CAIO EDUARDO VIEIRA ALCANTARA SILVA

EFEITO DA POSIÇÃO DO DOSSEL NA PREDIÇÃO DO ESTOQUE DE CARBONO EM FLORESTA NATIVA

LAVRAS – MG
2020

CAIO EDUARDO VIEIRA ALCANTARA SILVA

EFEITO DA POSIÇÃO DO DOSSEL NA PREDIÇÃO DO ESTOQUE DE CARBONO EM FLORESTA NATIVA

Monografia apresentada à Universidade Federal de Lavras, como parte das exigências do Curso de Engenharia Florestal, para a obtenção do título de Bacharel.

Dr. Kalill José Viana da Páscoa
Orientador

Dr. Lucas Rezende Gomide
Coorientador

LAVRAS – MG
2020

CAIO EDUARDO VIEIRA ALCANTARA SILVA

EFEITO DA POSIÇÃO DO DOSSEL NA PREDIÇÃO DO ESTOQUE DE CARBONO EM FLORESTA NATIVA

EFFECT OF CANOPY POSITION ON CARBON STOCK PREDICING IN NATIVE FOREST

Monografia apresentada à Universidade Federal de Lavras, como parte das exigências do Curso de Engenharia Florestal, para a obtenção do título de Bacharel.

APROVADA em 17 de agosto de 2020.

Lucas Rezende Gomide	UFLA
Evandro Nunes Miranda	UFLA

Dr. Kalill José Viana da Páscoa
Orientador

Dr. Lucas Rezende Gomide
Coorientador

LAVRAS – MG
2020

AGRADECIMENTOS

Gostaria de agradecer primeiramente aos meu pais Cristina e Magno, que são os principais responsáveis pela conclusão da minha Graduação.

Ao Lemaf, por todo apoio estrutural e principalmente dos seus técnicos sempre à disposição, Thiago, Kalill e Thiza.

Ao Professor Lucas Gomide e ao Doutorando Evandro Miranda Pela ajuda no desenvolvimento deste trabalho.

À Universidade Federal de Lavras por toda sua estrutura disponível e pela oportunidade de aprendizado.

À República Methiolate, na qual me acolheu inicialmente em Lavras e morei boa parte da graduação direcionando meus primeiros passos universitários.

Ao grupo de Maracatu Baque do Morro, que me proporcionou muito aprendizado étnico e cultural e distrações da vida acadêmica.

E no mais, a todos aqueles presentes em meu dia a dia, que foram responsáveis diretamente ou indiretamente para a conclusão deste.

Obrigado!

RESUMO

As florestas nativas são importantes para uma infinidade de serviços ecossistêmicos, entre eles o sequestro e estoque de carbono. Dessa forma, conhecer o estoque de carbono existente nos remanescentes florestais é de grande importância para ajudar a justificar sua preservação bem como a recuperação de áreas degradadas. O objetivo desse trabalho foi avaliar o uso de informações espectrais e índices de vegetação obtidos do satélite SENTINEL-2, variáveis hidrológicas (Precipitação interna e Armazenamento de água no solo) e variáveis geográficas em conjunto com informações dendrométricas do povoamento para a estimativa do estoque de carbono em diferentes estratos do dossel de um remanescente florestal pertencente a fitofisionomia Floresta Estacional Semidecidual. Para isso foram empregadas técnicas de aprendizado de máquinas (*Random forest* associado a meta-heurística Algoritmo genético - GARF) em comparação com a modelagem clássica por meio da Regressão Linear Múltipla p RLM utilizando o método *Stepwise*, para a seleção das variáveis mais indicadas para incrementar o poder preditivo dos modelos. Os resultados indicam que com exceção das variáveis dendrométricas, as variáveis analisadas apresentarem baixa correlação com estoque de carbono, apesar disso, todas elas contribuíram de alguma forma para a melhoria das estimativas de carbono nos diferentes estratos do dossel. O método GARF privilegiou o uso dos dados espectrais e dendrometricos para os percentis superiores e os dados dendrométricos e hidrológicos para os inferiores, enquanto o RLM privilegiou o uso conjunto das variáveis espectrais, dendrometricas, hidrológicas e geográficas para os percentis superiores e dendrométricos e espectrais para os inferiores. O método GARF, com exceção do percentil 90, foi capaz de produzir melhores resultados em comparação com o RLM, apesar de as diferenças não serem grandes. Dessa forma, o uso conjunto dessas variáveis se mostra promissor para a produção de estimativas mais confiáveis.

Palavras-chave: Modelagem, Seleção de variáveis, SENTINEL-2

.

ABSTRACT

The native forests remains are an important role for many ecological reasons and global services, including a carbon sink. Thus, knowing the carbon stock existing in the forest remnants is of great importance to help justify its preservation as well as the recovery of degraded areas. The objective of this work was to evaluate the use of spectral information and vegetation indexes obtained from the satellite SENTINEL-2, hydrological variables (internal precipitation and water storage in the soil) and geographic variables together with dendrometric information from the stand to estimate the stock of carbon in different strata of the canopy of a forest remnant belonging to phytophysiognomy Seasonal Semideciduous Forest. For that, machine learning techniques were used (Random forest associated with metaheuristic Genetic algorithm - GARF) in comparison with classical modeling using Multiple Linear Regression p RLM using the Stepwise method, for the selection of the most suitable variables to increase the predictive power of models. The results indicate that, with the exception of the dendrometric variables, the analyzed variables have a low correlation with carbon stock, despite all of which have contributed in some way to the improvement of carbon estimates in the different strata of the canopy. The GARF method favored the use of spectral and dendrometric data for the upper percentiles and dendrometric and hydrological data for the lower ones, while the RLM favored the combined use of spectral, dendrometric, hydrological and geographic variables for the upper and dendrometric and spectral percentiles for the lower ones. The GARF method, with the exception of the 90th percentile, was able to produce better results compared to the RLM, although the differences were not large. Thus, the joint use of these variables is promising for the production of more reliable estimates.

Keywords: Modeling, feature selection, SENTINEL-2

SUMÁRIO

CAPÍTULO I – INTRODUÇÃO GERAL

1 INTRODUÇÃO

As florestas nativas desempenham uma importante função ecológica de armazenamento do carbono em sua biomassa (TESFAYE, 2016), o que nas últimas décadas tem gerado um grande debate acerca da importância da preservação desses ecossistemas terrestres em termos de sequestro, ciclagem e estoque de carbono como forma de contribuir para a minimização dos efeitos das mudanças climáticas (DA SILVA FILHO et al., 2020). Nesse cenário, os estudos científicos que visam avaliar e quantificar o estoque de carbono por meio dos remanescentes florestais se tornaram frequentes. Normalmente são realizados empregando-se modelos alométricos, que estimam o estoque de carbono de forma individual ou para o povoamento, utilizando informações dendrométricas coletadas por meio do inventário florestal, como diâmetro medido a 1,30 do solo (DAP), área basal (G), entre outras (RIBEIRO et al., 2011). Contudo, os inventários florestais são operações que exigem grande volume de recursos financeiros, logísticos e de mão de obra, tanto na coleta, quanto no processamento das informações, principalmente para grandes áreas florestais.

Dessa forma, é de interesse científico desenvolver formas alternativas para realizar a estimativa do estoque de carbono contido nos remanescentes florestais, seja por formas híbridas com o método clássico de estimativa, ou mesmo sugerindo novos métodos aplicando geotecnologias. Dentro dessas geotecnologias, uma ferramenta que trouxe grandes benefícios aos estudos envolvendo os recursos florestais foi o Sensoriamento Remoto que permitiu um maior detalhamento de informações para caracterizar o povoamento. Através do estudo das imagens obtidas por sensores acoplados a satélites se tornou possível fazer inferências espaciais de maneira precisa sobre padrões das florestas como área foliar (LIBERATO, 2011), volumetria (MIGUEL et al., 2015), estado fitossanitário (ROSA et al., 2008), nutricional (NETO et al., 2002) e hídrico (SADER et al., 1995), bem como identificar alterações na cobertura vegetal (ROCHA et al., 2011), com base nas informações espectrais de reflectância das folhas do dossel.

Esses benefícios se tornaram possíveis devido ao fato de os pigmentos foliares, em especial as clorofilas, terem absorções preferenciais em determinados comprimentos de ondas do espectro óptico, o que permitiu o desenvolvimento de índices de vegetação, que podem ser correlacionados às mais diferentes informações sobre a distribuição dos recursos florestais, dentre elas, o estoque de carbono (BOLFE; BATISTELLA; FERREIRA, 2012). Outras informações comumente relacionadas com os padrões das florestas nativas são as informações

hidrológicas. Atributos como a precipitação interna e armazenamento da água no solo estão diretamente correlacionados com diversas funções eco-hidrológicas e consequentemente, influenciam os padrões de crescimento e dispersão dos indivíduos dentro de um povoamento, dessa forma, são capazes de influenciar os padrões de estoque de carbono (MELLO et al., 2008). Diante do exposto, o presente trabalho tem como objetivo principal, avaliar o uso de informações espectrais e hidrológicas em conjunto com informações dendrométricas do povoamento para a estimativa do estoque de carbono em um remanescente florestal pertencente a fitofisionomia Floresta Estacional Semidecidual, empregando técnicas de aprendizado de máquinas (*Random forest* associado a meta-heurística Algoritmo genético) em comparação com a modelagem clássica por meio da Regressão *Stepwise*, para a seleção das variáveis mais indicadas para incrementar o poder preditivo dos modelos.

2 REFERENCIAL TEÓRICO

2.1 Estoque de carbono pelas florestas

Nas últimas décadas, desmatamento, incêndios e a degradação das florestas naturais têm contribuído intensamente para o aumento das concentrações de dióxido de carbono (CO_2) na atmosfera, e consequentemente com as mudanças climáticas (FROLKING et al., 2009; HANSEN et al., 2013; TESFAYE et al., 2016). Por essa razão, a quantificação e o monitoramento do sequestro e estoque de carbono, vem despertando a atenção de vários cientistas, como pode ser observado em Seidel et al. (2011), Lu et al. (2004), Scolforo et al. (2015), Baccini et al. (2008), Silveira et al. (2019a, 2019b), Wang et al. (2011) e Dube (2018).

Esse grande apelo pela preservação e restauração das florestas, se deve ao fato de as mesmas fixarem grandes quantidades de carbono, captado da atmosfera, em sua biomassa (FANG et al. 2014; RIBEIRO et al., 2011). Por esse fato, a quantificação precisa do estoque de carbono contido em fragmentos florestais é essencial para estimativas sobre as emissões e sequestros de CO_2 da atmosfera para a implementação de políticas de mitigação das mudanças climáticas (HIGUCHI et al., 2004; SCOLFORO et al., 2015).

Estudos a nível mundial estimam que as florestas são responsáveis por cerca de 80% do estoque da biomassa acima do solo, desempenhando um papel importante no ciclo do carbono (HOUGHTON, 2005), sendo esses valores variáveis em relação ao tipo de vegetação em questão, a composição de espécies, tamanho populacional, idade, estágio de sucessão da floresta (WATZLAWICK et al., 2002).

De forma geral, as florestas tropicais são consideradas como o principal sumidouro de carbono (STEPHENS et al. 2007). Pan et al. (2011) estima que estoque atual de carbono nas florestas do mundo é de 861 Petagramas, onde aproximadamente 55% (473 Petagramas) está estocado nas florestas tropicais, principalmente devido a associação com os fatores edafoclimáticos locais que favorecerem a produção de biomassa.

No caso do Brasil, as principais fitofisionomias responsáveis por estoques médios de carbono que vão de 268 Mg ha^{-1} na Floresta Amazônica (BROWN; LUGO, 1992) até 5,3 Mg ha^{-1} na Floresta no Campo Cerrado (SCOLFORO et. al., 2008).

Scolforo et al. (2008) estimaram o estoque de carbono presente na biomassa aérea da vegetação nativa do estado de Minas Gerais e encontraram os seguintes valores médios por fitofisionomia: Floresta Ombrófila:196,7 Mg ha^{-1}, Floresta Estacional Semidecidual: 151,8

MgC ha^{-1}, Floresta Estacional Decidual: 38,1 MgC ha^{-1}, Cerradão: 31,8 MgC.ha^{-1}, Cerrado *Sensu Stricto*: 14,2 MgC ha^{-1} e Campo Cerrado: 5,3 MgC ha^{-1}.

Essa variação ainda pode ser notada entre fragmentos florestais pertencentes a uma mesma fitofisionomia, já que a capacidade de sequestro é muito dependente do clima local, da topografia, das espécies que compõem o ecossistema, das características do solo, da hidrologia, do histórico de distúrbios na área, entre outros fatores (DAVIDSON; JANSSENS, 2006).

2.2 Métodos de estimativa do estoque de carbono em florestas

Os inventários florestais são considerados uma ferramenta primordial para se conhecer informações sobre potencial produtivo de uma floresta (VIBRANS; GASPER; MÜLLER, 2012). No caso dos inventários em florestas nativas, entre as variáveis normalmente coletadas, o diâmetro medido a 1,30 metros do solo (DAP), a altura total (HT) e a identificação botânica da espécie merecem destaque.

Essas são consideradas as principais variáveis de interesse, pois diversos estudos já descreveram a relação empírica existente entre biomassa e as variáveis DAP, HT e densidade básica da madeira (DBM) (CHAVE et al. 2014). Como as quantidades de carbono estocadas são controladas principalmente pela biomassa, estimativas acuradas da biomassa são fundamentais para estimativas precisas de estoque de carbono pelas florestas.

As equações alométricas, sem dúvidas, estão entre as ferramentas mais utilizadas para a predição da biomassa florestal (RIBEIRO et al., 2011) já que, na maioria das vezes, a biomassa individual das árvores ou mesmo a biomassa do povoamento não pode ser medida diretamente no campo.

Contudo, a base para aplicação da modelagem da biomassa ainda é a amostragem destrutiva (HIGUCHI et al., 2004), onde indivíduos são abatidos com o objetivo de se quantificar, via métodos próprios, a biomassa existente que será utilizada como variável dependente para os ajustes dos modelos matemáticos. Porém, dados de amostragem destrutiva para florestas são caros e difíceis de serem obtidos (CUBAS et al., 2016).

