



**PAULA GOMIDES VITOR SCOLFORO**

**EVOLUÇÃO NA MODELAGEM DA DISTRIBUIÇÃO  
ESPACIAL DO ESTOQUE DE CARBONO ARBÓREO NO  
CERRADO BRASILEIRO**

**LAVRAS - MG**

**2021**

**PAULA GOMIDES VITOR SCOLFORO**

**EVOLUÇÃO NA MODELAGEM DA DISTRIBUIÇÃO ESPACIAL DO  
ESTOQUE DE CARBONO ARBÓREO NO CERRADO BRASILEIRO**

Tese apresentada ao Programa de Pós-Graduação em Engenharia Florestal como parte das exigências do curso de doutorado em Engenharia Florestal.

Prof. Dr. José Roberto Soares Scolforo  
Orientador

Prof. Dr. Cristian Montes  
Coorientador

**LAVRAS – MG  
2021**

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

Scolforo, Paula Gomides Vitor.

Evolução na modelagem da distribuição espacial do estoque de  
carbono arbóreo no Cerrado brasileiro / Paula Gomides Vitor

Scolforo. - 2021.

87 p.

Orientador(a): José Roberto Soares Scolforo.

Coorientador(a): Cristian Montes.

Tese (doutorado) - Universidade Federal de Lavras, 2021.

Bibliografia.

1. Manejo. 2. Modelagem. 3. Carbono. I. Scolforo, José Roberto  
Soares. II. Montes, Cristian. III. Título.

**PAULA GOMIDES VITOR SCOLFORO**

**EVOLUÇÕES NA MODELAGEM DA DISTRIBUIÇÃO ESPACIAL DO  
ESTOQUE DE CARBONO ARBÓREO NO CERRADO BRASILEIRO**

**EVOLUTION IN THE ABOVEGROUND CARBON STOCK SPATIAL  
DISTRIBUTION MODELING IN THE BRAZILIAN SAVANAH**

Tese apresentada ao Programa de  
Pós-Graduação em Engenharia  
Florestal como parte das exigências  
do curso de doutorado em  
Engenharia Florestal.

APROVADA em 31 de março 2021

Dr. José Roberto Soares Scolforo UFLA  
Dr. Lucas Rezende Gomide UFLA  
Dr. Otávio Camargo Campoé UFLA  
Dr. Fortunato Silva de Menezes UFLA  
Dra. Ximena Mendes de Oliveira Unicentro

Prof. Dr. José Roberto Soares Scolforo  
Orientador

Prof. Dr. Cristian Montes  
Coorientador

**LAVRAS – MG  
2021**

*Aos meus amados, Henrique e João Henrique.  
À minha tia Marina (in memoriam).  
Dedico*

## AGRADECIMENTOS

À Deus, pelo dom da vida, por ser presença constante no meu existir.

Ao meu marido Henrique, pelo amor, companheirismo, dedicação, carinho, amizade, trocas. Obrigada por ter me dado a mão nestes 11 anos de caminhada, por toda a ajuda concedida neste trabalho e em tantos outros, por ser também uma referência profissional. A você todo meu amor e a minha gratidão.

Ao meu amado filho, minha luz, que chegou para transformar as nossas vidas trazendo alegria infinita. Você ressignificou a vida da mamãe João Henrique, fez ela querer fazer este mundo melhor e também ser uma pessoa melhor todos os dias. Obrigada.

Aos meus pais, Eliana e Paulo Henrique, e a meu irmão Pedro Paulo, por serem meu ponto referência e porto seguro. Por todo incentivo e amor.

Aos meus avós, Ilza, Laura, Marinho e Nilson (in memoriam), pelo colinho e amor que todo neto merecia receber, sou sempre grata.

À minha família, onde estão minhas raízes e referências, especialmente as minhas tias Glauce, Rosana e Terezinha e família, e aos meus sogros José Roberto e Sandra, e a Roberta.

Ao professor e orientador, Dr. José Roberto Scolforo, por ser uma referência, por toda experiência compartilhada.

Ao professor e co-orientador, prof Dr. Cristian Montes, pela oportunidade oferecida de passar um período de muito aprendizado na University of Georgia, sou extremamente grata. Obrigada por nos acolher tão bem.

A todos os professores envolvidos na minha formação até aqui, desde os primeiros anos de estudo, especialmente aos professores do LEMAF.

A instituição UFLA e todos os funcionários que a fazem funcionar. Pelos inesquecíveis anos vividos aqui, sou eternamente grata, e tenho muito orgulho de ser UFLA.

Aos amigos e colegas que eu fiz durante toda essa jornada e aos que já estavam nela antes do doutorado começar, os momentos compartilhados estão eternizados na minha memória, assim como cada um de vocês. Um agradecimento a Thiza, Rafaella, Ximena, Luciane, Maria Eduarda, e todos os demais companheiros por todos estes anos passados no LEMAF.

Um agradecimento especial a minha amiga Rafa, por todas as ajudas ao longo de todos estes anos, pela amizade verdadeira, sou sempre grata.

A CAPES pela concessão de bolsa de estudos.

A jornada maravilhosa que trilhei até aqui, que me fez ser quem sou, a realizar meus sonhos e me fazer sonhar tantos outros, a todos os envolvidos.

**MUITO OBRIGADA!**

*“Quem acredita sempre alcança (Renato Russo)”*

*“Tudo é possível para quem crê (Bíblia Sagrada)”*

## RESUMO GERAL

A compreensão da distribuição espacial de estoque de carbono permite com que zonas prioritárias de conservação sejam definidas. Adicionalmente, a compreensão específica da distribuição de carbono dentro das zonas prioritárias tende a tornar a gestão pública florestal ainda mais eficiente. O desenvolvimento de uma técnica que permita mapear de forma precisa a distribuição do estoque de carbono no cerrado, assim como de uma técnica que permita a predição precisa e com baixo nível de incerteza para os diferentes fragmentos florestais tornam-se relevantes para o contexto florestal atual. Diante dessas perspectivas, essa tese apresenta uma abordagem híbrida para a modelagem espacial do estoque de carbono arbóreo no cerrado de Minas Gerais (Artigo 1), e uma abordagem de crescimento e predição (CP) para quantificação espacial do estoque de carbono com baixo grau de incerteza associado as estimativas (Artigo 2). Neste estudo, foi utilizada uma base de dados proveniente do Inventário Florestal de Minas Gerais. Além disso, variáveis de sensoriamento remoto foram extraídas da Landsat TM enquanto variáveis bioclimáticas relacionadas a precipitação e temperatura foram extraídas do WorldClim para todos os fragmentos de Cerrado avaliados. De maneira geral, o Artigo 1 indica como a abordagem híbrida, que combina Lasso (least absolute shrinkage selection operator) para seleção de variáveis preditoras, ajuste de modelo linear por máxima verossimilhança e modelagem dos resíduos por meio de krigagem ordinária, foi uma técnica substancialmente mais precisa que as técnicas convencionalmente aplicadas para este tipo de estudo. Em especial, a abordagem híbrida permite com que o mapeamento da distribuição de carbono no cerrado seja realizado de forma precisa, em média, ao longo de todo gradiente ambiental do mesmo. Já o Artigo 2, enfatiza o valor de todas as informações auxiliares disponíveis para uso, em especial dados de inventários passados. Nele, informações de inventário, de variáveis bioclimáticas e de sensoriamento remoto foram utilizadas para o ajuste de um modelo preditivo do estoque de carbono. A utilização de diversas variáveis auxiliares permitiu o desenvolvimento de uma técnica que não somente mapeia de forma precisa a distribuição do estoque de carbono, mas também mapeia de forma precisa ao nível fragmento florestal (sítio)-específica, combinando modelagem de crescimento com modelagem geoespacial. As diferentes técnicas apresentadas ao longo dessa tese agregam grande valor para quantificação do estoque de carbono em florestas nativas no Brasil.

**Palavras-chave:** Inventário Florestal, Krigagem, Lasso, Modelagem, Máxima Verossimilhança.

## GENERAL ABSTRACT

Understanding the carbon stock spatial distribution allows for the definition of priority areas for conservation. Furthermore, the understanding of the carbon stock distribution within priority areas is critically important for public forest management efficiency. The development of an approach that allows for the precise carbon stock distribution mapping in the Brazilian Savanah, as well as an approach that allows for the precise carbon stock estimates with a low level of uncertainty for different forest fragments is relevant to the current forest context. This dissertation presents a hybrid approach for the spatial modeling of aboveground carbon stock in the Savanah of Minas Gerais (Article 1), and a growth and prediction (CP) approach for the precise carbon stock spatial mapping with low degree of uncertainty (Article 2). A database from the Forest Inventory of Minas Gerais was used in this dissertation. In addition, remote sensing variables were extracted from Landsat TM while bioclimatic variables were extracted from WorldClim for all forest fragments. Overall, Article 1 highlights how the hybrid approach, which combines Lasso (least absolute shrinkage selection operator) for variable selection, maximum likelihood for fitting a linear model, and the ordinary kriging of the residuals based on the linear model, was substantially more precise than conventional approaches applied for this type of study. In particular, the hybrid approach allows for a precise carbon stock distribution mapping in the Savanah across the environmental gradient. Article 2 emphasizes the importance of all auxiliary variables, especially the past inventory results. Forest inventory, bioclimatic and remote sensing variables were combined and used to fit a predictive model of carbon stock distribution. The use of several auxiliary variables allowed for the development of an approach that not only precisely maps the carbon stock distribution, but also provide precise forest fragment (site)-specific carbon stock estimates. The different approaches presented in this dissertation add great value on the way to quantify the carbon stock in native forests in Brazil.

**Keywords:** Forest Inventory, Kriging, Lasso, Modeling, Maximum Likelihood.

## SUMÁRIO

<b>PRIMEIRA PARTE</b> .....	13
<b>1 INTRODUÇÃO</b> .....	14
<b>2 REVISÃO DE LITERATURA</b> .....	15
2.1 Cerrado.....	15
2.2 Inventário Florestal de Minas Gerais .....	16
2.3 Modelagem da distribuição do estoque de carbono .....	17
2.3.1 Random Forest (RF).....	18
2.3.2 Least absolute shrinkage selection operator (lasso) .....	19
2.3.3 Krigagem com regressão.....	21
2.3.4 Modelos de crescimento .....	22
2.3.5 Máxima verossimilhança .....	23
2.4 Quantificação da incerteza de estimativa.....	24
<b>3. CONSIDERAÇÕES GERAIS</b> .....	25
<b>REFERÊNCIAS</b> .....	26
<b>SEGUNDA PARTE – ARTIGOS</b> .....	32
<b>ARTIGO 1: UM NOVO MODELO PARA DISTRIBUIÇÃO DO ESTOQUE DE CARBONO NO CERRADO</b> .....	33
1. INTRODUÇÃO .....	35
2. MATERIAL E MÉTODOS .....	37
2.1. Caracterização da área de estudo .....	37
2.2. Caracterização da amostragem.....	40
2.3 Dados bioclimáticos .....	41
2.4 Técnicas de modelagem.....	42
2.4.1. Random Forest (RF).....	43
2.4.2. Lasso .....	43
2.4.3 Krigagem com Regressão (KR).....	44
2.4.4 Abordagem Híbrida .....	45
2.4.5 Abordagem Híbrida .....	46
2.5 Distribuição do estoque de carbono arbóreo no cerrado de Minas Gerais.....	47
<b>3. RESULTADOS</b> .....	47
3.1 Avaliação das diferenças técnicas .....	47

3.1.1 RF .....	47
3.1.2 KR.....	49
3.1.3 Lasso .....	50
3.1.4 Abordagem híbrida.....	51
3.2. Comparação entre as diferentes técnicas .....	53
3.2.1 Avaliação do ganho de precisão de cada etapa da abordagem híbrida.....	54
3.3 Distribuição do estoque de carbono arbóreo no cerrado.....	55
4. DISCUSSÃO .....	56
5. CONCLUSÃO.....	57
REFERÊNCIAS.....	59
<b>ARTIGO 2: REDUZINDO O EFEITO DA INCERTEZA NA MODELAGEM DA</b>	
<b>DISTRIBUIÇÃO DO ESTOQUE DE CARBONO NO CERRADO: UMA</b>	
<b>ABORDAGEM QUE COMBINA MODELAGEM DE CRESCIMENTO E</b>	
<b>MODELAGEM GEOESPACIAL.....</b>	
	<b>63</b>
1. INTRODUÇÃO .....	65
2. MATERIAL E MÉTODOS .....	66
2.1 Caracterização da área de estudo .....	66
2.2. Amostragem do estoque de carbono arbóreo do cerrado de Minas Gerais .....	67
2.3. Variáveis bioclimáticas e de sensoriamento remoto .....	69
2.4 Sistema de modelagem .....	69
2.4.1. Abordagem híbrida .....	70
2.4.2. Abordagem de crescimento e predição (CP).....	72
2.4.3. Precisão das técnicas de modelagem.....	74
3. RESULTADOS.....	75
3.1 Avaliação estatística das diferentes abordagens .....	75
3.1.1 Abordagem híbrida .....	75
3.1.2 Abordagem de CP.....	78
3.2 Comparação entre os diferentes métodos.....	82
3.2.1 Avaliação das abordagens aplicadas aos resultados de inventários atualizados .....	83
3.3 Incertezas associadas às estimativas geradas pelas diferentes abordagens.....	83
4. DISCUSSÃO .....	85

<b>5. CONCLUSÃO.....</b>	<b>86</b>
<b>REFERÊNCIAS.....</b>	<b>87</b>

**PRIMEIRA PARTE**

## 1 INTRODUÇÃO

O mapeamento da distribuição do estoque de carbono em florestas nativas através de modelagem matemática é cada dia mais popular (MORAIS *et al.*, 2020). Em especial, a dificuldade imposta, desde esforço de campo até mesmo ao dispêndio de recurso financeiro na realização de inventários em grande escala vem acarretando a construção de redes amostrais otimizadas e que naturalmente são extrapoladas através de modelagem para o conhecimento espacializado da variável de interesse.

Técnicas como krigagem ordinária, stepwise, e mais recentemente krigagem com regressão, lasso e random forest vêm sendo aplicadas para fins de modelagem da distribuição do estoque de carbono em florestas atlânticas, amazônica, cerrado, caatinga, etc (SCOLFORO *et al.*, 2016; CARVALHO *et al.*, 2019; DAVID, *et al.*, 2019; SILVEIRA *et al.*, 2019).

Essas técnicas, no entanto, podem apresentar modelos com certo desequilíbrio entre o poder preditivo x consistência biológica (SABATIA; BURKHART, 2014). Sendo assim, exemplos de modelos com poder preditivo, porém não necessariamente controlando o efeito de colinearidade entre as variáveis preditoras estão disponíveis na literatura, assim como modelos explicativos, mas com baixo poder preditivo (ALTOÉ, 2017; SILVEIRA *et al.*, 2019).

Diante dessa perspectiva, esta tese objetivou o desenvolvimento de abordagens que utilizam diferentes técnicas visando a geração de modelos explicativos e com boa capacidade preditiva para estimativa de carbono em fragmentos de Cerrado, oriundos do Inventário Florestal de Minas Gerais.

O artigo 1 apresenta uma abordagem que permite a geração de estimativas com consistência biológica e ao mesmo tempo com precisão média superior as técnicas usualmente aplicadas em estudos de modelagem de distribuição de estoque de carbono arbóreo, como random forest e krigagem com regressão.

O artigo 2 objetivou desenvolver uma abordagem que combina a modelagem de crescimento e a modelagem geoespacial através da combinação de todos os dados auxiliares disponíveis e correlacionados a variável de interesse. No artigo 2, fica claro como a nova abordagem, em especial pelo aproveitamento de inventários passados, reduz consideravelmente as incertezas das estimativas de estoque de carbono, o que permite a gestão fragmento florestal - específica para tomada de decisão pública.

## **2 REVISÃO DE LITERATURA**

### **2.1 Cerrado**

O cerrado é o segundo maior bioma do país, e é classificado como a maior savana neotropical do mundo (OLIVEIRA et al., 2006). Sua área original abrangia mais de 2 milhões de Km<sup>2</sup> (EITEN, 1993), ocupando cerca de 23,9% do território brasileiro, sendo encontrada em toda extensão do território, abrangendo mais de 15 estados, desde Roraima e amapá, ao norte, até São Paulo e Paraná, mais ao sul (BRASIL, 2009).

