.

WU, Chaofan et al. Comparison of machine-learning methods for above-ground biomass estimation based on Landsat imagery. **Journal of Applied Remote Sensing**, v. 10, n. 3, p. 035010-035010, 2016.

WULDER, Michael A. et al. The role of LiDAR in sustainable forest management. **The forestry chronicle**, v. 84, n. 6, p. 807-826, 2008.

XIONG, Zheng et al. Evaluating explorative prediction power of machine learning algorithms for materials discovery using k-fold forward cross-validation. **Computational Materials Science**, v. 171, p. 109203, 2020.

ZAHAWI, Rakan A.; REID, John L.; HOLL, Karen D. Hidden costs of passive restoration. **Restoration Ecology**, v. 22, n. 3, p. 284-287, 2014.

ZAHAWI, Rakan A. et al. Using lightweight unmanned aerial vehicles to monitor tropical forest recovery. **Biological Conservation**, v. 186, p. 287-295, 2015.

ZAMANI JOHARESTANI, Mehdi et al. PM2.5 prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data. **Atmosphere**, v. 10, n. 7, p. 373, 2019.

ZHANG, Lifu et al. An analysis of shadow effects on spectral vegetation indexes using a ground-based imaging spectrometer. **IEEE Geoscience and Remote Sensing Letters**, v. 12, n. 11, p. 2188-2192, 2015.

ZHAO, Qingxia et al. Estimating forest canopy cover in black locust (*Robinia pseudoacacia* L.) plantations on the Loess Plateau using random forest. **Forests**, v. 9, n. 10, p. 623, 2018.

APÊNDICE A – Variáveis preditoras selecionadas para cada modelo preliminar utilizando a importância relativa (IR). Em que: IR = Importância relativa; SR = Sensoriamento remoto. Nomenclatura das variáveis de índices de vegetação (IV) e de atributos texturais (AT) na base de dados: media_IV_planet e media_*_text_AT_planet, onde média = valor médio da variável extraída de cada transecto; IV = índice de vegetação avaliado (NDVI, SAVI, etc); * = bandas da imagem; text = Medida de textura; AT = atributos texturais avaliados (Energy, mean, etc); planet = fonte dos dados.

Descrição do modelo	Modelo	Seleção	Variáveis	IR
Modelo estimativo de riqueza de espécies utilizando as variáveis de observação de campo + variáveis de SR	IA	excluídas	uso_anterior	0.00540
			matriz	0.03281
			media_blue_text_var_planet	0.35583
			media_NIR_text_DiffVar_planet	0.56107
			media_blue_text_IC1_planet	0.70548
			SRTM_elev	1.03707
			media_NIR_text_IDM_planet	1.68468
			media_red_text_CProm_planet	2.30138
			media_NIR_text_DiffEnt_planet	2.89101
			media_green_text_IDM_planet	2.91782
			media_green_text_shade_planet	3.04491
			media_green_text_IC1_planet	3.14757
			media_NIR_text_corr_planet	4.00263
			media_green_text_corr_planet	5.15253
			media_NIR_text_CProm_planet	5.59213
			media_red_text_shade_planet	5.75373
			media_blue_text_energy_planet	6.78732
			media_NIR_text_shade_planet	7.16075
	media_green_text_energy_planet	8.53355		
	media_blue_text_corr_planet	10.11553		
	media_red_text_corr_planet	13.86669		
	IA	Mantidas	media_NIR_text_IC1_planet	28.04358
			media_NIR_text_energy_planet	29.06976
			media_red_text_energy_planet	32.46055
			media_GNDVI_planet	37.29665
			media_NIR_text_mean_planet	47.58397
			gado	50.72745
			Rocada	52.57323
tecnica_restauracao			57.79396	
rejeito			68.94125	
estagio			82.48564	

Continuação.