Por essa razão, muitos autores optam por utilizar fatores de expansão da biomassa, que convertem as estimativas de volume, cuja oferta de equações na literatura é consideravelmente maior, em estimativas de biomassa. Da mesma forma, essas estimativas de biomassa podem ser convertidas em estoque de carbono pelo uso de fatores de conversão da biomassa em carbono. Para esse fim, o Painel Intergovernamental de Mudanças Climáticas (*Intergovernmental Panel*

on Climate Change - IPCC) recomenda o uso do fator 0,47 para a conversão da biomassa em carbono de espécies arbóreas pertencentes a florestas tropicais e subtropicais (IPCC, 2006).

Nas últimas décadas, com o avanço das geotecnologias e da disponibilidade de dados, principalmente de satélites, novos métodos de estimativas do estoque de carbono em florestas vêm sendo desenvolvidos, permitido a incorporação aos modelos de predição usuais de parâmetros detectados remotamente que levam em conta a variabilidade espacial do carbono nos fragmentos florestais (WANG et al. 2009) ou mesmo avaliações indiretas usando unicamente técnicas de sensoriamento remoto (FUCHS et al. 2009).

2.3 Sensoriamento remoto e o estoque de carbono nas florestas

O uso das informações espectrais têm sido de grande importância para estimativa do estoque de carbono para áreas florestais (PONZONI et al., 2015; WERE et al., 2015; WU et al., 2016) pois fornece uma gama de informações distribuídas de forma espacial e temporal das áreas de interesse (CHEN, 2013; LU et al., 2012), são dinâmicos, menos onerosos e mais rápidos quando comparado a métodos tradicionais, com precisão aceitável, além de permitir um monitoramento sem a degradação da floresta (XING, HE e LI, 2014).

Permitem ainda, o uso de imagens multiespectrais e multitemporais, com processamento rápido de grandes quantidades de dados e compatibilidade com sistemas de informação geográfica (SIG), associando as estimativas à classificação de uso e cobertura do solo e a detecção de alterações da superfície em áreas florestais (LU et al., 2005).

Sua aplicação se baseia no uso dos valores de reflectância e de suas combinações que geram os índices de vegetação. Os índices se baseiam nas características espectrais da vegetação verde sadia, diretamente influenciados pelos pigmentos foliares, principalmente a clorofila, que enfatizam a diferença entre a forte absorção da radiação eletromagnética vermelha e a forte dispersão da radiação infravermelha. Assim, o comportamento da absorção da radiação fotossinteticamente ativa é fundamental para várias pesquisas que procuram associar essa informação as variáveis de interesse.

Basicamente os métodos para estimar a biomassa e o estoque de carbono com dados de sensoriamento remoto assumem que as informações do dossel, advindas dos sensores, estão fortemente correlacionadas com as variáveis de interesse (LU et al., 2012). Mas no caso das florestas algumas variáveis podem influenciar a reflectância do dossel, entre elas podem ser citadas a iluminação (ângulo de incidência solar), fechamento do dossel, índice de área foliar, deciduidade, ciclos fenológicos, entre outras.

Segundo Tan et.al. (2007) a integração entre dados de inventário florestal e imagens de satélite em resoluções espaciais consistentes, permite melhorar as estimativas de estoques de carbono para grandes áreas florestais. Sendo essa metodologia especialmente interessante quando associada aos conjuntos de dados de inventários florestais nacionais (WANG et al. 2009).

No Brasil, Lu et. al. (2004), que estimaram a biomassa acima do solo em uma grande área de Floresta Amazônica usando dados do sensor Landsat Thematic Mapper (TM). Seis bandas e vários índices de vegetação foram submetidos a análise da correlação de Pearson, segundo os autores a banda TM5 permitiu gerar índices fortemente correlacionados com os parâmetros da floresta, enquanto os obtidos pelas bandas TM4 e TM3 foram fracamente correlacionados, entre eles destacando o índice de vegetação atmosférica resistente, o índice de vegetação atmosférica e do solo e o índice de vegetação com diferença normalizada.

Nesse sentido, pesquisadores vem desenvolvendo uma variedade de índices de vegetação através de combinações das diferentes bandas dos satélites, principalmente os comprimentos de onda vermelho e infravermelho próximo, buscando uma melhor explicação da abundância ou a atividade da vegetação. Segundo Jensen (2011) são considerados como os principais índices de vegetação:

- SR - Razão Simples:

$$SR = \frac{\rho_{nir}}{\rho_{red}}$$

É dada pela razão entre a radiação refletida na banda do infravermelho próximo (ρ_{nir}) e a radiação refletida na banda do vermelho (ρ_{red}). O SR é sensível a variações da biomassa da vegetação e sobre o índice de área foliar (IAF) principalmente em vegetações de grande biomassa, como as florestas.

- NDVI - Índice de vegetação por diferença normalizada:

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}}$$

Muito semelhante ao SR, sendo mais sensível a mudanças sazonais na vegetação, capaz de reduzir os ruídos por diferenças de iluminação solar, sombras de nuvens, mas muito instável a cor do solo e as suas condições de umidade.

- SAVI - Índice de vegetação ajustado ao solo:

$$SAVI = \frac{(1+L)(\rho_{nir} - \rho_{red})}{\rho_{nir} + \rho_{red} + L}$$

O SAVI foi desenvolvido buscando solucionar o problema de instabilidade do NDVI as características do substrato do dossel, adicionando a formulação o fator L, que quando igual a 0,5 minimiza as variações de brilho dos solos.

- ARVI - Índice de vegetação resistente a atmosfera:

$$ARVI = \left(\frac{\rho^*_{nir} - \rho^*_{rb}}{\rho^*_{nir} + \rho^*_{rb}} \right)$$

$$\rho^*_{rb} = \rho^*_{red} - \gamma\left(\rho^*_{blue} - \rho^*_{red}\right)$$

O ARVI foi construído de forma que fosse menos sensível aos efeitos atmosféricos, corrigindo a reflectância na banda do vermelho pela diferença entre a reflectância na banda do azul e do vermelho. O valor de γ normalmente é igual a 1,0 e representa o efeito aerossol da atmosfera.

- SARVI - Índice de vegetação resistente a atmosfera e ao solo:

$$SARVI = \frac{\rho^*_{nir} - \rho^*_{rb}}{\rho^*_{nir} + \rho^*_{rb} + L}$$

O SARVI deriva de uma junção do SAVI e do ARVI, corrigindo tanto os ruídos devido aos solos como os derivados da atmosfera.

- EVI - Índice de vegetação realçado:

$$EVI = G\frac{\rho^*_{nir} - \rho^*_{red}}{\rho^*_{nir} + C_1\rho^*_{red} + C_2\rho^*_{blue} + L}(1+L)$$

O EVI é uma modificação do NDVI, contendo um fator de ajuste para solos (*L*), e dois coeficientes (C_1 e C_2), que descrevem o uso da banda azul para a correção da banda vermelha quanto ao espalhamento atmosférico por aerossóis. Esses fatores assumem os respectivos valores: 6,0; 7,5 e 1,0. Além desses, o EVI conta com um fator *G* com valor fixo de 2,5 que visa melhorar a sensibilidade para regiões de alta biomassa, permitindo um melhor desempenho do monitoramento da vegetação através da diminuição da influência do substrato abaixo do dossel e através da redução da influência atmosférica.

- TVI - Índice de vegetação triangular:

$$TVI = 0,5\left(120\left(\rho_{nir} - \rho^*_{green}\right)\right) - 200\left(\rho_{red} - \rho^*_{green}\right)$$

O TVI descreve a energia radiativa absorvida pelos pigmentos como a diferença relativa entre as reflectâncias na banda do vermelho e no infravermelho próximo com a reflectância na banda do verde, onde a absorção de luz pela clorofila é insignificante.

- VARI - Índice de vegetação resistente à atmosfera no visível:

$$VARI_{green} = \frac{\rho_{green} - \rho_{red}}{\rho_{green} + \rho_{red} - \rho_{blue}}$$

O VARI deriva do ARVI sendo um índice muito pouco sensível aos efeitos atmosféricos, utilizado principalmente para estudos sobre frações da vegetação.

2.4 Seleção de variáveis para a estimativa do estoque de carbono

O primeiro passo para o desenvolvimento de métodos alternativos para a estimativa do estoque de carbono é identificar variáveis que podem ser incorporadas ao conjunto de variáveis utilizadas classicamente. Essa identificação passa pelo estabelecimento da relação existente entre o estoque de carbono e as variáveis disponíveis, já que muitos dessas variáveis são totalmente irrelevantes ou redundantes.

A forma mais simples de realizar essa avaliação é utilizando a análise de correlação. Os coeficientes de correlação são métodos estatísticos utilizados para medir as relações entre variáveis e o que elas representam, procurando identificar se existe alguma relação entre a variabilidade de ambas e quantificando esse grau de correlação (MOORE et. al. 2007). Na área científica os coeficientes de correlação são muito importantes quando se pretende avaliar muitas variáveis, pois assim, é possível entender como a variabilidade de uma afeta a outra.

Entre os coeficientes de correlação, o coeficiente de correlação de Pearson (r) explica a relação entre duas variáveis através de valores situados entre -1 e 1. Quando o coeficiente se aproxima de 1, existe um aumento no valor de uma variável quando a outra também aumenta, ou seja, há uma relação linear positiva. Se o coeficiente se aproxima de -1, o valor de uma variável aumenta quando o da outra diminui, existindo uma correlação negativa ou inversa. E para coeficientes de correlação próximos de zero não há relação entre as duas variáveis.

Assim, modelos matemáticos que utilizam variáveis independentes altamente correlacionadas à variável dependente tendem a apresentar um melhor ajuste (FUJIWARA et al., 2009). Sendo que a forma tradicional de relacionar várias variáveis em um modelo se dá através da Regressão Linear Múltipla (RLM).

Como forma de selecionar as variáveis que melhor se adequa para a essa modelagem, normalmente é aplicando o procedimento *Stepwise*, que adota como critério de seleção a estatística F e o valor do Critério de Informação de Akaike (AIC). O método *Stepwise* é feito de forma iterativa, adicionando e removendo variáveis, a partir de um critério de seleção (ALVES; LOTUFO; LOPES, 2013), procedimento esse que permite avaliar a contribuição de cada variável independente dentre as existentes no modelo, selecionando aquelas significativas estatisticamente e que explicam da melhor forma a variável dependente.

Esse procedimento, contudo, não leva em consideração a existência de multicolinearidade entre as variáveis independentes, problema esse que afeta a estimativa dos parâmetros do modelo. Para solucionar esse problema, recomenda-se o diagnóstico por meio do VIF (*Variance Inflation Factor*), que indica a presença de multicolinearidade em uma variável quando o seu valor é maior que 10, indicando assim o descarte da variável em questão do modelo.

A grande limitação dos modelos lineares múltiplos é que eles só se adequam a variáveis que tem correlação linear entre si, o que na área biológica nem sempre ocorre. Diante do aumento da complexidade dos fenômenos a serem modelados, houve a necessidade do emprego de métodos mais robustos que lidasse com natureza linear ou não. Assim, outras técnicas também vêm sendo desenvolvidas, como por exemplo o uso de abordagens de inteligência artificial e aprendizado de máquinas, como o *Randon Forest* e o Algoritmo Genético, que apresentam maior flexibilidade quanto a relação existente entre as variáveis.

2.5 Algoritmo Genético

Criado por John Holland na década de 60, o algoritmo genético (AG) é um método de otimização que utiliza os princípios da teoria da evolução de Darwin na busca da solução ótima de um determinado problema (YU; XU, 2014).

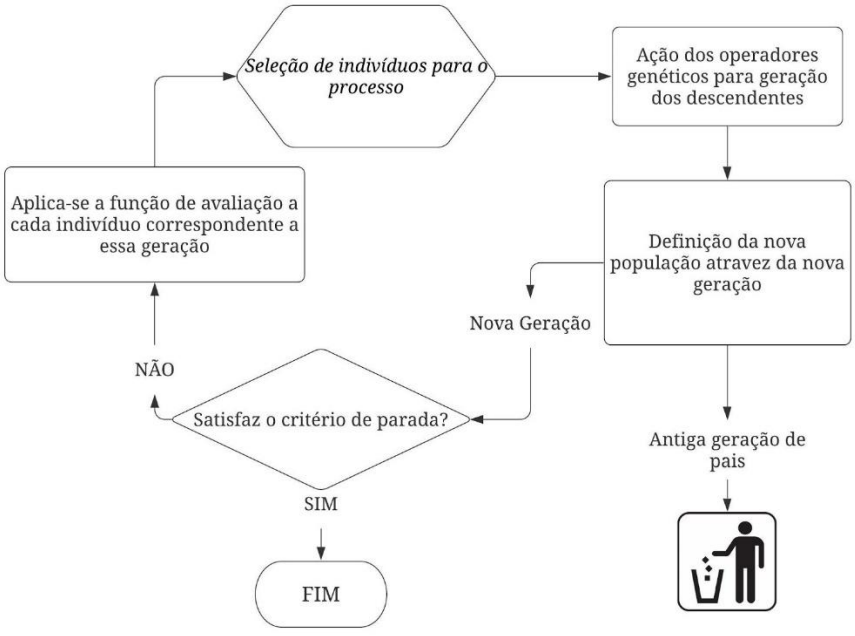
Segundo Garg (2016) e Yang et al. (2008) o AG é um algoritmo de busca heurística versátil, apesar de apresentar baixa taxa de convergência, autodidata e robusto na busca de soluções ótimas globais, particularmente em problemas multiobjectivo, funcionando

corretamente mesmo quando à em seus parâmetros de entrada apresentam ruídos ou estejam minimamente alterados (DAS et al. 2018).

Em sua busca pela solução ótima o AG considera inicialmente várias soluções individuais, gerando uma população dentro da qual serão realizados testes de convergência de acordo com o intuito da pesquisa. A partir dessa população inicial, e considerando a regra de sobrevivência/reprodução do mais apto, o algoritmo aplica operadores genéticos nas melhores soluções encontradas, afim de produzir melhores descendentes, realizando uma busca local na vizinhança em busca de novas e melhores soluções, eliminando as piores de geração em geração.

