



IAGO MENDES DE OLIVEIRA

**MAPEAMENTO DA BIOMASSA ACIMA DO SOLO NA
BACIA DO RIO GRANDE UTILIZANDO DADOS DE
SENSORIAMENTO REMOTO E APRENDIZADO DE
MÁQUINA**

**LAVRAS-MG
2022**

IAGO MENDES DE OLIVEIRA

**MAPEAMENTO DA BIOMASSA ACIMA DO SOLO NA BACIA DO RIO GRANDE
UTILIZANDO DADOS DE SENSORIAMENTO REMOTO E APRENDIZADO DE
MÁQUINA**

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

Prof. Dr. Fausto Weimar Acerbi Júnior
Orientador

**LAVRAS-MG
2022**

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

Oliveira, Iago Mendes de.

Mapeamento da biomassa acima do solo na bacia do Rio Grande utilizando dados de sensoriamento remoto e aprendizado de máquina / Iago Mendes de Oliveira. - 2022.

57 p.

Orientador(a): Fausto Weimar Acerbi Júnior.

Dissertação (mestrado acadêmico) - Universidade Federal de Lavras, 2022.

Bibliografia.

1. Biomassa acima do solo. 2. Rio Grande. 3. Sensoriamento remoto. I. Acerbi Júnior, Fausto Weimar. II. Título.

IAGO MENDES DE OLIVEIRA

**MAPEAMENTO DA BIOMASSA ACIMA DO SOLO NA BACIA DO RIO GRANDE
UTILIZANDO DADOS DE SENSORIAMENTO REMOTO E APRENDIZADO DE
MÁQUINA**

**MAPPING ABOVEGROUND BIOMASS IN THE RIO GRANDE BASIN USING
REMOTE SENSING DATA AND MACHINE LEARNING**

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

APROVADA em 29 de junho de 2022.

Dr. Fausto Weimar Acerbi Júnior

Dra. Marcela de Castro Nunes Santos Terra

Dr. Inacio Thomaz Bueno

UFLA

UFLA

UNICAMP

Prof. Dr. Fausto Weimar Acerbi Júnior
Orientador

**LAVRAS-MG
2022**

*À minha esposa, Laura.
À minha querida família.
Aos meus amigos.
Dedico*

AGRADECIMENTOS

Escrever esta seção não é uma tarefa muito simples, principalmente quando se trata de um trabalho da importância deste.

Esta dissertação é fruto do empenho de muitas pessoas. Assim, deixo o meu sincero agradecimento:

À Deus, o criador, pela vida.

Ao M. Gabriel e à UDV, por me mostrar o caminho a ser seguido.

Aos meus pais, José Célio e Luciléa, e à minha irmã, Lailah, que não mediram esforços para que eu pudesse chegar até aqui.

À minha querida esposa, Laura, pelo amor e suporte diário. Sem você, este trabalho não seria possível.

Aos meus amigos, pela convivência sadia e alegre, pelos conselhos e por me fazerem uma pessoa melhor.

Ao Prof. Fausto, por ter aceitado ser meu orientador, pela paciência e principalmente por confiar em mim e em meu trabalho.

À Universidade Federal de Lavras, ao Programa de Pós-Graduação em Engenharia Florestal e por terem me proporcionado tantos aprendizados.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro.

RESUMO

A biomassa acima do solo (BAS) é um dos reservatórios de carbono mais dinâmicos do ecossistema terrestre, mas pouco se sabe sobre sua distribuição espacial. Atualmente, os métodos mais precisos para a quantificação da BAS são caros, demorados e limitados à pequenas áreas. Neste cenário, o sensoriamento remoto (SR) aparece como uma ferramenta promissora, fornecendo cobertura espacial e temporal suficiente para o monitoramento contínuo e de baixo custo da BAS em grandes áreas. O objetivo deste trabalho foi realizar o mapeamento da BAS na bacia do Rio Grande, Minas Gerais, a partir de dados multissensores de SR, como informações das bandas B2-5 (reflectância de superfície) e B10 (temperatura de superfície) dos sensores OLI e TIRS do satélite Landsat 8 (L8), dados de radar de abertura sintética (SAR) do Sentinel-1A (S1) e dados do modelo digital de elevação SRTM. Nós também utilizamos dados derivados destes, como cinco índices de vegetação (NDVI, SAVI, DVI, ARVI e EVI) e 18 medidas de textura derivadas da matriz de coocorrência de níveis de cinza (GLCM), usando uma janela de 3 x 3 pixels. Para realizar a modelagem e o mapeamento da BAS, foi utilizado o algoritmo de aprendizado de máquina *Random Forest*. Utilizou-se o método de validação cruzada aninhada e a métrica *Root Mean Square Error* (RMSE) para medir o desempenho do modelo, que apresentou o valor de 48,79 Mg.ha⁻¹ (44,22%). A otimização dos hiperparâmetros m_{try} (22), *sample fraction* (0,22) e *min.node.size* (10) foi realizada utilizando o método da busca aleatória. Os resultados encontrados reforçaram a importância do uso de dados de diferentes sensores e domínios (espectral e espacial) na modelagem da BAS. O índice de vegetação ARVI foi a variável mais importante para o modelo. No entanto, a variável não conseguiu captar a alteração da BAS em áreas de vegetação mais densa, com valores de BAS próximos a 114 Mg.ha⁻¹. A elevação e as medidas de textura *dissimilarity* e *sum entropy*, associadas à polarização VH (S1) e ao índice de vegetação DVI (L8), foram essenciais para mitigar o efeito da saturação dos dados.

Palavras-chave: Biomassa acima do solo, Rio Grande, Sensoriamento remoto.

ABSTRACT

Aboveground biomass (AGB) is one of the most dynamic carbon reservoirs in the terrestrial ecosystem, but little is known about its spatial distribution. Currently, the most accurate methods for quantifying AGB are expensive, time-consuming, and limited to small areas. In this scenario, remote sensing (RS) appears as a promising tool, providing sufficient spatial and temporal coverage for continuous and low-cost monitoring of AGB in large areas. The objective of this work was to map the AGB in the Rio Grande basin, Minas Gerais, from multi-sensor RS data, such as information from the B2-5 (surface reflectance) and B10 (surface temperature) bands of the sensors OLI and TIRS from the Landsat 8 (L8) satellite, Synthetic Aperture Radar (SAR) data from Sentinel-1A (S1), and data from the SRTM digital elevation model. We also used data derived from these, such as five vegetation indices (NDVI, SAVI, DVI, ARVI, and EVI) and 18 texture measures derived from Gray-level Co-occurrence Matrix (GLCM), using a 3 x 3 window size. To perform the modeling and mapping of the AGB, we used the Random Forest machine learning algorithm. We used the nested cross-validation method and the Root Mean Square Error (RMSE) metric to measure the performance of the model, which presented a value of 48.79 Mg.ha⁻¹ (44.22%). We optimized the *m_{try}* (22), *sample.fraction* (0.22) and *min.node.size* (10) hyperparameters using the random search method. The results reinforced the importance of using data from different sensors and domains (spectral and spatial) in AGB modeling. The ARVI vegetation index was the most important variable for the model. However, the variable could not capture the change in AGB in areas of denser vegetation, with AGB values close to 114 Mg.ha⁻¹. The variables *elevation* and the texture metrics *dissimilarity* and *sum entropy* associated with VH polarization (S1) and DVI vegetation index (L8) were essential to mitigate the effect of data saturation.

Keywords: Aboveground biomass, Rio Grande, Remote sensing.

LISTA DE FIGURAS

Figura 1 – Processo de aquisição de dados através do sensoriamento remoto.....	17
Figura 2 – Espectro eletromagnético.....	18
Figura 3 – Método das matrizes de coocorrência de níveis de cinza.	22
Figura 4 – Localização da área de estudo.....	24
Figura 5 – Obtenção das variáveis preditoras.....	27
Figura 6 – Fluxo de trabalho.....	29
Figura 7 – Validação cruzada aninhada.....	31
Figura 8 – Partições espacialmente disjuntas geradas pela validação cruzada espacial.....	32
Figura 9 – Processo de otimização dos hiperparâmetros do modelo.....	33
Figura 10 – Biomassa acima do solo estimada nos fragmentos florestais amostrados.	35
Figura 11 – Variabilidade da biomassa acima do solo nos fragmentos amostrados.	36
Figura 12 – Distribuição dos dados antes e após a etapa de pré-processamento.....	37
Figura 13 – Resultado do ciclo externo da validação cruzada aninhada.	39
Figura 14 – Predições do modelo <i>Random Forest</i> e distribuição dos resíduos.....	40
Figura 15 – Distribuição dos resíduos por classes de densidade de biomassa acima do solo.	40
Figura 16 – Variáveis mais importantes para o modelo.	41
Figura 17 - Dependência parcial das variáveis mais importantes para o modelo.....	42
Figura 18 - Resultado da validação cruzada espacial utilizada na otimização dos hiperparâmetros do modelo <i>Random Forest</i>	43
Figura 19 – Biomassa acima do solo na bacia do Rio Grande.	44

LISTA DE TABELAS

Tabela 1 – Índices de vegetação utilizados na modelagem da biomassa acima do solo.	20
Tabela 2 – Espaço de busca dos hiperparâmetros.	32
Tabela 3 – Resumo estatístico da variável BAS antes e após a remoção dos <i>outliers</i>	37
Tabela 4 – Variáveis preditoras utilizadas na modelagem.	38

SUMÁRIO

1	INTRODUÇÃO	13
2	REFERENCIAL TEÓRICO	15
2.1	Métodos de quantificação da biomassa acima do solo	15
2.1.1	Métodos diretos	15
2.1.2	Métodos indiretos	16
2.2	Uso do sensoriamento remoto na modelagem da biomassa aérea.....	16
2.2.1	Tipos de sensores de sensoriamento remoto.....	17
2.2.2	Variáveis utilizadas na modelagem da biomassa acima do solo.....	19
3	MATERIAIS	24
3.1	Área de estudo.....	24
3.2	Bases de dados.....	24
3.2.1	Inventário florestal	24
3.2.2	Landsat 8 OLI/TIRS	25
3.2.3	Sentinel-1A	25
3.2.4	Análise de textura	25
3.2.5	SRTM.....	26
3.3	Integração dos dados	26
3.3.1	Obtenção das variáveis preditoras.....	26
3.3.2	Extração dos valores.....	26
4	MÉTODOS	28
4.1	Modelo Random Forest.....	28
4.2	Otimização e avaliação do desempenho do modelo	29
4.2.1	Seleção de variáveis	29
4.2.2	Validação cruzada k-fold	30
4.2.3	Validação cruzada aninhada	30
4.2.4	Estimativa da capacidade preditiva do modelo	31
4.2.5	Otimização dos hiperparâmetros do modelo	33
4.2.6	Determinação das variáveis mais importantes para o modelo	34
5	RESULTADOS	35
5.1	Biomassa nos fragmentos inventariados.....	35
5.2	Remoção de <i>outliers</i>	36
5.3	Variáveis selecionadas	37
5.4	Desempenho do modelo.....	38
5.5	Variáveis preditoras mais importantes.....	40
5.6	Otimização dos hiperparâmetros	42

5.7	Mapeamento da biomassa acima do solo.....	43
6	DISCUSSÃO	45
6.1	Estimativa da capacidade de generalização do modelo <i>Random Forest</i>	45
6.2	Problema da saturação e tendências nas estimativas	45
6.3	Variáveis mais importantes para o modelo	46
7	CONCLUSÃO	48
	REFERÊNCIAS.....	49

1 INTRODUÇÃO

As florestas cobrem aproximadamente 30% da superfície terrestre. Elas fornecem serviços ecológicos, econômicos e sociais aos sistemas naturais e à humanidade, incluindo refúgios para biodiversidade, provisão de alimentos, regulação dos ciclos hidrológicos e proteção dos solos. Além disso, influenciam no clima através da troca de energia, água, dióxido de carbono (CO₂) e outros elementos químicos com a atmosfera (BONAN, 2008). Cerca de 45% de todo carbono (C) terrestre estão armazenados nas florestas, que contribuem com aproximadamente 50% da produção líquida terrestre e podem sequestrar grandes quantidades de C anualmente (BONAN, 2008; MA et al., 2017; MCGARVEY, 2015). A biomassa, como peso seco, é 50% C. Nas florestas, cerca de 70-90% da biomassa está acima do solo, a maior parte em árvores (HOUGHTON; HALL; GOETZ, 2009). A biomassa acima do solo (BAS) é um dos reservatórios de C mais dinâmicos do ecossistema terrestre, uma vez que está em constante alteração devido à ocorrência de incêndios, desmatamentos e mudanças no uso do solo (KUMAR; MUTAMBA, 2017). Por esse motivo, seu monitoramento deve ser contínuo.

Quantificar a BAS é importante por diversas razões (HOUGHTON; HALL; GOETZ, 2009), contudo, o conhecimento da sua distribuição ainda é limitado (HEROLD et al., 2019). A BAS pode ser medida ou estimada por métodos diretos e indiretos. Os métodos diretos, embora mais precisos, são restritos à pequenas áreas, pois são métodos de natureza destrutiva, demorados e de custo elevado (LU et al., 2016). Por outro lado, os métodos indiretos buscam estimar a BAS através de equações alométricas (ROITMAN et al., 2018; ZIMBRES et al., 2021) ou por meio de técnicas de sensoriamento remoto (SR) (LU et al., 2016).

Estudos anteriores mostraram que o SR pode contribuir efetivamente com a quantificação e monitoramento da BAS em grandes áreas, utilizando tanto dados de sensores ativos quanto passivos (LI et al., 2020). Os dados de sensores ativos, como os dados de *radar* de abertura sintética (SAR, do inglês *Synthetic Aperture Radar*), têm a vantagem de serem relativamente imunes às condições atmosféricas e poderem penetrar na copa das árvores e interagir com os principais componentes da BAS, como troncos, galhos e folhas. Por outro lado, dados de sensores passivos (e.g., Landsat, Sentinel etc.), bem como suas derivações (e.g., índices espectrais, imagens de textura etc.), possuem forte relação com a BAS (LI et al., 2020). No entanto, a literatura mais recente tem mostrado que o uso integrado desses dois tipos de sensores resulta em estimativas mais acuradas de BAS quando comparadas com estimativas feitas por modelos calibrados com dados de um único tipo de sensor (BISPO et al., 2020; FENG

et al., 2017; GHOSH; BEHERA, 2018; HUANG et al., 2019; LI et al., 2020; PHUA et al., 2017; THEOFANOUS et al., 2021, ZIMBRES et al., 2021).

Embora as pesquisas tenham explorado o uso de dados de SR para estimar a BAS, os métodos de modelagem, ou algoritmos, ainda são pouco compreendidos (LU et al., 2016). Os algoritmos frequentemente utilizados para estimar a BAS podem ser classificados em paramétricos e não-paramétricos. Os algoritmos paramétricos são baseados em um modelo matemático pré-definido, que define a relação entre as entradas e as saídas (CAMPOS et al., 2018; WALLNER et al., 2015). Contudo, como a relação entre a BAS e os dados de SR normalmente é muito complexa para ser capturada por algoritmos paramétricos (LU et al., 2016), o uso de métodos mais flexíveis muitas vezes se faz necessário.

Os algoritmos não-paramétricos, não são baseados em modelos matemáticos pré-definidos; em vez disso, sua estrutura é definida a partir dos próprios dados (e.g., *Random Forest* (RF) e *Support Vector Machine* (SVM)). Diversos trabalhos utilizaram algoritmos não-paramétricos para estimar a BAS (CAO; WANG; TANG, 2020), mas as incertezas quanto às estimativas ainda continuam altas (BISPO et al., 2020; LU, 2005). Portanto, compreender e identificar as fontes de incertezas é fundamental para melhorar as estimativas de BAS (LU, 2005).

Neste estudo, nosso objetivo principal foi mapear a BAS na bacia hidrográfica do Rio Grande (BRG) utilizando dados coletados em campo, via inventário florestal, e dados de SR. Nós utilizamos o algoritmo RF para realizar as previsões e para classificar as variáveis de acordo com seu grau de importância, a fim de identificar as variáveis mais relevantes para o modelo.