A vegetação deste bioma é caracterizada por cinco fitofisionomias, sendo elas: Campo Limpo, Campo Sujo, Campo Cerrado, Cerrado Sensu Stricto e Cerradão (COUTINHO, 1978). Originalmente, o Cerrado Sensu Stricto dominava 65% desta formação, o Cerradão ocupava apenas 1%, e o restante (34%), abrangiam os demais tipos fisionômicos (JUNIOR et al, 2005). As diferenças nas fitofisionomias do Cerrado ocorrem principalmente devido a fatores edáficos, fatores climáticos, presença do fogo e ações antrópicas (EITEN 1972; OLIVEIRA-FILHO et al. 1990).

Estas diversas formações indicam que o Cerrado tem uma alta diversidade florestal, sendo que o número de plantas vasculares encontrados no mesmo é superior ao da grande maioria dos biomas do mundo, e o endemismo de sua flora chega a 44% (KLINK e MACHADO, 2005). No entanto, por ser uma vegetação relativamente baixa, com galhos tortuosos, o Cerrado era considerado de baixa prioridade para conservação, e os interesses para estudo desta região tiveram início apenas na década de 80 (PRIMAK e RODRIGUES, 2001).

A demora para conscientização de que o bioma Cerrado era diverso e rico fez com que grande parte de sua vegetação natural fosse perdida. Estima-se que 67% da mesma tenha sido desmatada, restando apenas fragmentos isolados em determinados pontos (BRASIL, 1999). Esta degradação ocorreu principalmente devido a pecuária, expansão da fronteira agrícola, queimadas, cortes seletivos e introdução de espécies exóticas (RIBEIRO e WALTER, 1998), além da ocupação humana desordenada (FELFILI et al., 2002).

O Cerrado, portanto, é hoje classificado como um hotspot mundial, e apesar de seu alto índice de endemismo e seu acelerado processo de fragmentação, apenas 2,61% de suas áreas são protegidas (ICMBio, 2020). Como agravante da situação, ainda há uma

falta de informações sobre a distribuição e organização de suas comunidades (ASSUNÇÃO e FELFILI, 2004), especialmente no estado de Minas Gerais, onde poucos estudos foram realizados (NERI et al., 2007).

O Cerrado ocupa cerca de 60% do estado de Minas Gerais, e sua fitofisionomia de maior expressão é o *Sensu Stricto* (RUFINI et al., 2010). Este é caracterizado por um estrato arbóreo-arbustivo e herbáceo-graminoso (EITEN, 1994), com sua cobertura vegetal podendo variar de 10 a 60% (FELFILI e FELFILI, 2001), e ocorre sobre Latossolos e Neossolos (HARIDASAN, 1992). Segundo Ribeiro e Walter (1998) suas árvores variam de tamanho entre 3 a 6 metros, são tortuosas e ocorrem em locais com evidências de queimadas.

Além desta fitofisionomia, Minas Gerais conta ainda com o Cerradão e o Campo Cerrado. Segundo o IBGE (2012), o Cerradão é uma vegetação florestal, que ocorre em locais com estações do ano bem definidas, com solos profundos (areníticos lixiviados), tendo uma vegetação com porte arbóreo chegando até 10 metros de altura, que difere das Florestas Estacionais apenas devido a sua composição florística. Já o Campo Cerrado encontra-se em solos rasos, normalmente com afloramentos rochosos (RIBEIRO E WALTER, 1998), com a presença de uma flora esclerófila adaptada (MENDONÇA e LINS, 2000).

## **2.2 Inventário Florestal de Minas Gerais**

A composição de plantas de uma região varia com relação a idade, tamanho, espécie e crescimento (SCOTT; GOVE 2014). No estado de Minas Gerais por exemplo, existem vegetações combinadas de forma extremamente rica, como áreas de campo, campo cerrado, campo rupestre, cerrado, cerradão, Floresta Estacional Decidual Montana, Floresta Estacional Decidual Submontana, Floresta Estacional semidecidual Montana, Floresta Estacional Semidecidual Submontana, Floresta ombrófila Alto Montana, Floresta Ombrofila Montana, Floresta Ombrofila Submontana, e regiões de vereda (CARVALHO *et al.*, 2008).

Cabe a execução de um inventário florestal então a tarefa de amostrar as áreas conforme a característica de interesse para o real conhecimento de diferentes parâmetros vegetacionais (SCOTT; GOVE 2014). O inventário florestal de Minas Gerais nasceu do convênio entre o Instituto Estadual de Florestas (IEF), a Universidade Federal de Lavras

(UFLA) e a Fundação de Apoio ao Ensino, Pesquisa e Extensão (FAEPE), em 2003, quando ocorreu o mapeamento da flora nativa e reflorestamento no estado (CARVALHO; SCOFORO, 2008). O inventário de campo ocorreu entre 2005 e 2008.

O Inventário Florestal de Minas Gerais gerou importantes informações com relação a qualidade dos fragmentos florestais, o estoque de carbono arbóreo, a quantidade e localização de áreas desmatadas, a riqueza de espécies, etc (CARVALHO *et al.*, 2008). Abrangeu 169 fragmentos de vegetação no estado, com 4.456 parcelas instaladas e inventariadas (ALTOÉ, 2017). Adicionalmente, em 2010 houve a remedição de parte dos fragmentos inventariados entre 2005-2008, além do levantamento do estoque de carbono de solos, raízes e serapilheira (ALTOÉ, 2017).

A partir do Inventário Florestal de Minas Gerais também foi possível a construção do Zoneamento Ecológico Econômico do estado de Minas Gerais. Este zoneamento possibilitou o aprimoramento do licenciamento ambiental do estado, além de garantir maior segurança na tomada de decisão político-sócio-econômica. Por fim, outra consequência relevante do Inventário Florestal de Minas Gerais foi a descoberta de 14 novas espécies arbóreas, além de outras 64 espécies que nunca haviam sido noticiadas no estado de Minas Gerais (SEMAD, 2008).

### **2.3 Modelagem da distribuição do estoque de carbono**

O sequestro de carbono tem sido visto como a principal forma de manter em níveis toleráveis a concentração dos gases CO<sub>2</sub> na atmosfera, e desta forma mitigar as mudanças climáticas globais. Diversos estudos têm sido feitos objetivando quantificar o estoque de carbono nas diversas formações vegetais ao redor do mundo, e nos diversos biomas brasileiros (FERREIRA *et al.*, 2012, MORAIS *et al.*, 2013, FRANÇA *et al.*, 2015, SCOLFORO *et al.*, 2015, PIRANI *et al.*, 2016, ALTOÉ 2017, MORAIS *et al.*, 2017, GARDON *et al.*, 2020, ZANINI ET AL 2021).

No cerrado, entretanto, apesar de ser o segundo maior bioma brasileiro (KLINK; MACHADO, 2005), e um dos mais ameaçados pelo desmatamento no país, o entendimento a respeito do estoque de carbono se encontra em estágio inicial. O primeiro estudo quantificando o estoque de carbono em áreas em processo de restauração do cerrado no Brasil foi publicado há poucos anos, em 2006 (GARDON *et al.*, 2020).

Adicionalmente, de todos os estudos publicados no tema estoque de carbono até hoje, apenas 7% são relacionados ao bioma cerrado (GARDON *et al.*, 2020).

Estudos sobre a modelagem do estoque de carbono do cerrado também são raros (RIBEIRO *et al.*, 2011; OLIVEIRA E SCOLFORO 2017). Citando alguns estudos, SCOLFORO *et al.*, (2015) e SCOLFORO *et al.*, (2016) modelaram a distribuição do estoque de carbono arbóreo no cerrado do estado de Minas Gerais através da krigagem com regressão. Morais *et al.* (2016) estudaram a distribuição espacial e vertical do carbono em serrapilheira para uma pequena área no município de Olhos D'água, Minas Gerais. Ribeiro *et al.* (2011) estudaram o estoque de carbono abaixo e acima do solo para uma pequena área na cidade de Curvelo, Minas Gerais. Outros estudos foram realizados modelando o estoque de carbono ou biomassa no cerrado fora de Minas Gerais, por exemplo, no Distrito federal (REZENDE *et al.*, 2006, FRANÇA; PAIVA 2015) e Bahia (OLIVEIRA *et al.*, 2019).

Desta forma, percebe-se que ainda há muito o que se estudar sobre a vegetação do bioma cerrado, considerando sua total extensão tanto em Minas Gerais, como no Brasil. Para uma modelagem sólida e confiável uma extensa base de dados é necessária. A modelagem precisa do estoque e incremento de carbono ao longo dos anos, de acordo com as condições ambientais e de manejo, pode ser entendida como uma ferramenta imprescindível na mitigação do aquecimento global.

### **2.3.1 Random Forest (RF)**

O RF é uma técnica introduzida a literatura por Breiman (2001). Esta técnica é amplamente utilizada em estudos de modelagem de índice de sítio (SABATIA; BURKHART, 2014), classificação de espécies (CARVALHO *et al.*, 2019), modelagem espacial de volume em plantios de eucalipto (REIS *et al.*, 2019), modelagem espacial de estoque de carbono em biomas nativos (SILVEIRA *et al.*, 2019) etc.

Cutler *et al.* (2007) aplicaram RF para classificação ecológica e quando compararam o RF às demais técnicas (classification trees, regressão logística aditiva e regressão logística) reportaram o RF como: acurado para classificação, consistente para determinação da importância das variáveis, flexível para modelagem de atributos com interações complexas, etc. Mutanga *et al.* (2012) reportaram como o RF acomodou bem tanto a seleção de variáveis, como a predição precisa da biomassa em vegetações alagadas, sendo esta técnica superior ao stepwise.

Por outro lado, apesar de todas as vantagens amplamente reportadas na literatura, alguns cuidados precisam ser tomados. Sabatia & Burkhart (2014) reportaram que apesar de estatísticas precisas na modelagem de índice de sítio em plantios de *Pinus taeda* no sudeste do EUA, os resultados a partir de RF podem apresentar certa inconsistência. Como exemplo, os autores avaliaram como o modelo resultante se comportaria quando simulado melhores regimes hídricos para a região, e notaram a ocorrência de estimativas biologicamente inconsistentes em diversos locais.

De maneira geral, o RF parte de uma amostra aleatória baseada nas variáveis estudadas e a partir dessa amostra aleatória faz crescer uma árvore. Repetindo-se a amostragem aleatória diversas vezes resulta em várias árvores. Por essa lógica de execução, a expectativa é que ao se selecionar as amostras de forma independente, os resultados tendem a serem menos correlacionados entre si e, portanto, a árvore ao final será combinada a partir de resultados que são de certa forma independentes.

### **2.3.2 Least absolute shrinkage selection operator (lasso)**

Dentre as diferentes técnicas de aprendizado de máquina, o lasso é uma que vem ganhando cada vez maior popularidade (ALTOÉ, 2017). Uma das principais vantagens é a possibilidade dos coeficientes serem explícitos, uma vez que está é uma técnica semiparamétrica. Por técnica semi-paramétrica entende-se que apesar dos coeficientes serem reportados, os mesmos sofrem encolhimento para mitigação de colinearidade e, assim, essa técnica introduz viés ao valor esperado de cada coeficiente (TIBSHIRANI, 1996). De maneira geral, o lasso é uma técnica que pelo encolhimento de parâmetros permite mitigar colinearidade entre as variáveis preditoras e, também selecionar as variáveis preditoras mais importantes para estimativa da variável de interesse (TIBSHIRANI, 1996).

Moreno et al. (2017) aplicaram lasso para o ajuste de modelos de crescimento ao nível de árvore individual para a variável DAP no Chile. Os autores reportaram diversas vantagens, uma vez que lasso permitiu a geração de modelos parcimoniosos e com bom grau de explicação. Os mesmos autores, entretanto, reportaram a dificuldade de avaliação dos parâmetros ajustados por lasso, uma vez que quando esta técnica não é utilizada numa perspectiva Bayesiana, ela não gera intervalos de confiança e mesmo erro padrão dos

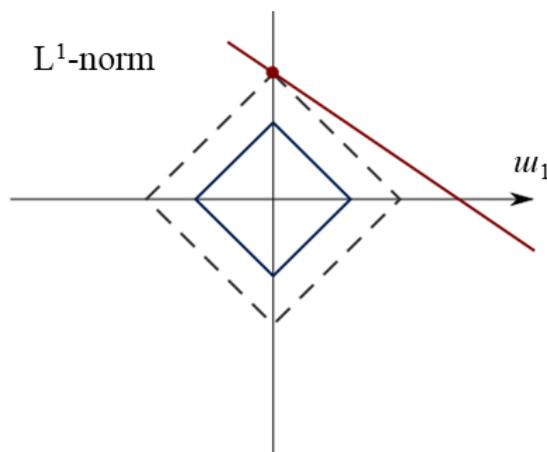
coeficientes o que de certa forma tende a comprometer sua utilização para fins de inferência.

Altoé (2017) aplicando lasso na modelagem da distribuição do estoque de carbono de raízes, solo, serrapilheira, arbóreo em diferentes biomas no Brasil reportou modelos com seleção adequada de variáveis, que naturalmente puderam ser utilizados para fins de entendimento dos principais drivers que impactam espacialmente a distribuição de carbono nos diferentes compartimentos florestais.

Historicamente, stepwise sempre foi a técnica utilizada para seleção de variáveis. Esta técnica, porém, só funciona de forma adequada quando somente poucas variáveis preditoras são altamente correlacionadas à variável de interesse. Caso contrário, stepwise tende a gerar modelos superparametrizados. Uma outra técnica popular é a regressão ridge, a qual evita ajuste de modelos superparametrizados pelo encolhimento dos parâmetros (o que mitiga colinearidade) a partir da soma dos quadrados dos mesmos ser inferior a um valor  $\lambda$  (SCOLFORO *et al.*, 2017). Lasso combina a propriedade de seleção de variáveis tal qual stepwise, mas garantindo a mitigação de colinearidade entre variáveis preditoras (tal qual regressão ridge) a fim de gerar um modelo parcimonioso.

Sendo assim, o lasso penaliza os parâmetros do modelo de forma que sua soma absoluta seja menor que o parâmetro de penalização ( $\lambda$ ). Em outras palavras, o Lasso penaliza os parâmetros do modelo como se fosse baseado em um método geométrico (Figura 1).

**Figura 1.** Representação geométrica do lasso.



### 2.3.3 Krigagem com regressão

Krigagem com regressão combina a aplicação de uma regressão baseada em variáveis bioclimáticas, geográficas e/ou de sensoriamento remoto para predição da variável de interesse, com a consequente modelagem da autocorrelação espacial presente nos resíduos advindos do modelo de regressão (SCOLFORO *et al.*, 2015).

O modelo de regressão pode ser realizado de diversas maneiras, como por exemplo através de stepwise. Morais *et al.* (2020), utilizaram a técnica de krigagem com regressão para estimar o estoque de carbono nos solos do cerrado brasileiro, baseando o modelo de regressão em stepwise para a seleção das variáveis preditoras mais correlacionadas à variável de interesse e que maximizassem o teste de F. Scolforo *et al.* (2016), por outro lado, utilizaram krigagem com regressão para o mapeamento do carbono arbóreo de diferentes biomas brasileiros, partindo do ajuste de um modelo linear em função de variáveis dummy e geográficas.

A krigagem ordinária dos resíduos advindos do modelo de regressão tornou-se uma prática comum para geração de estimativas mais precisas e com menores incertezas, em especial para as zonas de extremo (MELLO *et al.*, 2013). O erro das predições dos modelos de regressão por vezes eram tratados como aleatórios. Muitas vezes, porém, uma fração destes apresentam autocorrelação espacial, o que quer dizer que parte do resíduo até então tratado como aleatório pode ser explicado e assim tornar o processo de estimativa mais preciso.