Assim, a estrutura geral do AG pode ser descrita da seguinte maneira: codificação, seleção da população inicial, avaliação da melhor resposta, operadores genéticos (*crossover* e mutação), geração de descendentes e análise da nova geração de acordo com o critério de parada, caso a nova geração não seja satisfatória, aplica-se novamente a função de avaliação para cada indivíduo (Figura 1).

Figura 1 - Fluxograma da estrutura de um Algoritmo Genético Simples.



Fonte: Adaptado de Linden (2012).

Latifi, Nothdurft e Koch (2010) encontraram resultados superiores para o AG quando comparado ao método *stepwise* na seleção de seleção de variáveis, para a modelagem da biomassa e volume de uma floresta, a partir do sensoriamento remoto.

2.6 Random Forest

O *Random Forest* (RF) é um Algoritmo de aprendizado de máquina criado pelo matemático Leo Breiman, muito utilizado para problemas de regressão e classificação, (BREIMAN, 2001). Ele permite uma modelagem flexível de interações em conjuntos de dados de altas dimensões, criando um grande número de árvores de regressão, a partir de um subconjunto de amostras de treinamento, e calculando a média de suas previsões (WAGER; ATHEY, 2018)

Kinoshita et al. (2016) compararam o algoritmo RF com dois interpoladores geoestatísticos para a estimativa do estoque de carbono no solo em função de dados de reflectância, com base no erro médio reduzido (RME) o RF se mostrou o método capaz de gerar as melhores estimativas espaciais do estoque de carbono, apoiando as afirmações da literatura sobre o seu bom desempenho.

O algoritmo necessita de três parâmetros para a produção das árvores de decisão, que são: *ntree* (número de árvores treinadas na floresta), *nodesize* (tamanho do nó do terminal de destino) e *mtry* (número de saídas aleatórias usadas para dividir um nó da árvore) (O'BRIEN; ISHWARAN, 2019).

Com base nesses valores, o algoritmo divide a base de dados em dois subconjuntos por *bootstrap*, um subconjunto aleatório de dois terços das observações usadas para treinar as árvores (ensacamento), e o terço restante dos dados (fora do saco) são usados para a validação (WOZNICKI et al., 2019).

Após isso, se define de forma randômica o número de nós das árvores de decisão, onde são selecionados números de variáveis aleatoriamente, e a variável que apresentar a melhor divisão é selecionada para aumentar a árvore nesse nó (WOZNICKI et al., 2019).

Para a avaliação de cada nó são utilizadas métricas. No caso dos problemas de regressão a métrica de avaliação é o *mean squared error* (MSE, erro quadrático médio), já para problemas de classificação é utilizado o critério de Gini (WOZNICKI et al., 2019). Assim, a árvore de decisão que recebeu o menor valor de MSE ou que recebeu mais votos na classificação é a resposta final do algoritmo (BELGIU; DRĂGU, 2016).

Apesar do RF apresentar resultados satisfatórios em relação a bancos de dados volumoso e com grande número de variáveis, ele tem pouca precisão quando se trata de conjunto de dados complexos (grandes e com interações variáveis complexas) (SPEISER et al.,

2019), incluindo na maioria das vezes variáveis com características distintas, que influencia o desempenho da regressão ou dos classificadores (KUMAR; SHAIKH, 2017).

Por isso, a importância de fornecer o melhor conjunto de atributos possível para maximizar o desempenho do algoritmo e minimizar a exigência computacional (KUMAR; SHAIKH, 2017). Outras formas de reduzir essa complexidade são através de usos dos métodos de *feature selection* (GHAEMI; FEIZI-DERAKHSHI, 2016), métodos recursivos de redução de variáveis (ABDOH; ABO RIZKA; MAGHRABY, 2018) ou Análise dos componentes principais - PCA (GEETHA et al., 2019).

Mais recentemente a hibridização entre algoritmo genético e *Random Forest* (GARF), que veem trazendo ótimos resultados na seleção de variáveis e melhorando a capacidade preditiva e de classificação do RF (ALIČKOVIĆ; SUBASI, 2017; CERRADA et al., 2016; HONG et al., 2018; NAGHIBI; AHMADI; DANESHI, 2017; PAUL et al., 2017).

2.7 Hibridação do Algoritmo Genético e Random Forest (GARF)

Como apresentado anteriormente, o RF permite retirar informações mais relevantes na predição da variável resposta, podendo ser inclusive implementado para a seleção de variáveis. Esse ponto é especialmente interessante visando avaliar em grandes bancos de dados quais informações são importantes para a predição da variável de interesse e quais não tem potencial.

Por sua vez, o AG é um método versátil na busca de soluções ótimas, funcionando de maneira adequada mesmo quando os parâmetros de entrada apresentam ruídos. Assim, uma alternativa que vêm sendo difundida é a hibridação desses dois métodos.

Bader-El-Den e Gaber, (2012) evidenciaram o potencial da hibridação na melhoria da performance do RF, uma vez que, associado ao algoritmo genético para mudar dinamicamente as árvores na floresta, resultou no aumento da precisão do modelo. A excelência do GARF para seleção de características também é comprovada nos estudos de Cerrada et al., (2016), Crisman et al. (2016), Ma e Fan (2017), Paing e Choomchuay (2018).

3 CONSIDERAÇÕES FINAIS

Dessa forma, a avaliação do método híbrido GARF em comparação com os métodos clássicos de regressão para a estimativa do estoque de carbono, pode produzir novas formas de avaliar esse serviço ecológico, permitindo a incorporação de variáveis espectrais, geográficas e ou hidrológicas, visando ganhos na precisão de tais estimativas.

REFERÊNCIAS

ABDOH, S. F.; ABO RIZKA, M.; MAGHRABY, F. A. Cervical cancer diagnosis using random forest classifier with SMOTE and feature reduction techniques. **IEEE Access**, v. 6, p. 59475–59485, 2018.

ALIČKOVIĆ, E.; SUBASI, A. Breast cancer diagnosis using GA feature selection and Rotation Forest. **Neural Computing and Applications**, v. 28, n. 4, p. 753–763, 2017.

ALVES, M. F.; LOTUFO, A. D. P.; LOPES, M. L. M. Seleção de variáveis stepwise aplicadas em redes neurais artificiais para previsão de demanda de cargas elétricas. **Proceeding Series of the Brazilian Society of Computational and Applied Mathematics**, v. 1, n. 1, 2013.

BACCINI, A. et al. A first map of tropical Africa's above-ground biomass derived from satellite imagery. **Environmental Research Letters**, Bristol, v. 3, n. 4, 2008.

BADER-EL-DEN, M.; GABER, M.. Garf: towards self-optimised random forests. In: **International conference on neural information processing**. Springer, Berlin, Heidelberg, p. 506-515, 2012.

BELGIU, M.; DRĂGU, L. Random forest in remote sensing: A review of applications and future directions. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 114, p. 24–31, 2016.

BOLFE, É. L.; BATISTELLA, M.; FERREIRA, M. C.. Correlação de variáveis espectrais e estoque de carbono da biomassa aérea de sistemas agroflorestais. **Pesquisa Agropecuária Brasileira**, v. 47, n. 9, p. 1261–1269, 2012.

BREIMAN, L. Random forest. **Machine Learning**, Boston, v. 45, p. 5-32, 2001.

BROWN, S.; LUGO, A. E. Aboveground biomass estimates for tropical moist forests of the Brazilian Amazon. **Interciencia**. Caracas, v. 17, n. 1, p. 8-18, 1992.

CERRADA, M. et al. Fault diagnosis in spur gears based on genetic algorithm and random forest. **Mechanical Systems and Signal Processing**, v. 70–71, p. 87–103, 2016.

CHAVE, Jérôme et al. Improved allometric models to estimate the aboveground biomass of tropical trees. **Global Change Biology**, v. 20, n. 10, p. 3177–3190, 2014.

CHEN, Q. Lidar remote sensing of vegetation biomass. In: WENG, Q.; WANG, G. (Ed.). **Remote sensing of natural resources**. Boca Raton: CRC Press; Taylor &Francis Group, 2013. p. 399-420.

CRISMAN, T. J. et al. Identification of an efficient gene expression panel for glioblastoma classification. **PLoS ONE,** v. 11, n. 11, p. 1–19, 2016.

CUBAS, R.; et al. Carbon contents and modelling of total organic carbon for Pinus taeda L. from natural regeneration. **Revista Árvore**, v. 40, n. 4, p. 661-668, 2016.

DA SILVA FILHO, R. et al. Representação matemática do comportamento intra-anual do NDVI no Bioma Caatinga**. Ciência Florestal**, v. 30, n. 2, p. 473-488, 2020

DAS, A. K. et al. Prediction of fine particulate matter chemical components with a spatio-temporal model for the Multi-Ethnic Study of Atherosclerosis cohort. **Applied Energy**, v. 26, n. 5, p. 499–523, 2018.

DAVIDSON, E. A.; JANSSENS, I. A. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. **Nature**, v. 440, n. 7081, p. 165-173, 2006.

DUBE, T.; MUTANGA, O. The impact of integrating WorldView-2 sensor and environmental variables in estimating plantation forest species aboveground biomass and carbon stocks in uMgeni Catchment, South Africa. ISPRS **Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 119, p. 415-425, Sept. 2016.

FANG, J. et al. Forest biomass carbon sinks in East Asia, with special reference to the relative contributions of forest expansion and forest growth. **Global Change Biology**, v. 20, n. 6, p. 2019–2030, 2014.

FROLKING, S. et al. Forest disturbance and recovery: A general review in the context of spaceborne remote sensing of impacts on aboveground biomass and canopy structure. **Journal of Geophysical Research: Biogeosciences**, v. 114, n. 3, 2009.

FUCHS, H. et al. Estimating aboveground carbon in a catchment of the Siberian forest tundra: Combining satellite imagery and field inventory. **Remote Sensing of Environment**, v. 113, n. 3, p. 518-531, 2009.

FUJIWARA, Koichi et al. Soft-sensor development using correlation-based just-in-time modeling. **AIChE Journal**, v. 55, n. 7, p. 1754-1765, 2009.

GARG, H. A hybrid GSA-GA algorithm for constrained optimization problems. **Information Sciences**, v. 478, p. 292–305, 2016.

GEETHA, R. et al. Cervical cancer identification with synthetic minority oversampling technique and PCA analysis using random forest classifier. **Journal of medical systems**, v. 43, n. 9, p. 286, 2019.

GHAEMI, M.; FEIZI-DERAKHSHI, M. R. Feature selection using Forest Optimization Algorithm**. Pattern Recognition**, v. 60, p. 121–129, 2016.

HANSEN, M C et al. Supplementary Materials for High-Resolution Global Maps of 21st-Century Forest Cover Change. **Science**, v. 342, n. 6160, p. 850–853, nov. 2013.

HIGUCHI, N., et al.. Dinâmica e balanço do carbono da vegetação primária da Amazônia Central. **Floresta**, v. 34, n. 3, 2004.

HONG, H. et al. Applying genetic algorithms to set the optimal combination of forest fire related variables and model forest fire susceptibility based on data mining models. The case of Dayu County, China. **Science of the Total Environment**, v. 630, p. 1044–1056, 2018.

HOUGHTON, R. A. Aboveground forest biomass and the global carbon balance. **Global Change Biology**, v. 11, n. 6, p. 945–958, 2005.

INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE - IPCC. **Guidelines for national greenhouse gas inventories:** agriculture, forestry and other land use. Japan: Institute for global environmental strategies (IGES), 2006. v. 4. Disponível em: <http://www.ipcc-nggip.iges.or.jp/public/2006gl/vol4.html>. Acesso em: 1 Junho 2020.

JENSEN, J. R. **Sensoriamento remoto do ambiente:** uma perspectiva em recursos terrestres. Parêntese, 2011.

KINOSHITA, Rintaro et al. Large topsoil organic carbon variability is controlled by Andisol properties and effectively assessed by VNIR spectroscopy in a coffee agroforestry system of Costa Rica. **Geoderma**, v. 262, p. 254-265, 2016.

KUMAR, S. S.; SHAIKH, T. Empirical Evaluation of the Performance of Feature Selection Approaches on Random Forest. **2017 International Conference on Computer and Applications**, ICCA 2017, p. 227–231, 2017

LATIFI, Hooman; NOTHDURFT, Arne; KOCH, Barbara. Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: application of multiple optical/LiDAR-derived predictors. **Forestry**, v. 83, n. 4, p. 395-407, 2010.

LIBERATO, A. Marcolino. Estimativa do albedo e índice de área foliar na Amazônia. Revista Brasileira de Geografia Física, v. 4, n. 1, p. 22-32, 2011.

LINDEN, R. **Algoritmos Genéticos**. 3a Edição ed. Rio de Janeiro: Ciência Moderna, 2012.

LU, D. et al. Relationships between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon Basin. **Forest Ecology and Management**, Amsterdam, v. 198, p. 149-167, 2004.

LU, D. et al. Aboveground Forest Biomass Estimation with Landsat and LiDAR Data and Uncertainty Analysis of the Estimates, **International Journal of Remote Sensing**, v. 26, n. 12, p. 2509–2525, 2005.

LU, D. et al. Aboveground forest biomass estimation with Landsat and LiDAR data and uncertainty analysis of the estimates. **International Journal of Forestry Research**, v. 2012, 2012.

MA, L.; FAN, S. CURE-SMOTE algorithm and hybrid algorithm for feature selection and parameter optimization based on random forests. **BMC Bioinformatics**, v. 18, n. 1, p. 1–18, 2017.

MELLO, Juliana Milesi et al. Dinâmica dos atributos Físico-químicos e variação sazonal dos estoques de carbono no solo em diferentes fitofisionomias do Pantanal Norte Mato-grossense. **Revista Árvore** p. 325–336, 2008.

MIGUEL, Eder Pereira et al. Redes neurais artificiais para a modelagem do volume de madeira e biomassa do cerradão com dados de satélite. **Pesquisa Agropecuária Brasileira**, v. 50, n. 9, p. 829–839, 2015.

MOORE, David S., and Stephane Kirkland. The basic practice of statistics. **New York: WH Freeman**, v. 2, p. 100-101, 2007.

NAGHIBI, S. A.; AHMADI, K.; DANESHI, A. Application of Support Vector Machine, Random Forest, and Genetic Algorithm Optimized Random Forest Models in Groundwater Potential Mapping. **Water Resources Management**, v. 31, n. 9, p. 2761–2775, 2017.