2 REFERENCIAL TEÓRICO

2.1 Métodos de quantificação da biomassa acima do solo

Biomassa significa a massa de matéria de origem biológica, viva ou morta, animal ou vegetal (SILVEIRA et al., 2008). Segundo Salati (1994), a quantificação da biomassa florestal divide-se em métodos diretos e indiretos. Os métodos diretos implicam determinações, pelo fato de as árvores serem cortadas e seus componentes separados e pesados. *Determinação* significa uma medição real feita diretamente na biomassa, por exemplo, a pesagem de um fuste inteiro por meio de um dinamômetro ou uma balança. Todas as árvores de uma determinada parcela são derrubadas e pesadas, sendo feita em seguida a extrapolação da avaliação amostrada para a área total de interesse (SILVEIRA et al., 2008). Nos métodos indiretos, são feitas estimativas baseadas principalmente em dados advindos dessas determinações, assim como de dados originados de inventários florestais. Neste método, também entram as estimativas por meio de dados de sensoriamento remoto (KUMAR; MUTANGA, 2017).

2.1.1 Métodos diretos

Os métodos diretos de amostragem de biomassa podem ser enquadrados em duas grandes categorias (SANQUETTA, 2002): método de árvore individual e método da parcela.

O método da árvore individual é feito mediante a seleção de uma árvore média, sendo que, para se conhecer essa árvore, é preciso realizar um inventário florestal piloto e calcular o diâmetro ou a área seccional ou transversal dessa árvore. Após a definição dessa árvore, procede-se a derrubada e a determinação da biomassa de um número de indivíduos, que se constituem na amostra.

O método da parcela é feito cortando-se e pesando-se toda a biomassa de uma área pré-definida, podendo ser implementado pelo procedimento denominado corte total (*harvest method*), que consiste em se determinar toda a biomassa da floresta pelo corte e pesagem de todas as frações (inclusive as raízes) passo a passo, ou por amostragem em múltiplos estágios (*multi-stage sampling*), no qual cortam-se todos os indivíduos contidos na parcela, pesando-se todos os caules. Para as demais porções, são feitas determinações apenas para alguns indivíduos, estabelecendo-se uma relação de quantidades entre os dois grupos.

2.1.2 Métodos indiretos

A quantificação da biomassa de forma indireta pode ser feita por dois métodos quando se trabalha em campo (SOMOGYI, 2006): (1) utilizando dados de volume de árvores ou talhões e multiplicando-os por um fator apropriado, denominado fator de biomassa, que convertem (expandem ou reduzem) as estimativas de volume para estimativas de biomassa; e (2) realizando o ajuste de equações pelo uso de técnicas de regressão. Esse é o procedimento mais comum, no qual algumas árvores são amostradas, o peso de cada componente é determinado e relacionado por meio de regressão com variáveis dendrométricas, sendo escrita da seguinte forma: $B = f(P1, P2, p1, p2, \dots)$, onde B representa a biomassa em kg ou Mg; P representa as variáveis dendrométricas disponíveis (e.g., diâmetro e altura); e p representa os parâmetros do modelo (SILVEIRA et al., 2008).

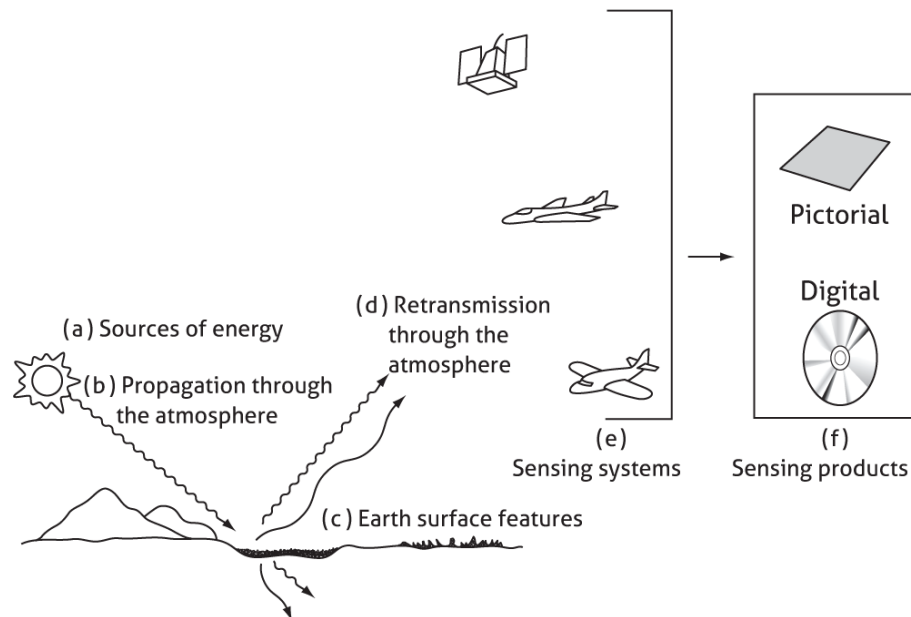
As técnicas de SR têm sido amplamente utilizadas em estudos na área florestal, uma vez que é possível estimar parâmetros biofísicos, como a biomassa, carbono e volume de madeira, pelas propriedades espectrais dos componentes da vegetação, como troncos, galhos, folhas etc.

2.2 Uso do sensoriamento remoto na modelagem da biomassa aérea

O SR, segundo Lillesand, Kiefer e Chipman (2015), é a ciência e a arte de obter informações sobre um objeto (alvo), área ou fenômeno através da análise de dados adquiridos por um dispositivo (sensor) que não está em contato com o objeto, área ou fenômeno sob investigação.

Os sensores aos quais nos referimos neste trabalho, são sensores operados por plataformas orbitais que captam a energia eletromagnética emitida ou refletida por um alvo na superfície terrestre. A Figura 1 ilustra o processo de aquisição de dados por meio do SR.

Figura 1 – Processo de aquisição de dados através do sensoriamento remoto.



Os elementos do processo de aquisição de dados são: as fontes de energia (a); propagação da energia pela atmosfera (b); interação da energia com os alvos na superfície (c); retransmissão da energia pela atmosfera (d); sensores aerotransportados e/ou orbitais (e); resultando na geração de dados dos sensores em forma pictórica e/ou digital (f). Fonte: Adaptado de Lillesand, Kiefer e Chipman (2015).

O SR é uma tecnologia capaz de superar as principais limitações dos métodos tradicionais de quantificação da biomassa, como a abrangência espacial, tempo, custo e mão de obra (KUMAR; MUTAMBA, 2017).

2.2.1 Tipos de sensores de sensoriamento remoto

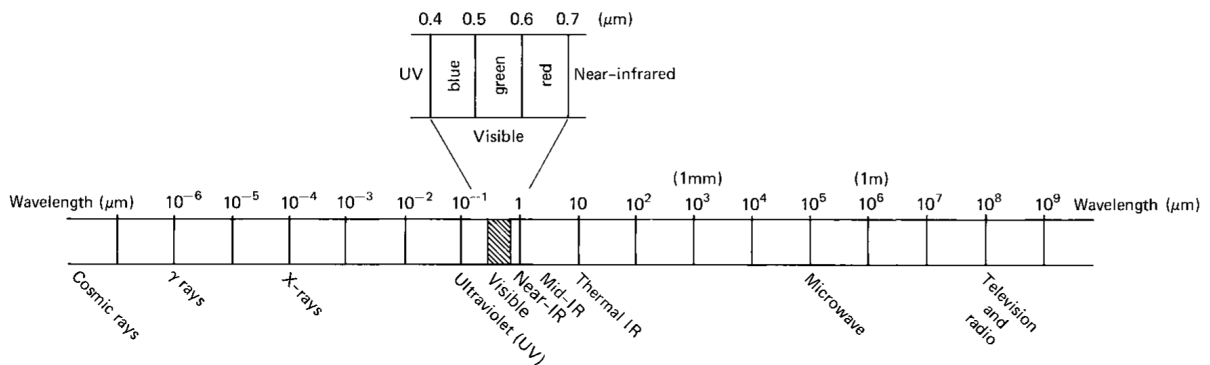
Atualmente, há uma grande variedade de dados de SR que podem ser utilizados para estimar a BAS, tais como dados de sensores ativos (*óptico*) e passivos (*radar*), disponibilizados em diferentes resoluções – espacial, espectral e temporal –, podendo ser comerciais ou não. A seleção dos dados a serem utilizados depende do escopo do projeto, das características da área de estudo e da vegetação, e do orçamento disponível.

Tanto o SR óptico quanto por *radar* seguem abordagens similares para análise, modelagem, mapeamento e monitoramento da BAS. As imagens de SR óptico fornecem informações espectrais sobre a estrutura horizontal da vegetação, enquanto as imagens de SR por *radar* fornecem informações sobre a estrutura vertical da vegetação, devido à capacidade de penetrar na copa das árvores até uma certa profundidade.

Por serem sensores passivos, os sensores ópticos necessitam de uma fonte externa de luz, como o sol. Nesse caso, a qualidade da informação adquirida pelo sensor é fortemente

influenciada pelas condições climáticas. Esses sensores captam frequências de radiação do espectro eletromagnético (FIGURA 2), especialmente nos comprimentos de onda da região do visível (0,4 – 0,7 μm) e do infravermelho (0,7 μm – 1 mm). As faixas do visível e do infravermelho, em especial do infravermelho próximo (NIR, do inglês *near infrared*), são os comprimentos de onda mais sensíveis às características da vegetação.

Figura 2 – Espectro eletromagnético.



Fonte: Lillesand, Kiefer e Chipman (2015).

Os dados de SR óptico são amplamente utilizados para estimar a BAS de diferentes tipos de florestas (SORIANO-LUNA et al., 2018). Diferentes tipos de sensores, com resolução espacial variando entre 10 e 1000 m, têm sido usados na modelagem da BAS em escala global, regional e local, tais como MODIS (BACCINI et al., 2008), Landsat ETM+ (LU; WENG, 2004), SPOT (NICHOL; SARKER, 2011) e Sentinel-2 (CHEN et al., 2019). Apesar do custo, sensores com resolução espacial menor que 1 m, como IKONOS, QuickBird-2 e WorldView-2 têm vantagem sobre os sensores de baixa e média (10 – 1000 m) resolução espacial (GÓMEZ et al., 2012; MARSHALL; THENKABAIL, 2015).

O monitoramento e avaliação das estimativas de BAS ao longo do tempo podem ser feitos a partir de dados de SR óptico devido ao grande acervo de dados existentes e à ampla cobertura espacial. No entanto, apesar do uso consolidado de dados de SR óptico na modelagem da BAS, estes estão limitados à sua fraca capacidade de penetração nas copas das árvores e nas nuvens, além de apresentarem problemas de saturação em altas densidades de BAS (AVITABILE et al., 2012).

Por outro lado, a interpretação dos dados de *radar*, especificamente dados SAR e *lidar* (*light detection and ranging*), nem sempre é direta, pois o sinal é responsivo às características da superfície, como a estrutura e umidade. Por ser um sensor ativo, o *radar* tem a vantagem de

fornecer sua própria fonte de energia, o que possibilita sua operação a qualquer hora do dia, independentemente das condições climáticas. Os dados SAR têm sido utilizados como complemento das características de reflectância espectral da vegetação na estimativa da BAS e têm se mostrado muito úteis em áreas frequentemente coberta por nuvens (SINHA et al., 2015).

A regressão baseada nos coeficientes de retroespalhamento (σ^0), também chamado de *backscattering*, é um dos métodos mais utilizados para estimar a BAS a partir de dados SAR (SANDBERG et al., 2011). Os principais fatores a serem considerados ao utilizar o coeficiente de retroespalhamento são o comprimento de onda (e.g., K, X, C, L, P), polarização (e.g., HH, VV, HV, VH), ângulo de incidência, cobertura do solo, e propriedades do terreno, como rugosidade e constante dielétrica.

A polarização pode ser linear ou cruzada. A polarização linear é obtida pela transmissão e recepção linear do sinal, horizontal (H) e vertical (V), HH e VV, respectivamente (ZEBKER; VAN ZYL; HELD, 1987). Na polarização cruzada, a transmissão e a recepção do sinal são diferentes (e.g., HV e VH). O retroespalhamento de polarização cruzada VH tem sido considerado o mais correlacionado com a BAS (LE TOAN et al., 2011). Os sensores SAR têm sido utilizados em diversos estudos de modelagem da BAS (CHEN et al., 2019; ENGLHART et al., 2012; FORKUOR et al., 2020; GHASEMI; SAHEBI; MOHAMMADZADEH, 2013; HAYASHI; MOTOHKA; SAWADA, 2019; LIU et al., 2019; MORIN et al., 2019; PEREGON; YAMAGATA, 2013).

2.2.2 Variáveis utilizadas na modelagem da biomassa acima do solo

Em estudos de modelagem da BAS, é importante integrar diferentes tipos de variáveis preditoras derivadas de sensores passivos e ativos, como as bandas espectrais, índices de vegetação e medidas de textura, a fim de obter modelos cada vez mais robustos.

2.2.2.1 Bandas espectrais

As bandas espectrais refletem a estrutura, textura e sombra da vegetação, relacionadas com a estrutura celular da folha e com os pigmentos fotossintetizantes, os quais estão correlacionados com a BAS (BACCINI et al., 2008). Os comprimentos de onda da região do visível e do NIR, também chamados de VNIR (*visibe and near infrared*), são mais sensíveis aos pigmentos e à saúde da planta de maneira geral. A faixa do infravermelho de ondas curtas (SWIR, *short wave infrared*), por sua vez, está relacionada com processos bioquímicos, massa

foliar por área e conteúdo de umidade da vegetação e do solo (BACCINI et al., 2008). Na região do vermelho ocorre a absorção pela clorofila, enquanto na região do NIR as células mesofílicas proporcionam maior reflexão (CHAO et al., 2019). A região do verde está relacionada com o verdor da vegetação.

2.2.2.2 Índices de vegetação

Os índices de vegetação são transformações utilizadas para realçar as propriedades espectrais da vegetação em uma imagem, especialmente nas regiões do visível e do NIR (PONZONI; SHIMABUKURO; KUPLICH, 2012). Esses índices são relacionados a parâmetros biofísicos da cobertura vegetal, como biomassa e índice de área foliar, além de minimizarem os efeitos de iluminação da cena, declividade da superfície e geometria, que influenciam os valores de reflectância da vegetação. Diversos índices de vegetação são utilizados na modelagem da BAS (Tabela 1).

Tabela 1 – Índices de vegetação utilizados na modelagem da biomassa acima do solo.

Índice de vegetação		Fórmula	Referência
<i>Atmospherically Resistant Vegetation Index</i>	ARVI	$\frac{NIR - (RED - \gamma \times (RED - BLUE))}{NIR + (RED - \gamma \times (RED - BLUE))}$	Kaufman e Tanre (1992)
<i>Difference Vegetation Index</i>	DVI	$NIR - RED$	Roujean e Breon (1995)
<i>Enhanced Vegetation Index</i>	EVI	$G \times \frac{NIR - RED}{NIR - C1 \times RED - C2 \times BLUE + L}$	Huete (1997)
<i>Modified Simple Ratio</i>	MSR	$\frac{\frac{NIR}{RED - 1}}{\sqrt{\frac{NIR}{RED} + 1}}$	Chen (1996)
<i>Normalized Difference Vegetation Index</i>	NDVI	$\frac{NIR - RED}{NIR + RED}$	Rouse et al. (1974)
<i>Normalized Ratio Vegetation Index</i>	NRVI	$\frac{RVI - 1}{RVI + 1}$	Baret e Guyot (1991)
<i>Ratio Vegetation Index</i>	RVI	$\frac{RED}{NIR}$	Richardson e Wiegand (1977)
<i>Soil-Adjusted Vegetation Index</i>	SAVI	$(1 + L) \times \frac{(NIR - RED)}{NIR + RED}$	Huete (1988)

Fonte: Adaptado de Lourenço (2021).

Segundo Lourenço (2021), os dois IV mais utilizados na modelagem da biomassa, são o NDVI e o SR. O SR, ou índice de vegetação da razão simples, foi o primeiro IV a ser usado (JORDAN, 1969). O SR é obtido pela divisão de valores de reflectância referentes à região do NIR por valores de reflectância correspondentes à região do vermelho. Rousset et al. (1974) normalizaram a razão simples para o intervalo de -1 a +1, propondo o índice de vegetação da diferença normalizada (NDVI). Para alvos terrestres, o limite inferior é de aproximadamente zero (0), e o limite superior, de aproximadamente 0,80 (PONZONI; SHIMABUKURO; KUPLICH, 2012).