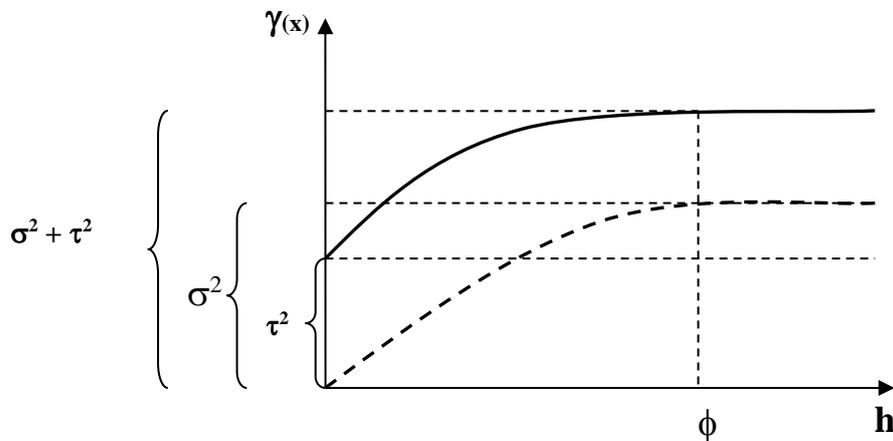
De maneira geral, a existência de autocorrelação espacial é verificada através do semivariograma, sendo este definido como:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x) - Z(x+h)]^2 \quad (1)$$

Em que:  $\hat{\gamma}(h)$  é a semivariância estimada para cada distância ( $h$ ),  $N(h)$  é o número de pares de ponto separados pela distância  $h$ ,  $Z(x)$  é o valor da variável regionalizada no ponto  $x$  e  $Z(x+h)$  é o valor no ponto  $x+h$ .

O semivariograma permite avaliar a existência ou não da autocorrelação espacial dos resíduos (Figura 2). Se verificada a autocorrelação espacial da variável de interesse, então um modelo explicativo é ajustado, como por exemplo, o exponencial que é o mais utilizado para fins florestais (SCOLFORO *et al.*, 2016).

**Figura 2.** Semivariograma típico para dados sem tendência e com ruído.  $\tau^2$  é o efeito pepita;  $\sigma^2$  é o patamar; e.  $\phi$  é o alcance.



Uma vez verificada a autocorrelação espacial da variável de interesse com o consequente ajuste do modelo exponencial, torna-se possível através dos parâmetros deste modelo a interpolação residual com base na sua posição geográfica através do uso da krigagem.

### 2.3.4 Modelos de crescimento

McTague et al. (2008) reportaram modelos de crescimento para florestas de harwood no sudeste dos EUA. O modelo que os autores reportaram no seu formato de projeção foi:

$$Y_2 = Y_1 + a^t \times e^{b+cX} \quad (2)$$

Em que:  $Y_2$  é variável de interesse projetada – estoque de carbono, por exemplo;  $Y_1$  é a variável de interesse no momento atual;  $t$  é o intervalo de tempo para a projeção em anos;  $X$  é a variável (ou conjunto de variáveis) independentes;  $a, b$  e  $c$  são os coeficientes a serem ajustados.

Para facilitar a convergência do modelo, assim como para verificação dos pressupostos de normalidade residual, o mesmo pode ser facilmente linearizado:

$$\ln(Y_2 - Y_1) = a + b \times t + c \times X \quad (3)$$

Em que: variáveis foram previamente definidas.

Enquanto no hemifério norte a existência de modelos de crescimento ou mesmo sistemas de crescimento para espécies ou fragmentos florestais são amplamente disponíveis na literatura (MCTAGUE *et al.*, 2008; GÁLVEZ *et al.*, 2014), o mesmo não pode ser dito para florestas nativas brasileiras. Scolforo *et al.* (2017), inclusive desenvolveu uma abordagem para modelagem de crescimento de DAP ao nível de espécie para florestas atlânticas brasileiras que permite a extrapolação de projeções pela falta de dados históricos.

### 2.3.5 Máxima verossimilhança

Máxima verossimilhança é um método de estimativa de parâmetros a partir da maximização da função da verossimilhança de forma que os dados observados sejam os mais prováveis no modelo estatístico definido (MONTES, 2012). De forma a computar a verossimilhança para cada observação, a função de densidade probabilística normal geralmente tende a ser usada:

$$\lambda(\theta, \sigma | x_i, y_i) = \sum_{i=1}^n \log \left[ \frac{1}{\sqrt{2\pi}\sigma} \exp \left( \frac{-y_i + f(x_i, \theta)}{\sigma^2} \right) \right] \quad (4)$$

Em que:  $\lambda$  é a verossimilhança em logaritmo neperiano;  $f(\cdot)$  é a função do modelo com a expectativa do parâmetro  $\theta$ ;  $y_i$  é o vetor para a variável dependente de  $i = 1 \dots n$ ;  $x_i$  é o vetor com as variáveis independentes de  $i = 1 \dots n$ ;  $\sigma^2$  representa o parâmetro da variância;  $\pi$  é uma constante.

A máxima verossimilhança é um método flexível de ajuste, uma vez que permite especificações mais complexas de modelos em relação aos mínimos quadrados ordinários (SCOLFORO *et al.*, 2020). Mínimos quadrados ordinários é um caso especial da máxima verossimilhança onde o  $\varepsilon \text{ iid} \sim N(0, \sigma^2)$ . O problema, no entanto, é que se qualquer pressuposto não for atendido no ajuste através dos mínimos quadrados ordinários,

qualquer nível de inferência (análise de incerteza) pode ficar comprometida (MONTES, 2012).

#### **2.4 Quantificação da incerteza de estimativa**

Mapeamento do estoque de carbono é mais desafiador do que o apresentado na literatura muitas vezes. Um dos maiores problemas nos mapeamentos de estoque de carbono é que por vezes a tendência do estoque apresentada na escala espacial é coerente, entretanto, quando avaliada as estimativas de forma local as mesmas apresentam alto grau de erro (incerteza). Nyström et al. (2015) reportaram que abordagens de assimilação de dados são ferramentas poderosas para o aumento de precisão das estimativas locais, o que naturalmente restringe o nível de incerteza das mesmas. Os autores aplicaram a técnica de assimilação de dados de Kalman filter, o qual permitiu com que dados de inventários atuais, assim como de inventários passados, sensoriamento remoto, clima, etc, fossem utilizados para quantificação de parâmetros florestais.

Kinane et al. (2021) propôs a redução da incerteza das estimativas de índice de área foliar em plantios de *Pinus taeda* no sudeste dos EUA a partir de combinação de diversas variáveis auxiliares, entre elas a série temporal de NDVI de imagens Landsat 5 e 7, além dos índices de áreas foliares reportados em medições anteriores.

Percebe-se que por mais que incerteza não seja retratada como um problema na quantificação de algum atributo florestal nos diversos biomas brasileiros, esse efeito tem relevância considerável, porém por enquanto não vem sendo retratado nos variados modelos propostos na literatura.

### **3. CONSIDERAÇÕES GERAIS**

A evolução computacional nos últimos anos vem permitindo com que novas técnicas sejam apresentadas a literatura, assim como vem permitindo o teste exaustivo destas vs as técnicas tradicionalmente aplicadas na modelagem de atributos naturais. Desta forma, a precisão requerida na modelagem do estoque de carbono ganha maior relevância a cada dia, em especial, pelos vários estudos que se baseiam nessa variável para compreensão dos eventos de longo prazo de mudanças climáticas.

Sendo assim, técnicas de modelagem que permitam a captura precisa da tendência de distribuição espacial do estoque de carbono se fazem relevantes para a definição de zonas prioritárias de conservação. Ao mesmo tempo, porém, o desenvolvimento de novas técnicas que também reflitam com a menor incerteza possível a distribuição de estoque carbono num gradiente espacial se fazem necessárias para permitir tomadas de decisão dos gestores públicos ao nível de fragmento florestal.

## REFERÊNCIAS

- ALTOÉ, T. F. **Modelagem e espacialização do estoque de carbono em fragmentos nativos de Minas Gerais em função de fatores edafoclimáticos**. 2017. 130 p. Tese. (Doutorado em Engenharia Florestal) – Universidade Federal de Lavras, Lavras, 2017.
- ASSUNÇÃO, S. L.; FELFILI, J. M. Fitossociologia de um fragmento de cerrado sensu stricto na APA de Paranoá, DF, Brasil. **Acta Botanica Brasilica**, São Paulo, v. 18, n. 4, p. 903-910, 2004.
- BRASIL. Ministério do Meio Ambiente, dos Recursos Hídricos e da Amazônia Legal. **Ações prioritárias para a conservação da biodiversidade do Cerrado e Pantanal**. Brasília, DF: Ventura Comunicações e Cultura, 1999. 24 p.
- BREIMAN, L. **Random forest**. Machine Learning, Boston, v. 45, p. 5-32, 2001.
- CARVALHO, L. M. T. de *et al.* Análises espaciais e estatísticas da flora nativa. In: CARVALHO, L. M. T.; SCOLFORO, J. R. (Ed.). **Inventário florestal de Minas Gerais: monitoramento da flora nativa 2005-2007**. Lavras: Ed. UFLA, 2008. cap. 2, p.25-316.
- CARVALHO, L. M. T.; SCOLFORO, J. R. (Org.). **Inventário florestal de Minas Gerais: monitoramento da flora nativa 2005-2007**. Lavras: Ed. UFLA, 2008. v. 1. 357p.
- CARVALHO, M.C. *et al.* Algoritmos de aprendizagem de máquina na modelagem da distribuição potencial de habitats de espécies arbóreas. **Nativa**, Sinop, v. 7, n. 5, p. 600-606, 2019. DOI: <http://dx.doi.org/10.31413/nativa.v7i5.7214>.
- COUTINHO, L. M. **O conceito de Cerrado**. Revista Brasileira de Botânica, São Paulo, v. 1, p. 17-24, 1978.
- CUTLER, R. D. *et al.* **Random forest for classification in ecology**. Ecology, Durham, n. 11, p. 2783-2792, 2007.
- EITEN, G. Vegetação do cerrado. In: PINTO, M. N. **Cerrado: caracterização, ocupação e perspectivas**. Brasília, DF: UnB, p. 11-73. 1994
- FELFILI, J. M.; NOGUEIRA, P. E.; SILVA-JUNIOR, M. C.; MARIMON, B. S.; DELITTI, W. B. C. Composição florística e fitossociológica do Cerrado sentido restrito

no município de Água Boa, MT. **Acta Botanica Brasilica**, São Paulo, v. 16, n. 1, p.103-112, 2002.

FELFILI, M. C.; FELFILI, J. M. Diversidade alfa e beta no cerrado sensu stricto da Chapada Pratinha, Brasil. **Acta Botanica Brasilica**, São Paulo, v. 15, n. 2, p. 243-254, 2001.

FRANÇA, A. M. D. S.; PAIVA, R. J. D. O. Estimativa e modelagem dos estoques de carbono em solos sob áreas de campo limpo úmido do Distrito Federal. **Sociedade & Natureza**, Uberlândia, v. 27, n. 1, p.171-184, jan-abr, 2015. DOI: <http://dx.doi.org/10.1590/1982-451320150112>

GÁLVEZ, F.B. *et al.* Using climate-FVS to project landscape-level forest carbon stores for 100 years from field and LiDAR measures of initial conditions. **Carbon Balance Manage.** Moscow. v.9, n. 1, p.1-13. 2014. DOI: 10.1186/1750-0680-9-1

GARDON, F. R.; DOS SANTOS, R. F.; RODRIGUES, R. R. Brazil's forest restoration, biomass and carbon stocks: A critical review of the knowledge gaps. **Forest Ecology and Management**, Amsterdam, v. 462, p.9 . 2020. DOI: 10.1016/j.foreco.2020.117972.

HARIDASAN, M. Observation sonsoils, foliar nutriente concentration and floristic composition of cerrado sensu stricto and cerradão communities in central Brazil. Pp.171-184. In: P.A. Furley; J. Proctor & J.A. Ratter (eds.). **Nature and Dynamics of Forest-Savanna Boundaries**. London, Chapman & Hall Publishing. 1992.

JUNIOR, B. H. M.; MUNDAYATAN, H. Comparação da vegetação arbórea e características edáficas de um cerradão e um cerrado sensu stricto em áreas adjacentes sobre solo distrófico no leste de Mato Grosso, Brasil. **Acta Botanica Brasilica**. Alta Floresta. v. 19. v. 4, p. 913-926. 2005.

KINANE, S.M.; MONTES, C.R.; ALBAUGH, T.J.; MISHRA, D.R. A Model to Estimate Leaf Area Index in Loblolly Pine Plantations Using Landsat 5 and 7 Images. **Remote Sens.** Bristol. v. 13, p. 1140. DOI: 10.3390/rs13061140

KLINK, C. A.; MACHADO, R. B. A conservação do cerrado brasileiro. **Mega diversidade**, Belo Horizonte, v. 1, n. 1, p. 147-155, jul. 2005.

MACHADO, R.B. *et al.* Estimativas de perda da área do Cerrado brasileiro. Relatório técnico não publicado. **Conservação Internacional**, Brasília, DF. 2004.

MANUEL E. *et al.* Modeling landscape dynamics in the central Brazilian savanna biome: future scenarios and perspectives for conservation, **Journal of Land Use Science**, v.8, n. 4, p. 403-421, Berlin. 2013. DOI: 10.1080/1747423X.2012.675363

MCTAGUE, J.P.; O'LOUGHLIN, D.; ROISE, J.P.; ROBISON, D.J.; KELLISON, R.C. The SOHARC model system for growth and yield of southern hardwoods. **Southern Journal of Applied Forestry**. Virginia. v.32, n. 4, p. 173–183. 2008.

MELLO C. R., VIOLA M. R., BESKOW S., NORTON L. D. Multivariate models for annual rainfall erosivity in Brazil. **Geoderma**, Amsterdam, v. 202/203, p. 88-102, jul. 2013.

MENDONÇA, M. P.; LINS, L. V. **Lista vermelha das espécies ameaçadas de extinção da flora de Minas Gerais**. Belo Horizonte: Fundação Biodiversitas; Fundação Zoológica de Belo Horizonte, 2000. 160 p.

MONTES, C. A **Resource Driven Growth and Yield Model for Loblolly Pine Plantations**. Ph.D. Thesis, North Carolina State University, Raleigh, NC, USA, 2012

MORAIS, V. A. *et al.* Spatial and vertical distribution of litter and belowground carbon in a Brazilian Cerrado vegetation. **Cerne**, Lavras. v. 23, n.1, p. 43-52. 2017. DOI: DOI: 10.1590/01047760201723012247.

MORAIS, V. A. *et al.*, Spatial distribution of soil carbon stocks in the Cerrado biome of Minas Gerais, Brazil. **Catena**, [S.l.], v.185, feb. 2020. DOI: 10.1016/j.catena.2019.104285

MORAIS, V. A., *et al.*, Modelagem do teor de carbono orgânico em solos de fragmentos de cerrado de Januária e Bonito de Minas, Minas Gerais. **Pesquisa Florestal Brasileira**, Colombo. v. 33, n. 76, p. 343-354. out./dez. 2013. DOI: 10.4336/2013.pfb.33.76.507.

MORENO, P.C.; PALMAS, S.; ESCOBEDO, F.J.; CROPPER, W.P.; GEZAN, S.A. Individual-Tree Diameter Growth Models for Mixed *Nothofagus* Second Growth Forests in Southern Chile. **Forests**, Basel, v. 8, n. 12, p. 506. 2017. DOI: 10.3390/f8120506

MUTANGA, O.; ADAM, E., CHO, M. A. High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. **International Journal of Applied Earth Observation and Geoinformation**. v.18, p. 399-406. aug. 2012.

NERI, A. V. *et al.* Análise da estrutura de uma comunidade lenhosa em área de cerrado sensu stricto no município de Senador Modestino Gonçalves, Norte de Minas Gerais, Brasil. **Revista Árvore**, Viçosa, v. 31, n. 1, p. 123-134, 2007.

NYSTRÖM, M. *et al.*, Data Assimilation in Forest Inventory: First Empirical Results. **Forests**, Basel, v. 6, p. 4540–4557, 2015. DOI: 10.3390/f6124384.

OLIVEIRA FILHO, A. T.; SCOLFORO, J. R. S. **Inventário florestal de Minas Gerais: espécies arbóreas da flora nativa**. Lavras: Editora da UFLA, 2008. 619 p.

OLIVEIRA, C. P. D.; FRANCELINO, M. R.; DAHER M., LELES; P. S. D. S. ANDRADE, F. C. D. Comparação de modelos estatísticos para estimativa da biomassa de árvores, e estimativa do estoque de carbono acima do solo em Cerrado. **Ciência Florestal**, Santa Maria, v. 29, n.1, p. 255-269. 2019.

OLIVEIRA, M.C. *et al.* Avaliação de diferentes níveis de intervenção na florística, diversidade e similaridade de uma área de Cerrado stricto sensu. **Cerne**, Lavras, v. 12, n.4, p. 342-349. 2006.