NETO, P.P. Utilização da videografia aérea na detecção de áreas com deficiências nutricionais em plantios de eucalipto. 2002. 86 f. **Dissertação (Mestrado em Ciências Florestais)** - Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba, SP, 2002.

O’BRIEN, R.; ISHWARAN, H. A random forests quantile classifier for class imbalanced data. **Pattern Recognition**, v. 90, p. 232–249, 2019

PAING, M. P.; CHOOMCHUAY, S. Improved Random Forest (RF) classifier for imbalanced classification of lung nodules. **ICEAST 2018 - 4th International Conference on Engineering, Applied Sciences and Technology**: Exploring Innovative Solutions for Smart Society, n. i, p. 1–4, 2018.

PAN, Y.et al. A large and persistent carbon sinks in the world's forests. **Science**, v. 333, n. 6045, p. 988-993, 2011.

PAUL, D. et al. Feature selection for outcome prediction in oesophageal cancer using genetic algorithm and random forest classifier. **Computerized Medical Imaging and Graphics**, v. 60, p. 42–49, 2017.

PONZONI, F. J. et al. Caracterização espectro-temporal de dosséis de *Eucalyptus* spp. mediante dados radiométricos TM/Landsat 5. **Cerne**, Lavras, v. 21, n. 2, p. 267-275, 2015.

RIBEIRO, S. C. et al. Above-and belowground biomass in a Brazilian Cerrado. **Forest Ecology and Management,** v. 262, n. 3, p. 491-499, 2011.

ROCHA, Genival Fernandes et al. Detecção de desmatamentos no bioma Cerrado entre 2002 e 2009; Padrões, tendências e impactos. [S.l.]: **Sociedad Brasileira de Cartografia, Geodésia, Fotogrametria e Sensoriamento Remoto**, 2011. v. 0.

ROSA, Anderson Melo et al. Monitoramento de *Thaumastocoris peregrinus* (Hemiptera: thamascoridae) em plantios de *Eucalyptus* usando sensoriamento remoto e abordagem PLS-DA, **Anais do XIX Simpósio Brasileiro de Sensoriamento Remoto**, p. 3629–3632, 2008.

SADER, Steven A.; AHL, Douglas; LIOU, Wen-Shu. Accuracy of Landsat-TM and GIS rule-based methods for forest wetland classification in Maine. **Remote Sensing of Environment**, v. 53, n. 3, p. 133-144, 1995.

SCOLFORO, J. R. OLIVEIRA. A.D.; ACERBI JUNIOR, F.W.A. Equações para estimar o volume de madeira das fisionomias, em Minas Gerais. In: SCOLFORO, J. R.; OLIVEIRA, A. D.; ACERBI JÚNIOR, F. W. **Inventário florestal de Minas Gerais: equações de volume, peso de matéria seca e carbono para diferentes fisionomias da flora nativa**. Lavras: UFLA,. p. 1-65, 117-128, 181-194, 2008.

SCOLFORO, H. F. et al. Spatial distribution of aboveground carbon stock of the arboreal vegetation in Brazilian biomes of Savanna, Atlantic Forest and Semi-Arid Woodland. **Plos One**, San Francisco, v. 10, n. 6, p. 1-20, 2015.

SEIDEL, D. et al. Review of ground-based methods to measure the distribution of biomass in forest canopies. **Annals of Forest Science**, Les Ulis, v. 68, n. 2, p. 225- 244, 2011.

SILVEIRA, E. M. O. et al. Estimating aboveground biomass loss from deforestation in the savanna and semi-arid biomes of Brazil between 2007 and 2017. In: **Tropical forests in transition: the role of deforestation and impacts from community composition to regional climate change**. London: IntechOpen, 2019a.

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**, Enschede, v. 78, p. 175-188, June 2019b.

SPEISER, J. L. et al. A comparison of random forest variable selection methods for classification prediction modeling. **Expert Systems with Applications**, v. 134, p. 93–101, 2019.

STEPHENS, B. B. et al. Weak northern and strong tropical land carbon uptake from vertical profiles of atmospheric CO2. **Science**, v. 316, n. 5832, p. 1732-1735, 2007.

TAN, K. et al. Satellite-based estimation of biomass carbon stocks for northeast China's forests between 1982 and 1999. **Forest ecology and management**, v. 240, n. 1-3, p. 114-121, 2007.

TESFAYE, M. A. et al. Impact of changes in land use, species and elevation on soil organic carbon and total nitrogen in Ethiopian Central Highlands. **Geoderma**, v. 261, p. 70-79, 2016.

VIBRANS, Alexander Christian et al. **Diversidade e conservação dos remanescentes florestais**. Blumenau: Edifurb, 2012.

XING, M. et al. An extended approach for biomass estimation in a mixed vegetation area using ASAR and TM data. **Photogrammetric Engineering and Remote Sensing**, v. 80, n. 5, p. 429–438, 2014.

WAGER, S.; ATHEY, S. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. **Journal of the American Statistical Association**, v. 113, n. 523, p. 1228–1242, 2018.

WALTZLAWICK, L.F. et al. Fixação de carbono em floresta ombrófila mista em diferentes estágios de regeneração. In: SANGUETA, C. R. et al. (Ed.). **As florestas e o carbono.** Curitiba: 2002. p. 153-174.

WANG, G. et al. Mapping and spatial uncertainty analysis of forest vegetation carbon by combining national forest inventory data and satellite images. **Forest Ecology and Management**, v. 258, n. 7, p. 1275-1283, 2009.

WANG, G. et al. Uncertainties of mapping aboveground forest carbon due to plot locations using national forest inventory plot and remotely sensed data. **Scandinavian Journal of Forest Research**, Stockholm, v. 26, n. 4, p. 360-373, 2011.

WERE, K. et al. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. **Ecological Indicators**, London, v. 52, p. 394-403, 2015.

WOZNICKI, S. A. et al. Development of a spatially complete floodplain map of the conterminous United States using random forest. **Science of the Total Environment**, v. 647, p. 942–953, 2019.

WU, C. et al. Comparison of machine-learning methods for above-ground biomass estimation based on Landsat imagery. **Journal of Applied Remote Sensing**, Orlando, v. 10, n. 3, p. 1-18, 2016.

YANG, H. et al. Optimal sizing method for stand-alone hybrid solar-wind system with LPSP technology by using genetic algorithm. **Solar Energy**, v. 82, n. 4, p. 354–367, 2008.

YU, F.; XU, X. A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network. **Applied Energy**, v. 134, p. 102–113, 2014.

CAPÍTULO II – ARTIGO

Modelagem do estoque de carbono a partir de variáveis dendrométricas, hidrológicas e do sensor Sentinel em um fragmento de floresta semidecídua

Artigo formatado conforme a NBR 6022 (ABNT, 2003) e adaptado às exigências do manual de normalização de trabalhos acadêmicos da UFLA

RESUMO

As florestas nativas são importantes para uma infinidade de serviços ecossistêmicos, entre eles o sequestro e estoque de carbono. Dessa forma, conhecer o estoque de carbono existente nos remanescentes florestais é de grande importância para ajudar a justificar sua preservação bem como a recuperação de áreas degradadas. O objetivo desse trabalho foi avaliar o uso de informações espectrais e índices de vegetação obtidos do satélite SENTINEL-2, variáveis hidrológicas (Precipitação interna e Armazenamento de água no solo) e variáveis geográficas em conjunto com informações dendrométricas do povoamento para a estimativa do estoque de carbono em diferentes estratos do dossel de um remanescente florestal pertencente a fitofisionomia Floresta Estacional Semidecidual. Para isso foram empregadas técnicas de aprendizado de máquinas (*Random forest* associado a meta-heurística Algoritmo genético - GARF) em comparação com a modelagem clássica por meio da Regressão Linear Múltipla p RLM utilizando o método *Stepwise*, para a seleção das variáveis mais indicadas para incrementar o poder preditivo dos modelos. Os resultados indicam que com exceção das variáveis dendrométricas, as variáveis analisadas apresentarem baixa correlação com estoque de carbono, apesar disso, todas elas contribuíram de alguma forma para a melhoria das estimativas de carbono nos diferentes estratos do dossel. O método GARF privilegiou o uso dos dados espectrais e dendrometricos para os percentis superiores e os dados dendrométricos e hidrológicos para os inferiores, enquanto o RLM privilegiou o uso conjunto das variáveis espectrais, dendrometricas, hidrológicas e geográficas para os percentis superiores e dendrométricos e espectrais para os inferiores. O método GARF, com exceção do percentil 90, foi capaz de produzir melhores resultados em comparação com o RLM, apesar de as diferenças não serem grandes. Dessa forma, o uso conjunto dessas variáveis se mostra promissor para a produção de estimativas mais confiáveis.

Palavras-chave: Modelagem, Seleção de variáveis, SENTINEL-2

1 INTRODUÇÃO

Devido ao importante papel que as florestas desempenham no armazenamento de carbono através da sua biomassa (TESFAYE, 2016), esse tema tem sido considerado de grande relevância pelos estudos ambientais e utilizados como argumentos para justificar a necessidade da preservação desses ecossistemas para fins de minimizar os efeitos de mudanças climáticas (DA SILVA FILHO et al., 2020).

Nesse cenário, os estudos científicos que visam quantificar o estoque de carbono existentes nos remanescentes florestais se tornaram frequentes. Normalmente eles são realizados empregando-se modelos alométricos, que estimam o estoque de carbono de forma individual ou para o povoamento, utilizando informações dendrométricas coletadas por meio do inventário florestal, como diâmetro medido a 1,30 do solo (DAP), área basal (G), entre outras (RIBEIRO et al., 2011).

Contudo, os inventários florestais são operações que exigem grande volume de recursos financeiros e logísticos, principalmente para grandes áreas florestais. Dessa forma, é de interesse científico desenvolver formas alternativas para realizar a estimativa do estoque de carbono contido nos remanescentes florestais, associando informações obtidas remotamente que permitam uma redução da intensidade amostral ou mesmos reduzindo os erros produzindo estimativas mais confiáveis.

Nesse caso, os dados de sensoriamento remoto são sem dúvida os mais explorados e que trouxeram grandes benefícios aos estudos envolvendo os recursos florestais. Através do estudo das imagens obtidas por sensores acoplados a satélites é possível fazer inferências espaciais de maneira precisa sobre padrões das florestas com base em informações do dossel.

Isso é possível devido ao fato de os pigmentos foliares, em especial as clorofilas, terem absorções preferenciais em determinados comprimentos de ondas do espectro óptico, o que permitiu o desenvolvimento de vários índices de vegetação, que podem ser correlacionados às mais diferentes informações sobre a distribuição dos recursos florestais contribuindo para a sua estimativa, dentre elas, o estoque de carbono (BOLFE; BATISTELLA; FERREIRA, 2012).

Da mesma forma, o uso de informações hidrológicas relacionadas ao interior da floresta, como o armazenamento de água no solo (ARM) e a precipitação interna (PI), vem sendo associadas a padrões ecológicos e de dinâmica do dossel, influenciando os padrões de crescimento e dispersão dos indivíduos dentro de um povoamento, sendo dessa forma, capazes de influenciar variáveis dendrométricas (MELLO et al., 2008).

Assim, sendo a dinâmica do estoque de carbono influenciada por uma grande gama de variáveis, da mesma forma, sua estimativa pode ser influenciada por um grande número de variáveis de forma isolada ou conjunta. Variáveis dendrometricas, espectrais, hidrológicas e geográficas podem auxiliar na melhoria das estimativas do estoque de carbono em diferentes estratos do dossel da floresta, captando a heterogeneidade da cobertura vegetal. Logo, a seleção das variáveis a serem utilizadas pelos modelos de predição pode ser um processo complexo que exija o uso de métodos iterativos.

Diante do exposto, o presente trabalho tem como objetivo, avaliar o uso de informações espectrais e hidrológicas em conjunto com informações dendrométricas para a estimativa do estoque de carbono em diferentes estrados do dossel de um remanescente florestal pertencente a fitofisionomia Floresta Estacional Semidecidual, empregando técnicas de aprendizado de máquinas (*Random forest* associado a meta-heurística Algoritmo genético) em comparação com a modelagem clássica por meio da Regressão Linear Múltipla com a seleção das variáveis pelo métido *Stepwise*.

2 MATERIAL E MÉTODOS

2.1 Área de estudo

O estudo foi realizado em um remanescente de Mata Atlântica, pertencente a fitofisionomia Floresta Estacional Semidecidual. Localizada no município de Lavras, Minas Gerais - Brasil, nas coordenadas 21°13'40" S e 44°57'50" W, o fragmento em questão apresenta área total de 6,1 ha. O relevo local é levemente ondulado, com declividade variando ente 5 e 15% (JUNIOR et al., 2017), onde a altitude varia entre 942 e 958 m. O clima da região é do tipo Cwb (Köppen), com inverno seco e verão temperado. As médias anuais de precipitação e temperatura são de, respectivamente, 1.529,5 mm e 19,3°C (mínima de 15,5 °C em julho e máxima de 21,5°C em janeiro), com 80% das chuvas concentradas de outubro a março, enquanto a estação seca se estende de abril a setembro (ALVARES et al., 2013).