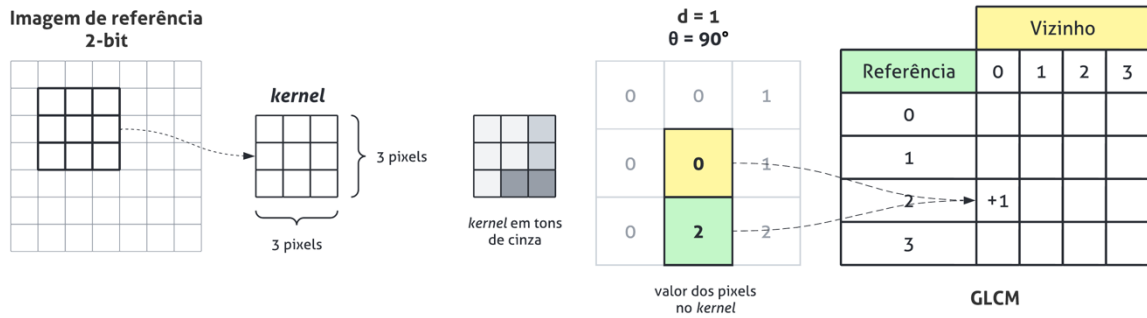
2.2.2.3 Medidas de textura

No âmbito do SR, *textura* refere-se à variação espacial dos níveis de cinza (valores) dos pixels de uma imagem. A caracterização da textura de uma imagem de SR pode ser feita através da análise de textura, que consiste na utilização de métodos estatísticos como meio de extrair informações sobre a distribuição espacial dos níveis de cinza em uma imagem.

Um dos métodos mais utilizados na análise de textura em trabalhos de modelagem da BAS é o método das *matrizes de coocorrência de níveis de cinza* ou GLCM (do inglês *gray level co-occurrence matrix*), proposto por Haralick, Shanmugam e Dinstein (1973).

A GLCM é uma matriz quadrada, simétrica e com o número de linhas e colunas igual ao nível de quantização (resolução radiométrica) da imagem. Segundo Hall-Beyer (2017), a GLCM é uma tabulação de quantas combinações diferentes de níveis de cinza ocorrem em uma imagem. O método é regido pelas seguintes propriedades: (a) tamanho da janela deslizante (usaremos o termo *kernel* daqui em diante), (b) direção θ (0° , 45° , 90° e 135°) e (c) distância de deslocamento d . A Figura 3 ilustra de maneira geral como a GLCM é criada.

Figura 3 – Método das matrizes de coocorrência de níveis de cinza.



Esquema representando a criação da matriz de coocorrência a partir de uma imagem de referência com resolução radiométrica de 2-bit, na qual cada pixel pode assumir 2^2 (4) valores (0, 1, 2 e 3). Neste exemplo, a direção θ considerada foi 90° (norte), e a distância de deslocamento d considerada foi de um (1) pixel. A figura mostra a contabilização de uma ocorrência (+1) do par de pixels cujos valores são 2 e 0. O *kernel* percorre todos os pixels da imagem da esquerda para a direita e de cima para baixo. Fonte: Do autor (2022).

Após contabilizar a ocorrência dos pares de pixels no *kernel* considerando uma distância d e uma direção θ , a GLCM (A) é então adicionada à sua transposta (A^t), a fim torná-la simétrica (S). Na sequência, a matriz simétrica é então normalizada, e seus valores expressos em probabilidade (P), que corresponde à probabilidade de ocorrência de um par de níveis de cinza em uma imagem. Considerando o exemplo da Figura 3, temos:

$$A = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad A^t = \begin{bmatrix} 3 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad S = \begin{bmatrix} 6 & 0 & 1 & 0 \\ 0 & 2 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Normalizando e expressando S em termos de probabilidade:

$$S = \begin{bmatrix} 6/12 & 0 & 1/12 & 0 \\ 0 & 2/12 & 1/12 & 0 \\ 1/12 & 1/12 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} .50 & 0 & .08 & 0 \\ 0 & .17 & .08 & 0 \\ .08 & .08 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

A notação $P_{i,j}$ é utilizada para representar a probabilidade P da linha i e da coluna j da matriz S normalizada. A partir desse ponto, diversas métricas (ou medidas) de textura podem ser calculadas, como o *contraste* (CO), pela seguinte equação:

$$CO = \sum_{i,j=0}^{N-1} P_{i,j} (i - j)^2 \quad (1)$$

A medida é calculada para cada pixel e, ao final, uma nova imagem é gerada. Na literatura é comum encontrarmos o termo *imagens de textura* para representar essas imagens. Diversos estudos têm utilizado as medidas de textura para estimar a BAS (KAYITAKIRE; HAMEL; DEFOURNY, 2006; LU; BATISTELLA, 2005). Trabalhos como Lu e Batistella

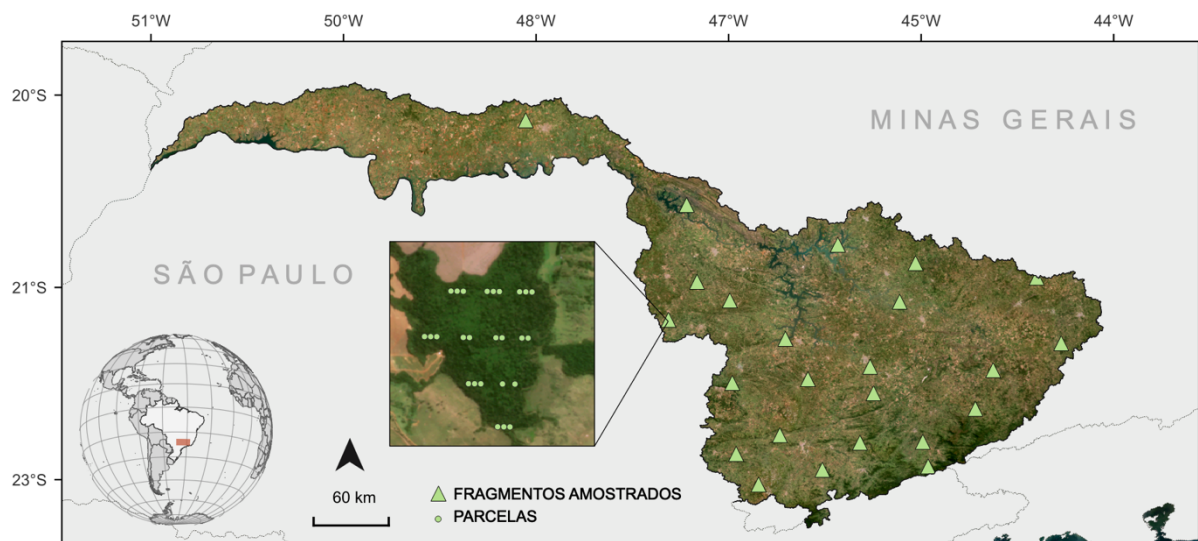
(2005) e Dube e Mutamba (2015) fornecem uma boa base para pesquisadores que pretendem utilizar as medidas de textura na modelagem da BAS.

3 MATERIAIS

3.1 Área de estudo

O estudo foi conduzido na bacia hidrográfica do Rio Grande (BRG), que está localizada na divisa entre os estados de Minas Gerais e São Paulo (FIGURA 4). A bacia está situada em uma área de transição entre os biomas Cerrado e Mata Atlântica e possui uma área de drenagem de aproximadamente 143.400 km². Devido à sua extensão, a BRG é diversa em termos florísticos e climáticos, apresentando, de acordo com a classificação de Köppen, os tipos climáticos *Aw*, *Cwb*, *Cwa*, *Cfb*, *Cfa* e *Cwc* (ALVARES et al., 2013). A temperatura média anual na bacia é de aproximadamente 23°C, com a precipitação média anual variando entre 1.250 mm nas regiões mais secas e 2.200 mm nas regiões mais úmidas.

Figura 4 – Localização da área de estudo.



Fonte: Do autor (2022).

3.2 Bases de dados

3.2.1 Inventário florestal

Neste trabalho, nós utilizamos dados do inventário florestal de 24 fragmentos de floresta estacional semidecidual (FES) realizado durante os anos de 2013 e 2015. Nesse período, 852 parcelas de 250 m² (10 x 250 m) foram alocadas na área de estudo. Em cada parcela, mediu-se

o diâmetro à altura do peito (DAP) e a altura total de todos os indivíduos com $DAP \geq 17,5$ cm. Em cada parcela, a BAS foi estimada utilizando a seguinte equação:

$$\ln (BAS) = -2,5375 + 2,2325 \ln (DAP) + 0,3127 \ln (HT) \quad (2)$$

3.2.2 Landsat 8 OLI/TIRS

Utilizamos dados de reflectância de superfície das bandas B2 (*Blue*), B3 (*Green*), B4 (*Red*), B5 (NIR) do sensor OLI e dados de temperatura de superfície da banda B10 (*Thermal infrared*) do sensor TIRS do satélite Landsat 8 (L8), bem como os índices de vegetação NDVI, SAVI, DVI, ARVI e EVI (TABELA 1).

3.2.3 Sentinel-1A

Os dados SAR utilizados neste trabalho foram adquiridos pelo satélite Sentinel-1A (S1). O S1 opera na banda C (frequência central de 5,404 GHz) e coleta dados em diferentes polarizações e resoluções. Neste trabalho, utilizamos o coeficiente de retroespalhamento (σ^0 , dB) de imagens GRD adquiridas no modo IW para compor a base de dados.

3.2.4 Análise de textura

Na análise da textura de uma imagem, o valor de uma medida de textura é calculado a partir do nível de cinza (valor) de todos os pixels localizados dentro de uma determinada unidade de análise (ou *kernel*). Medidas de textura de primeira ordem são resumos estatísticos dos valores dos pixels localizados dentro da unidade de análise. Medidas de textura de segunda ordem são derivadas da matriz de coocorrência de níveis de cinza (GLCM). A GLCM contém as frequências normalizadas com as quais os valores de um par de pixels adjacentes coocorrem dentro da unidade de análise. Neste trabalho, utilizamos a média (*mean*) e o desvio padrão (*stdDev*) como medidas de textura de primeira ordem, e as medidas propostas Haralick et al. (1973) e Connors, Trivedi e Harlow (1984) como medidas de segunda ordem. O tamanho do *kernel* utilizado foi de 3 x 3 pixels.

3.2.5 SRTM

O modelo digital de elevação com resolução espacial de 30 metros, resultado da missão SRTM (FARR et al., 2007), foi utilizado para obter a informação de elevação média em cada parcela.

3.3 Integração dos dados

3.3.1 Obtenção das variáveis preditoras

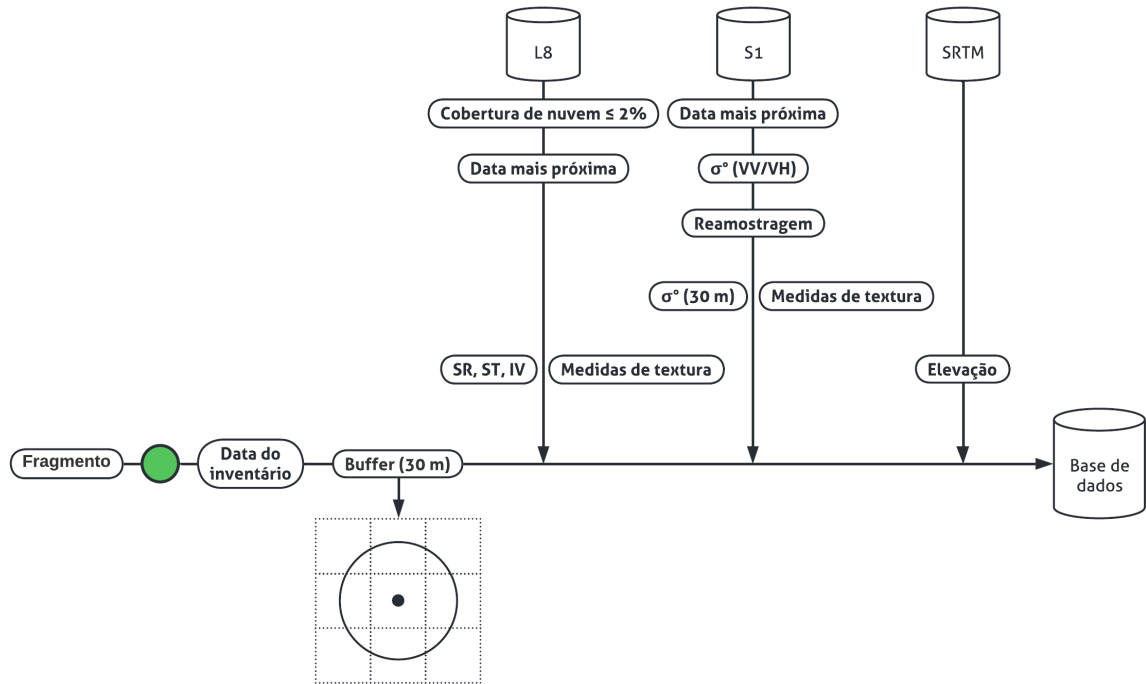
Considerando a localização dos fragmentos amostrados e a data de realização dos inventários florestais, as seguintes etapas foram executadas (FIGURA 5):

- 1) Inicialmente, selecionamos uma imagem L8 com até 2% de cobertura de nuvem e com data de aquisição mais próxima possível da data de realização do inventário florestal. O mesmo procedimento foi realizado utilizando os dados S1, com exceção da aplicação do filtro por cobertura de nuvem. Optamos por utilizar imagens com dupla polarização (VV/VH) adquiridas em órbitas descendentes.
- 2) Após essa etapa, as imagens S1 foram reamostradas para 30 m utilizando o método de reamostragem *nearest neighbor*.
- 3) Na sequência, os índices de vegetação e as medidas de textura foram calculados.

3.3.2 Extração dos valores

A partir do ponto central de cada parcela, um *buffer* de 30 m de raio foi criado com o objetivo de extrair as informações calculadas na seção anterior. Dentro de cada *buffer*, o valor médio de cada variável preditora foi extraído e incorporado à respectiva parcela do inventário florestal.

Figura 5 – Obtenção das variáveis preditoras.



Fonte: Do autor (2022).

4 MÉTODOS

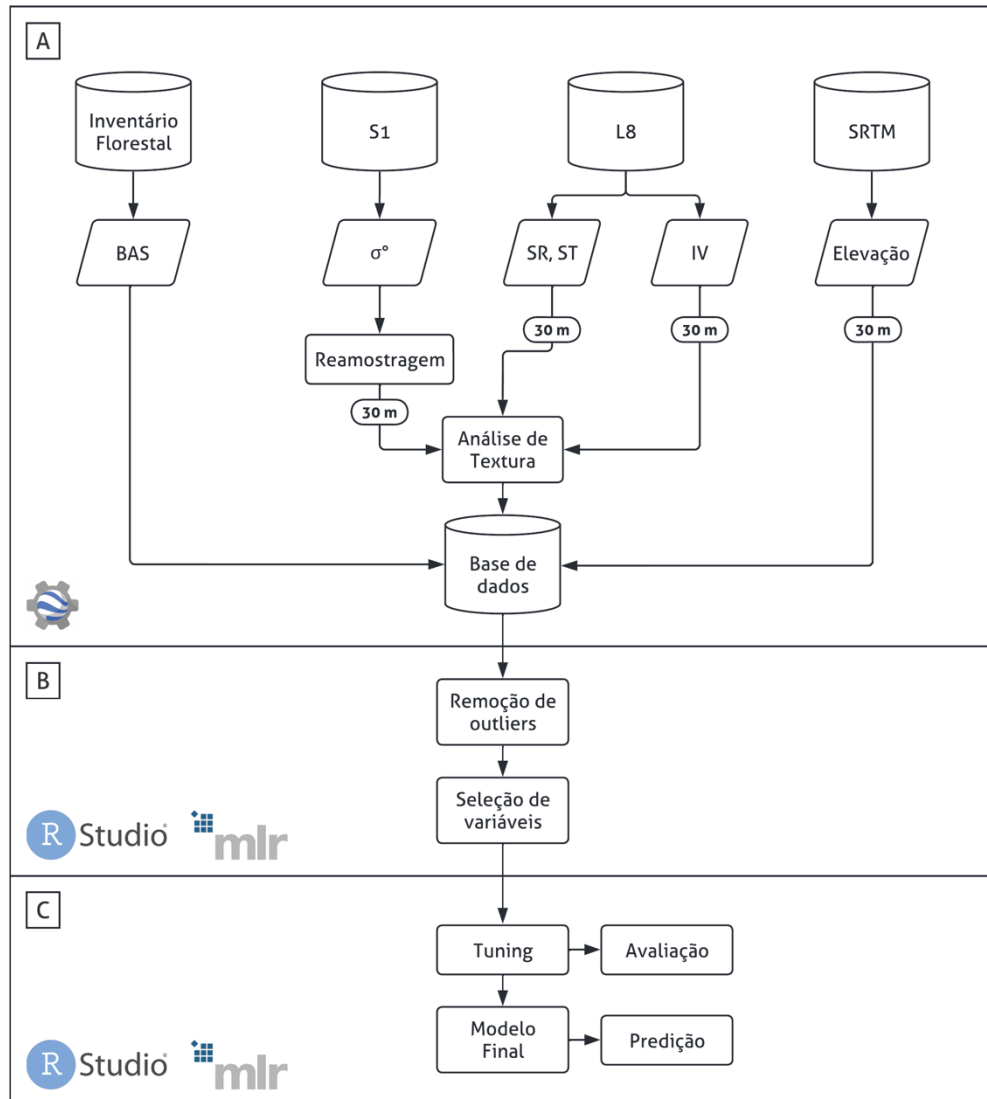
Neste trabalho, o modelo *Random Forest* (RF, BREIMAN, 2001) foi utilizado para estimar a BAS na BRG utilizando os dados obtidos na seção 3.2. O fluxo de trabalho adotado (FIGURA 6) pode ser dividido em três etapas: aquisição dos dados (FIGURA 6-A), pré-processamento dos dados (FIGURA 6-B) e modelagem (FIGURA 6-C). A etapa de aquisição de dados foi desenvolvida na plataforma Google Earth Engine (GEE, GORELICK et al., 2017).