Descrição do modelo	Modelo	Seleção	Variáveis	IR	
Modelo estimativo de riqueza de espécies utilizando apenas variáveis de SR	IB	excluídas	media_blue_text_var_planet	0.24757	
			media_blue_text_IC1_planet	0.40286	
			media_NIR_text_DiffVar_planet	0.45197	
			media_NIR_text_IDM_planet	0.93806	
			SRTM_elev	2.32099	
			media_red_text_CProm_planet	2.52314	
			media_NIR_text_DiffEnt_planet	2.63305	
			media_green_text_IDM_planet	3.34465	
			media_green_text_IC1_planet	3.59198	
			media_red_text_shade_planet	4.34388	
			media_NIR_text_corr_planet	4.57124	
			media_NIR_text_CProm_planet	5.73321	
			media_blue_text_energy_planet	7.49125	
			media_green_text_shade_planet	8.93803	
	media_green_text_corr_planet	9.95608			
	media_blue_text_corr_planet	10.11332			
	media_NIR_text_shade_planet	14.34211			
		IB	Mantidas	media_green_text_energy_planet	15.38451
				media_red_text_corr_planet	19.56980
				media_NIR_text_energy_planet	55.23269
	media_red_text_energy_planet			66.97953	
	media_NIR_text_IC1_planet			82.85486	
	media_GNDVI_planet			91.37736	
			media_NIR_text_mean_planet	186.53533	

Continuação.

Descrição do modelo	Modelo	Seleção	Variáveis	IR
Modelo estimativo de cobertura de dossel utilizando as variáveis de observação de campo + variáveis de SR	IIA	excluídas	matriz	1.27275
			uso_anterior	1.33844
			media_NIR_text_IDM_planet	11.62245
			tecnica_restauracao	13.09415
			rejeito	25.11732
			media_blue_text_IC1_planet	26.23944
			Rocada	31.28611
			media_NIR_text_DiffEnt_planet	38.73073
			gado	46.34526
			estagio	52.21757
			media_blue_text_IDM_planet	54.84504
			media_NIR_text_corr_planet	80.34224
			media_blue_text_shade_planet	100.64493
			media_green_text_IC1_planet	105.20415
		media_green_text_corr_planet	108.98441	
		media_NIR_text_CProm_planet	121.05478	
		media_NIR_text_IC1_planet	152.83853	
		idade_plantio	154.81914	
		media_red_text_corr_planet	166.38268	
		media_red_text_IC1_planet	206.71440	
		media_blue_text_corr_planet	225.44552	
		media_NIR_text_shade_planet	270.17872	
		media_blue_text_energy_planet	284.55526	
		SRTM_elev	295.61799	
		media_NIR_text_mean_planet	312.24892	
		media_NIR_text_energy_planet	331.84501	
		media_blue_text_var_planet	332.43698	
		media_green_text_energy_planet	397.65544	
media_green_text_IDM_planet	810.07557			
media_NDVI_planet	1596.10673			
IIA	Mantidas			

Continuação.

Descrição do modelo	Modelo	Seleção	Variáveis	IR	
Modelo estimativo de cobertura de dossel utilizando apenas variáveis de SR	IIB	Excluídas	media_blue_text_IC1_planet	34.91512	
			media_blue_text_shade_planet	35.97740	
			media_NIR_text_IDM_planet	36.46621	
			media_NIR_text_DiffEnt_planet	52.51161	
			media_green_text_IC1_planet	63.82446	
			media_NIR_text_corr_planet	70.81998	
			media_red_text_corr_planet	80.62250	
			media_blue_text_IDM_planet	86.60318	
			media_green_text_corr_planet	88.16710	
			media_NIR_text_CProm_planet	102.15486	
	media_NIR_text_IC1_planet	114.99102			
	media_blue_text_energy_planet	163.30486			
	media_red_text_IC1_planet	238.97274			
	media_blue_text_corr_planet	322.82893			
	media_NIR_text_shade_planet	334.29506			
	media_blue_text_var_planet	343.48439			
				SRTM_elev	346.25829
				media_green_text_energy_planet	369.58516
				media_NIR_text_mean_planet	406.02642
		IIB	Mantidas	media_NIR_text_energy_planet	455.63996
			media_green_text_IDM_planet	830.86049	
			media_NDVI_planet	1868.19800	