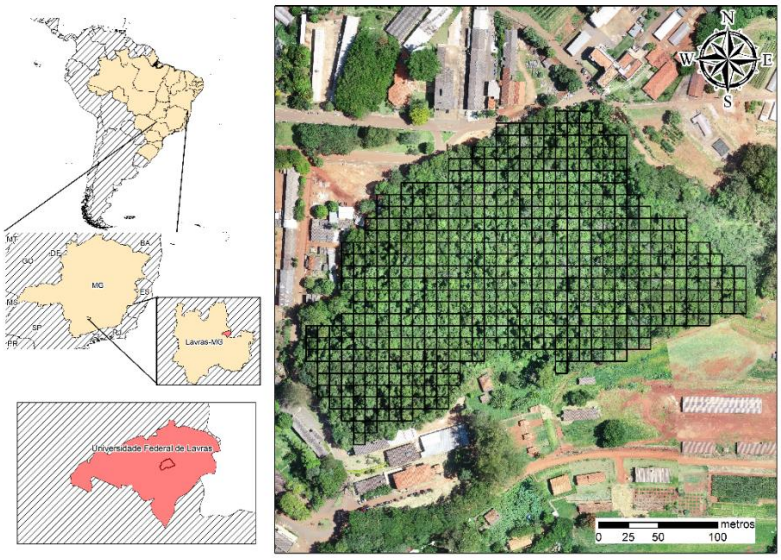
A área se encontra em estágio avançado de regeneração, sem históricos de corte raso desde a década de 1920. Após sofrer perturbações no passado, como exploração seletiva de madeira e pastoreio por gado no interior da mata, a área foi declarada como Reserva Florestal no ano de 1986 e desde então se mantém sem maiores perturbações (NUNES et al. 2003; OLIVEIRA-FILHO et al., 1997). O dossel da floresta é bastante denso, formado por um estrato

superior (copas de árvores isoladas com mais de 20 m de altura), um estrato médio (copas das árvores entre 12 e 15 m de altura) e o sub-bosque (copas das árvores com menos de 10 m de altura). Algumas aberturas no dossel (clareiras) são encontradas na área, causadas pela queda de árvores que morrem. A natureza semidecídua da floresta resulta em um comportamento sazonal do dossel, onde aproximadamente 50% das árvores perdem folhas no período mais seco do ano. Este comportamento deve ser levado em consideração na análise de imagens de satélite, pois podem afetar a observação da variável de interesse ou mesmo o cálculo dos índices de vegetação.

2.2 Distribuição espacial dos dados de carbono

A área da foi dividida em 546 parcelas de 100m², (10 × 10m), cobrindo 86,5% da área desconsiderando as bordas da mata (Figura 1). As parcelas foram submetidas a inventário florestal censitário em 2017, onde todos os indivíduos arbóreos com diâmetro medido a 1,30 metros do solo - DAP maior que 5 cm foram mensurados, identificadas botanicamente e coletadas as coordenadas (X e Y), permitindo a localização espacial dos mesmos dentro da floresta.

Figura 1 - Mapa com delimitação e distribuição das parcelas na área de estudo.



Fonte: Do autor (2020).

Os dados de carbono foram calculados utilizando os dados coletados pelo inventário, primeiramente foi calculado a biomassa acima do solo por indivíduo - AGB, utilizando a equação desenvolvida por Chave et al. (2014), conhecida como equação pantropical. Por ela,

AGB (kg) foi estimada em função da densidade básica da madeira - DBM (g cm^{-3}), DAP (cm) e do parâmetro de estresse bioclimático - E (AIC = 47, RSE = 0,243, df = 3998).

Esses cálculos foram realizados utilizando o pacote BIOMASS no software estatístico R (REJOU-MECHAIN et al., 2018). Nele, a determinação da DBM é feita com base no valor médio da espécie, gênero ou família, a partir do banco de dados global compilado por Chave et al. (2009). Por sua vez o parâmetro *E* foi extraído a partir das coordenadas geográficas do local, com o propósito de estimar por meio da rotina do pacote BIOMASS a altura das árvores em função da sazonalidade da temperatura e da precipitação.

As estimativas de biomassa foram convertidas em estoque de carbono pelo uso do fator de conversão para espécies arbóreas pertencentes a florestas tropicais e subtropicais proposto pelo Painel Intergovernamental de Mudanças Climáticas (*Intergovernmental Panel on Climate Change* - IPCC) de 0,47 (IPCC, 2006). Dessa forma, foram calculadas para cada parcela as variáveis estoque de carbono - C ($Mg.ha^{-1}$) e área basal - G ($G.ha^{-1}$), sendo C a variável de interesse na modelagem.

2.3 Variáveis explicativas

As variáveis espectrais utilizadas foram obtidas a partir das imagens do satélite MSI/Sentinel-2 do ano de 2017 (mesmo ano do inventário florestal) e adquiridas junto ao serviço geológico dos Estados Unidos (*United States Geological Survey*). Adotou-se o uso de duas imagens da área, a primeira referente ao mês de julho de 2017, período mais seco do ano e marcado pela deciduidade de parte das espécies existentes no fragmento. E a segunda imagem do mês de novembro de 2017, período mais úmido e onde existe o maior fechamento do dossel devido aos maiores valores de índice de área foliar. Em cada uma das imagens foram extraídos para o centroíde das parcelas do inventário os valores espectrais das bandas: Azul (Resolução 10 m); Verde (Resolução 10 m); Vermelha (Resolução 10 m); *Red Edge* 1 - Borda do vermelho* (Resolução 20 m); *Red Edge* 2 (Resolução 20 m); *Red Edge* 3 (Resolução 20 m); *Red Edge* 4 (Resolução 20 m); SWIR 1 - Infravermelho de onda curta (Resolução 20 m); SWIR 2 (Resolução 20 m) e NIR - Infravermelho próximo (Resolução 10 m), totalizando 10 bandas.

Os valores espectrais foram calculados para cada uma das parcelas do inventário em cada uma das datas os seguintes índices de vegetação, seguindo formulação proposta por Jensen (2011): SR - Razão Simples; NDVI - Índice de vegetação da diferença normalizada; SAVI - Índice de vegetação ajustado ao solo; ARVI - Índice de vegetação resistente a atmosfera; SARVI - Índice de vegetação resistente a atmosfera e ao solo; EVI - Índice de vegetação

realçado; TVI - Índice de vegetação triangular e VARI - Índice de vegetação resistente à atmosfera no visível, totalizando 8 índices.

Os dados hidrológicos foram obtidos através de 32 pluviômetros/sondas instalados no interior da floresta, espaçados aproximadamente 40 m um do outro, conforme metodologia descrita por Junior et al. (2017), e se referem a precipitação dos anos de 2015 e 2017. As florestas trazem imensos benefícios aos recursos hídricos, pois diminuem a velocidade da movimentação da água da chuva em direção aos cursos d'água e os efeitos da lixiviação do solo e assoreamento dos rios (LORENZON; DIAS; LEITE, 2013). As variáveis hidrológicas Precipitação interna - PI (mm), que se refere a fração da precipitação total que chega diretamente ao solo, e o Armazenamento de água no perfil do solo até 1 metro de profundidade – ARM ($m^3.m^{-3}$), foram consideradas como possíveis variáveis preditivas para o estoque do carbono, buscando refletir a variabilidade espacial existente do dossel. De posse dos dados e das coordenadas geográficas dos pontos de coleta, foram utilizadas técnicas geoestatísticas (ajuste do semivariograma esférico e Krigagem ordinária) que permitiram a especialização dos dados e obtenção dos valores das variáveis hidrológicas para cada parcela do inventário (centroíde). As análises geoestatísticas foram realizadas com o *software* R, utilizando o pacote g*eoR* (RIBEIRO JUNIOR E DIGGLE, 2001).

O carbono da vegetação ($Mg.ha^{-1}$) foi ainda associado ao percentil diamétrico, correspondente a estrutura vertical da floresta. Essa análise teve como objetivo avaliar se a resposta da modelagem para os diferentes estratos do dossel (em função dos percentis) apresentava alguma particularidade em relação a modelagem do dossel como um todo, o que poderia trazer uma maior correlação com os sensores. A distribuição dos percentis diamétricos foram trabalhadas dentro de cada parcela, nas posições 0%, 30%, 60% e 90%. O percentil 0% representa o valor das variáveis calculado considerando 100% as árvores existentes na parcela, enquanto o percentil 90% representa as variáveis calculadas considerando apenas as árvores mais grossas da parcela (10% em relação ao número total em ordem decrescente), os demais percentis foram calculados seguindo a mesma lógica. Ao final, um conjunto de 43 variáveis disponíveis (Tabela 1) foram avaliadas para explicar a capacidade preditiva do estoque de carbono.

Tabela 1: Variáveis explicativas utilizadas na modelagem do estoque de carbono por parcela.

Grupo	Variáveis	Sigla	Unidade
Geográficas	Latitude do centroide da parcela	Y	m
	Longitude do centroide da parcela	X	m
Povoamento	Área basal da parcela e seus percentis	G/ha	$m^2\ ha^{-1}$
Espectrais (Reflectância)	Azul	B	-
	Verde	G	-
	Vermelho	R	-
	Red Edge 1	RE1	-
	Red Edge 2	RE2	-
	Red Edge 3	RE3	-
	Red Edge 4	RE4	-
	Infravermelho de onda curta 1	SWIR1	-
	Infravermelho de onda curta 2	SWIR2	-
	Infravermelho próximo	NIR	-
Índices de vegetação	Razão Simples	SR	-
	Índice de vegetação da diferença normalizada	NDVI	-
	Índice de vegetação ajustado ao solo	SAVI	-
	Índice de vegetação resistente a atmosfera	ARVI	-
	Índice de vegetação resistente a atmosfera e ao solo	SARVI	-
	Índice de vegetação realçado	EVI	-
	Índice de vegetação triangular	TVI	-
	Índice de vegetação resistente à atmosfera no visível	VARI	-
Hidrológicas	Armazenamento de água no solo	ARM	m^3m^{-3}
	Precipitação interna	PI	mm

Fonte: Do Autor (2020).

2.4 Modelagem do estoque de carbono da vegetação

Inicialmente, após uma análise exploratória de dados, utilizou-se a correlação de Pearson (r) e a sua significância estatística, como indicativo da capacidade das variáveis, identificadas quais teriam maiores chances de serem incorporadas ao conjunto de variáveis a serem utilizadas e quais seriam irrelevantes ou redundantes. Posteriormente foi utilizada a Regressão Linear Múltipla (RLM) para descrever a relação empírica existente entre as variáveis preditivas e o estoque de carbono. Durante o ajuste dos modelos de RLM ainda se utilizou o procedimento *Stepwise*, adotando como critério de seleção a estatística F. Esse procedimento visou avaliar a contribuição de cada variável independente dentre as existentes no modelo, selecionando aquelas significativas estatisticamente e que explicam da melhor forma o estoque de carbono. Associado ao *Stepwise,* visando contornar o problema da existência de multicolinearidade entre as variáveis independentes, foi aplicado o diagnóstico por meio do VIF (*Variance Inflation Factor*), que indica a presença de multicolinearidade em uma variável quando o seu valor é maior que 10, indicando assim o descarte da variável em questão do

modelo (MIDI & BAGHERI, 2010). Durante a implementação do método, a base de dados foi dividida entre base de treino (80%) e validação (20%), com o objetivo de evitar uma avaliação enviesada dos modelos gerados, os ajustes ainda foram realizados para os percentis 0%, 30%, 60% e 90% em relação ao diâmetro das árvores da floresta.

Diante da complexidade que envolve a modelagem do estoque de carbono e da facilidade de se obter variáveis correlacionadas, que permitam uma análise investigativa na busca por explicações ecológicas e melhoria das estimativas, a forma alternativa a RLM para a modelagem do carbono foi a aplicação de uma técnica de seleção de variáveis baseadas no aprendizado de máquinas em uma metodologia híbrida conhecida como GARF (*Genetic Algorithm + Random Forest*) desenvolvida por Carvalho (2019). O método associa a capacidade do algoritmo genético na seleção de variáveis para a solução ótima de um determinado problema multiobjetivo onde é gerada uma "população" seguindo os princípios da teoria de Darwin (YU; XU, 2014) com a capacidade do *Random Forest* de gerar uma modelagem flexível de interações em conjuntos de dados de altas dimensões, criando um grande número de árvores de regressão, a partir de um subconjunto de amostras de treinamento, e calculando a média de suas previsões (WAGNER; ATHEY, 2018).

O procedimento metodológico se caracterizou como um problema de otimização multiobjetivo, que buscou aumentar a precisão das estimativas utilizando o menor número de variáveis possível, em função da razão entre o erro *out-of-bag* – *erroOOB* e o erro *out-of-bag* máximo possível - $erroOOB_{max}$ (calculado através de testes preliminares), e da razão entre o número de variáveis habilitadas pelo AG - *n*; e pelo número total de variáveis testadas no experimento – NVT, conforme Equação 1.

$$fitness = \left(\frac{erroOOB}{erroOOB_{\max}} + \frac{n}{NVT} \right) \quad [1]$$

Os ajustes foram realizados seguindo a mesma divisão da base de dados por percentis do dossel da floresta (0%, 30%, 60% e 90%), sendo cada uma dessas dividida entre base de treino (80%) e validação (20%). Para a implementação do GARF com base em testes de parametrização foram definidos os seguintes parâmetros para o Algoritmo Genético: tamanho da população (100 indivíduos); taxa de seleção (0,5); taxa de mutação (0,1); operador de seleção (torneio), operadores de *crossover* (1 ponto de corte) e critério de parada (10 gerações). Já para o *Random Forest:* número de árvores cultivadas para regressão (ntree: 50); número de variáveis preditoras amostradas aleatoriamente a cada divisão da árvore (mtry: 2) e número mínimo de amostras dentro dos nós terminais (*nodesize:* 5). Foram geradas 10 repetições da rotina do GARF, onde a cada repetição foram obtidas as seguintes estatísticas de ajuste: Coeficiente de

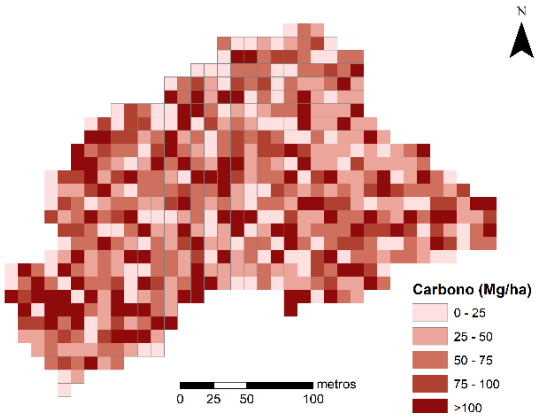
determinação (R^2), Bias (%) e Raiz do erro quadrado médio percentual (*Root Mean Square Error* – RMSE%), sendo esses valores relativos a média dos modelos gerados dentro de cada *fold*.