O *framework* utilizado para o pré-processamento dos dados e para a modelagem foi implementado pelo pacote *mlr3* (LANG et al., 2019) utilizando a linguagem de programação estatística *R* (R CORE TEAM, 2022). O pacote *ranger* foi utilizado para implementar o algoritmo RF (WRIGHT; ZIEGLER, 2017).

4.1 Modelo Random Forest

O RF é um algoritmo de aprendizado de máquina não-paramétrico. Nesse algoritmo, uma fração dos dados de treinamento é escolhida aleatoriamente para construir uma árvore de decisão. A fração dos dados de treinamento não escolhidos é utilizada para estimar o erro em cada árvore. Em cada nó da árvore de decisão, um conjunto de variáveis preditoras é escolhido aleatoriamente para determinar a divisão dos dados. Centenas de árvores são construídas dessa forma. A predição de novos dados é feita pela agregação (e.g., média) das predições de todas as árvores construídas (LIAW; WIENER, 2002). O RF é insensível a ruídos na base de dados e não sobreajusta (BREIMAN, 2001). Ainda, é capaz de classificar as variáveis importantes para o modelo e produzir uma medida independente do erro de predição (ADAM et al., 2014).

Figura 6 – Fluxo de trabalho.



BAS = Biomassa acima do solo; σ° = Coeficiente de retroespalhamento; **SR** = Reflectância de superfície; **ST** = Temperatura de superfície; **IV** = Índices de vegetação. Fonte: Do autor (2022).

4.2 Otimização e avaliação do desempenho do modelo

4.2.1 Seleção de variáveis

Muitas vezes, os conjuntos de dados incluem um grande número de variáveis. A técnica de extrair um subconjunto de variáveis relevantes é conhecida como seleção de variáveis, ou *feature selection*. A seleção de variáveis pode melhorar a interpretabilidade, melhorar o desempenho e acelerar o processo de treinamento de um modelo. Três métodos de seleção de variáveis são enfatizados na literatura: os métodos *filter*, *wrapper* e *embedded* (CHANDRASHEKAR; SAHIN, 2014). Neste trabalho, nós utilizamos a correlação de Pearson

(ρ) – um método de filtragem (*filter*) – como critério de seleção das variáveis. O algoritmo *findCorrelation* do pacote *caret* (KUHN, 2008) foi utilizado para essa finalidade. O algoritmo considera o valor absoluto da correlação entre duas variáveis para decidir quais variáveis serão removidas da base de dados. O limiar de 0,80 foi utilizado, ou seja, as variáveis que forem altamente correlacionadas entre si ($|\rho| \geq 0,80$), serão removidas.

4.2.2 Validação cruzada k-fold

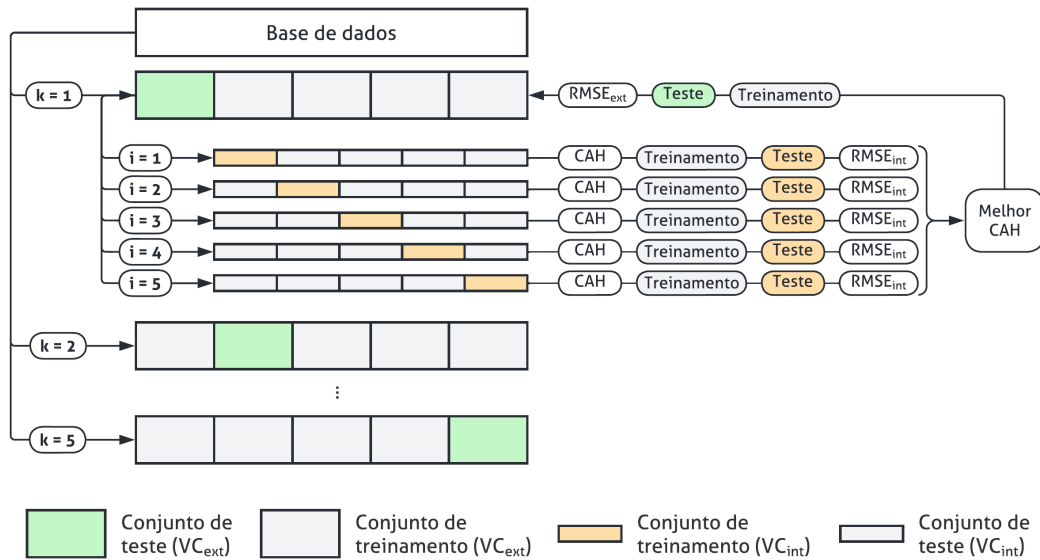
A validação cruzada (VC) é um método de reamostragem que utiliza diferentes partições de um conjunto de dados para testar e avaliar um modelo em diferentes iterações. O objetivo da VC é avaliar a capacidade preditiva de um modelo, bem como indicar problemas (e.g., *overfitting*). Um ciclo da validação cruzada consiste em (1) particionar um conjunto de dados em subconjuntos complementares, (2) realizar a análise em um subconjunto (conjunto de treinamento) e (3) validar a análise no outro (conjunto de teste). Para reduzir a variabilidade, na maioria dos métodos de VC, vários ciclos são executados usando partições diferentes, e os resultados da VC são combinados ao longo dos ciclos para fornecer uma estimativa mais robusta da capacidade preditiva do modelo.

Na validação cruzada *k-fold*, o conjunto de dados é dividido aleatoriamente em k partições de tamanhos (aproximadamente) iguais. De k partições, uma é utilizada na validação e $k-1$ no treinamento do modelo. Esse processo (*dividir-treinar-validar*) é repetido k vezes. Cada partição k é utilizada uma vez como conjunto de teste. A média dos resultados das k iterações é utilizada para produzir uma única estimativa.

4.2.3 Validação cruzada aninhada

A validação cruzada aninhada (VCA) é um conceito estatístico utilizado para produzir uma estimativa mais robusta da capacidade de generalização de um modelo. A VCA é composta por dois níveis de VC: a VC externa (VC_{ext}) e a VC interna (VC_{int}). A VC_{ext} é utilizada para estimar a capacidade preditiva do modelo, enquanto a VC_{int} é utilizada para otimizar os hiperparâmetros do modelo. A VCA elimina o viés introduzido pela VC tradicional na medida em que os conjuntos de teste da VC_{ext} não são utilizados na otimização dos hiperparâmetros do modelo na VC_{int} . Neste trabalho, usaremos k para representar as partições da VC_{ext} e i para as partições do VC_{int} (FIGURA 7).

Figura 7 – Validação cruzada aninhada.

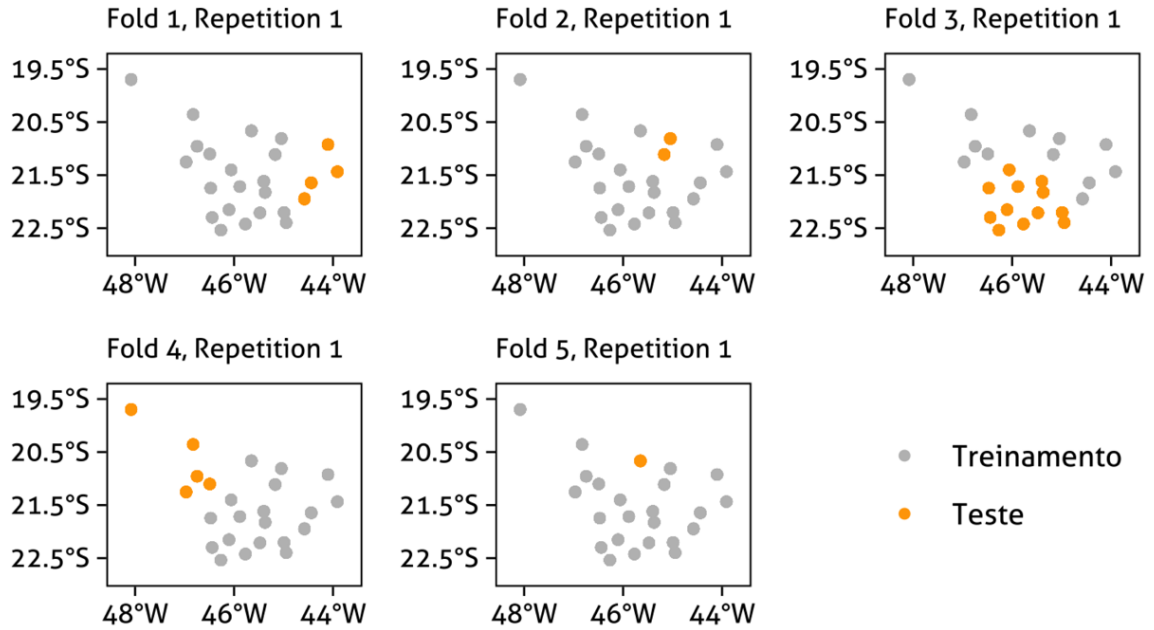


CAH = Configuração aleatória de hiperparâmetros; **RMSE_{int}** = RMSE do ciclo interno; **RMSE_{ext}** = RMSE do ciclo externo. Fonte: Do autor (2022).

4.2.4 Estimativa da capacidade preditiva do modelo

Neste estudo, nós utilizamos a VCA para estimar a capacidade preditiva do modelo RF. Na VC_{ext} , a estratégia utilizada foi a validação cruzada espacial (VC_{esp} , BRENNING, 2012) com 10 repetições, que utiliza agrupamentos *k-means* para criar partições espacialmente disjuntas com o objetivo reduzir o efeito da autocorrelação (FIGURA 8). O mesmo procedimento foi adotado na VC_{int} , mas sem repetições.

Figura 8 – Partições espacialmente disjuntas geradas pela validação cruzada espacial.



Este procedimento foi repetido 10 vezes, totalizando 50 (5 *folds* x 10 repetições) conjunto de dados de treinamento e de teste espacialmente independentes. Fonte: Do autor (2022).

Na VC_{int} , a otimização dos hiperparâmetros em cada partição i foi realizada pelo método da busca aleatória. Nesse método, n configurações aleatórias de hiperparâmetros (CAH) são avaliadas, sendo necessário definir um critério de parada, um espaço de busca e uma métrica para avaliar o desempenho de cada configuração. Neste trabalho, o critério de parada utilizada foi o número de avaliações ($n_{evals} = 50$). O espaço de busca de cada hiperparâmetro otimizado é mostrado na Tabela 2. A métrica utilizada para avaliar o desempenho das configurações foi a RMSE.

Tabela 2 – Espaço de busca dos hiperparâmetros.

Hiperparâmetro	Descrição ¹	LI	LS
m_{try}	Número de variáveis testadas na divisão de cada nó.	\sqrt{p}	p
<i>sample.fraction</i>	Fração do número de observações utilizada em cada árvore.	0,2	0,9
<i>min.node.size</i>	Número mínimo de observações em um nó terminal.	1	10

¹ Probst et al., 2019; **LI** = Limite inferior; **LS** = Limite superior; **p** = Número de variáveis na base de dados. Fonte: Do autor (2022).

Após cada ciclo da VC_{int} , a melhor combinação de hiperparâmetros encontrada foi utilizada para treinar um segundo modelo. Esse modelo foi treinado e avaliado utilizando os

conjuntos de treinamento e de teste da VC_{ext} correspondente. Esse processo se repetiu em todas as k partições 10 vezes. A capacidade preditiva do modelo (CPM) foi estimada pela equação:

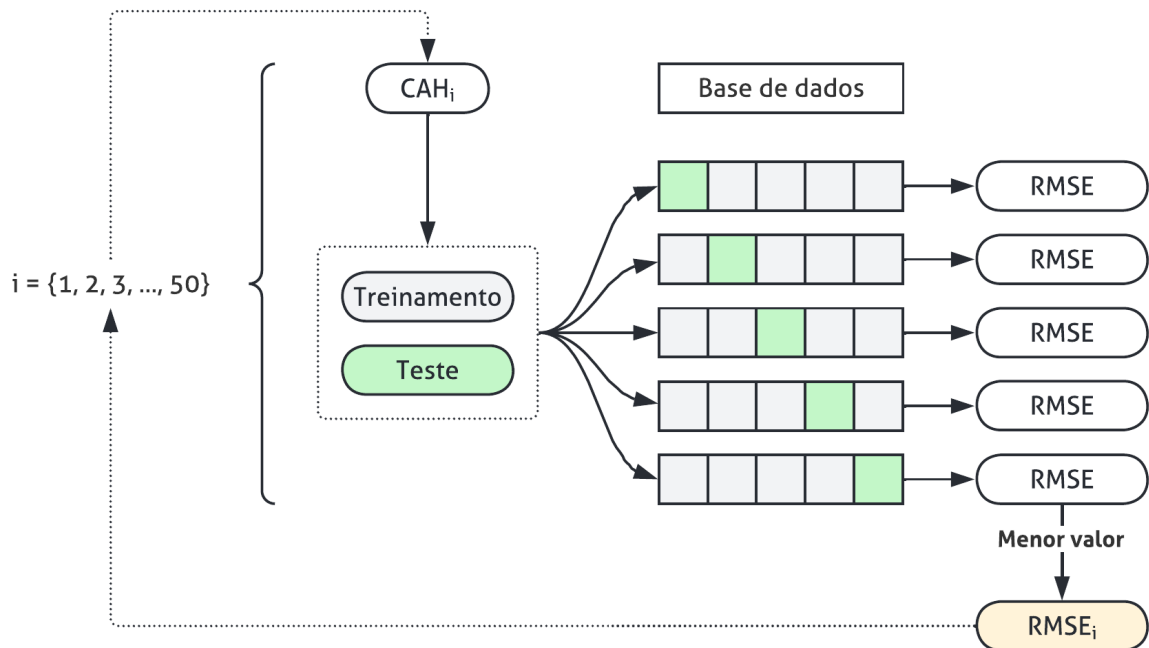
$$CPM = \frac{1}{nk} \sum_{i=1}^n \sum_{j=1}^k RMSE_{e_{ij}} \quad (3)$$

onde n corresponde ao número de repetições e k ao número de partições.

4.2.5 Otimização dos hiperparâmetros do modelo

A otimização dos hiperparâmetros foi feita pelo método da busca aleatória, dentro do espaço de busca definido na Tabela 2. O processo foi conduzido da seguinte forma:

Figura 9 – Processo de otimização dos hiperparâmetros do modelo.



CAH = Configuração aleatória de hiperparâmetros. Fonte: Do autor (2022).

- 1) O algoritmo de busca aleatória sugere uma configuração aleatória de hiperparâmetros (CAH) para ser avaliada.
- 2) O modelo, configurado com a CAH sugerida, é treinado e avaliado de acordo com um método e métrica de avaliação definidos pelo usuário. Neste trabalho, nós utilizamos o método da VC_{esp} ($k = 5$) e a métrica RMSE para avaliar cada configuração.
- 3) As etapas 1 e 2 se repetem até que um critério de parada seja alcançado. Aqui, o critério de parada utilizado foi o número de CAH a serem testadas ($n_{evals} = 50$).