OLIVEIRA-FILHO, A.T.; SHEPHERD, G.J.; MARTINS, F.R; STUBBLEBINE, W.H. Environmental factors affecting physiognomic and floristic variation in an area of cerrado in central Brazil. **Journal of Tropical Ecology**. Cambridge, v.5, p. 413-451. 1990.

PIRANI, F. R. **Dinâmica e modelagem do estoque de carbono no fuste da vegetação arbustivo/arbórea de duas fitofisionomias do bioma Cerrado em Mato Grosso-Brasil**. 2016. Tese (Doutorado em Ciências Florestais) - Universidade de Brasília, Brasília, 2016.

PRIMACK, R.; RODRIGUES, E. 2001. **Biologia da Conservação**. p. 328. Efraim Rodrigues, Londrina.

REIS, A.A., FRANKLIN, S.E., MELLO, J.M., ACERBI JR, F.W. Volume estimation in a Eucalyptus plantation using multi-source remote sensing and digital terrain data: a case study in Minas Gerais State, Brazil. **International Journal of Remote Sensing**, Londres v. 40, n. 7, p. 1-2018. DOI: 10.1080/01431161.2018.1530808

REZENDE, A. V.; Vale, A. D.; Sanquetta, C. R.; Figueiredo Filho, A.; Felfili, J. M. Comparação de modelos matemáticos para estimativa do volume, biomassa e estoque de carbono da vegetação lenhosa de um cerrado sensu stricto em Brasília, DF. **Scientia Forestalis**, Piracicaba, v.71, n.2, p. 65-73. 2006.

RIBEIRO, J. F.; WALTER, B. M. T. Fitofisionomias do bioma cerrado. In: SANO, S.; ALMEIDA, S. (Eds.). **Cerrado ambiente e flora**. Planaltina: Embrapa Cerrados, 1998. p. 89-166.

RIBEIRO, S. C.; FEHRMANN, L.; SOARES, C. P. B., JACOVINE, L. A. G., KLEINN, C., DE OLIVEIRA GASPAR, R. Above-and belowground biomass in a Brazilian Cerrado. **Forest Ecology and Management**, Amsterdam, v. 262, n.3, p. 491-499. 2011. DOI : 10.1016/j.foreco.2011.04.017

RUFINI, A. L.; SCOLFORO, J. R. S.; OLIVEIRA, A. D. D.; MELLO, J. M. D. Equações volumétricas para o Cerrado Sensu Stricto, em Minas Gerais. **Cerne**, Lavras, v. 16, n. 1, p. 1-11, jan./mar. 2010. DOI: 10.1590/S0104-77602010000100001.

SABATIA, C., BURKHART, H.E. Predicting site index of plantation loblolly pine from biophysical variables. **Forest Ecology and Management**, Amsterdam, v. 326, p. 142-156, aug, 2014. DOI: 10.1016/j.foreco.2014.04.019

SCOLFORO, H. F.; SCOLFORO, J. R. S.; DE MELLO, J. M.; DE MELLO, C. R.; MORAIS, V. A. Spatial interpolators for improving the mapping of carbon stock of the arboreal vegetation in Brazilian biomes of Atlantic forest and Savanna. **Forest Ecology and Management**, Amsterdam, v. 376, p. 24-35, sep., 2016. DOI: 10.1016/j.foreco.2016.05.047

SCOLFORO, H. F.; SCOLFORO, J. R. S.; MELLO, C. R.; MELLO, J. M.; FERRAZ FILHO, A. C. 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. e0128781, 2015. DOI: 10.1371/journal.pone.0128781

SCOLFORO, H.F. *et al.* Incorporating rainfall data to better plan Eucalyptus clones deployment in eastern Brazil. **Forest Ecology and Management**, Amsterdam, v. 391, p. 145–153, may, 2017. DOI: 10.1016/j.foreco.2017.02.025.

SCOLFORO, H.F. *et al.*, A New Approach for Modeling Volume Response from Mid-Rotation Fertilization of *Pinus taeda* L. Plantations. **Forests**, Basel, v. 11, n. 646, 2020 DOI: 10.3390/f11060646

SCOTT, C. T.,; Gove, J. H. **Forest inventory**. Wiley StatsRef: Statistics Reference Online. 2014.

SEMAD - Secretaria de Estado de Meio Ambiente e Desenvolvimento Sustentável. 2008. Disponível em: <http://www.meioambiente.mg.gov.br/noticias/491-2o-combio-novas-especies-sao-destaque-de-inventario-florestal-de-minas-gerais>.

SILVEIRA, E. *et al.*, Object-based random forest modelling of aboveground forest biomass outperforms a pixel-based approach in a heterogeneous and mountain tropical environment. **Forest Ecology and Management**, Amsterdam, v. 445, n. 1, p. 96-109, 2019b. DOI: 10.1016/j.jag.2019.02.004.

TIBSHIRANI, R., Regression Shrinkage and Selection via the Lasso. **Journal of the Royal Society**. Series B (Methodological), [S.I.], v. 58, n. 1, p. 267-288. 1996.

ZANINI, A. M.; MAYRINCK, R. C.; VIEIRA, S. A.; DE CAMARGO, P. B.; RODRIGUES, R. R. The effect of ecological restoration methods on carbon stocks in the Brazilian Atlantic Forest. **Forest Ecology and Management**, Amsterdam, v. 481, p. 118734, 2021. DOI: 10.1016/j.foreco.2020.118734.

**SEGUNDA PARTE – ARTIGOS**

Os artigos foram redigidos conforme a norma para publicação periódica científica NBR  
6022 (ABNT, 2003)

## **ARTIGO 1: UM NOVO MODELO PARA DISTRIBUIÇÃO DO ESTOQUE DE CARBONO NO CERRADO**

### **RESUMO**

A evolução da tecnologia computacional permite cada dia mais o uso de técnicas sofisticadas para prever diferentes atributos naturais. Este estudo objetivou propor uma nova abordagem para estimar o estoque de carbono arbóreo, bem como contrastar e comparar esta abordagem a diferentes técnicas comumente aplicadas na literatura. Um total de 26 fragmentos nativos de Cerrado em Minas Gerais foram utilizados, onde o estoque de carbono arbóreo (Mg/ha) foi quantificado em cada um desses. Técnicas como krigagem com regressão (KR), least absolute shrinkage selection operator (Lasso) e random forest (RF) foram ajustados para este conjunto de dados. Adicionalmente uma nova abordagem é apresentada, sendo que a mesma combina: (1) lasso para seleção de variáveis preditoras que mitiguem multi-colinearidade, (2) ajuste de um modelo linear baseado nas variáveis resultantes de (1) por meio da máxima verossimilhança, e (3) modelagem da autocorrelação espacial dos resíduos de (2) por meio de krigagem ordinária. Os resultados apresentados destacam a nova abordagem como superior às demais, tanto pela precisão apresentada quanto por respeitar o comportamento biológico requerido na estimativa do estoque de carbono arbóreo. A nova abordagem (abordagem híbrida) contribui como uma técnica precisa para quantificação da distribuição do estoque de carbono em florestas tropicais.

**PALAVRAS-CHAVE:** lasso, máxima verossimilhança, modelagem, krigagem.

## **A NEW APPROACH FOR MODELING CARBON STOCK DISTRIBUTION IN THE BRAZILIAN SAVANNAH**

### **ABSTRACT**

The evolution of computational technology allows for the use of more sophisticated techniques to predict different attributes. This study aimed to propose a new approach to estimate the aboveground carbon stock, as well as to contrast and compare this approach to different techniques commonly applied in the literature. A total of 26 forest fragments in the Savannah of Minas Gerais were used in the study. Techniques such as regression kriging (KR), least absolute shrinkage selection operator (Lasso) and random forest (RF) were fit for this dataset. Additionally, a new approach is presented, which combines: (1) lasso for variable selection, (2) linear model fitting based on the selected variables resulting from (1) by means of maximum likelihood, and (3) ordinary kriging of the residuals of (2). The results highlighted the new approach as superior to the others, both for the precision presented and for respecting the biologically soundness required when estimating the aboveground carbon stock. The new approach (hybrid approach) is precise for quantifying the carbon stock distribution in tropical forests.

**KEYWORDS:** lasso, maximum likelihood, modeling, kriging.

## 1. INTRODUÇÃO

O cerrado é um dos exemplos da grande diversidade das florestas brasileiras, porém sua presença em regiões de maiores perturbações antrópicas destaca o desafio governamental para a manutenção da sua cobertura florestal e estoque de carbono, sendo estes componentes fundamentais contra as mudanças climáticas (ZANINI *et al.*, 2021).

Em especial, apesar da quantificação precisa da cobertura vegetal dos diversos remanescentes florestais ser possível através de imagens de satélite (PEREIRA *et al.*, 2019), a quantificação do estoque de carbono é desafiadora dada a variabilidade florestal existente no cerrado brasileiro (SILVEIRA *et al.*, 2019a).

De maneira geral uma das formas mais efetivas e consolidadas para controlar a variabilidade e permitir a inferência apropriada do estoque de carbono é através de uma rede de amostragem devidamente estratificada (SCOLFORO *et al.*, 2016). A partir da presença de uma rede amostral que contemple o gradiente ambiental (estágio sucessional, fitofisionomia, etc) do cerrado brasileiro ou qualquer outro bioma torna-se possível modelar e espacializar a distribuição do estoque de carbono em locais não amostrados. A espacialização do estoque de carbono, inclusive, é essencial por permitir a tomada de decisão pública de forma que esta garanta a proteção das maiores fontes de sumidouro de carbono (SCOLFORO *et al.*, 2015).

Diversas técnicas vêm sendo aplicadas para modelagem da distribuição do estoque de carbono de florestas nativas ao redor do mundo. A mais comum é a geração de modelos que correlacionem o estoque de carbono com variáveis bioclimáticas (SILVEIRA *et al.*, 2019B), as quais estão geralmente disponíveis em escalas refinadas e com grande precisão (ALTOÉ, 2017).

Um problema recorrente é que, de maneira geral, as variáveis bioclimáticas são colineares entre elas (ALTOÉ, 2017). Sendo assim, a modelagem da distribuição do estoque de carbono em função dessas variáveis pelos mínimos quadrados ordinários ou mesmo pela máxima verossimilhança podem levar a estimativas imprecisas, uma vez que essas técnicas por não terem quaisquer propriedades que mitiguem a colinearidade entre as variáveis preditoras, além de não serem aptas para seleção de variáveis, podem apresentar sinais inadequados para os coeficientes ajustados (SCOLFORO *et al.*, 2017). Scolforo *et al.* (2017), inclusive, preponderou tal efeito pela modelagem do índice de sítio de plantios de eucalipto no sudeste brasileiro.

A técnica de krigagem com regressão (KR) é uma das formas mais comuns e aceitas para modelagem da distribuição do estoque de carbono em função de variáveis bioclimáticas. Essa técnica geralmente se baseia no uso de regressão por stepwise. Richardson et al. (2017), por exemplo, reportaram resultados precisos de modelagem de carbono em função de variáveis geográficas selecionadas a partir do uso de stepwise. Lima et al. (2020) concluíram que a técnica de stepwise foi bem sucedida para espacialização do estoque de carbono em florestas tropicais brasileiras. De maneira geral a técnica de stepwise confere bons resultados, especialmente pela seleção de variáveis que maximizem o teste de F do modelo resultante. Geralmente seguido de stepwise, faz-se a krigagem para modelagem da autocorrelação espacial dos resíduos (Morais et al., 2020). Morais et al. (2020), por exemplo, reportaram resultados precisos para o mapeamento do estoque de carbono nos solos do cerrado a partir do uso da KR. Na KR, no entanto, vale ressaltar que tanto a krigagem ordinária dos resíduos quanto stepwise incorrem na ausência de propriedades que mitiguem a colinearidade entre as variáveis predictoras para quantificar a distribuição espacial do estoque de carbono.

Uma outra técnica cada vez mais utilizada e reportada na literatura é o random forest (RF). Silveira et al. (2019c), inclusive, discutiram a qualidade dos resultados da modelagem de distribuição espacial de carbono em florestas brasileiras a partir do RF. Mascaro et al. (2014) também estimaram o estoque de carbono na Amazônia Ocidental através de RF e reportaram precisão desejável na estimativa da variável de interesse. Autores como Sabatia & Burkhardt (2014), por outro lado, reportaram e discutiram que mesmo com filtragem das variáveis mais importante para ajuste do RF, o mesmo ainda assim pode gerar estimativas inconsistentes. Esse fato se deve a presença de colinearidade no ajuste, o qual dificilmente é mitigado.

Uma abordagem que vem ganhando popularidade pela sua propriedade de mitigar multicolinearidade via encolhimento dos parâmetros ajustados assim como para seleção de variáveis é a técnica “Least Absolut Shrinkage Selection Operator” (lasso) (ALTOÉ, 2017)). Badgery et al. (2013), por exemplo, através de lasso reportaram um modelo ajustado com seleção das variáveis-chave e mitigação de colinearidade entre parâmetros que garantiu estimativas consistentes de carbono do solo na Austrália. Altóe (2017) apresentou modelos biologicamente consistentes para estimar diferentes compartimentos de estoque de carbono em biomas tipicamente brasileiros. Cabe ressaltar, no entanto, que nos exemplos supracitados, o lasso serviu principalmente para geração de modelos que

retornassem variáveis relevantes para quantificação da distribuição espacial da variável de interesse e não necessariamente que retornassem o melhor modelo ajustado, em muito justificado pelo encolhimento dos parâmetros (ALTOÉ, 2017).

Apesar de diversos estudos e técnicas aplicadas para modelagem da distribuição do estoque de carbono em florestas tropicais, a literatura demonstra que ainda há carência de uma abordagem que possibilite ao mesmo tempo: (1) selecionar as variáveis mais relevantes e ao mesmo independentes entre si na estimativa de carbono, (2) ajuste que se aproveite dessas variáveis sem encolher os coeficientes e garantir o poder preditivo desejado ao modelo, e (3) possibilitar ao fim com que qualquer viés de estimativa seja retirado via correção residual espacial. Tais propriedades são relevantes quando estudos de estoque de carbono da parte arbórea no cerrado brasileiro são reportados.

Este estudo teve como objetivo contrastar diferentes técnicas de aprendizagem supervisionada, como lasso, RF e KR a uma nova abordagem híbrida. O contraste buscou demonstrar as limitações de cada técnica, além da definição da melhor para estimava da distribuição do estoque de carbono arbóreo no cerrado brasileiro.

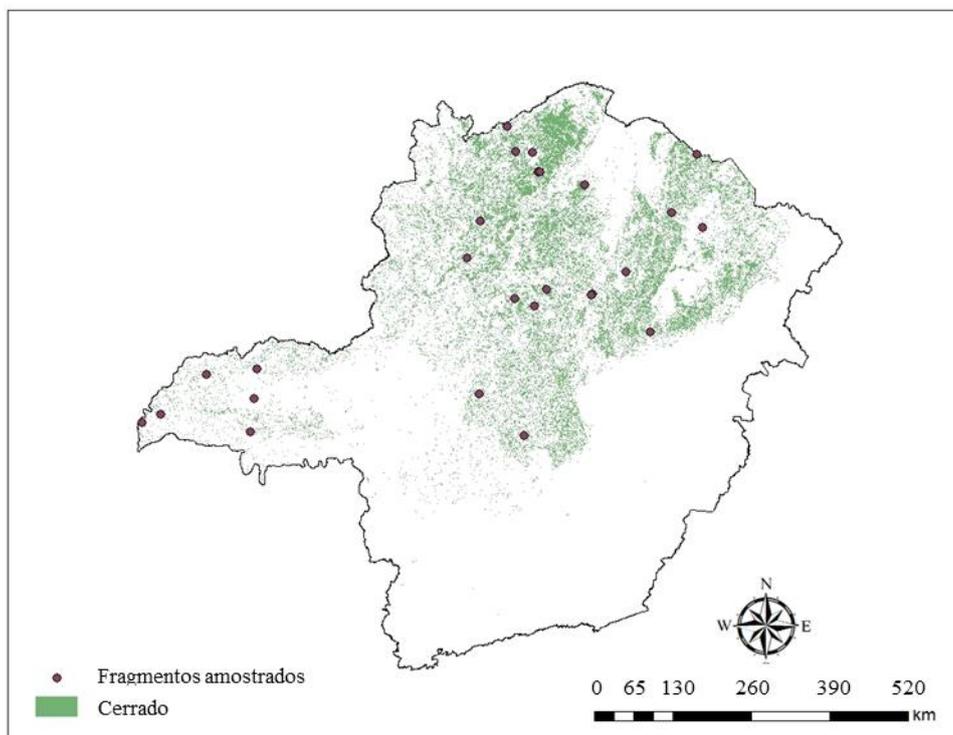
A abordagem híbrida consiste em: (1) utilização de lasso para seleção das variáveis preditoras, (2) ajuste do modelo com as variáveis selecionadas em (1) pela máxima verossimilhança, e (3) krigagem ordinária dos resíduos. Por fim, espera-se que a abordagem híbrida além de apresentar estimativas mais precisas, elimine lacunas de outras técnicas como presença de colinearidade, aptidão de seleção de variáveis e correção residual.