A implementação do GARF foi realizada no *software* R (*Version* 3.6.3 – © 2020 RStudio, Inc.), utilizando o pacote *randomForest* (LIAW; WIENER, 2002). A metodologia foi processada em uma CPU com processador Intel (R) Core™ i7-7500U CPU @ 2,90 GHz, com memória instalada (RAM) de 8,0 GB. Ao final do processo, com base nas estatísticas de ajuste por repetição, foi selecionado o melhor modelo entre as repetições, sendo esse utilizado para a comparação da performance com os modelos de RLM, utilizando a análise do gráfico de resíduos e a dispersão gráfica dos valores estimados versus os observados. Para cada modelo/método foram obtidas as seguintes estatísticas de ajuste: Coeficiente de determinação (R^2), Bias (%) e Raiz do erro quadrado médio percentual (*Root Mean Square Error* – RMSE%) e realizada a análise do gráfico de resíduos e da dispersão gráfica dos valores estimado versus observados.

3 RESULTADOS

Neste fragmento florestal foram catalogados 4.997 indivíduos sendo as espécies *Xylopia brasiliensis, Copaifera langsdorffii, Amaioua intermedia, Trichilia emarginata* e *Ocotea odorífera* as mais abundantes, de um total de 188 espécies e 47 famílias. A floresta apresentou a clássica distribuição exponencial negativa de florestas nativas, com predominância dos indivíduos nas menores classes, com até 20 cm de DAP. Essa constatação contribui para o entendimento da variabilidade do estoque de carbono, já que o estoque de carbono individual apresentou estreita ligação com o diâmetro da árvore, seguindo uma tendência exponencial (Figura 2). Dessa forma, a variabilidade existente em relação ao número de indivíduos - N existente na parcela e o DAP desses indivíduos é fundamental para essa avaliação.

Figura 2 – Comportamento do estoque de carbono ($Mg.ha^{-1}$) da floresta (a) e sua espacialização (b) na área de estudo.



Fonte: Do Autor (2020).

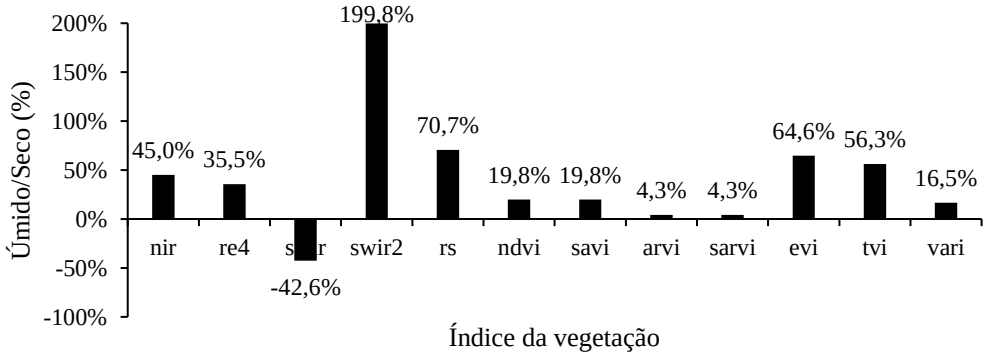
Notou-se um alto coeficiente de variação para o estoque de carbono existente nas parcelas (78,2%), com valor médio de 77,4 $Mg.ha^{-1}$. Nota-se que as parcelas com maiores valores de área de basal são aquelas que apresentam os maiores valores de estoque de carbono, indicando estreita ligação entre as variáveis. Contudo, o ajuste do semivariograma não indicou a existência de uma clara dependência espacial das variáveis C/ha e G/ha para nenhum dos percentis, onde o Grau de dependência espacial – GDE médio para essas variáveis foi de respectivamente 35,9% e 29,3%, indicando segundo a classificação de Cambardella et al. (1994), uma baixa dependência espacial, já que os valores de GDE são menores que 45%.

A análise da correlação de Pearson entre as 43 potenciais variáveis preditivas do estoque de carbono, indicou uma alta capacidade explicativa da variável G, para todos os percentis. Para o percentil 0 (100% das árvores da parcela) e 30 foi calculada uma correlação de 0,97; já para o percentil 60 e 90 a correlação foi de 0,98; todas altamente significativas estatisticamente. Em relação as variáveis espectrais, as correlações foram muito baixas (abaixo de 0,15). De forma geral, os valores de reflectância das bandas e dos índices de vegetação no período úmido (novembro - 11) foram levemente superiores ao período seco (julho - 07). Para o percentil 0 e 30 somente a variável SWIR1 no período úmido foi significativa estatisticamente, ambas com o valor de -0,15. Já para o percentil 60 apenas o NDVI para o período úmido foi significativo estatisticamente (0,13). Para o percentil 90 nenhuma variável espectral foi significativa. Entre as variáveis hidrológicas as correlações foram menores ainda, onde nenhuma variável foi considerada estatisticamente significativa.

Quanto ao efeito sazonal sobre os valores dos índices de vegetação (Figura 3) notou-se, com exceção para o SWIR1 (reduziu 42,6%), os índices de vegetação tiveram maiores valores

no período úmido (novembro). Destaca-se o valor do SWIR2 que foi em média 199,8% maior no período úmido em comparação com a média dos valores no período seco, sendo uma das variáveis que mais captaram a sazonalidade do dossel.

Figura 3 – Relação entre os índices de vegetação para o período seco e úmido no fragmento florestal estudado.



Fonte: Do Autor (2020).

Os resultados das estatísticas de ajuste dos modelos por percentil para a base de treino e validação, número de variáveis e variáveis selecionadas, utilizando a RLM associada ao método de seleção de variáveis *Stepwise* e para o método hibrido GARF são apresentadas na Tabela 2.

Tabela 2: Estatísticas de ajuste por percentil para a base de treino e validação, número de variáveis e variáveis selecionadas, utilizando o método RLM *Stepwise* e o método hibrido GARF.

Método	Percentil	Base	RMSE	RMSE(%)	Bias(%)	R^2 (%)	Nº Variáveis	Variáveis selecionadas	Tempo (min)
RLM*	0	Treino	13,17	17,35	6,65E-10	94,72%	4	G_0 + ARM_2017 +Y + swir2_11	0,013
		Validação	16,55	19,80	-0,50	94,71%			
GARF	0	Treino	7,33	9,66	0,17	98,36%	3	G_0 + re4_07+ ndvi_07	3,426
		Validação	15,80	18,91	-0,17	95,18%			
RLM*	30	Treino	12,54	16,98	7,49E-10	95,09%	5	G_30 + vari_07 + re2_11 + ARM_2017 + Y	0,006
		Validação	16,39	20,15	-0,70	94,87%			
GARF	30	Treino	6,57	8,90	0,16	98,65%	2	G_30 + swir_07	3,696
		Validação	13,83	17,00	-2,23	96,35%			
RLM*	60	Treino	11,50	17,36	7,83E-15	95,47%	2	G_60 + swir2_11	0,003
		Validação	14,25	19,06	-1,00	95,70%			
GARF	60	Treino	6,02	9,09	0,28	98,76%	2	G_60 + X	3,714
		Validação	12,39	16,57	0,03	96,75%			
RLM*	90	Treino	7,21	18,20	1,55E-10	95,98%	3	G_90 + green_07 + Y	0,002
		Validação	8,17	19,07	0,67	97,31%			
GARF	90	Treino	3,74	9,43	0,12	98,92%	2	G_90 + P_INT_2017	4,501
		Validação	20,72	48,36	4,0997	82,68%			

*Variáveis ignificativas ao nível de 5% de probabilidade e VIF abaixo de 10; RLM: Regressão Linear Múltipla; GARF: Randon Forest + Algoritmo Genético; RMSE: *Root-mean-square error* (Raiz quadrada do erro médio); Bias: variação do dado medido em relação a um valor de referência; R^2: coeficiente de determinação percentual.

Fonte: Do Autor (2020).

O percentil 0 (todas as árvores das parcelas) o modelo ajustado pelo método RLM, para a base de treino apresentou um RMSE de 17,35% e um R^2 de 94,72%, a base de validação apresentou resultados bem semelhantes com RMSE de 19,80% e eficiência de 94,71%. Para esse método foram selecionadas 4 variáveis para o modelo, uma do povoamento (G_0), uma hidrológica (ARM_2017), uma espectral (SWIR2_11) e uma geográfica (Y), indicando uma importância de todas as classes de variáveis. Ainda para o percentil 0, mas agora utilizando o método GARF a base de treino apresentou um RMSE de 9,66% (80% menor que a base de treino para o RLM), enquanto para a base de validação o RMSE foi de 18,91% (5% menor que a base de treino para o RLM), já os valores para R^2 foram respectivamente de 98,36% e 95,18%. Considerando os resultados da base de treino e validação houve uma queda da precisão de aproximadamente 50%. Nesse método foram selecionadas 3 variáveis, uma do povoamento (G_0), uma espectral (RE4_07) e um índice de vegetação (NDVI_07).

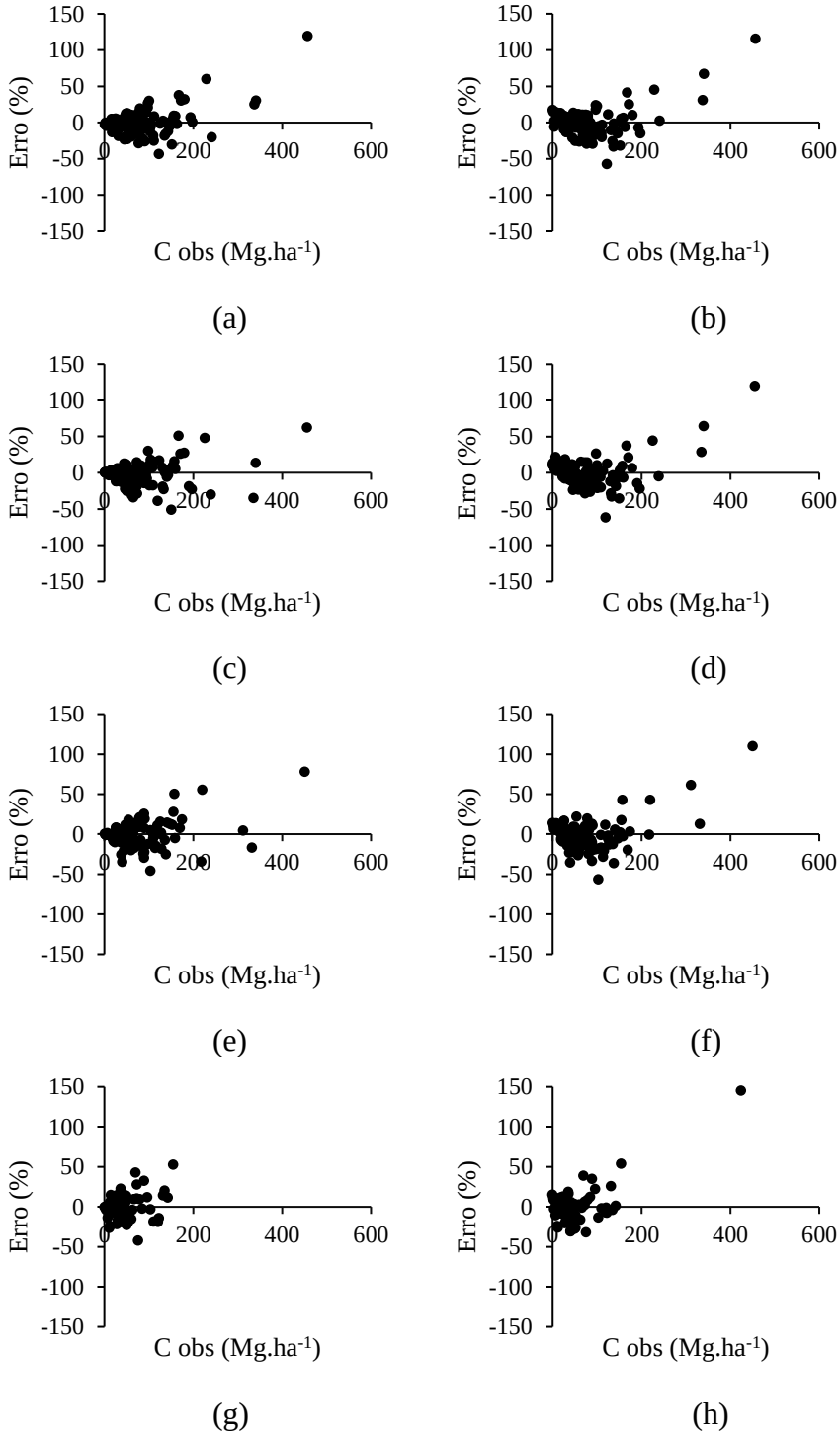
No percentil 30 (60% das maiores árvores) pelo método RLM o RMSE para a base de treino e validação foram respectivamente 16,98% e 20,15%, enquanto R^2 foi de 95,09% e 94,87%, resultados que diferiram pouco dos resultados encontrados para o percentil 0 utilizando o mesmo método. Nesse método foram selecionadas 5 variáveis, uma do povoamento (G_30), um índice de vegetação (VARI_07), uma espectral (RE2_11), uma hidrológica (ARM_2017) e uma geográfica (Y). Utilizando o GARF para o percentil 30 a base de treino apresentou um RMSE de 8,90% bem menor que o encontrado para a base de validação de 17,00% (50% maior), já os valores de R^2 foram bem semelhantes (respectivamente para a base de treino e validação 98,65% e 96,35%). Nesse método foram selecionadas 2 variáveis, uma do povoamento (G_30) e um índice de vegetação (SWIR_07). Comparando os métodos RLM e GARF para o percentil 30 foi notado um ganho de 91% de precisão na base de treinamento e de 13% na base de validação pelo uso método GARF, resultados próximos aos encontrados no percentil 0.

Para o percentil 60 (40 % das maiores árvores) pelo método RLM o RMSE para a base de treino e validação foram respectivamente 17,36% e 19,06%, enquanto R^2 foi de 95,47% e 95,70%, resultados que diferiram pouco dos resultados encontrados para o percentil 0 e 30 utilizando o mesmo método. Nesse método foram selecionadas 2 variáveis, uma do povoamento (G_60), um índice de vegetação (SWIR2_11). Utilizando o GARF para o percentil 60 a base de treino apresentou um RMSE de 9,09%, bem menor que o encontrado para a base de validação de 16,57% (45% maior), os valores de R^2 foram bem semelhantes (respectivamente para a base de treino e validação 98,76% e 98,75%). Nesse método foram selecionadas 2 variáveis, uma do povoamento (G_60) e uma variável geográfica (X). Comparando os métodos RLM e GARF

para o percentil 60 foi notado um ganho de 91% de precisão na base de treinamento e de 15% na base de validação pelo uso método GARF.