- 4) Após testar as 50 CAH, aquela que retornar o menor valor da métrica de avaliação é utilizada para treinar o modelo final.

4.2.6 Determinação das variáveis mais importantes para o modelo

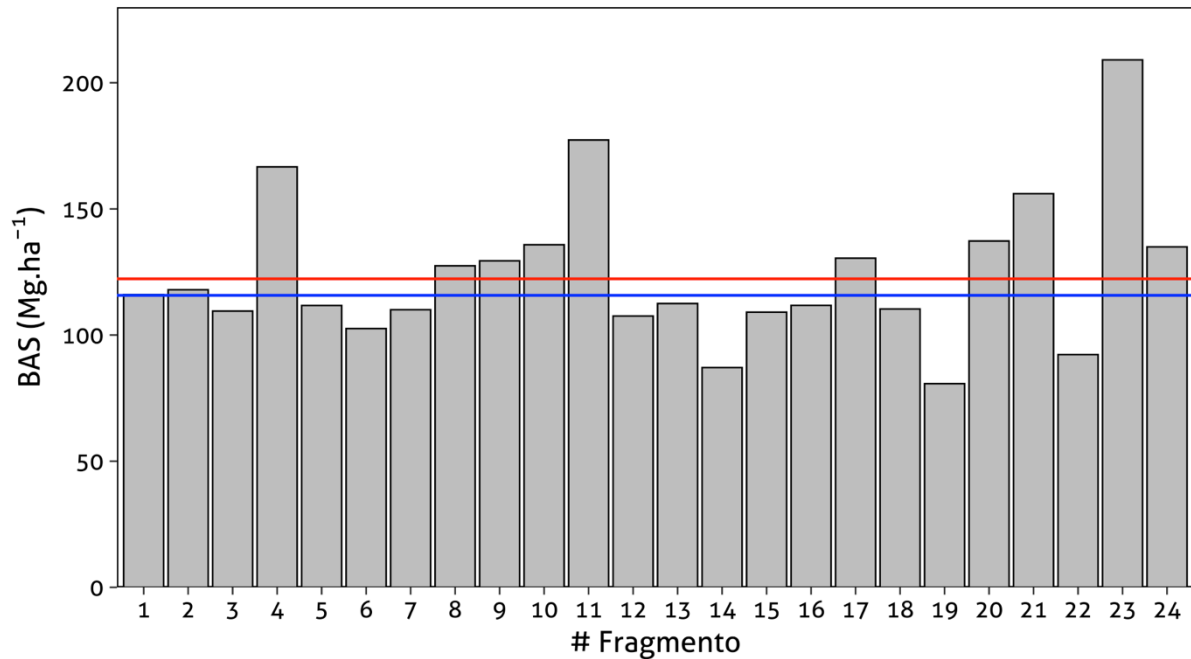
A fim de identificar as variáveis mais importantes para o modelo, nós utilizamos o método de avaliação de importância baseado em permutações (BREIMAN, 2001). O método se baseia na medida de importância do algoritmo RF e é descrito com mais detalhes por Fisher, Rudin e Dominici (2019). A ideia principal do método é medir o quanto o desempenho de um modelo altera se o efeito de uma variável é removido. A remoção do efeito de uma variável pode ser feita de diferentes formas; neste trabalho, nós utilizamos a permutação para esse fim. Assim, a importância de uma variável é medida pela redução do desempenho do modelo quando os valores dessa variável são aleatoriamente permutados (BREIMAN, 2001). Ou seja, se uma variável é importante, é esperado que o desempenho do modelo piore ao permutá-la.

5 RESULTADOS

5.1 Biomassa nos fragmentos inventariados

A distribuição da BAS estimada nos fragmentos amostrados é apresentada na Figura 10.

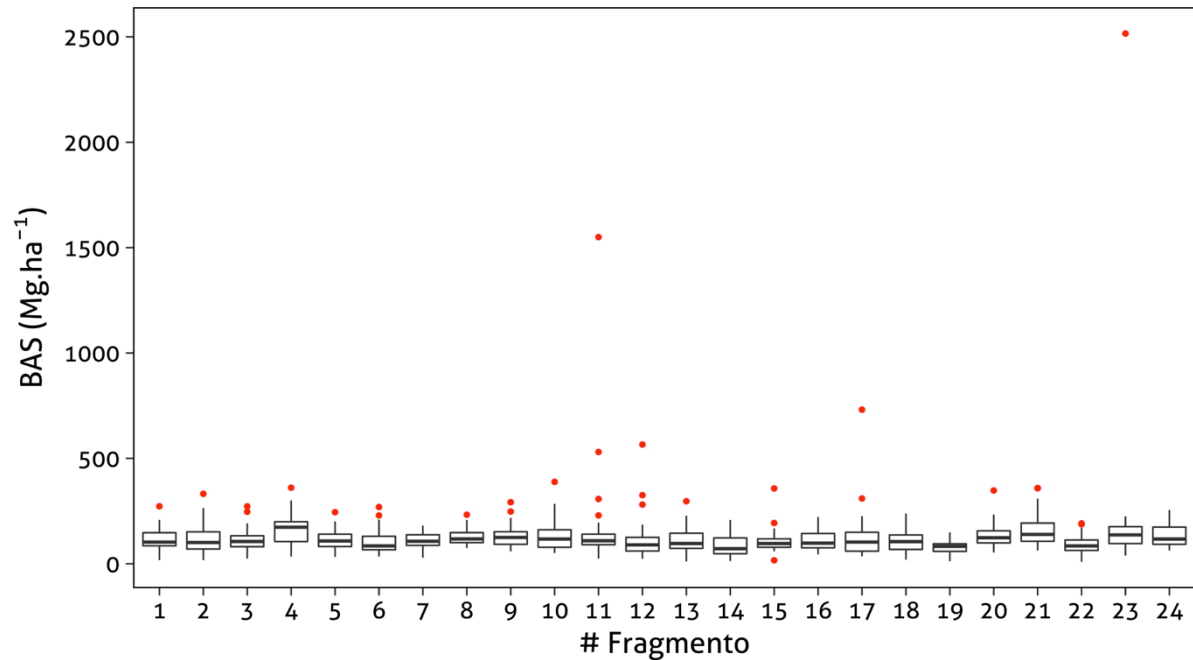
Figura 10 – Biomassa acima do solo estimada nos fragmentos florestais amostrados.



Média (linha vermelha) = 122,30 Mg.ha⁻¹; Desvio padrão (linha azul) = 115,74 Mg.ha⁻¹. Fonte: Do autor (2022).

As estimativas de BAS apresentaram grande variabilidade em sua distribuição com média = 122,30 Mg.ha⁻¹ e desvio padrão = 115,74 Mg.ha⁻¹. Essa variabilidade fica mais evidente ao observar os *boxplots* da Figura 11.

Figura 11 – Variabilidade da biomassa acima do solo nos fragmentos amostrados.

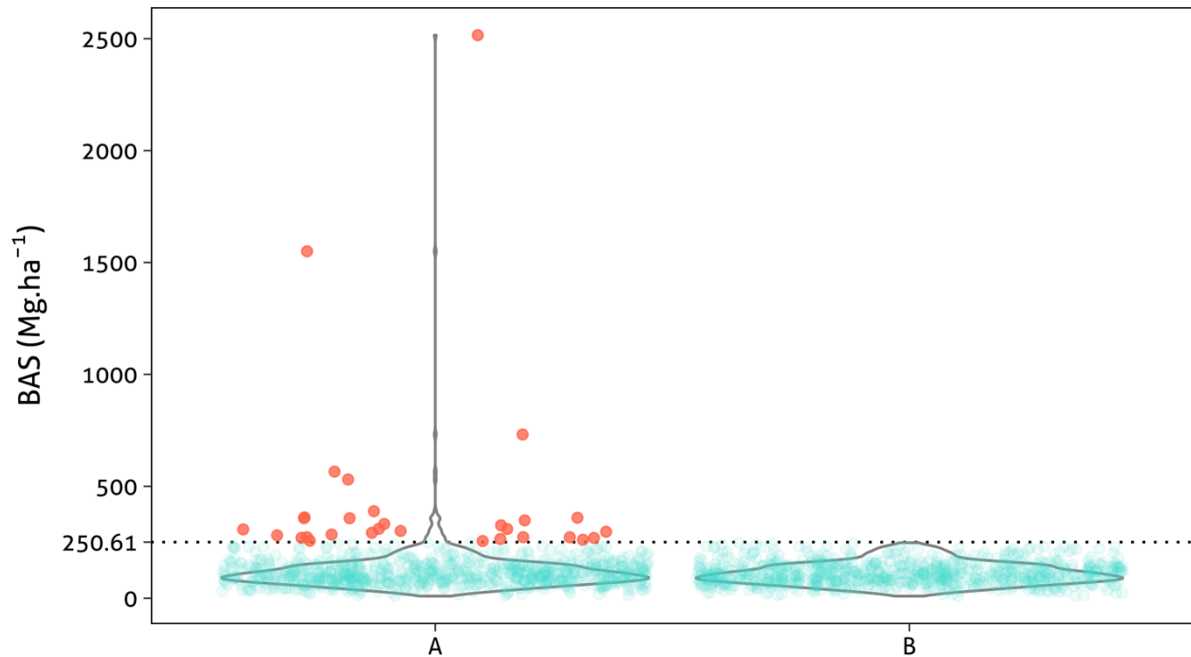


Os pontos vermelhos representam os valores discrepantes (*outliers*) estimados em cada fragmento. Fonte: Do autor (2022).

5.2 Remoção de *outliers*

Praticamente todos os fragmentos apresentaram valores discrepantes (*outliers*). Os fragmentos 11, 17 e 24 chamam a atenção por apresentarem valores estimados até 20x maiores do que a média observada neste estudo e em estudos anteriores (SCOLFORO; MELLO; SILVA, 2008). A hipótese considerada é a de que esses valores resultaram de erros sistemáticos ocorridos durante o processo de aquisição e/ou processamento dos dados do inventário florestal, e não de valores reais de BAS. O método de Tukey (TUKEY, 1977) foi utilizado para remover os valores discrepantes observados nas Figuras Figura 10 e Figura 11. A Figura 12 mostra o resultado da aplicação do método, que resultou na remoção de 30 observações.

Figura 12 – Distribuição dos dados antes e após a etapa de pré-processamento.



A = Base de dados com *outliers*; **B** = Base de dados sem *outliers*. A linha pontilhada representa o limite superior do método de Tukey ($250,61 \text{ Mg.ha}^{-1}$), que corresponde a $Q3 + (1,5 \times \text{IQR})$, onde: $Q3$ = terceiro quartil; IQR = amplitude interquartil. Todas as observações localizadas acima do limite superior foram consideradas *outliers*, e, portanto, removidas. Fonte: Do autor (2022).

A nova distribuição dos dados (FIGURA 12-B) está de acordo com os achados de Scolforo, Mello e Silva (2008), que encontraram valores de BAS variando entre 15,10 e 303,4 Mg.ha^{-1} , com média de 102,9 Mg.ha^{-1} em fragmentos de FES na BRG. A Tabela 3 mostra o resumo estatístico antes (FIGURA 12-A) e após (FIGURA 12-B) a remoção dos *outliers* da base de dados.

Tabela 3 – Resumo estatístico da variável BAS antes e após a remoção dos *outliers*.

	Mínimo ¹	Máximo ¹	Média ¹	DP ¹
A	9,42	2.515,70	122,30	115,74
B	9,42	247,82	110,34	48,31

¹ Valores em Mg.ha^{-1} ; **DP** = Desvio padrão. Fonte: Do autor (2022).

5.3 Variáveis selecionadas

O método de filtragem de seleção de variáveis resultou na remoção de 217 variáveis preditoras. A Tabela 4 mostra as variáveis preditoras que permaneceram na base de dados ($n = 47$) e foram utilizadas na modelagem.

Tabela 4 – Variáveis preditoras utilizadas na modelagem.

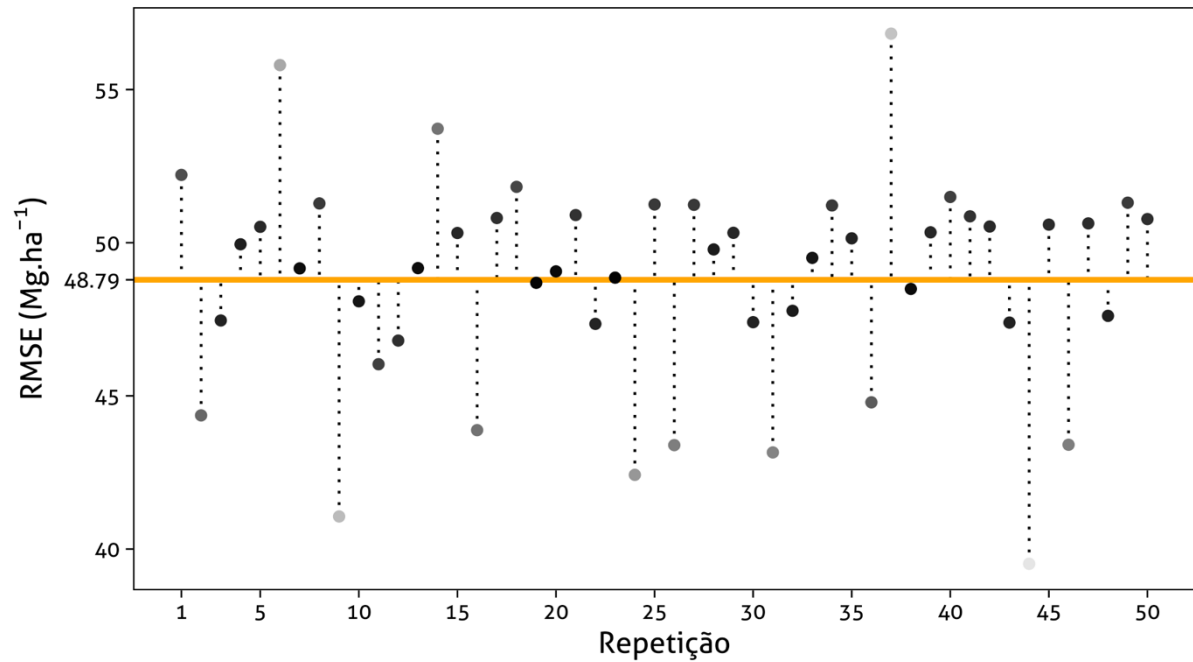
Fonte	Descrição	Nome da variável
L8	SR	SR_B2
	SR_*_text	SR_B2_corr, SR_B2_prom, SR_B2_shade, SR_B3_corr, SR_B5_mean, SR_B5_shade.
	ST_*_text	ST_B10_asm, ST_B10_corr, ST_B10_dent, ST_B10_dvar, ST_B10_mean, ST_B10_prom, ST_B10_shade
	IV	ARVI, RVI
S1	IV_text	ARVI_corr, ARVI_imcorr2, ARVI_prom, DVI_SENT, EVI_corr, EVI_sent, NDVI_imcorr2, RVI_corr, RVI_prom, RVI_shade, RVI_stddev, SAVI_imcorr2
	POL	VV, VH
SRTM	POL_text	VH_asm, VH_corr, VH_diss, VH_idm, VH_imcorr1, VH_imcorr2, VH_shade, VV_corr, VV_diss, VV_idm, VV_imcorr1, VV_imcorr2, VV_shade
		elevation

SR = Reflectância de superfície; **ST** = Temperatura de superfície; * = bandas da imagem L8; **text** = Medida de textura; **IV** = Índice de vegetação; **POL** = Polarização; **corr** = *Correlation*; **prom** = *Cluster prominence*; **shade** = *Cluster shade*; **mean** = Média; **asm** = *Angular second moment*; **dent** = *Difference entropy*; **dvar** = *Difference variance*; **imcorr2** = *Information measure of correlation 1*; **sent** = *Sum entropy*; **imcorr2** = *Information measure of correlation 2*; **stddev** = Desvio padrão; **diss** = *Dissimilarity*; **idm** = *Inverse difference moment*. Fonte: Do autor (2022).

5.4 Desempenho do modelo

O desempenho do modelo foi medido pela métrica RMSE, que segundo a VCA foi de 48,79 Mg.ha⁻¹ [44.22%]. O resultado do ciclo externo (VC_{ext}) da VCA é mostrado na Figura 13.