## **2. MATERIAL E MÉTODOS**

### **2.1. Caracterização da área de estudo**

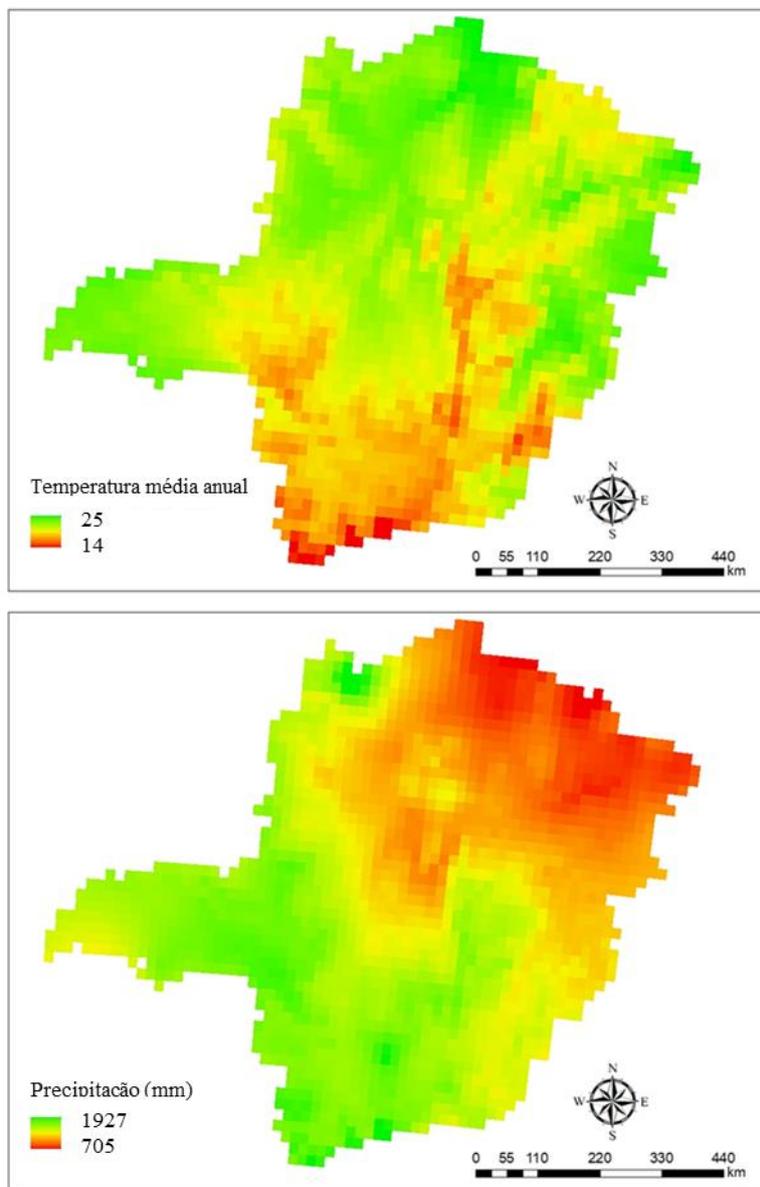
A área de estudo compreende o bioma cerrado do estado de Minas Gerais (Figura 1). Vale ressaltar a importância desse bioma no Brasil, uma vez que a ampla ocorrência do mesmo se concentra nas regiões mais populosas do Brasil, ou seja, em locais com os maiores níveis de antropização.

**Figura 1.** Distribuição do cerrado em Minas Gerais.



Os fragmentos de cerrado estão distribuídos em parte considerável de Minas Gerais, assim como os fragmentos selecionados para amostragem. Verifica-se um gradiente de pluviosidade anual negativo do sul para o norte de Minas Gerais, enquanto o mesmo gradiente torna-se positivo para temperatura média anual (Figura 2). Também é possível perceber como o padrão dos fragmentos de cerrado se diferencia de norte a sul no estado, sendo notadamente influenciados, dentre outros fatores, pelo gradiente climático (Tabela 1).

**Figura 2.** Gradiente climático (temperatura média anual em  $^{\circ}\text{C}$  e precipitação em mm) em Minas Gerais.



A Tabela 1 expressa os valores médios de estoque de carbono arbóreo (Mg/ha), assim como as médias de temperatura e pluviosidade anual referente a cada fragmento amostrado. De maneira geral fica evidente a grande variação no estoque de carbono no cerrado de Minas Gerais, assim como uma correlação positiva entre estoque de carbono e distribuição pluviométrica (76%) e negativa com temperatura média anual (-55%).

**Tabela 1.** Estoque de carbono arbóreo (Mg/ha) e clima para os diferentes fragmentos amostrados no cerrado de Minas Gerais.

ID	Longitude	Latitude	Área (ha)	Altitude (m)	Temperatura média anual (°C)	Precipitação anual (mm)	Estoque de Carbono (Mg/ha)
13	-42,77	-17,74	488,3	951	20,4	1089	9,5
21	-42,12	-16,14	591,3	837	22,7	914	5,0
34	-44,63	-19,40	355,2	720	20,9	1350	41,0
46	-44,68	-17,54	677,1	512	22,9	1233	24,5
47	-45,01	-17,47	271,0	566	22,6	1269	7,0
53	-44,52	-17,30	859,5	561	22,7	1205	12,6
56	-45,86	-16,94	236,9	583	23,0	1286	9,2
67	-45,69	-16,40	725,8	509	23,5	1217	22,2
70	-44,97	-15,34	253,1	593	23,6	1073	13,1
73	-44,85	-15,61	56,8	509	24,2	1076	15,0
74	-44,83	-15,60	74,8	510	24,2	1074	12,1
75	-45,24	-15,33	488,3	629	23,5	1098	15,2
84	-44,07	-15,71	241,0	817	22,3	1042	5,7
86	-42,63	-15,99	284,6	906	21,8	950	4,5
89	-45,41	-15,00	336,2	757	22,6	1102	9,8
94	-42,35	-15,10	201,1	997	21,9	867	12,5
96	-43,26	-16,92	307,3	926	20,5	1081	14,8
98	-45,44	-18,88	234,5	619	21,8	1376	22,7
102	-49,13	-18,79	312,3	722	20,9	1301	32,9
103	-50,97	-19,67	150,3	672	20,9	1400	33,2
104	-49,14	-19,21	301,4	687	21,4	1328	27,8
106	-49,15	-19,70	89,4	871	20,3	1347	37,7
107	-49,96	-18,92	285,1	710	21,0	1352	30,9
121	-43,88	-17,40	223,1	873	21,0	1173	22,6
122	-43,89	-17,39	320,6	869	20,9	1172	26,9
173	-50,66	-19,53	157,2	803	20,7	1374	40,6

## 2.2. Caracterização da amostragem

Os dados deste estudo são originados do Inventário Florestal de Minas Gerais (IFMG), o qual é detalhadamente descrito em Scolforo et al. (2015). De maneira geral uma rede de parcelas foi instalada no cerrado de Minas Gerais em 2010 através da utilização da amostragem em multi-estágio.

O desenho amostral do IFMG consistiu primeiramente na estratificação de Minas Gerais em subbacias hidrográficas, sendo que todas foram devidamente amostradas. Dentro de cada os fragmentos de cerrado tiveram suas respectivas fitofisionomias classificadas (1º estágio de amostragem). O segundo estágio de amostragem consistiu na seleção aleatória dos fragmentos a serem amostrados na combinação subbacia

hidrográfica/fitofisionomia. Por fim o terceiro estágio de amostragem consistiu no lançamento de parcelas sistematicamente distribuídas dentro de cada um dos fragmentos selecionados.

Sendo assim, 26 fragmentos de cerrado foram amostrados, sendo 19 fragmentos de cerrado sensu stricto, 2 de campo cerrado e 5 de cerradão. Conforme observado na tabela 1, os fragmentos de cerrado apresentam grande variação em termos de estoque de carbono, o que por consequência valida a necessidade de uma boa distribuição de fragmentos amostrados e valida a necessidade da amostragem multi-estágio.

Nos 26 fragmentos foram lançadas 244 parcelas de 1.000 m<sup>2</sup>. As parcelas foram sistematicamente distribuídas dentro de cada fragmento amostrado.

Variáveis dendrométricas foram mensuradas em cada uma das parcelas. Todas as árvores com diâmetro a 1,30 metros acima do solo (DAP)  $\geq 5$  cm tiveram seus DAPs e alturas (em metros) medidas. Adicionalmente, todas as árvores mensuradas foram botanicamente identificadas e plaqueteadas para fins de remedição. Por fim, as coordenadas geográficas do ponto central de cada parcela foram coletadas com GPS (SCOLFORO *et al.*, 2008).

### 2.3 Dados bioclimáticos

Dados bioclimáticos foram extraídos do *WorldClim* (HIJMAN *et al.*, 2015). As variáveis têm resolução de 1km e foram extraídas para cada um dos fragmentos do cerrado de Minas Gerais. As variáveis bioclimáticas são descritas abaixo:

Bio1: temperatura média anual (°C)

Bio2: média da amplitude de temperatura mensal (°C)

Bio3: isotermalidade (Bio2/Bio7) x 100

Bio4: desvio padrão da temperatura multiplicado por 100

Bio5: temperatura máxima (°C) do mês mais quente

Bio6: temperatura mínima (°C) do mês mais frio

Bio7: diferença entre Bio5 e Bio6

Bio8: temperatura média (°C) do trimestre mais úmido

Bio9: temperatura média (°C) do trimestre mais seco

Bio10: temperatura média (°C) do trimestre mais quente

Bio11: temperatura média (°C) do trimestre mais frio

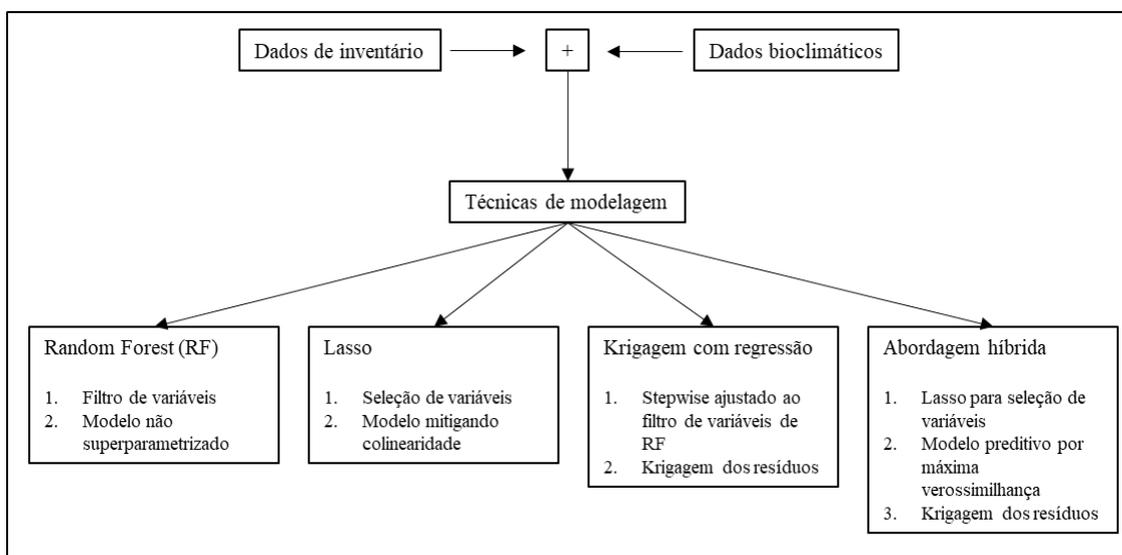
Bio12: precipitação anual (mm)

- Bio13: precipitação (mm) do mês mais úmido  
 Bio14: precipitação (mm) do mês mais seco  
 Bio15: coeficiente de variação da precipitação  
 Bio16: precipitação (mm) do trimestre mais úmido  
 Bio17: precipitação (mm) do trimestre mais seco  
 Bio18: precipitação (mm) do trimestre mais quente  
 Bio19: precipitação (mm) do trimestre mais frio

## 2.4 Técnicas de modelagem

A ideia central deste estudo foi definir a melhor técnica de modelagem da distribuição do estoque de carbono arbóreo para o cerrado de Minas Gerais. A seguir, técnicas já apresentadas na literatura como KR, RF e lasso foram descritas, assim como uma nova abordagem híbrida, apresentada para estimar a variável de interesse (Figura 3).

**Figura 3.** Diagrama das técnicas de modelagem.



Todas as análises foram realizadas no software R (R Core Team, 2016), a partir da utilização dos seguintes pacotes: glmnet (FRIEDMAN; HASTIE; TIBSHIRANI, 2010), randomForest (LIAW; WIENER, 2002), gdata (WARNES *et al.* 2018), reshape (WICKHAM, 2005), gstat (PEBESMA, 2004), optimx (NASH; VARADHAN, 2011), sp (PEBESMA; BIVAND, 2005), raster (HIJMANS; van ETTEN, 2012) e rgdal (KEITH *et al.*, 2010).

### 2.4.1. Random Forest (RF)

Random Forest (RF) é uma técnica não paramétrica amplamente aplicada em problemas de classificação e regressão (SILVEIRA *et al.*, 2019c). Como qualquer técnica de ajuste, o RF carece de algumas parametrizações iniciais de forma que o modelo ajustado ao final carregue as variáveis preditoras mais importantes.

Neste estudo o desenvolvimento do RF se deu em duas etapas. Na primeira etapa, buscou-se identificar as variáveis mais importantes para predição do estoque de carbono. Já a segunda etapa, com um conjunto reduzido de variáveis independentes, o modelo RF foi ajustado para prever o estoque de carbono.

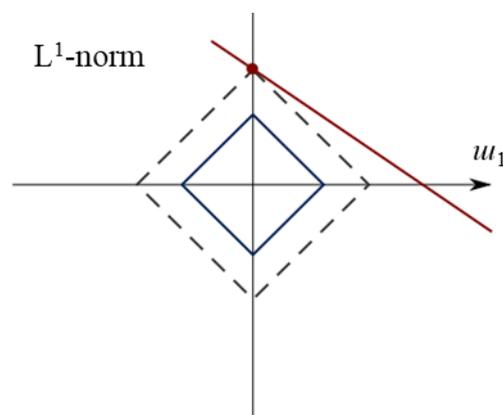
Apesar de diferentes abordagens propostas na literatura, este estudo decidiu por fixar o número de árvores em 2.000 unidades, sendo este número considerável para garantir a estabilidade das estimativas geradas pelo RF.

De maneira geral a seleção das variáveis mais importantes foi realizada através da exclusão de variável a variável de acordo com o ranking de importância das mesmas. Sendo assim, o ajuste do modelo final baseou-se apenas em um conjunto restrito de variáveis que minimizassem o erro quadrático médio. Com esse conjunto de variáveis definidas o modelo final foi ajustado e levado para aplicação.

### 2.4.2. Lasso

O lasso (TIBSHIRANI, 1996) é uma técnica de aprendizado de máquina com foco em selecionar variáveis preditoras e mitigar colinearidade no modelo resultante pela penalização dos parâmetros do modelo (Figura 4).