Para o percentil 90 (10 % das maiores árvores) pelo método RLM o RMSE para a base de treino e validação foram respectivamente 18,20% e 19,07%, enquanto R^2 foi de 95,98% e 97,31%, resultados que diferiram pouco dos resultados encontrados para o demais percentis utilizando o mesmo método. Nesse método foram selecionadas 2 variáveis, uma do povoamento (G_90), um índice espectral (GREEN_07). Utilizando o GARF para o percentil 90 a base de treino apresentou um RMSE de 9,43% bem menor que o encontrado para a base de validação de 48,36% (80% maior), já os valores de R^2 foram bem diferentes, sendo respectivamente para a base de treino e validação 98,92% e 82,68%. Nesse método foram selecionadas 2 variáveis, uma do povoamento (G_90) e uma hidrológica (P-INT_2017). Comparando os métodos RLM e GARF para o percentil 90 foi notado um ganho de 93% de precisão na base de treinamento e uma perda de 61% na base de validação pelo uso método GARF. De forma geral, a modelagem GARF ganha dos modelos via RLM, com exceção do percentil 90, onde o método GARF não conseguiu manter a precisão das estimativas, por outro lado, o RLM apresentou nesse percentil os melhores resultados entre todos. As Figura 4 apresenta os gráficos de resíduos para os modelos ajustados por percentil.

Figura 4: Gráficos de resíduos (direita) para os modelos ajustados por GARF e RLM para os diferentes percentis: a) GARF – Percentil 0; b) RLM – Percentil 0; c) GARF – Percentil 30; d) RLM – Percentil 30; e) GARF – Percentil 60; f) RLM – Percentil 60; g) GARF – Percentil 90 e h) RLM – Percentil 90.

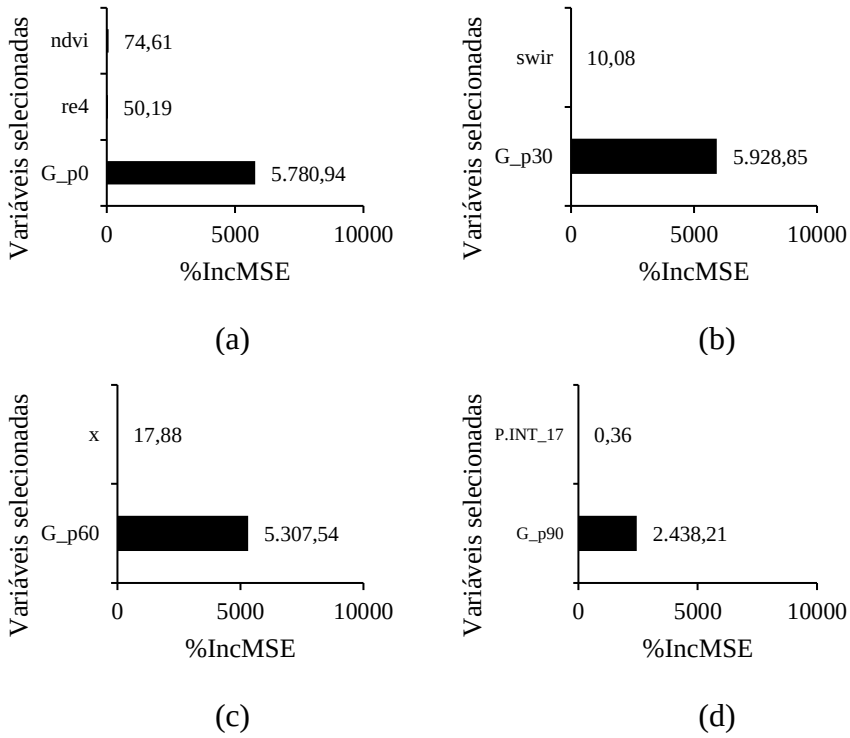


Do Autor (2020).

Pode ser observado uma tendência subestimativa dos modelos para as parcelas com estoque de carbono acima de 200 $Mg.ha^{-1}$. Nos gráficos de resíduos ainda se nota a presença de alguns *outliers* relativos as subestimativas dos modelos (valores de erro > 100%).

A Figura 5 apresenta os valores de importância das variáveis selecionadas (%IncMSE) pelo GARF para os percentis analisados. Percebe-se que a contribuição da variável área basal (G) foi predominante para as estimativas de estoque de carbono independente do percentil, já as variáveis espectrais, geográficas e hidrológicas foram responsáveis por contribuições menores, mas mesmo assim importantes para a assertividade das estimativas, justificando a sua inclusão nos modelos.

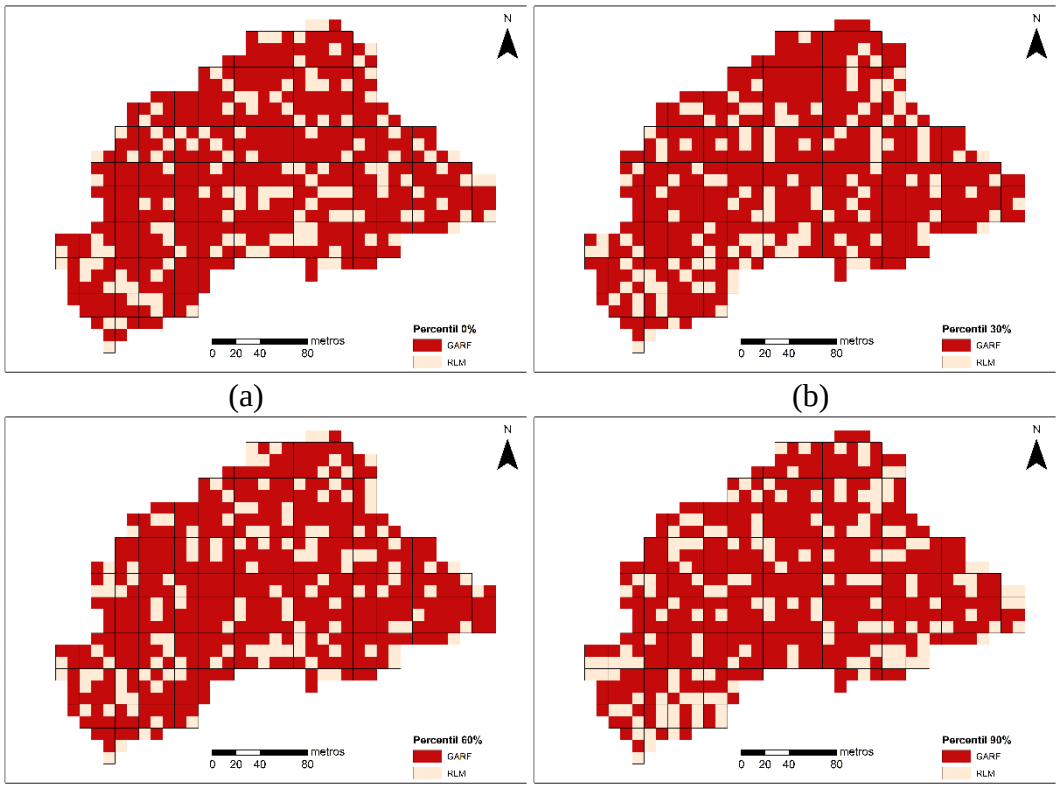
Figura 5: Valores de importância (%IncMSE) das variáveis selecionadas pelo *Random Forest* para a estimativa do estoque de carbono para: a) percentil 0, b) percentil 30, c) percentil 60 e d) percentil 90.



Fonte: Do Autor (2020).

A Figura 6 apresenta a distribuição espacial dos resultados do melhor método para a estimativa do estoque de carbono em cada parcela do inventário.

Figura 6: Distribuição espacial do métodos com menores erros na estimativa do estoque de carbono em cada parcela do inventário para: a) percentil 0%, b) percentil 30%, c) percentil 60% e d) percentil 90%.



Fonte: Do Autor (2020).

Observa-se que o método GARF apresentou melhores resultados para a maior parte das parcelas inventariadas, totalizando 75,3% no percentil 0; 76,1% no percentil 30; 75,5% no percentil 60 e 71,5% no percentil 90. De forma geral, houve um predomínio dos erros até 7% e 12% para os métodos GARF e RLM em todos os percentis avaliados, indicando uma boa precisão dos modelos produzidos.

O valor total do estoque de carbono do fragmento para o percentil 0 foi de 422,62 Mg de carbono, onde a estimativa pelo RLM foi de 423,08 Mg (0,11% em relação ao observado) e pelo GARF 422,44 Mg (-0,04% em relação ao observado), ambos com diferenças menores que 0,5% em relação ao total, os demais percentis tiveram assertividades semelhantes, com diferenças de 0,32% e 0,15% respectivamente pelos métodos GARF e RLM para o percentil 30 (411,32 Mg); -0,18% e 0,22% pelos métodos GARF e RLM para o percentil 60 (370,80 Mg) e -0,88% e -0,14% pelos métodos GARF e RLM para o percentil 90 (219,85 Mg).

4 DISCUSSÃO

Nunes et al. (2003) analisando os dados do mesmo fragmento florestal, referentes ao período 1986 e 1996, encontraram diferenças estatística entre os valores de área basal para 2 estratos, chamados de alto e baixo, onde respectivamente os valores médios encontrados eram de 26,15 m²/ha e 20,2 m²/ha, segundo os autores essa diferença sugeriria a existência de uma comunidade arbórea mais jovem no estrato baixo, indicando a existência de diferenças de estágio de regeneração do fragmento naquele momento. Nossa análise us7ando a modelagem geoestatística não foi capaz de observar essa estratificação, sendo que os dados apresentaram baixa dependência espacial segundo a classificação proposta por Cambardella et al. (1994). Isso indica que atualmente a floresta apresenta um estágio de regeneração mais avançado. Por outro lado, a estágio de regeneração mais avançado sugere a existência de árvores mais maduras, com maiores valores de estoque de carbono, indicando uma maior heterogeneidade entre os estoques em função da distribuição diamétrica das árvores no interior da floresta.

Alvarenga et al. (2012) avaliaram a heterogeneidade entre o volume de madeira existentes nas parcelas de inventário de fragmentos florestais de Cerrado *Sensu Stricto* e encontraram valores de coeficiente de variação de 72,2%, valor muito próximo ao encontrado na área desse estudo, confirmando o alto grau de variação existente as variáveis dendrométricas. Como essa distribuição é o resultado da união de diferentes fatores edafoclimáticos que se refletem no fechamento do dossel, consequentemente seus efeitos poderiam ser captados pelos índices espectrais da vegetação. Contudo, a detecção remota quantitativa do dossel da vegetação é complexa devido ao tamanho, forma, propriedades de reflectância espectral das folhas e aberturas existentes devido à queda de árvores (GALEANA-PIZAÑA et al., 2014).

Dessa forma, a correlação encontrada entre os valores espectrais e índices de vegetação foi muito baixa (próxima a 10%). Reis et al. (2020) avaliando diferentes abordagens e conjuntos de dados para modelar distribuição espacial do volume de madeira em um fragmento de Cerrado *Sensu Stricto*, utilizando bandas espectrais e índices de vegetação oriundos das imagens do satélite Landsat 5 TM, encontraram correlações levemente superior para o NDVI, SAVI e EVI (próximas a 20%), ressaltando, nesse caso, a diferença existente entre a variável objeto e o sensor utilizado. A área desse estudo devido a natureza semidecídua da floresta apresenta um comportamento sazonal do dossel, onde aproximadamente 50% das árvores perdem folhas no período mais seco do ano. Este comportamento foi levado em consideração na análise de imagens de satélite, pois podem afetar a observação da variável de interesse ou mesmo o cálculo dos índices de vegetação.

Almeida et al. (2014) demonstram em seus trabalhos a influência da deciduidade sobre o índice de vegetação NDVI, o que pode explicar a diferença observada entre os valores de reflectância e dos índices de vegetação existentes entre o período úmido (novembro de 2017) e período seco (julho de 2017), onde os valores no período úmido foram levemente superiores (20%) que os encontrados no período seco, em que ocorre a deciduidade (perda) de parte das folhas. Embora vários estudos utilizem dados espectrais para explicar variáveis dendrométricas, quando se trabalha em áreas pequenas e com alta diversidade espacial da vegetação a análise da correlação indica uma baixa capacidade explicativa. Isso ocorre porque fragmentos/parcelas com valores dendrométricos semelhantes podem apresentar características espectrais diferentes devido a composição das espécies, abertura do dossel, índice de área foliar, densidade de indivíduos, entre outros (REIS et al., 2020). Contudo, o grande diferencial dos dados remotos é que eles podem ser importantes quando associados a outras variáveis, preenchendo lacunas de informações gerada pelos dados coletados apenas no campo (MENG et al., 2009; VIANA et al., 2012; ALMEIDA et al., 2014; PONZONI et al., 2015) contribuindo para uma maior assertividade dos modelos.

Miguel et al. (2015) encontraram melhores resultados da estimativa do volume de um fragmento de savana arborizada quando comparado com Reis et al. (2020), associando o uso de dados de área basal aos índices de vegetação, assim, o uso de dados de campo associados a outras informações podem ser uma metodologia promissora para a modelagem de informações dendrométricas em casos onde a vegetação apresenta alta variabilidade espacial, caso das florestas nativas. Da mesma forma que os índices espectrais, as informações hidrológicas são capazes de representar atributos relacionados heterogeneidade da abertura do dossel, já que tanto a precipitação interna quanto o armazenamento de água no solo são influenciados pelo índice de área foliar e pela deciduidade dos indivíduos arbóreos da floresta. Essa influência foi documentada por Junior et al. (2017) analisando os valores de armazenamento de água no solo para a mesma área desse estudo, onde os autores associaram a variabilidade espacial dessa variável ao particionamento da precipitação gerado pelo dossel da floresta, que consequentemente está associado ao índice de área foliar.