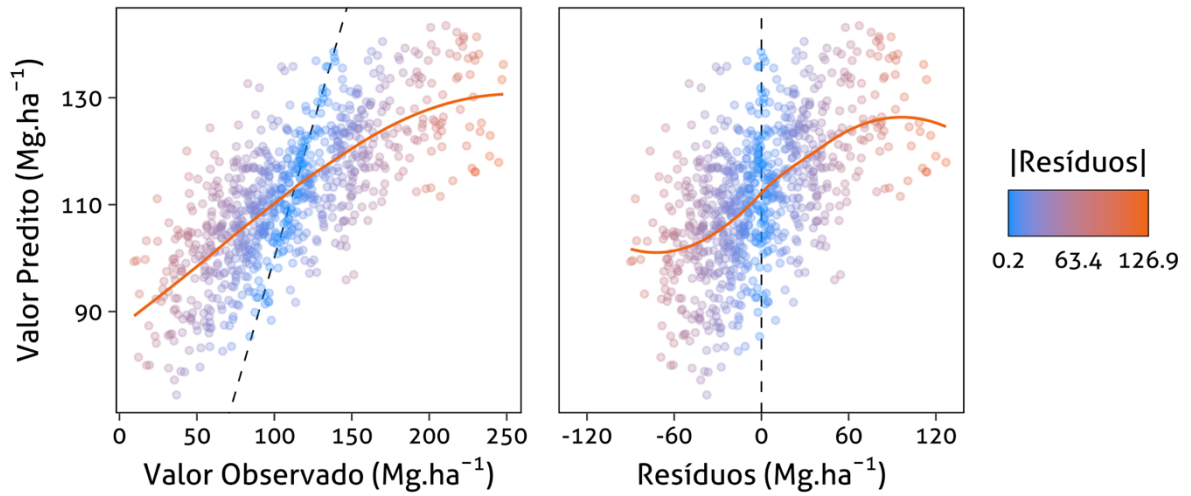
Figura 13 – Resultado do ciclo externo da validação cruzada aninhada.



O desempenho estimado do modelo (48,79 Mg.ha⁻¹, linha em destaque) corresponde ao desempenho médio (RMSE média) de todas as repetições (5 partições x 10 repetições = 50 repetições) do ciclo externo da validação cruzada aninhada. Fonte: Do autor (2022).

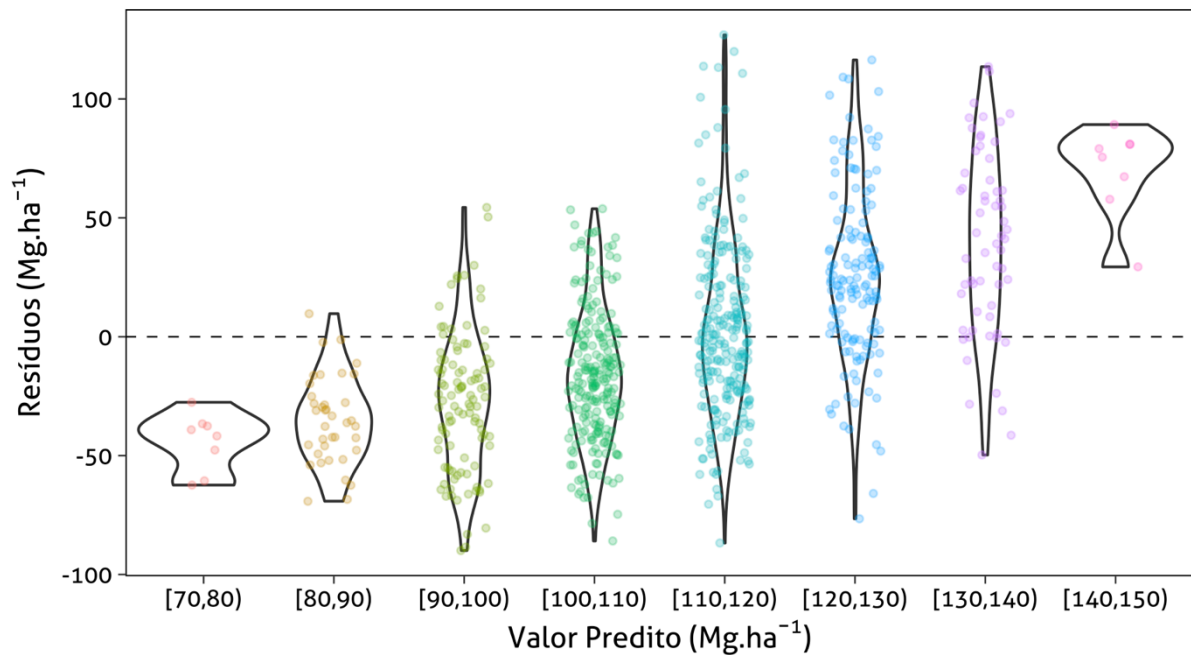
A análise dos valores preditos, bem como dos resíduos é mostrado na Figura 14. Os resultados mostram uma tendência na distribuição dos resíduos em oposição aos valores preditos. De maneira geral, o modelo subestimou a BAS em áreas mais densas (resíduos positivos) e superestimou em áreas menos densas (resíduos negativos). O comportamento dos resíduos por classe de densidade de BAS é mostrado na Figura 15.

Figura 14 – Predições do modelo *Random Forest* e distribuição dos resíduos.



Fonte: Do autor (2022).

Figura 15 – Distribuição dos resíduos por classes de densidade de biomassa acima do solo.

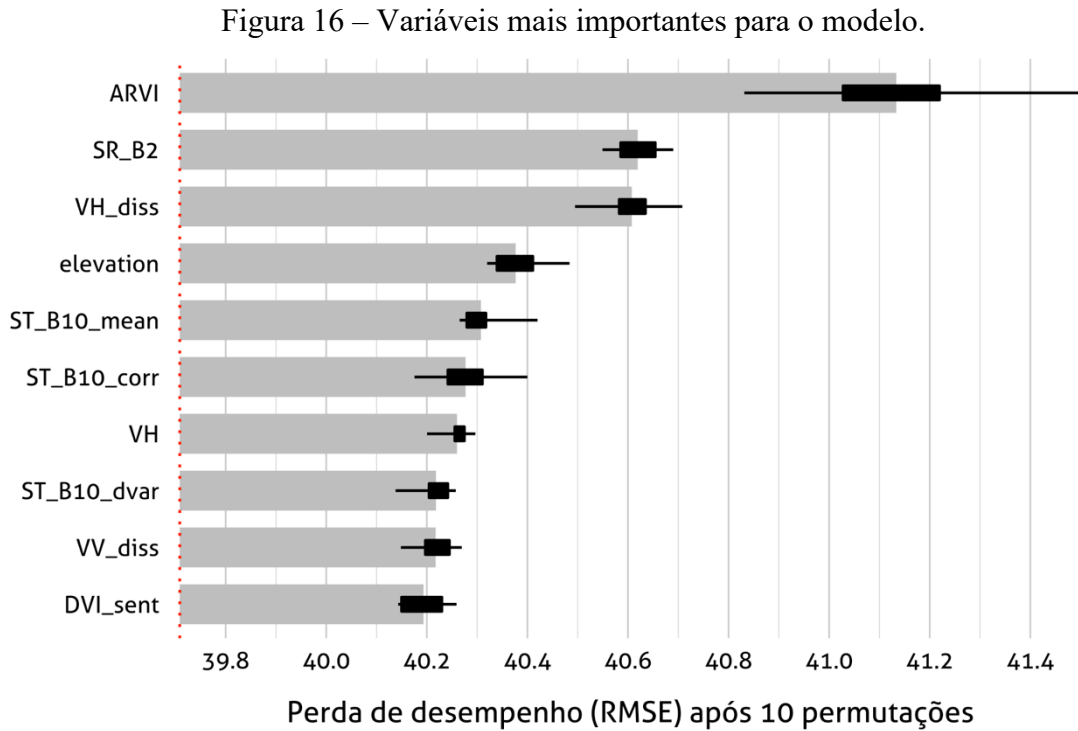


Fonte: Do autor (2022).

5.5 Variáveis preditoras mais importantes

As dez variáveis mais importantes para o modelo são apresentadas na Figura 16. O índice de vegetação ARVI foi a variável mais importante no modelo, seguida das variáveis *SR_B2*, *VH_diss* e *elevation*. O aumento no valor da métrica RMSE causado pela permutação dos valores dessas variáveis foi de 41,13 [37,29%], 40,62 [36,81%], 40,61 [36,80%] e 40,38 Mg.ha⁻¹ [36,59%], respectivamente. No que diz respeito à análise de textura, as medidas

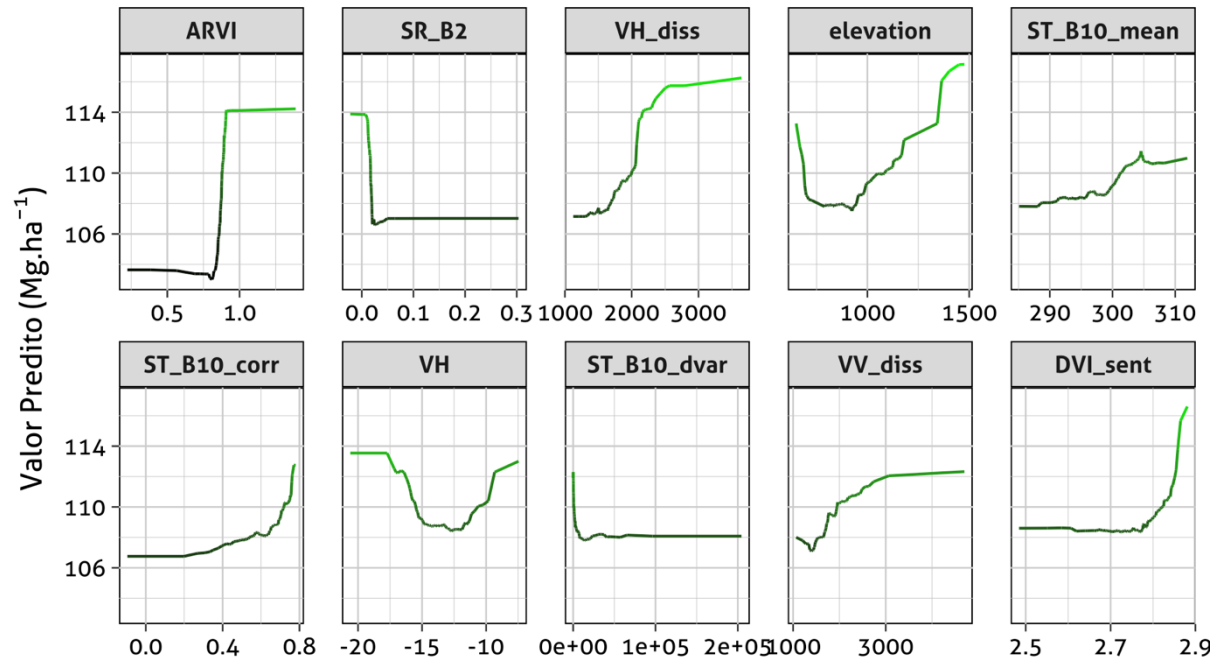
dissimilarity (diss), média (mean), *correlation* (corr), *difference variance* (dvar) e *sum entropy* (sent), derivadas dos dados L8 e S1, foram as mais relevantes.



Fonte: Do autor (2022).

A Figura 17 mostra os perfis de dependência parcial das principais variáveis do modelo. Os resultados mostram que as variáveis *elevation*, *VH_diss* e *DVI_sent* conseguiram captar a variação da BAS em áreas mais densas ($BAS \geq 114 \text{ Mg.ha}^{-1}$). O índice de vegetação ARVI, embora tenha sido a variável mais importante para o modelo, não conseguiu captar a variação da BAS nessas áreas. Por outro lado, a variável foi capaz de captar a variação da BAS em locais menos densos ($BAS \leq 106 \text{ Mg.ha}^{-1}$).

Figura 17 - Dependência parcial das variáveis mais importantes para o modelo.

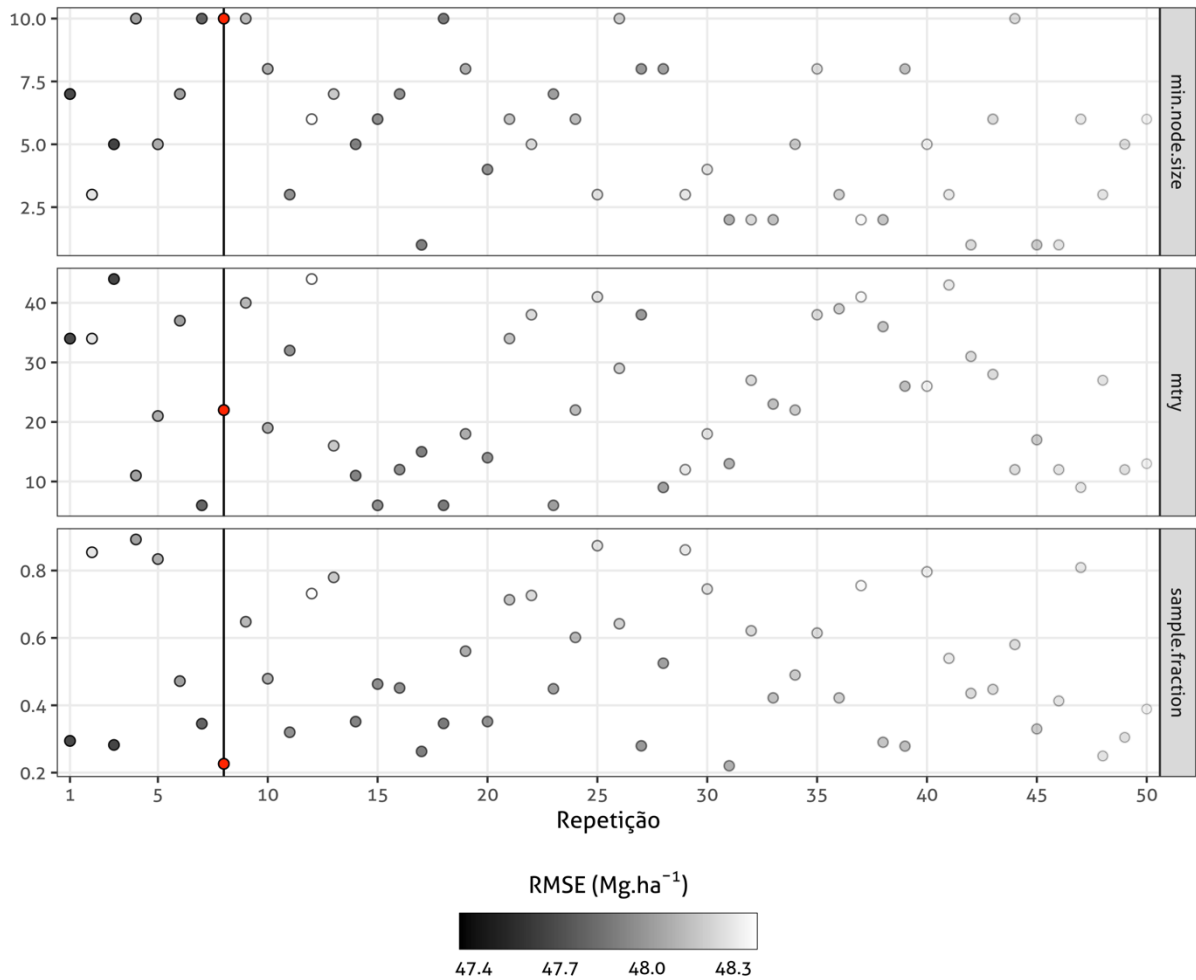


Fonte: Do autor (2022).

5.6 Otimização dos hiperparâmetros

A melhor configuração de hiperparâmetros foi: $min.node.size = 10$; $m_{try} = 22$ e $sample.fraction = 0,22$. O resultado da avaliação de todas as configurações de hiperparâmetros testadas ($n_{evals} = 50$) é apresentado na Figura 18.

Figura 18 - Resultado da validação cruzada espacial utilizada na otimização dos hiperparâmetros do modelo *Random Forest*.