**Figura 4.** Representação geométrica do lasso.



O lasso penaliza os parâmetros ajustados para que sua soma absoluta seja menor que o parâmetro de penalização ( $\lambda$ ), satisfazendo-se com a minimização da soma dos quadrados dos erros e gerando um modelo parcimonioso (1). Nesse caso, o lasso tem uma forma de penalização que trabalha com o módulo do valor das estimativas fazendo-o ser tão restritivo a ponto de promover a exclusão de variáveis com pouco impacto. O  $\lambda$  mínimo, que garante a retirada de coeficientes menos correlacionados a variável de interesse, foi obtido por meio da validação cruzada, conforme Simon, Friedman e Hastie (2013).

$$\min_{\beta_0, \beta} \{ \sum_{i=1}^N (y_i - \beta_0 - x_i^T \beta)^2 \} \text{ subject to } \sum_{i=1}^N |\beta_j| \leq t \quad (1)$$

Em que:  $\beta$ 's correspondem ao vetor de coeficientes; min é a função de minimizar;  $y_i$  é o vetor para a variável dependente de  $i = 1 \dots n$ ;  $x_i$  é o vetor com as variáveis independentes de  $i = 1 \dots n$ ;  $t$  é um parâmetro livre que determina o grau de regularização;  $\sum |\beta_i|$  é a restrição que garante com que os coeficientes não ultrapassem o valor de  $t$ .

Sendo assim, o estoque de carbono foi modelado em função das 19 variáveis bioclimáticas através de lasso.

### 2.4.3 Krigagem com Regressão (KR)

#### 2.4.3.1 Stepwise

Stepwise é uma técnica amplamente utilizada no meio florestal pela sua capacidade de selecionar variáveis, sendo essa seleção baseada no teste de F. De maneira geral, stepwise combina os métodos “forward” e “backward”, iniciando a seleção de variáveis pela variável de maior correlação com a variável de interesse (“forward”). Sequencialmente variáveis são inseridas uma a uma até o momento em que o teste de F não difere entre um dado modelo e o modelo anterior (ou seja, sem a última variável introduzida no modelo – “backward”).

Nesse estudo, o stepwise foi iniciado apenas com as variáveis selecionadas pelo RF, uma vez que buscou-se evitar a geração de um modelo superparametrizado. Sendo assim, o stepwise foi utilizado para gerar o melhor modelo possível a partir de uma seleção prévia de variáveis. Essa lógica converge com demais estudos que utilizaram

análise de componentes principais para essa filtragem inicial de variáveis (Sabatia & Burkhart, 2014).

#### 2.4.3.2 Krigagem ordinária dos resíduos

Para krigagem ordinária dos resíduos o principal passo é determinar via análise variográfica qual o melhor modelo espacial a ser utilizado para a interpolação via krigagem (JOURNAL; HUIJBREGTS, 1978). Diferentes modelos estão disponíveis na literatura, entretanto, para fins florestais o modelo exponencial tende a apresentar resultados satisfatórios, conforme demonstrado no estudo de Scolforo et al. (2015).

A krigagem ordinária permite que parte do erro antes assumido como aleatório, seja modelado de acordo com a sua posição geográfica. O estimador da krigagem ordinária é apresentado abaixo:

$$\hat{Z}_{X_0} = \sum_{i=1}^n \lambda_i Z_{x_i} \quad (2)$$

Em que:  $\hat{Z}_{X_0}$  é a estimativa do valor na posição  $X_0$ ;  $n$  é o número de pontos amostrais vizinhos utilizados para predição dos pontos não amostrados  $\hat{Z}_{X_0}$ ;  $\lambda_i$  é o  $i$ -th peso associado para cada  $i$ -th observação da variável de interesse na posição  $X$ ; e  $Z_{x_i}$ , é o valor observado.

#### 2.4.4 Abordagem Híbrida

A abordagem híbrida consiste em três etapas para modelagem da distribuição do estoque de carbono arbóreo no cerrado de Minas Gerais. Etapas descritas abaixo:

##### **Etapa 1: lasso**

Ajusta-se o modelo de estoque de carbono em função das 19 variáveis bioclimáticas. Espera-se que o lasso selecione as variáveis mais relevantes do ponto de vista de predição. Adicionalmente, espera-se que as variáveis selecionadas sejam as mais independentes possíveis entre elas.

## Etapa 2: modelagem por máxima verossimilhança

A partir dos resultados da Etapa 1, ajusta-se um modelo por meio da máxima verossimilhança de forma que os parâmetros ajustados para as variáveis selecionadas não sejam penalizados (encolhidos):

$$\lambda(\theta, \sigma | x_i, y_i) = \sum_{i=1}^n \log \left[ \frac{1}{\sqrt{2\pi}\sigma} \exp \left( \frac{-y_i + f(x_i, \theta)}{\sigma^2} \right) \right] \quad (3)$$

Em que:  $\lambda$  é a verossimilhança em logaritmo neperiano;  $f(\cdot)$  é a função do modelo com a expectativa do parâmetro  $\theta$ ;  $y_i$  é o vetor para a variável dependente de  $i = 1 \dots n$ ;  $x_i$  é o vetor com as variáveis independentes de  $i = 1 \dots n$ ;  $\sigma^2$  representa o parâmetro da variância;  $\pi$  é uma constante.

Como o método dos mínimos quadrados ordinários é um caso especial da máxima verossimilhança, onde os resíduos são independentes, seguem uma distribuição normal e apresentam variância desconhecida. Para fins de inferência, análise de incerteza, etc, se alguma das pressuposições não é atendida, então o uso dos mínimos quadrados é inviabilizado. No entanto, pela máxima verossimilhança é possível corrigir tal efeito via modelagem explícita da variância residual.

## Etapa 3: krigagem ordinária dos resíduos

Modelos que quantifiquem a distribuição de algum atributo florestal por vezes tendem a apresentar maiores erros de estimativa para as regiões de extremos. Sendo assim, a Etapa 3 consistiu na krigagem ordinária residual do estoque de carbono arbóreo. Espera-se que ganhos em termos de precisão da estimativa sejam obtidos.

### 2.4.5 Abordagem Híbrida

Dada a quantidade de fragmentos utilizados para modelagem da distribuição do estoque de carbono arbóreo ( $n = 26$  fragmentos), preferiu-se pela validação das técnicas a partir de *bootstrapping* não paramétrico.

*Bootstrapping* não paramétrico permitiu com que 1.000 bancos de dados (com repetição) fossem simulados e então os modelos ajustados pelas diferentes técnicas foram

aplicados para cada um destes. Assim foi possível extrair informações como erro médio (T), erro absoluto (MAE) e mesmo erro quadrático médio (RMSE) dos diferentes modelos, além de gráficos residuais.

$$T \text{ (Mg/ha)} = \frac{\sum_{i=1}^n (obs-est)}{n} \quad (4)$$

$$MAE \text{ (Mg/ha)} = \frac{\sum_{i=1}^n (|obs-est|)}{n} \quad (5)$$

$$RMSE \text{ (Mg/ha)} = \sqrt{\frac{\sum_{i=1}^n (obs-est)^2}{n}} \quad (6)$$

Em que:  $i$  é o número total de fragmentos simulados; obs: valores observados de estoque de carbono arbóreo (Mg/ha); valores estimados de estoque de carbono arbóreo (Mg/ha).

## 2.5 Distribuição do estoque de carbono arbóreo no cerrado de Minas Gerais

A diferença na estimativa de estoque de carbono arbóreo para os fragmentos de cerrado em Minas Gerais pode ser significativa a depender do modelo/técnica selecionado.

Sendo assim, este item buscou retratar o quanto as diferentes técnicas podem sub ou superestimar o estoque de carbono arbóreo em relação a técnica definida como a mais adequada para gerar tal informação.

## 3. RESULTADOS

### 3.1 Avaliação das diferenças técnicas

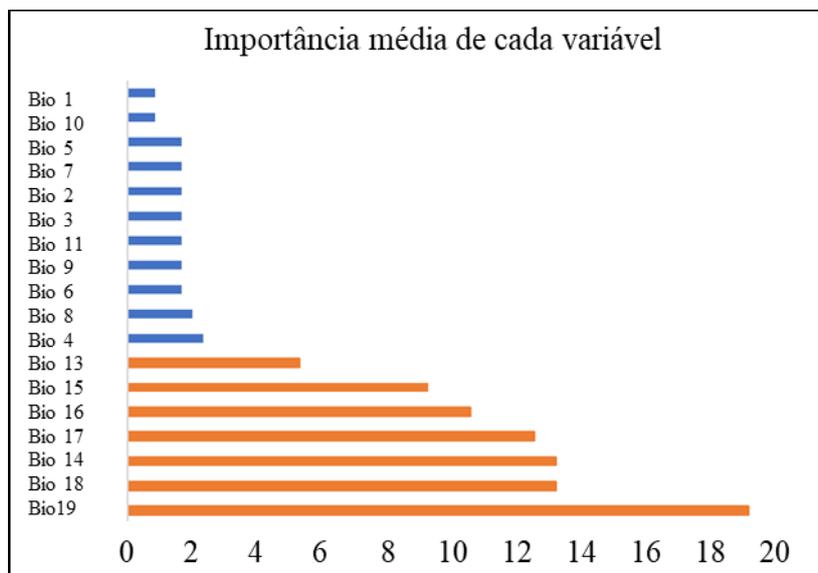
#### 3.1.1 RF

A primeira avaliação foi a respeito do entendimento da importância de cada variável bioclimática se todas fossem utilizadas na modelagem da distribuição do estoque de carbono. Vale ressaltar que, apesar de todas variáveis terem um grau de importância, as variáveis bio19, bio18, bio17, bio16, bio15, bio14 e bio 13 expressam 83% da importância do total (Figura 5).

Estas sete variáveis são relacionadas ao regime pluviométrico de Minas Gerais. Cada variável expressa uma particularidade quanto ao regime pluviométrico local, sendo

essas expressões importantes para o entendimento da distribuição do estoque de carbono arbóreo no cerrado de Minas Gerais.

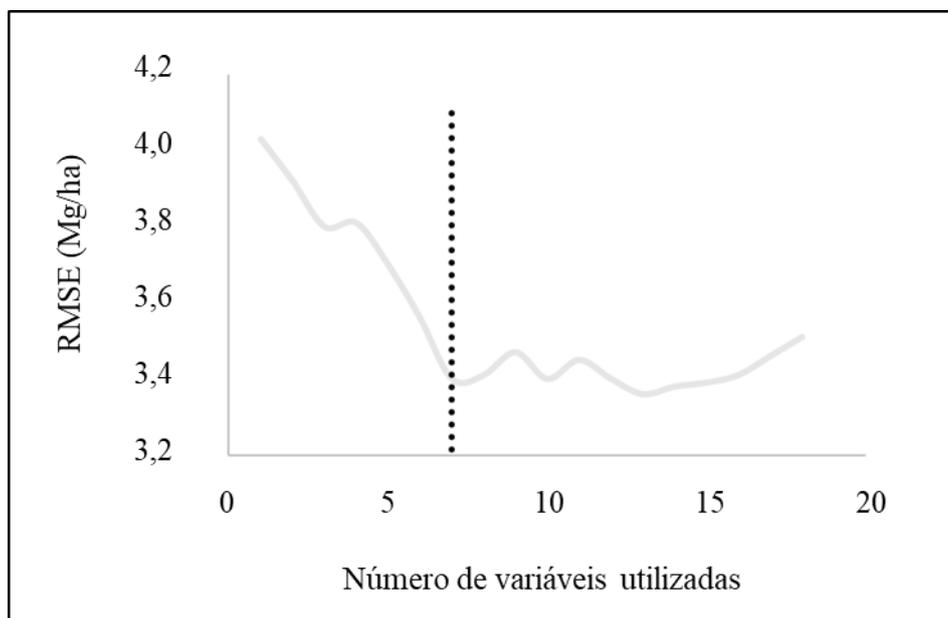
**Figura 5.** Importância de cada variável bioclimática com relação a modelagem do estoque de carbono arbóreo (Mg/ha) no cerrado de Minas Gerais.



Após o entendimento da importância de cada variável o próximo passo foi a seleção das variáveis necessárias para a modelagem precisa do estoque de carbono arbóreo no cerrado. Desta forma, buscou-se a seleção de variáveis, evitando o ajuste de um modelo superparametrizado.

Uma série de modelos foram ajustados, iniciando-se pelo modelo completo (com todas as variáveis bioclimáticas) e a partir de então os demais modelos foram ajustados com a remoção recursiva das variáveis (sempre iniciando pelas menos importantes). Com todos os 19 possíveis modelos ajustados, avaliou-se o ponto ótimo do RMSE com as sete variáveis: bio19, bio18, bio17, bio16, bio15, bio14 e bio 13 (Figura 6).

**Figura 6.** Comportamento do erro quadrático médio (RMSE, Mg/ha) com a remoção recursiva de variáveis, considerando ordem crescente para remoção.



### 3.1.2 KR

O primeiro passo foi o ajuste do modelo global com base em stepwise, lembrando que stepwise partiu do uso das variáveis bio13, bio14, bio15, bio16, bio17, bio18 e bio19, conforme seleção via RF:

$$\hat{C} = -79,0 - 3,6Bio14 + 0,8Bio15 - 1,9Bio17 + 0,01Bio18 + 1,1Bio19 \quad (7)$$

Em que:  $\hat{C}$  = estoque de carbono predito/estimado (Mg/ha); Bio's são as variáveis bioclimáticas previamente descritas.

Mesmo com a filtragem inicial das variáveis Bio13 - Bio19, a regressão stepwise não pode ser utilizada para fins de estimativa da distribuição do estoque de carbono. Um passo crítico para validação de um modelo preditivo passa pela verificação do fornecimento de estimativas biologicamente consistentes. A tabela 2 expressa a correlação linear entre o estoque de carbono e as variáveis bioclimáticas selecionadas por stepwise. Vale ressaltar que apesar da correlação negativa entre Bio15 e estoque de carbono, o coeficiente da equação (7) expressa sinal positivo. Situação semelhante de inversão de sinal ocorreu no coeficiente Bio17.

**Tabela 2.** Correlação entre o estoque de carbono arbóreo do cerrado de Minas Gerais e as variáveis bioclimáticas.

Variável	Correlação
Bio14	0,68
Bio15	-0,69
Bio17	0,67
Bio18	0,70
Bio19	0,77

### 3.1.3 Lasso

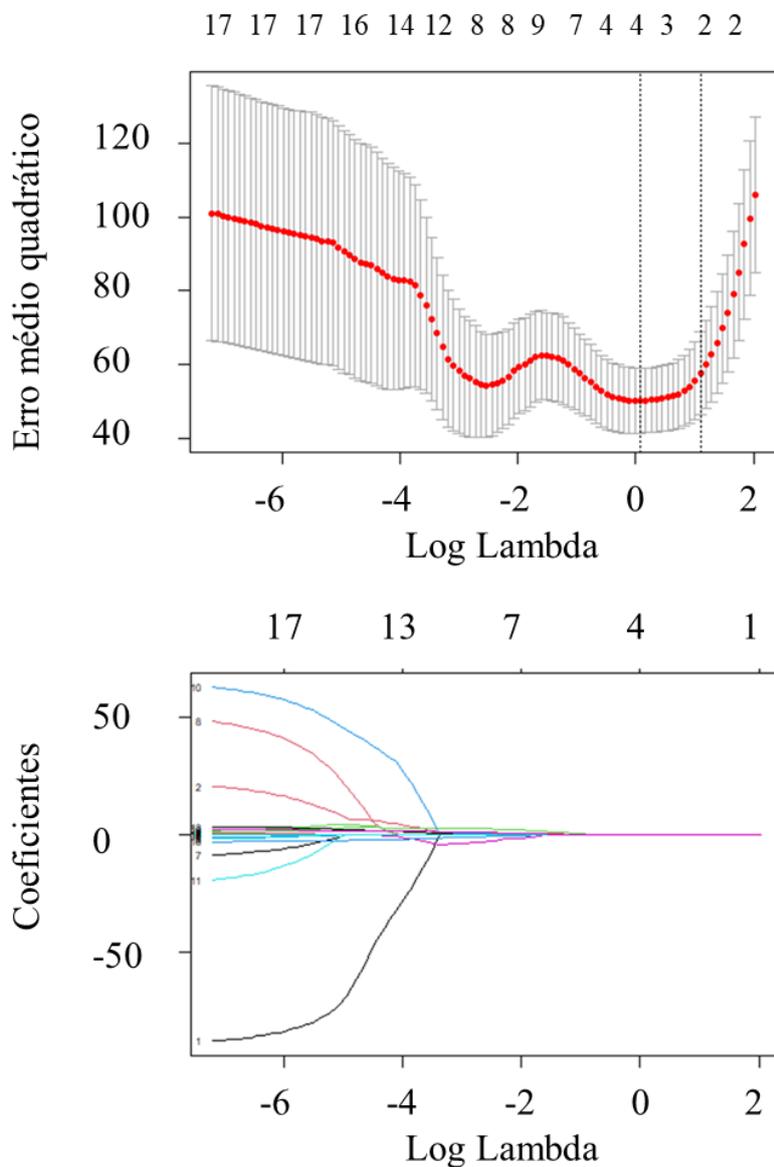
Lasso selecionou duas variáveis, considerando o  $\lambda$  a um desvio: Bio16 e Bio19. As duas variáveis são correlacionadas com o estoque de carbono (Bio16 x C = 0,68 e Bio19 x C = 0,77) e não apresentam alta correlação entre si (Bio16 x Bio19 = 0,5).

$$\hat{C} = -4,42 + 0,026Bio16 + 0,1451Bio19 \quad (8)$$

Em que: todas as variáveis foram previamente descritas.