Quanto a modelagem, tanto o método de regressão linear múltipla- RLM com seleção via *Stepwise*, quanto o GARF produziram bom modelos para a estimativa do estoque de carbono. Considerando todos os percentis avaliados os valores de RMSE para a base de treinamento variaram entre 8,90% (GARF para o percentil 30) e 18,20% (RLM para o percentil 90), já para a base de validação os valores de RMSE variaram entre 16,57% (GARF para o percentil 60) e 48,39% (GARF para o percentil 90). Considerando apenas o percentil 0 (todas

as árvores) o RMSE para a validação via RLM e GARF foram respectivamente de 19,80% e 18,91%. Comparado aos métodos testados por Reis et al. (2020) para a fisionomia Cerrado que utilizaram unicamente variáveis espectrais, a Krigagem ordinária (36,23%), a Krigagem com Regressão (37,68%) e a Regressão linear múltipla (36,42%) apresentaram erros consideravelmente maiores.

O método GARF mostrou a capacidade de gerar menores erros na estimativa do estoque de carbono que a RLM. Esse desempenho se deve ao fato do GARF ser um método baseado no algoritmo *Random Forest*, que é conhecido por retirar informações mais relevantes para predição da variável resposta, com destacada implementação para a seleção de variáveis. O *Random Forest* é um algoritmo muito sensível a quantidade de observações inseridas nele, não se beneficiando quando utiliza uma quantidade de amostras pequena (MA et al., 2017), se beneficiando muito de um tamanho de amostra maior (FASSNACHT et al., 2014; LI et al., 2016).

RLM com seleção via *Stepwise* é um método altamente reconhecido pelas suas qualidades, assim, modelos matemáticos que utilizam variáveis independentes altamente correlacionadas à variável dependente tendem a apresentar ótimos ajustes (FUJIWARA et al., 2009). Nesse caso, atenção especial deve ser dada, a existência de multicolinearidade entre as variáveis independentes, um problema comum em regressões, no qual as variáveis independentes estão altamente correlacionadas entre si, fato esse que indica que uma ou mais variáveis podem ser desnecessárias no modelo, uma vez que uma das premissas de regressão é que nenhuma relação linear pode existe entre quaisquer variáveis independentes (MONTGOMERY et al., 2012).

Para contornar esse problema, recomenda-se o diagnóstico por meio do uso do VIF (*Variance Inflation Factor*), entre os modelos gerados o teste não mostrou ocorrência de multicolinearidade entre as variáveis independentes selecionadas (VIF <10), com distribuição uniforme dos erros. Analisando as variáveis selecionadas pela RLM, para os percentis 0 e 30, nota-se a predominância de um padrão formado por modelos que utilizam variáveis dendrométricas associadas a variáveis de sensoriamento, variáveis hidrológicas e variáveis geográficas.

O modelo via RLM para o percentil 0 selecionou as variáveis Área basal, ARM_2017, Y e a banda SWIR2_11. SWIR2_11 é a banda do infravermelho de ondas curtas no período mais úmido do ano, apesar de o seu uso não ser comum no mapeamento da vegetação, essa região do espectro eletromagnético fornece importantes informações sobre elementos biofísicos e bioquímicos da vegetação (por exemplo, lignina, celulose). Segundo Vicente et al., 2007, o

SWIR demonstrou ser capaz de delinear diferenças nas etapas de senescência, assim como presença de vegetação não fotossinteticamente ativa no mapeamento de fitofisionomias em ambiente tropical utilizando dados do sensor ASTER. A variável ARM_2017 (Armazenamento de água no solo) pode retratar indiretamente a abertura do dossel ou mesmo uma condição que proporcionou um maior desenvolvimento dos indivíduos. Já a variável Y (longitude) pode explicar alguma tendência de dependência espacial do estoque de carbono.

Já no modelo via GARF para o percentil 0, houve um predomínio das variáveis dendrométricas associadas a variáveis de sensoriamento. A variável RE4_07 (Reflectância da banda Red edge no período mais seco) é um espectro eletromagnético que está na faixa do infravermelho muito utilizado para avaliar os teores de clorofila das folhas, vigor de plantas, detectar e determinar focos de estresse (doenças, pragas, entre outros), apresentando alta sensibilidade de resposta ao teor de clorofila presente nas folhas. Da mesma forma, o NDVI_07 (*Normalized Difference Vegetation Index* ou Índice de Vegetação da Diferença Normalizada para o período mais seco) é frequentemente usado para medir a intensidade de atividade da clorofila, sendo muito sensível a mudanças sazonais na vegetação, caso da fisionomia Semidecidua em relação a perda de folhas no período mais seco do ano. Dessa forma, ao seleciona-lo o modelo considerou o teor de clorofila existente no dossel no período com limitações hídricas como uma importante variável preditiva do estoque de carbono.

Considerando todos os modelos de todos os percentis destaca-se as variáveis G (selecionado pelos 8 modelos); Y (3 modelos); SWIR2_11 e ARM_2017 (2 modelos) e P-INT_2017, RE4_07, NDVI_07, VARI_07, RE2_11, SWIR_07, X e GREEN_07 (1 modelo). O interessante é que mesmo apresentando baixas correlações essas variáveis foram significativas e importantes para gerar modelos mais assertivos. Miguel et al. (2015) estimou o volume de savana arborizada usando dados de área basal e índices de vegetação (EVI, NDVI, SAVI e SR) derivados de imagens do sensor LISS-III. Almeida et al. (2014) ajustaram modelos à estimativa do volume de madeira na Caatinga brasileira usando imagens do Landsat 5 TM utilizando NDVI e SAVI e banda TM3. Reis et al. (2020) estimaram o volume para um fragmento de Cerrado *Sensu Stricto* utilizando dados da banda TM2, TM5 e EVI da imagem Landsat 5 TM. Esses resultados confirmam a capacidade dos dados espectrais de contribuir para a estimativas de variáveis dendrométricas.

5 CONCLUSÕES

Tanto o método via RLM quanto o GARF demostram que, para a área de estudo, a integração de dados espectrais e hidrológicos aos dados de inventário de campo foram capazes de melhorar as estimativas do estoque de carbono, mas não foram capazes de substitui-los ao ponto de realizar boas estimativas sem informações dendrométricas. Os diferentes estratos do dossel demostraram seguir um comportamento de modelagem que diferiram entre si em relação a seleção das variáveis preditivas. O método GARF foi capaz de produzir melhores resultados que o RLM, mesmo considerando a maior complexidade de implementação e tempo de processamento do GARF, contudo, a Regressão Linear Múltipla com seleção das variáveis via *Stepwise* ainda é um método que apresenta boa precisão e fácil implementação, sendo também adequado para estimativas de estoque de carbono em florestas nativas.

REFERÊNCIAS

ALMEIDA, André Quintão et al. Relações empíricas entre características dendrométricas da Caatinga brasileira e dados TM Landsat 5. **Pesquisa Agropecuária Brasileira**, v. 49, n. 4, p. 306-315, 2014.

ALVARENGA, Luiz Henrique Victor et al. Desempenho da estratificação em um fragmento de cerrado *stricto sensu* utilizando interpolador geoestatístico. **Cerne**, v. 18, n. 4, p. 675-681, 2012.

ALVARES, Clayton Alcarde et al. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711-728, 2013.

CAMBARDELLA, C. A., MOORMAN, T. B., PARKIN, T. B., KARLEN, D. L., NOVAK, J. M., TURCO, R. F., & KONOPKA, A. E. (1994). Field-scale variability of soil properties in central Iowa soils. **Soil science society of America journal,** 58(5), 1501-1511.

CARVALHO, M. C. **Inteligência computacional na modelagem florestal: teor de carbono e distribuição geográfica de espécies**. 2019.148 p. Tese (Doutorado em Engenharia Florestal) - Universidade Federal de Lavras, Lavras, 2019.

CHAVE, J., COOMES, D., JANSEN, S., LEWIS, S. L., SWENSON, N. G., & ZANNE, A. E.. Towards a worldwide wood economics spectrum. **Ecology letters**, 12(4), 351-366, 2009.

CHAVE, Jérôme et al. Improved allometric models to estimate the aboveground biomass of tropical trees. **Global change biology**, v. 20, n. 10, p. 3177-3190, 2014.

BOLFE, Édson Luis; BATISTELLA, Mateus; FERREIRA, Marcos César. Correlação de variáveis espectrais e estoque de carbono da biomassa aérea de sistemas agroflorestais. **Pesquisa Agropecuária Brasileira**, v. 47, n. 9, p. 1261–1269, 2012.

DA SILVA FILHO, Rivaildo et al. Representação matemática do comportamento intra-anual do NDVI no Bioma Caatinga. **Ciência Florestal**, v. 30, n. 2, p. 473-488, 2020

FASSNACHT, F. E. et al. Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass. **Remote Sensing of Environment,** v. 154, n. 1, p. 102–114, 2014.

FUJIWARA, Koichi et al. Soft-sensor development using correlation-based just-in-time modeling. **AIChE Journal**, v. 55, n. 7, p. 1754-1765, 2009.

GALEANA-PIZAÑA, J. Mauricio et al. Modeling the spatial distribution of above-ground carbon in Mexican coniferous forests using remote sensing and a geostatistical approach. **International Journal of Applied Earth Observation and Geoinformation**, v. 30, p. 179-189, 2014.

INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE - IPCC. **Guidelines for national greenhouse gas inventories:** agriculture, forestry and other land use. Japan: Institute for global environmental strategies (IGES), 2006. v. 4. Disponível em: <http://www.ipcc-nggip.iges.or.jp/public/2006gl/vol4.html>. Acesso em: 1 Junho 2020.

JENSEN, J. R. **Sensoriamento remoto do ambiente**: uma perspectiva em recursos terrestres. Parêntese, 2011.

JUNIOR, JA Junqueira et al. Time-stability of soil water content (SWC) in an Atlantic Forest-Latosol site. **Geoderma**, v. 288, p. 64-78, 2017.

LI, M. et al. A systematic comparison of different object-based classification techniques using high spatial resolution imagery in agricultural environments. **International Journal of Applied Earth Observation and Geoinformation**, v. 49, p. 87–98, 2016.

LIAW, A.; WIENER, M. Classification and Regression by randomForest. **R news**, v. 2, p. 18–22, 2002.

LORENZON, A. S.; DIAS, H. C. T.; LEITE, H. G. Precipitação efetiva e interceptação da chuva em um fragmento florestal com diferentes estágios de regeneração. Revista Árvore, Viçosa, MG, v. 37, n. 4, p. 619-627, jul./ago. 2013.

NUNES, Yule Roberta Ferreira et al. Variações da fisionomia, diversidade e composição de guildas da comunidade arbórea em um fragmento de floresta semidecidual em Lavras, MG. **Acta botanica brasilica**, v. 17, n. 2, p. 213-229, 2003.

MA, L.; FAN, S. CURE-SMOTE algorithm and hybrid algorithm for feature selection and parameter optimization based on random forests. **BMC Bioinformatics**, v. 18, n. 1, p. 1–18, 2017.

MELLO, Juliana Milesi et al. Dinâmica dos atributos Físico-químicos e variação sazonal dos estoques de carbono no solo em diferentes fitofisionomias do Pantanal Norte Mato-grossense. **Revista Árvore** p. 325–336, 2008.

MENG, Qingmin; CIESZEWSKI, Chris; MADDEN, Marguerite. Large area forest inventory using Landsat ETM+: a geostatistical approach. **ISPRS Journal of Photogrammetry and Remote Sensing,** v. 64, n. 1, p. 27-36, 2009.

MIDI, H., & BAGHERI, A. . Robust multicollinearity diagnostic measure in collinear data set. In **Proceedings of the 4th international conference on applied mathematics, simulation, modeling** (2010, July) (pp. 138–142).

MIGUEL, Eder Pereira et al. Redes neurais artificiais para a modelagem do volume de madeira e biomassa do cerradão com dados de satélite. **Pesquisa Agropecuária Brasileira**, v. 50, n. 9, p. 829-839, 2015.

MONTGOMERY, Douglas C.; PECK, Elizabeth A.; VINING, G. Geoffrey. **Introduction to linear regression analysis**. John Wiley & Sons, 2012.

OLIVEIRA-FILHO, Ary T.; DE MELLO, José Márcio; SCOLFORO, José Roberto S. Effects of past disturbance and edges on tree community structure and dynamics within a fragment of tropical semideciduous forest in south-eastern Brazil over a five-year period (1987–1992). **Plant Ecology**, v. 131, n. 1, p. 45-66, 1997.

PONZONI, Flávio Jorge et al. Caracterização espectro-temporal de dosséis de Eucalyptus spp. mediante dados radiométricos TM/Landsat5. **Cerne**, v. 21, n. 2, p. 267-275, 2015.

REIS, Aliny Aparecida et al. Modeling the spatial distribution of wood volume in a Cerrado Stricto Sensu remnant in Minas Gerais state, Brazil. **Scientia Forestalis**, n. 125, 2020.

RÉJOU-MÉCHAIN, Maxime et al. biomass: An r package for estimating above-ground biomass and its uncertainty in tropical forests. **Methods in Ecology and Evolution**, v. 8, n. 9, p. 1163-1167, 2017.

RIBEIRO, S. C. et al. Above-and belowground biomass in a Brazilian Cerrado. **Forest Ecology and Management**, v. 262, n. 3, p. 491-499, 2011.

RIBEIRO JR, P. J.; DIGGLE, Peter J. geoR: a package for geostatistical analysis, R News 1/2: 15–18. **Find this article online**, 2001.

FAYE, M. A. et al. Impact of changes in land use, species and elevation on soil organic carbon and total nitrogen in Ethiopian Central Highlands. **Geoderma**, v. 261, p. 70-79, 2016.

VIANA, H. et al. Estimation of crown biomass of Pinus pinaster stands and shrubland above-ground biomass using forest inventory data, remotely sensed imagery and spatial prediction models. **Ecological Modelling**, v. 226, p. 22-35, 2012.

VICENTE, L. E.; SOUZA FILHO, C. R.; PEREZ FILHO, A. O uso do infravermelho de ondas curtas (SWIR) no mapeamento de fitofisionomias em ambiente tropical por meio de classificação hiperespectral de dados do sensor ASTER. **Anais XIII Simpósio Brasileiro de Sensoriamento Remoto,** Florianópolis, Brasil, INPE, 2007.

YU, Feng; XU, Xiaozhong. A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network. **Applied Energy**, v. 134, p. 102-113, 2014.

WAGER, Stefan; ATHEY, Susan. Estimation and inference of heterogeneous treatment effects using random forests. **Journal of the American Statistical Association**, v. 113, n. 523, p. 1228-1242, 2018.