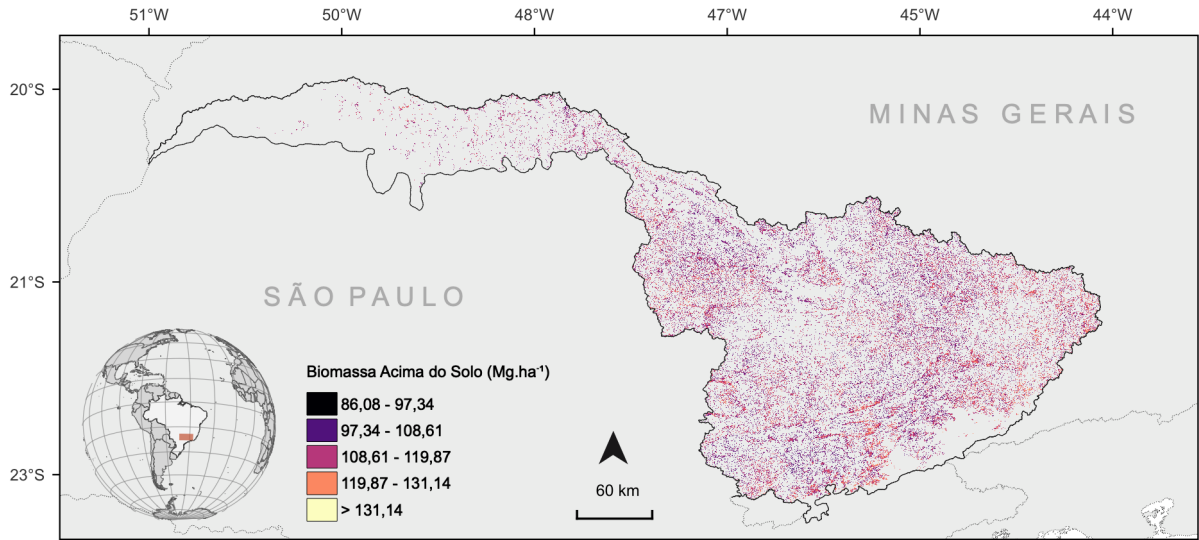


A melhor configuração de hiperparâmetros foi encontrada na sétima repetição (RMSE = 47,35 Mg.ha⁻¹, pontos vermelhos). Fonte: Do autor (2022).

5.7 Mapeamento da biomassa acima do solo

O mapeamento da BAS foi realizado utilizando o modelo RF treinado com a melhor combinação de hiperparâmetros (FIGURA 18). O resultado é mostrado na Figura 19. Os valores de BAS mapeados variaram de 86 a 142 Mg.ha⁻¹.

Figura 19 – Biomassa acima do solo na bacia do Rio Grande.



Fonte: Do autor (2022).

6 DISCUSSÃO

6.1 Estimativa da capacidade de generalização do modelo *Random Forest*

O desempenho do modelo RF utilizado para realizar o mapeamento da BAS na bacia do Rio Grande, estimado pela métrica RMSE, foi de 48,79 Mg.ha⁻¹ [44,22%]. O modelo, cuja performance pode ser considerada satisfatória, apresentou desempenho inferior a modelos de predição de BAS anteriormente reportados na literatura. No bioma Mata Atlântica, por exemplo, Barbosa et al. (2014) encontraram RMSE de 35 Mg.ha⁻¹. Silveira et al. (2019a, 2021), utilizando o modelo RF para estimar a BAS na bacia do Rio Doce, encontraram RMSE de 20,08 e 33,43 Mg.ha⁻¹, respectivamente.

Muitos estudos, no entanto, não consideram o efeito da autocorrelação espacial (AE) dos dados ao avaliar o desempenho de um modelo estatístico (SCHRATZ et al., 2019). A AE é um problema que, quando não controlado, pode levar a estimativas superotimistas (enviesadas) da capacidade preditiva de um modelo (BRENNING, 2005). Neste estudo, nós utilizamos o método de reamostragem baseado em agrupamentos *k-means* nos dois níveis da VCA a fim de mitigar esse problema. Dessa forma, todos os modelos utilizados para estimar a capacidade de generalização do modelo final foram treinados, avaliados e otimizados em conjuntos de dados espacialmente independentes. Por esse motivo, é comum que o modelo apresente métricas de desempenho ligeiramente piores quando comparado a modelos que não consideraram o efeito da AE (MEYER et al., 2018; MICHELETTI et al., 2013; ROBERTS et al., 2017; SCHRATZ et al., 2019).

6.2 Problema da saturação e tendências nas estimativas

A avaliação dos valores estimados mostrou que há uma limitação no modelo, que é a superestimativa nos menores valores de BAS e a subestimativa nos maiores valores de BAS. A modelagem da BAS utilizando dados de SR (ópticos e de radar) está sujeita à diminuição da sensibilidade à BAS à medida que a vegetação vai se tornando mais densa. Assim, mudanças na BAS acima de um nível de saturação resultam em mudanças na variável de SR que são pequenas em comparação com a variabilidade do sinal.

De acordo com Rodríguez-Veiga et al. (2019), esse comportamento pode ser explicado pelo número insuficiente de amostras de referência nos extremos, o que está de acordo com nossos resultados. As classes de densidade 70-80 Mg.ha⁻¹ e 140-150 Mg.ha⁻¹, que representam

os valores extremos da faixa de valores de BAS (FIGURA 15), são compostas por 8 observações cada, o que representa menos que 2% do número total de observações utilizadas na modelagem ($n = 822$). Avitabile et al. (2016) também encontraram o mesmo comportamento – superestimativa nos menores valores de BAS e subestimativa nos maiores valores de BAS – ao comparar os modelos desenvolvidos por Saatchi et al. (2011) e Baccini et al. (2012) com dados de referência.

Outros estudos, que também utilizaram o algoritmo RF, se depararam com o mesmo problema, e o relatam como um efeito do processo de agregação que ocorre nos algoritmos de aprendizado de máquina baseados em árvores de decisão, conhecido como *bagging* (BREIMAN, 2001).

Blackard et al. (2008), por outro lado, sugeriram que o problema de saturação dos dados poderia explicar a superestimativa em áreas de vegetação menos densas; e que os problemas relacionados à diferença entre a escala dos dados de referência (dados de inventário florestal, na maioria das vezes) e a escala do pixel (resolução espacial) poderiam explicar a subestimativa em áreas de vegetação mais densas.

Esse problema, no entanto, não parece ser um comportamento exclusivo dos algoritmos baseados em árvores de decisão. Tsui et al. (2013) relatou o mesmo problema utilizando a krigagem; Del Frate e Solimini (2004), por sua vez, percebeu o mesmo comportamento utilizando *Neural Networks*; Kattenborn et al. (2015) também reportaram o mesmo comportamento em algoritmos semi e não-paramétricos, *Generalized Additive Model*, *Generalized Boosted Regression Models* e *Boosted Generalized Additive Model*, e reforçaram a hipótese de que a principal causa desse comportamento foi a insuficiência de dados de referência.

6.3 Variáveis mais importantes para o modelo

Os resultados encontrados neste trabalho ressaltam a importância do uso sinérgico de dados multissensores na modelagem da BAS. As duas variáveis mais importantes para o modelo são dados ópticos derivados do satélite L8. Esse tipo de informação é amplamente utilizado para estimar a BAS (LU, 2005; SILVEIRA et al., 2019a). Por outro lado, as medidas de textura utilizadas neste trabalho desempenharam um papel fundamental na mitigação do problema da saturação dos dados. Diversos trabalhos anteriormente publicados sustentam essa constatação (CUTLER et al., 2012; ECKERT, 2012; KELSEY; NEFF, 2014; KUPLICH; CURRAN; ATKINSON, 2005; SARKER; NICHOL, 2011; SIBANDA et al., 2017). Dentre as dez

variáveis mais importantes para o modelo, seis são medidas de textura derivadas tanto de dados L8 (*ST_B10_mean*, *ST_B10_corr*, *ST_B10_dvar* e *DVI_sent*) quanto de dados S1 (*VH_diss* e *VV_diss*). As medidas de textura *mean*, *correlation* (*corr*), *different variance* (*dvar*), *sum entropy* (*sent*) e *dissimilarity* (*diss*) já foram utilizadas com sucesso em outros estudos de modelagem da BAS a partir de uso de dados de SR (BASTIN et al., 2014; CUTLER et al., 2012; DUBE; MUTANGA, 2015; ECKERT, 2012; SARKER; NICHOL, 2011; SARKER et al., 2013). Embora essas medidas de textura tenham desempenhado um papel importante na maioria desses estudos, a literatura indica que a identificação e seleção das medidas de textura requerem uma seleção adicional dos parâmetros que controlam o método GLCM, como tamanho de janela mais adequado, direção e distância deslocamento. Neste estudo, nós utilizamos a configuração padrão da função *gcmTexture* no GEE (seção 3.2.4), e a avaliação de diferentes configurações não foi realizada. Entretanto, trabalhos como o de Dube e Mutanga (2015) são capazes de fornecer um bom ponto de partida para a definição desses parâmetros.

A elevação (*elevation*) também foi uma variável importante na mitigação do efeito da saturação dos dados. A variável foi capaz de acompanhar as alterações em praticamente todas as faixas de valores da BAS, influenciando principalmente as estimativas entre 106-118 Mg.ha⁻¹ (FIGURA 17). Os trabalhos de Baccini (2004), Castillo et al. (2017), Gallaun et al. (2010) e López-Serrano et al. (2019) encontraram resultados semelhantes. Para os autores, a elevação tem a capacidade de controlar a distribuição espacial da BAS. Gallaun *et al.*, por exemplo, perceberam que as florestas de coníferas da Europa estocam mais carbonos em elevadas altitudes. Por outro lado, nas florestas de folhosas, o comportamento observado foi o oposto.

7 CONCLUSÃO

Este estudo utilizou o algoritmo RF, dados de SR e análise de textura para mapear BAS na BRG. O modelo apresentou desempenho satisfatório, principalmente em locais onde a densidade de BAS está entre 100-130 Mg.ha⁻¹. Entretanto, o modelo superestimou os menores valores de BAS e subestimou os maiores. Os resultados encontrados reforçaram a importância do uso das medidas de textura associadas a dados de diferentes tipos de sensores para a mitigação do efeito da saturação dos dados, comum quando se trabalha com dados de SR na modelagem da BAS. Estudos futuros devem avaliar diferentes configurações do método GLCM, principalmente o tamanho da janela, além de incluir variáveis biofísicas/climáticas e outros índices de vegetação, com a finalidade de melhorar as estimativas de BAS na área de estudo.

REFERÊNCIAS

- ADAM, E. 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, 19 maio 2014.
- ALVARES, C. A. et al. **Köppen's climate classification map for Brazil**. *Meteorologische Zeitschrift*, v. 22, n. 6, p. 711–728, 1 dez. 2013.
- AVITABILE, V. et al. **An integrated pan-tropical biomass map using multiple reference datasets**. *Global Change Biology*, v. 22, n. 4, p. 1406–1420, 10 abr. 2016.
- AVITABILE, V. et al. **Capabilities and limitations of Landsat and land cover data for aboveground woody biomass estimation of Uganda**. *Remote Sensing of Environment*, v. 117, p. 366–380, fev. 2012.
- BACCINI, A. et al. **A first map of tropical Africa's above-ground biomass derived from satellite imagery**. *Environmental Research Letters*, v. 3, n. 4, p. 045011, out. 2008.
- BACCINI, A. et al. **Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps**. *Nature Climate Change*, v. 2, n. 3, p. 182–185, 29 mar. 2012.
- BACCINI, A. et al. **Forest biomass estimation over regional scales using multisource data**. *Geophysical Research Letters*, v. 31, n. 10, maio 2004.
- BARBOSA, J. M. et al. **Remotely sensed biomass over steep slopes: An evaluation among successional stands of the Atlantic Forest, Brazil**. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 88, p. 91–100, fev. 2014.
- BARET, F.; GUYOT, G. **Potentials and limits of vegetation indices for LAI and APAR assessment**. *Remote Sensing of Environment*, v. 35, n. 2–3, p. 161–173, fev. 1991.
- BASTIN, J.-F. 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, dez. 2014.
- BISPO, P. DA C. et al. **Woody Aboveground Biomass Mapping of the Brazilian Savanna with a Multi-Sensor and Machine Learning Approach**. *Remote Sensing*, v. 12, n. 17, p. 2685, 19 ago. 2020.
- BLACKARD, J. et al. **Mapping U.S. forest biomass using nationwide forest inventory data and moderate resolution information**. *Remote Sensing of Environment*, v. 112, n. 4, p. 1658–1677, 15 abr. 2008.
- BONAN, G. B. **Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests**. *Science*, v. 320, n. 5882, p. 1444–1449, 13 jun. 2008.
- BREIMAN, L. **Random Forests**. *Machine Learning*, v. 45, n. 1, p. 5–32, 2001.

BRENNING, A. **Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: The R package sperrorest**. 2012 IEEE International Geoscience and Remote Sensing Symposium. Anais...IEEE, jul. 2012.

BRENNING, A. **Spatial prediction models for landslide hazards: review, comparison and evaluation**. Natural Hazards and Earth System Sciences, v. 5, n. 6, p. 853–862, 7 nov. 2005.

CAMPOS, V. E. et al. **Remote sensing data to assess compositional and structural indicators in dry woodland**. Ecological Indicators, v. 88, p. 63–70, maio 2018.

CAO, L.; WANG, Y.; TANG, H. **Special Section Guest Editorial: Advances in Remote Sensing for Forest Structure and Functions**. Journal of Applied Remote Sensing, v. 14, n. 02, p. 1, 27 maio 2020.

CARTUS, O. et al. **A National, Detailed Map of Forest Aboveground Carbon Stocks in Mexico**. Remote Sensing, v. 6, n. 6, p. 5559–5588, 16 jun. 2014.

CASTILLO, J. A. A. et al. **Estimation and mapping of above-ground biomass of mangrove forests and their replacement land uses in the Philippines using Sentinel imagery**. ISPRS Journal of Photogrammetry and Remote Sensing, v. 134, p. 70–85, dez. 2017.

CHANDRASHEKAR, G.; SAHIN, F. **A survey on feature selection methods**. Computers & Electrical Engineering, v. 40, n. 1, p. 16–28, jan. 2014.

CHAO, Z. et al. **Estimation methods developing with remote sensing information for energy crop biomass: A comparative review**. Biomass and Bioenergy, v. 122, p. 414–425, mar. 2019.

CHEN, J. M. **Evaluation of Vegetation Indices and a Modified Simple Ratio for Boreal Applications**. Canadian Journal of Remote Sensing, v. 22, n. 3, p. 229–242, set. 1996.

CHEN, L. et al. **Optimal Combination of Predictors and Algorithms for Forest Above-Ground Biomass Mapping from Sentinel and SRTM Data**. Remote Sensing, v. 11, n. 4, p. 414, 18 fev. 2019.

CONNERS, R. W.; TRIVEDI, M. M.; HARLOW, C. A. **Segmentation of a high-resolution urban scene using texture operators**. Computer Vision, Graphics, and Image Processing, v. 25, n. 3, p. 273–310, mar. 1984.

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, jun. 2012.

DEL FRATE, F.; SOLIMINI, D. **On Neural Network Algorithms for Retrieving Forest Biomass From SAR Data**. IEEE Transactions on Geoscience and Remote Sensing, v. 42, n. 1, p. 24–34, jan. 2004.

DUBE, T.; MUTANGA, O. **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, out. 2015.

ECKERT, S. **Improved Forest Biomass and Carbon Estimations Using Texture Measures from WorldView-2 Satellite Data.** Remote Sensing, v. 4, n. 4, p. 810–829, 26 mar. 2012.

ENGLHART, S. et al. **Aboveground biomass estimation of tropical peat swamp forests using SAR and optical data.** 2012 IEEE International Geoscience and Remote Sensing Symposium. Anais...IEEE, jul. 2012. Disponível em: <<http://ieeexplore.ieee.org/document/6352092/>>

FENG, Y. et al. **Examining effective use of data sources and modeling algorithms for improving biomass estimation in a moist tropical forest of the Brazilian Amazon.** International Journal of Digital Earth, v. 10, n. 10, p. 996–1016, 3 out. 2017.

FISHER, A.; RUDIN, C.; DOMINICI, F. **All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously.** Journal of machine learning research, v. 20, n. 177, 2019.

FORKUOR, G. et al. **Above-ground biomass mapping in West African dryland forest using Sentinel-1 and 2 datasets - A case study.** Remote Sensing of Environment, v. 236, p. 111496, jan. 2020.

GALLAUN, H. 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, jun. 2010.