A Figura 7 apresenta a localização de  $\lambda$  ( $\lambda = 1,003$ ), assim como o comportamento dos coeficientes em relação à penalização de  $\lambda$ .

**Figura 7.** Relação do RMSE (Mg/ha) e dos coeficientes com o valor de penalização  $\lambda$ .



### 3.1.4 Abordagem híbrida

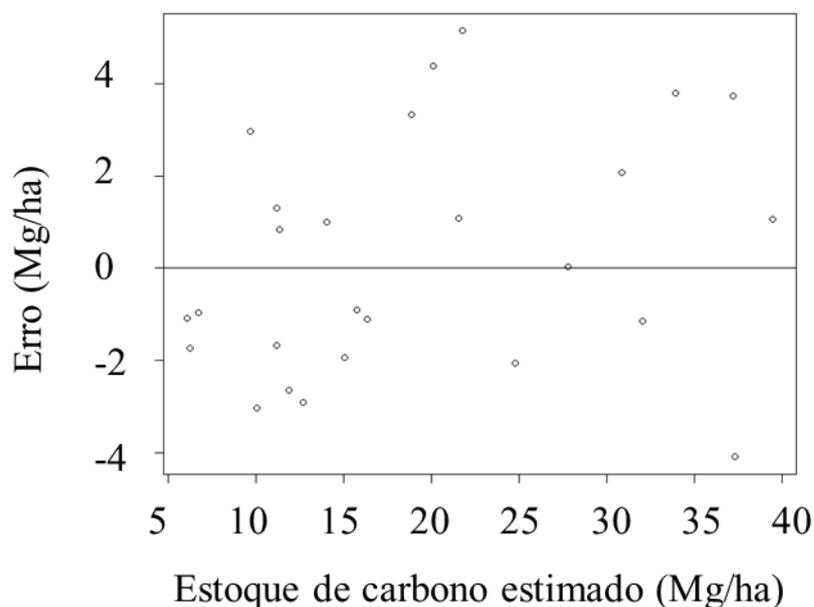
A primeira etapa expressa anteriormente por lasso selecionou as variáveis preditoras Bio16 e Bio19. Estas variáveis foram utilizadas para modelagem preditiva por meio da máxima verossimilhança na Etapa 2:

$$\hat{C} = -28,8 + 0,061Bio16 + 0,253Bio19 \quad (9)$$

Em que: todas as variáveis foram previamente descritas.

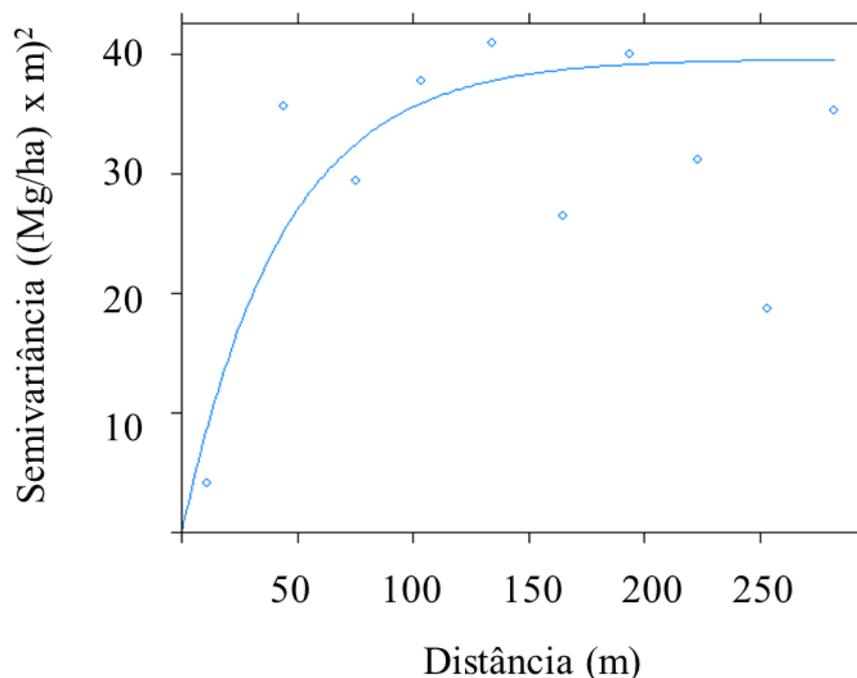
A seleção das variáveis por lasso resultou em um ajuste adequado na Etapa 2, pois os sinais dos coeficientes de Bio16 e Bio19 apresentaram consistência biológica e as estatísticas foram satisfatórias ( $R^2 = 68\%$  e erro padrão residual = 6,8 Mg/ha), e a distribuição residual foi não tendenciosa (Figura 8).

**Figura 8.** Distribuição residual da equação 9.



Já a etapa 3 consistiu na krigagem ordinária dos resíduos. Pela modelagem semivariográfica (Figura 9), percebe-se claramente a correlação espacial dos resíduos. Esse é um fator crítico para a Etapa 3, uma vez que tal semivariograma deixa explícito como a krigagem ordinária tende a contribuir para melhoria das estimativas. O semivariograma dos resíduos demonstra estrutura de continuidade espacial com alcance de 135 m e efeito pepita próximo de zero, o que retrata um grau de dependência espacial superior a 80%.

**Figura 9.** Semivariograma teórico e experimental para os resíduos da variável estoque de carbono arbóreo (Mg/ha) do cerrado de Minas Gerais.



### 3.2. Comparação entre as diferentes técnicas

A Tabela 3 expressa como RF, lasso e abordagem híbrida apresentam em média resultados sem qualquer viés de estimativa (T). A abordagem híbrida apresentou os resultados mais precisos comparada às demais técnicas. Os resultados de lasso chamaram atenção, mostrando que por mais que essa técnica possibilite uma apropriada seleção de variáveis para mitigar a colinearidade, o encolhimento dos parâmetros prejudica a qualidade das estimativas, conforme observado por MAE e RMSE.

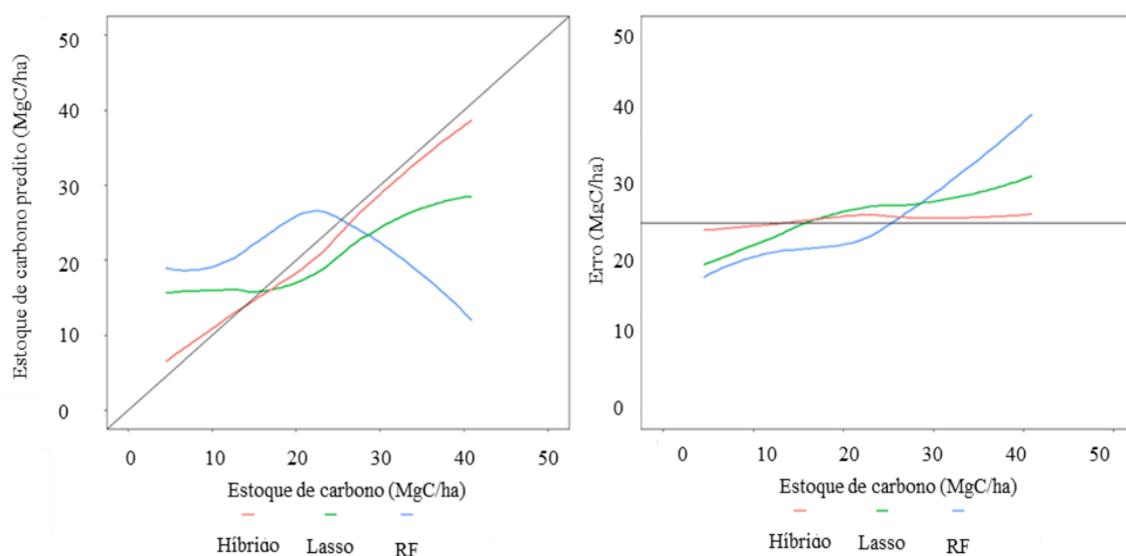
**Tabela 3.** Precisão das estimativas geradas pelas diferentes técnicas de modelagem de distribuição do estoque de carbono arbóreo do cerrado de Minas Gerais.

Estatística	Random Forest	Lasso	Abordagem Híbrida
T (Mg/ha)	-0,18	-0,02	0,23
MAE (Mg/ha)	2,97	10,14	2,66
RMSE (Mg/ha)	3,57	11,97	2,91

A Figura 10 expressa como o RF apesar de apresentar estimativas, em média, sem viés, tende a concentrar estas em relação ao valor médio de estoque de carbono do cerrado de Minas Gerais. Isso quer dizer que, apesar da estratégia de mitigação de colinearidade e superparametrização, o modelo resultante apresentou estimativas com baixa precisão para áreas com maiores e menores estoques de carbono.

Por fim, vale a pena ressaltar a eficiência da abordagem híbrida. Além dos resultados médios expressados por T, MAE e RMSE, essa abordagem também foi flexível o suficiente para estimar adequadamente o estoque de carbono em Minas Gerais, não importando o local, além da consistência biológica de toda modelagem conforme mencionado anteriormente.

**Figura 10.** Aderência das estimativas x observações do estoque de carbono (a), tendência média residual do estoque de carbono arbóreo do cerrado de Minas Gerais.



### 3.2.1 Avaliação do ganho de precisão de cada etapa da abordagem híbrida

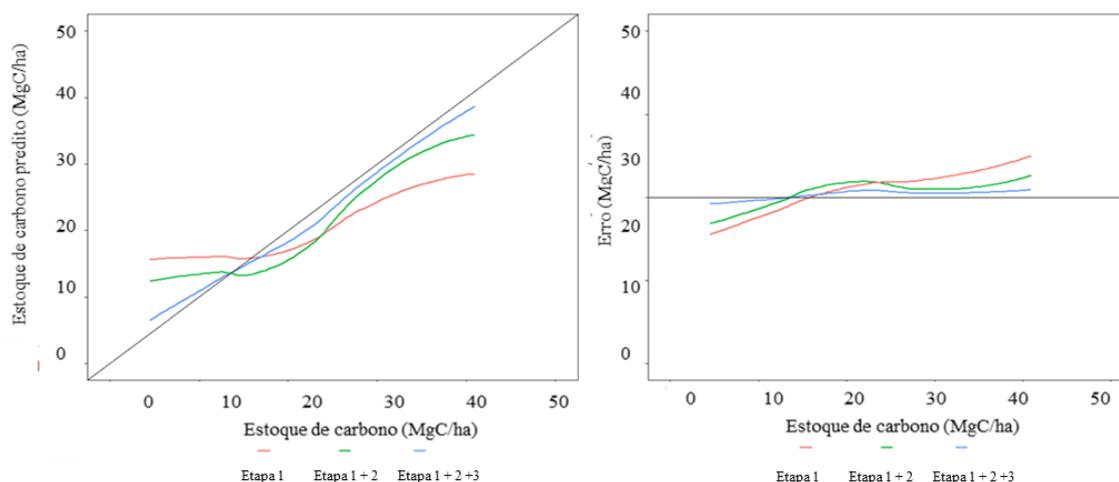
A partir da Tabela 4 fica evidente que a modelagem por meio da máxima verossimilhança reduziu o erro das estimativas pela metade, sendo o efeito da krigagem residual semelhante.

**Tabela 4.** Precisão das estimativas geradas a partir da inclusão das diferentes etapas da abordagem híbrida.

Estatística	Etapa 1	Etapa 1 + 2	Etapa 1 + 2 + 3
MAE (Mg/ha)	10,14	5,38	2,66
RMSE (Mg/ha)	11,97	6,31	2,91

Mesmo comportamento é observado na Figura 11. A combinação lasso + máxima verossimilhança corrigiu parte da tendência de estimativa apresentado por lasso e a posterior combinação destes com a krigagem dos resíduos permitiu a geração de resultados ainda mais precisos, especialmente para os extremos, como esperado.

**Figura 11.** Aderência das estimativas x observações do estoque de carbono (a), tendência média residual do estoque de carbono arbóreo do cerrado de Minas Gerais.



### 3.3 Distribuição do estoque de carbono arbóreo no cerrado

As estimativas da abordagem híbrida são mais precisas que das demais técnicas. Isso quer dizer que a abordagem híbrida é a técnica ideal para expressar a distribuição do estoque de carbono arbóreo do cerrado. A pergunta que resta, no entanto, tem relação a qual diferença na estimativa global de estoque de carbono das demais técnicas em relação a abordagem híbrida (Tabela 5).

**Tabela 5.** Diferença em % na estimativa global de estoque de carbono da abordagem híbrida vs as demais técnicas para o cerrado de Minas Gerais.

Técnica	%
Random Forest x Abordagem Híbrida	6,8
Lasso x Abordagem Híbrida	2,2

De forma objetiva pode-se responder que as demais técnicas superestimam o estoque de carbono arbóreo do cerrado de Minas Gerais em relação a abordagem híbrida em até 7%. Esses valores ligam o sinal de alerta relacionado a escolha de qual técnica utilizar na estimativa de carbono em florestas nativas e como as mesmas podem interferir em possíveis tomadas de decisão dos gestores públicos.

#### 4. DISCUSSÃO

Uma nova abordagem para quantificação de estoque de carbono arbóreo no cerrado de Minas Gerais foi apresentada neste estudo. Esta técnica gerou resultados acurados e consistentes do ponto de vista biológico, além disso, foi superior em relação à outras técnicas comumente reportadas na literatura.

A combinação de lasso e modelagem por máxima verossimilhança foi reportada por Koirala et al. (2020) para modelagem de índice de sítio de *Pinus taeda* no sudeste dos EUA. Os autores reportaram a vantagem do uso do lasso para seleção de variáveis para modelagem da variável resposta, sendo depois ajustado o modelo por meio da máxima verossimilhança e sem o encolhimento dos parâmetros.

Por mais que o lasso seja extremamente eficaz na seleção de variáveis, o encolhimento dos parâmetros ao final do processo iterativo de ajuste do modelo garante com que o mesmo seja interpretativo a respeito da variável resposta, porém não garante poder preditivo desejável. Altoé (2017), por exemplo, apresentou resultados de modelagem de carbono via lasso e utilizou os mesmos para ressaltar o entendimento dos principais fatores que interferem na distribuição espacial do estoque de carbono em raízes, solo e serrapiheira em diferentes biomas no Brasil.

A abordagem híbrida, além da modelagem com consistência biológica biológica (SCOLFORO *et al.*, 2020), também contempla o efeito da modelagem espacial dos resíduos através da krigagem ordinária. Mello et al. (2013), por exemplo, reportaram como utilização da krigagem residual tende a corrigir distorções de estimativas da variável de interesse, especialmente nos extremos da distribuição.

Por mais que a KR tenha ampla utilização na literatura, o uso para a modelagem de estoque de carbono arbóreo em função de um vetor de variáveis correlacionadas entre si é desafiador. Morais et al. (2020), por exemplo, modelaram o estoque de carbono de solo no cerrado brasileiro e se depararam ao final com um modelo com grau satisfatório de precisão, porém com forte presença de colinearidade. Para mitigar o efeito de colinearidade, Sabatia & Burkhart (2014) mencionaram a seleção de variáveis via RF. Os autores pontuaram que um filtro de variáveis poderia ser utilizado no stepwise e assim gerar um modelo biologicamente consistente, o que não foi verificado no estudo em questão.