GHASEMI, N.; SAHEBI, M. R.; MOHAMMADZADEH, A. **Biomass Estimation of a Temperate Deciduous Forest Using Wavelet Analysis.** IEEE Transactions on Geoscience and Remote Sensing, v. 51, n. 2, p. 765–776, fev. 2013.

GHOSH, S. M.; BEHERA, M. D. **Aboveground biomass estimation using multi-sensor data synergy and machine learning algorithms in a dense tropical forest.** Applied Geography, v. 96, p. 29–40, 1 jul. 2018.

GÓMEZ, C. et al. **Modeling Forest Structural Parameters in the Mediterranean Pines of Central Spain using QuickBird-2 Imagery and Classification and Regression Tree Analysis (CART).** Remote Sensing, v. 4, n. 1, p. 135–159, 10 jan. 2012.

GORELICK, N. et al. **Google Earth Engine: Planetary-scale geospatial analysis for everyone.** Remote Sensing of Environment, v. 202, p. 18–27, dez. 2017. FARR, T. G. et al. The Shuttle Radar Topography Mission. Reviews of Geophysics, v. 45, n. 2, p. RG2004, 19 maio 2007.

HARALICK, R. M.; SHANMUGAM, K.; DINSTEN, I. **Textural Features for Image Classification.** IEEE Transactions on Systems, Man and Cybernetics, v. SMC-3, n. 6, p. 610–621, 1973.

HAYASHI, M.; MOTOHKA, T.; SAWADA, Y. **Aboveground Biomass Mapping Using ALOS-2/PALSAR-2 Time-Series Images for Borneo's Forest**. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, v. 12, n. 12, p. 5167–5177, dez. 2019.

HEROLD, M. et al. **The Role and Need for Space-Based Forest Biomass-Related Measurements in Environmental Management and Policy**. Surveys in Geophysics, v. 40, n. 4, p. 757–778, 11 jul. 2019.

HOUGHTON, R. A.; HALL, F.; GOETZ, S. J. **Importance of biomass in the global carbon cycle**. Journal of Geophysical Research: Biogeosciences, v. 114, n. G2, p. n/a-n/a, jun. 2009.

HUANG, H. et al. **Integration of multi-resource remotely sensed data and allometric models for forest aboveground biomass estimation in China**. Remote Sensing of Environment, v. 221, p. 225–234, fev. 2019.

HUETE, A. **A comparison of vegetation indices over a global set of TM images for EOS-MODIS**. Remote Sensing of Environment, v. 59, n. 3, p. 440–451, mar. 1997.

HUETE, A. **A soil-adjusted vegetation index (SAVI)**. Remote Sensing of Environment, v. 25, n. 3, p. 295–309, ago. 1988.

JIANG, F. et al. **Estimating the aboveground biomass of coniferous forest in Northeast China using spectral variables, land surface temperature and soil moisture**. Science of The Total Environment, v. 785, p. 147335, set. 2021.

JORDAN, C. F. **Derivation of Leaf-Area Index from Quality of Light on the Forest Floor**. Ecology, v. 50, n. 4, p. 663–666, jul. 1969.

KATTENBORN, T. et al. **Mapping forest biomass from space – Fusion of hyperspectral EO1-hyperion data and Tandem-X and WorldView-2 canopy height models**. International Journal of Applied Earth Observation and Geoinformation, v. 35, p. 359–367, mar. 2015.

KAUFMAN, Y. J.; TANRE, D. **Atmospherically resistant vegetation index (ARVI) for EOS-MODIS**. IEEE Transactions on Geoscience and Remote Sensing, v. 30, n. 2, p. 261–270, mar. 1992.

KAYITAKIRE, F.; HAMEL, C.; DEFOURNY, P. **Retrieving forest structure variables based on image texture analysis and IKONOS-2 imagery**. Remote Sensing of Environment, v. 102, n. 3–4, p. 390–401, jun. 2006.

KELSEY, K.; NEFF, J. **Estimates of Aboveground Biomass from Texture Analysis of Landsat Imagery**. Remote Sensing, v. 6, n. 7, p. 6407–6422, 9 jul. 2014.

KUHN, M. **Building predictive models in R using the caret package**. Journal of Statistical Software, v. 28, n. 5, p. 1–26, 2008.

KUMAR, L.; MUTANGA, O. **Remote Sensing of Above-Ground Biomass**. Remote Sensing, v. 9, n. 9, p. 935, 10 set. 2017.

- KUPLICH, T. M.; CURRAN, P. J.; ATKINSON, P. M. **Relating SAR image texture to the biomass of regenerating tropical forests**. *International Journal of Remote Sensing*, v. 26, n. 21, p. 4829–4854, 10 nov. 2005.
- LANG, M. et al. **mlr3: A modern object-oriented machine learning framework in R**. *Journal of Open Source Software*, v. 4, n. 44, p. 1903, 11 dez. 2019.
- LE TOAN, T. et al. **The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle**. *Remote Sensing of Environment*, v. 115, n. 11, p. 2850–2860, nov. 2011.
- LI, C.; ZHOU, L.; XU, W. **Estimating Aboveground Biomass Using Sentinel-2 MSI Data and Ensemble Algorithms for Grassland in the Shengjin Lake Wetland, China**. *Remote Sens.*, v. 13, p. 1595, 2021.
- LI, Y. et al. **Forest aboveground biomass estimation using Landsat 8 and Sentinel-1A data with machine learning algorithms**. *Scientific Reports*, v. 10, n. 1, p. 9952, 19 dez. 2020.
- LIAW, A.; WIENER, M. **Classification and Regression by randomForest**. *R News*, v. 2, n. 3, p. 18–22, 2002.
- LILLESAND, T.; KIEFER, R. W.; CHIPMAN, J. **Remote Sensing and Image Interpretation, 7th Edition**. 7. ed. [s.l.] Wiley, 2015.
- LÓPEZ-SERRANO, P. M. et al. **Modeling of Aboveground Biomass with Landsat 8 OLI and Machine Learning in Temperate Forests**. *Forests*, v. 11, n. 1, p. 11, 19 dez. 2019.
- LOURENÇO, P. **Biomass Estimation Using Satellite-Based Data**. In: *Forest Biomass - From Trees to Energy*. [s.l.: s.n.].
- LU, D. **Aboveground biomass estimation using Landsat TM data in the Brazilian Amazon**. *International Journal of Remote Sensing*, v. 26, n. 12, p. 2509–2525, 20 jun. 2005.
- LU, D. et al. **A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems**. *International Journal of Digital Earth*, v. 9, n. 1, p. 63–105, 2 jan. 2016.
- LU, D.; BATISTELLA, M. **Exploring TM image texture and its relationships with biomass estimation in Rondônia, Brazilian Amazon**. *Acta Amazonica*, v. 35, n. 2, p. 249–257, jun. 2005.
- LU, D.; WENG, Q. **Spectral Mixture Analysis of the Urban Landscape in Indianapolis with Landsat ETM+ Imagery**. *Photogrammetric Engineering & Remote Sensing*, v. 70, n. 9, p. 1053–1062, 1 set. 2004.
- MA, L. et al. **Ecosystem carbon storage in forest fragments of differing patch size**. *Scientific Reports*, v. 7, n. 1, p. 13173, 13 dez. 2017.
- MARSHALL, M.; THENKABAIL, P. **Advantage of hyperspectral EO-1 Hyperion over multispectral IKONOS, GeoEye-1, WorldView-2, Landsat ETM+, and MODIS vegetation indices in crop biomass estimation**. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 108, p. 205–218, out. 2015.

- MCGARVEY, J. C. et al. **Carbon storage in old-growth forests of the Mid-Atlantic: toward better understanding the eastern forest carbon sink.** Ecology, v. 96, n. 2, p. 311–317, fev. 2015.
- MEYER, H. et al. **Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation.** Environmental Modelling & Software, v. 101, p. 1–9, mar. 2018.
- MICHELETTI, N. et al. **Machine Learning Feature Selection Methods for Landslide Susceptibility Mapping.** Mathematical Geosciences, v. 46, n. 1, p. 33–57, 20 jan. 2014.
- MORIN, D. et al. **Estimation and Mapping of Forest Structure Parameters from Open Access Satellite Images: Development of a Generic Method with a Study Case on Coniferous Plantation.** Remote Sensing, v. 11, n. 11, p. 1275, 29 maio 2019.
- NI-MEISTER, W. et al. **Assessing general relationships between aboveground biomass and vegetation structure parameters for improved carbon estimate from lidar remote sensing.** Journal of Geophysical Research: Biogeosciences, v. 115, n. G2, p. n/a-n/a, jun. 2010.
- NICHOL, J. E.; SARKER, M. L. R. **Improved Biomass Estimation Using the Texture Parameters of Two High-Resolution Optical Sensors.** IEEE Transactions on Geoscience and Remote Sensing, v. 49, n. 3, p. 930–948, mar. 2011.
- PANDIT, S.; TSUYUKI, S.; DUBE, T. **Exploring the inclusion of Sentinel-2 MSI texture metrics in above-ground biomass estimation in the community forest of Nepal.** Geocarto International, v. 35, p. 1832–1849, 2020.
- PEREGON, A.; YAMAGATA, Y. **The use of ALOS/PALSAR backscatter to estimate above-ground forest biomass: A case study in Western Siberia.** Remote Sensing of Environment, v. 137, p. 139–146, out. 2013.
- PHUA, M. H. et al. **Synergistic use of Landsat 8 OLI image and airborne LiDAR data for above-ground biomass estimation in tropical lowland rainforests.** Forest Ecology and Management, v. 406, p. 163–171, dez. 2017.
- PONZONI, F. J.; SHIMABUKURO, Y. E.; KUPLICH, T. M. **Sensoriamento remoto da vegetação.** 2. ed. São Paulo: Oficina de Textos, 2012.
- R CORE TEAM. **R: A Language and Environment for Statistical Computing.** Vienna, Austria. R Foundation for Statistical Computing, 2022. Disponível em: <https://www.r-project.org/>
- RICHARDSON, A. J.; WIEGAND, C. L. **Distinguishing vegetation from soil background information.** Photogrammetric engineering and remote sensing, v. 43, n. 12, p. 1541-1552, 1977.
- ROBERTS, D. R. et al. **Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure.** Ecography, v. 40, n. 8, p. 913–929, ago. 2017.

RODRÍGUEZ-VEIGA, P. 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, maio 2019.

ROITMAN, I. et al. **Optimizing biomass estimates of savanna woodland at different spatial scales in the Brazilian Cerrado: Re-evaluating allometric equations and environmental influences**. *PLOS ONE*, v. 13, n. 8, p. e0196742, 1 ago. 2018.

ROUJEAN, J. L.; BREON, F. M. **Estimating PAR absorbed by vegetation from bidirectional reflectance measurements**. *Remote Sensing of Environment*, v. 51, n. 3, p. 375–384, mar. 1995.

ROUSE, J. W. et al. **Monitoring vegetation systems in the Great Plains with ERTS**. *Third Earth Resources Technology Satellite-1 Symposium. Anais...* 1 jan. 1974.

SAATCHI, S. S. et al. **Benchmark map of forest carbon stocks in tropical regions across three continents**. *Proceedings of the National Academy of Sciences*, v. 108, n. 24, p. 9899–9904, 14 jun. 2011.

SALATI, E. **Emissão x seqüestro de CO₂ – uma nova oportunidade de negócios para o Brasil**. In: SEMINÁRIO EMISSÃO X SEQUESTRO DE CO₂ – UMA NOVA OPORTUNIDADE DE NEGÓCIOS PARA O BRASIL, Rio de Janeiro. Anais... Rio de Janeiro: CVRD, p. 15-37, 1994.

SANDBERG, G. et al. **L- and P-band backscatter intensity for biomass retrieval in hemiboreal forest**. *Remote Sensing of Environment*, v. 115, n. 11, p. 2874–2886, nov. 2011.

SANQUETTA, C. R. **Métodos de determinação de biomassa florestal**. In: SANQUETTA, C. R. et al. (Eds.). *As florestas e o carbono*. Curitiba: [s.n.], 2002, p. 119-140.

SARKER, L. R.; NICHOL, J. E. **Improved forest biomass estimates using ALOS AVNIR-2 texture indices**. *Remote Sensing of Environment*, v. 115, n. 4, p. 968–977, 15 abr. 2011.

SARKER, M. L. R. et al. **Forest Biomass Estimation Using Texture Measurements of High-Resolution Dual-Polarization C-Band SAR Data**. *IEEE Transactions on Geoscience and Remote Sensing*, v. 51, n. 6, p. 3371–3384, jun. 2013.

SCHRATZ, P. et al. **Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data**. *Ecological Modelling*, v. 406, p. 109–120, 24 ago. 2019.

SCOLFORO, J. R. S.; MELLO, J. M. DE; SILVA, C. P. DE C. **Inventário Florestal de Minas Gerais: Floresta Estacional Semidecidual e Ombrófila - Florística, Estrutura, Diversidade, Similaridade, Distribuição Diamétrica e de Altura, Volumetria, Tendências de Crescimento e Áreas Aptas para Manejo Florestal**. Lavras: Editora UFLA, 2008.

SIBANDA, M. et al. **Estimating Biomass of Native Grass Grown under Complex Management Treatments Using WorldView-3 Spectral Derivatives**. *Remote. Sens.*, v. 9, p. 55, 2017.

- SILVEIRA, E. M. O. et al. **Modelling aboveground biomass in forest remnants of the Brazilian Atlantic Forest using remote sensing, environmental and terrain-related data.** *Geocarto International*, v. 36, n. 3, p. 281–298, 7 fev. 2021.
- SILVEIRA, E. M. O. et al. **Object-based random forest modelling of aboveground forest biomass outperforms a pixel-based approach in a heterogeneous and mountain tropical environment.** *International Journal of Applied Earth Observation and Geoinformation*, v. 78, p. 175–188, jun. 2019a.
- SILVEIRA, E. M. O. 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, ago. 2019b.
- SILVEIRA, P. et al. **O estado da arte na estimativa de biomassa e carbono em formações florestais.** *Floresta*, v. 38, n. 1, 27 mar. 2008.
- SINHA, S. et al. **A review of radar remote sensing for biomass estimation.** *International Journal of Environmental Science and Technology*, v. 12, n. 5, p. 1779–1792, 20 maio 2015.
- SOMOGYI, Z. et al. **Indirect methods of large-scale forest biomass estimation.** *European Journal of Forest Research*, v. 126, n. 2, p. 197–207, 5 abr. 2007.
- SORIANO-LUNA, M. et al. **Determinants of Above-Ground Biomass and Its Spatial Variability in a Temperate Forest Managed for Timber Production.** *Forests*, v. 9, n. 8, p. 490, 11 ago. 2018.
- THEOFANOUS, N. et al. **Aboveground Biomass Estimation in Short Rotation Forest Plantations in Northern Greece Using ESA’s Sentinel Medium-High Resolution Multispectral and Radar Imaging Missions.** *Forests*, v. 12, n. 7, p. 902, 11 jul. 2021.
- TSUI, O. W. et al. **Integrating airborne LiDAR and space-borne radar via multivariate kriging to estimate above-ground biomass.** *Remote Sensing of Environment*, v. 139, p. 340–352, dez. 2013.
- WALLNER, A. et al. **Estimation of forest structural information using RapidEye satellite data.** *Forestry*, v. 88, n. 1, p. 96–107, 1 jan. 2015.
- WOOD, E. M. et al. **Image texture as a remotely sensed measure of vegetation structure.** *Remote Sensing of Environment*, v. 121, p. 516–526, jun. 2012.
- WRIGHT, M. N.; ZIEGLER, A. **ranger : A Fast Implementation of Random Forests for High Dimensional Data in C++ and R.** *Journal of Statistical Software*, v. 77, n. 1, 2017.
- ZEBKER, H. A.; VAN ZYL, J. J.; HELD, D. N. **Imaging radar polarimetry from wave synthesis.** *Journal of Geophysical Research*, v. 92, n. B1, p. 683, 1987.
- ZHAO, P. et al. **Examining Spectral Reflectance Saturation in Landsat Imagery and Corresponding Solutions to Improve Forest Aboveground Biomass Estimation.** *Remote Sens.*, v. 8, p. 469, 2016.

ZIMBRES, B. et al. **Mapping the stock and spatial distribution of aboveground woody biomass in the native vegetation of the Brazilian Cerrado biome.** *Forest Ecology and Management*, v. 499, p. 119615, 1 nov. 2021.