Por fim, a técnica RF com larga aplicação em estudos de classificação de espécies (CARVALHO, 2018), modelagem espacial de volume em plantios de eucalipto (REIS *et al.*, 2019), etc, apresentou resultados com bom nível médio de precisão. Tal precisão média, entretanto, não se confirmou quando avaliada a performance do modelo ajustado ao longo do gradiente ambiental do cerrado em Minas Gerais. Sabatia e Burkhart (2014) reportaram precisão satisfatória via modelagem por RF para índice de sítio em florestas de *Pinus taeda* nos EUA. Entretanto os mesmos autores reportaram que os resultados quando extrapolados para quantificar o índice de sítio de forma espacializada no sudeste dos EUA apresentavam resultados enviesados.

A abordagem híbrida estimou um menor estoque de carbono arbóreo no cerrado quando comparada as demais técnicas. Este resultado é relevante para a tomada de decisão a respeito de estratégias públicas para proteção ambiental. Como próximos passos, deve-se aplicar a abordagem híbrida nos demais biomas brasileiros. Desta maneira, os modelos ajustados não seriam somente baseados em função de variáveis bioclimáticas, como vem sendo feito atualmente.

## 5. CONCLUSÃO

Uma nova abordagem foi apresentada neste estudo para estimar carbono do cerrado no estado de Minas Gerais. Esta nova abordagem, chamada de abordagem híbrida, possui propriedades quanto a mitigação de colinearidade, seleção de variáveis com maior aptidão preditiva e correção de tendência residual.

A abordagem híbrida apresentou resultados mais precisos que as demais técnicas e se mostra como uma alternativa viável de ser replicada em estudos de mapeamento da

distribuição do estoque de carbono em florestas nativas. Adicionalmente, este estudo reforça a necessidade de modelagem com adequado grau de precisão, uma vez que tendências equivocadas nas estimativas podem comprometer estratégias de decisão da gestão pública.

## REFERÊNCIAS

ALTOÉ, T. F. **Modelagem e espacialização do estoque de carbono em fragmentos nativos de Minas Gerais em função de fatores edafoclimáticos**. 2017. 130 p. Tese. (Doutorado em Engenharia Florestal) – Universidade Federal de Lavras, Lavras, 2017.

BADGERY W. B. *et al.*, Relationship between environmental and land-use variables on soil carbon levels at the regional scale in central New South Wales, Australia. **Soil Research**, Clayton South, v. 51, p. 645–656, dez., 2013. DOI: 10.1071/SR12358.

CARVALHO, M.C. *et al.* Algoritmos de aprendizagem de máquina na modelagem da distribuição potencial de habitats de espécies arbóreas. **Nativa**, Sinop, v. 7, n. 5, p. 600-606, 2019. DOI: <http://dx.doi.org/10.31413/nativa.v7i5.7214>.

FRIEDMAN J; HASTIE T; TIBSHIRANI R. **glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models**. 2009. R package version 1.1-4, URL <http://CRAN.R-project.org/package=glmnet>. (acesso em 10 de agosto de 2021).

HIJMANS R J.; CAMERON S. E.; PARRA J.L.; JONES P.G.; JARVIS A. Very high-resolution interpolated climate surfaces for global land areas. **International Journal Climatology**, Chichester. v. 25, n. 15, p. 1965-1978, nov. 2005. DOI: 10.1002/joc.1276/full.

HIJMANS, R. J.; VAN ETTEN; J. **raster: Geographic analysis and modeling with raster data**. 2012. R package version 2.0-12. URL <http://CRAN.R-project.org/package=raster>. (acesso em 10 de agosto de 2021).

JOURNEL, A. G.; HUIJBREGTS, C. J. **Mining geostatistics**. London: Ed. Academic, 1978. 600 p.

KEITT, T.H.; BIVAND, R.; PEBESMA, E.; ROWLINGSON, B. **rgdal: bindings for the Geospatial Data Abstraction Library**. 2010. R package version 0.6-27. URL: <http://CRAN.R-project.org/package=rgdal>. (acesso em 10 de agosto de 2021).

KOIRALA, A.; MONTES, C.; BULLOCK, B.P. Modeling dominant height using stand and water balance variables for loblolly pine in the Western Gulf, US. **Forest Ecology**

**and Management**, Amsterdam, v. 479, p. 118610, jan., 2020. DOI: 10.1016/j.foreco.2020.118610

LIAW, A.; WIENER, M. **Classification and Regression by randomForest**. R News, v. 2, n.3, p. 18-22. 2002.

LIMA R. B. *et al.* Modeling and Spatialization of Biomass and Carbon Stock Using Lidar Metrics in Tropical Dry Forest, Brazil: Preliminary Results. **Research Square**, v.11, aug., 2020. DOI: 10.21203/rs.3.rs-55277/v1 PPR: PPR199325.

MASCARO, J. *et al.* A Tale of Two “Forests”: Random Forest Machine Learning Aids Tropical Forest Carbon Mapping. **PLOS ONE**, San Francisco, v. 9, n. 1, p. e85993, jan., 2014. DOI: 10.1371/journal.pone.0085993.

MELLO C. R., VIOLA M. R., BESKOW S., NORTON L. D. Multivariate models for annual rainfall erosivity in Brazil. **Geoderma**, Amsterdam, v. 202/203, p. 88-102, jul. 2013.

MORAIS, V. A. *et al.*, Spatial distribution of soil carbon stocks in the Cerrado biome of Minas Gerais, Brazil. **Catena**, [S.l.], v.185, feb. 2020. DOI: 10.1016/j.catena.2019.104285

NASH, J.C.; VARADHAN, R., Unifying Optimization Algorithms to Aid Software System Users: optimx for R. **Journal of Statistical Software**, Innsbruck, v. 43, n.9, p. 1–14, 2011. Disponível em <http://www.jstatsoft.org/v43/i09/>. (acesso em 10 de agosto de 2021).

PEBESMA E. J. Multivariable geostatistics in S: the gstat package. **Computers & Geosciences**, Vandœuvre-les-Nancy, v. 30, n. 7, p. 683-691, aug., 2004.

PEBESMA, E.; BIVAND, R. S. S classes and methods for spatial data: the sp package. **R news**, v. 5, n. 2, p. 9-13, 2005.

PEREIRA, E.A.; SILVEIRA, E.M.O.; BUENO, I.T.; ACERBI JR, F.W., Spatial and spectral remote sensing features to detect deforestation in Brazilian Savannas. **Advances in Forestry Science**, Cuiabá, v.6, n.4, p.775-782, 2019. DOI: 10.34062/afs.v6i4.7525.

R CORE TEAM. R: **A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. 2019. Disponível em <https://www.R-project.org/>. (acesso em 10 de agosto de 2021).

REIS, A.A., FRANKLIN, S.E., MELLO, J.M., ACERBI JR, F.W. Volume estimation in a Eucalyptus plantation using multi-source remote sensing and digital terrain data: a case study in Minas Gerais State, Brazil. **International Journal of Remote Sensing**, Londres v. 40, n. 7, p. 1-2018. DOI: 10.1080/01431161.2018.1530808

RICHARDSON, H.J.; HILL, D.J.; DENESIUK, D.R.; FRASER, L.H. A comparison of geographic datasets and field measurements to model soil carbon using random forests and stepwise regressions (British Columbia, Canada). **GIScience & Remote Sensing**, [SI], v. 54, n. 4, p. 573-591, 2017. DOI: 10.1080/15481603.2017.1302181

SABATIA, C., BURKHART, H.E. Predicting site index of plantation loblolly pine from biophysical variables. **Forest Ecology and Management**, Amsterdam, v. 326, p. 142-156, aug, 2014. DOI: 10.1016/j.foreco.2014.04.019

SCOLFORO, H. F.; SCOLFORO, J. R. S.; DE MELLO, J. M.; DE MELLO, C. R.; MORAIS, V. A. Spatial interpolators for improving the mapping of carbon stock of the arboreal vegetation in Brazilian biomes of Atlantic forest and Savanna. **Forest Ecology and Management**, Amsterdam, v. 376, p. 24-35, sep., 2016. DOI: 10.1016/j.foreco.2016.05.047

SCOLFORO, H. F.; SCOLFORO, J. R. S.; MELLO, C. R.; MELLO, J. M.; FERRAZ FILHO, A. C. 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. e0128781, 2015. DOI: 10.1371/journal.pone.0128781

SCOLFORO, H.F. *et al.*, A New Approach for Modeling Volume Response from Mid-Rotation Fertilization of Pinus taeda L. Plantations. **Forests**, Basel, v. 11, n. 646, 2020 DOI: 10.3390/f11060646

SCOLFORO, J. R.; OLIVEIRA, A. D.; ACERBI JÚNIOR, F. W. A. (Eds.). **Inventário florestal de Minas Gerais: equações de volume, peso de matéria seca e carbono para diferentes Fisionomias da flora nativa**. Lavras: Ed. UFLA, 2008. 216 p.

SILVEIRA, E. *et al.*, Object-based random forest modelling of aboveground forest biomass outperforms a pixel-based approach in a heterogeneous and mountain tropical environment. **Forest Ecology and Management**, Amsterdam, v. 445, n. 1, p. 96-109, 2019b. DOI: 10.1016/j.jag.2019.02.004.

SILVEIRA, E. *et al.* Pre-stratified modelling plus residuals kriging reduces the uncertainty of aboveground biomass estimation and spatial distribution in heterogeneous savannas and forest environments. **International Journal of Applied Earth Observation and Geoinformation**, v. 78, p. 175-188, aug., 2019c. DOI: 10.1016/j.foreco.2019.05.016.

SILVEIRA, E. *et al.* Spatial distribution of wood volume in Brazilian savannas. **Academia Brasileira de Ciências**, Rio de Janeiro, v. 91, n. 4, 2019a. DOI: 10.1590/0001-3765201920180666

SIMON, N.; FRIEDMAN, J.; HASTIE, T. A blockwise descent algorithm for group-penalized multiresponse and multinomial regression. **Journal of Statistical Software**, [Linz], v. 20, n. 2, p. 1-15, nov. 2013.

TIBSHIRANI, R., Regression Shrinkage and Selection via the Lasso. **Journal of the Royal Society**. Series B (Methodological), [S.I.], v. 58, n. 1, p. 267-288. 1996.

WARNES, G.R. *et al.* **gdata: Various R programming tools for data manipulation**. 2014. R package version 2.18.0. URL <https://CRAN.R-project.org/package=gdata>. (acesso em 10 de agosto de 2021).

WICKHAM, H., **reshape: Flexibly reshape data**. 2005. R package version 0.71.2. URL: <http://had.co.nz/reshape/>. (acesso em 10 de agosto de 2021).

ZANINI, A. M.; MAYRINCK, R. C.; VIEIRA, S. A.; DE CAMARGO, P. B.; RODRIGUES, R. R. The effect of ecological restoration methods on carbon stocks in the Brazilian Atlantic Forest. **Forest Ecology and Management**, Amsterdam, v. 481, p. 118734, 2021. DOI: 10.1016/j.foreco.2020.118734.

**ARTIGO 2: REDUZINDO O EFEITO DA INCERTEZA NA MODELAGEM DA DISTRIBUIÇÃO DO ESTOQUE DE CARBONO NO CERRADO: UMA ABORDAGEM QUE COMBINA MODELAGEM DE CRESCIMENTO E MODELAGEM GEOESPACIAL**

**RESUMO**

A modelagem da distribuição do estoque de carbono arbóreo nas florestas nativas brasileiras sempre esteve atrelada as variáveis bioclimáticas e de sensoriamento remoto. Por mais que modelos preditivos possam mapear com boa precisão a tendência da distribuição dessa variável de interesse, ainda há predominância de considerável incerteza nas estimativas geradas. Sendo assim, este estudo objetivou trazer como novidade o desenvolvimento de uma abordagem que combina modelagem de crescimento e modelagem geoespacial. Um total de 58 fragmentos nativos de cerrado em Minas Gerais foram utilizados neste estudo, considerando dados de inventários realizados entre 2005-2007, com a posterior remedição de 26 fragmentos entre 2010-2011. A nova abordagem apresentada neste estudo combina: (1) lasso para seleção de variáveis que mitiguem multi-colinearidade no modelo de crescimento, (2) ajuste de um modelo de crescimento por meio da máxima verossimilhança, (3) lasso para seleção de variáveis que mitiguem multi-colinearidade no modelo preditivo, (4) ajuste de um modelo preditivo que combine informações de inventário, de variáveis bioclimáticas e de sensoriamento remoto, e (5) correção residual por meio de krigagem ordinária. A nova abordagem (abordagem de crescimento e predição) foi contrastada à abordagem híbrida apresentada no artigo 1 desta tese. Os resultados apresentados destacam que na existência de informações passadas de inventário florestal, a aplicação da abordagem de crescimento e predição traz ganhos significativos tanto em termos de precisão das estimativas como numa sensível redução de incerteza nos mapas globais. A nova abordagem é uma ferramenta inovadora para gestão pública específica ao nível de fragmento florestal.

**PALAVRAS-CHAVE: sensoriamento remoto; inventário florestal; máxima verossimilhança.**

**REDUCING THE UNCERTAINTY EFFECT IN THE CARBON STOCK****DISTRIBUTION MODELING IN THE SAVANNAH: AN APPROACH THAT****COMBINES GROWTH AND GEOSPATIAL MODELING****ABSTRACT**

Modeling the aboveground carbon stock distribution in native Brazilian forests is conventionally based on bioclimatic and remote sensing as independent variables. Although predictive models usually offer carbon stock maps with good precision, there is still considerable uncertainty associated with the estimates generated from these models. Hence, this study aimed to develop an approach that combines growth and geospatial modeling. A total of 58 forest fragments in the Savannah of Minas Gerais were used in this study, where these fragments were first measured between 2005-2007 and 26 fragments were later remeasured between 2010-2011. The new approach presented in this study combines: (1) lasso for variable selection that mitigate multi-collinearity in the growth model, (2) fit of a growth model by means of maximum likelihood, (3) lasso for variable selection that mitigates multi-collinearity in the predictive model, (4) fit of a predictive model that combines past and current forest inventory, bioclimatic and remote sensing information, and (5) ordinary kriging of the residuals generated in (4). The new approach (growth and prediction approach) was compared with the hybrid approach presented in Article 1 of this dissertation. The results highlighted that in the existence of past forest inventory information, the application of the growth and prediction approach brings significant gains both in terms of precision of the estimates and in the reduction of the uncertainty in the carbon stock global maps. The new approach is an innovative tool for management at the forest fragment level.

**KEYWORDS: remote sensing, forest inventory, maximum likelihood.**

## 1. INTRODUÇÃO

O cerrado é o segundo maior bioma do Brasil e é classificado como a maior savana neotropical do mundo (OLIVEIRA et al., 2006). Sua área original abrangia mais de 2 milhões Km<sup>2</sup> (EITEN, 1994), ocupando cerca de 22% do território brasileiro (BRASIL, 2009). O cerrado apresenta relevante riqueza em diversidade de espécies. O número de plantas vasculares encontrados no mesmo é superior ao da grande maioria dos biomas do mundo (Mendonça et al, 1998). Além disso, o endemismo de sua flora que chega a 44% (KLINK e MACHADO, 2005). O bioma é classificado como um hotspot mundial de biodiversidade, e apesar de seu alto índice de endemismo apresenta acelerado processo de fragmentação (KLINK e MACHADO, 2005).

A riqueza que envolve o cerrado indica que seu monitoramento é essencial do ponto de vista de gestão pública (SCOLFORO *et al.*, 2015). Sendo assim em 2003 foi iniciado um inventário em larga escala com foco no monitoramento do cerrado no estado de Minas Gerais (SCOLFORO *et al.*, 2008). O inventário florestal de Minas Gerais (IFMG) objetivou, dentre demais objetivos, a quantificação da riqueza de espécies do cerrado de Minas Gerais, assim como a quantificação do estoque de carbono arbóreo deste bioma e a definição de regiões prioritárias de conservação (SCOLFORO *et al.*, 2008).

Ao longo do tempo, o conhecimento acerca dos estoques de carbono arbóreo no cerrado foi gerado pelo IFMG, assim como por estudos de modelagem da distribuição dessa variável (ALTOÉ, 2016; SILVEIRA *et al.*, 2019). Normalmente, em florestas nativas, a modelagem da distribuição de estoque de carbono, assim como nicho ecológico, estoque volumétrico, riqueza de espécies, etc, é baseada na utilização de variáveis bioclimáticas e de sensoriamento remoto (ALTOÉ, 2016; CARVALHO *et al.*, 2019; SILVEIRA *et al.*, 2019), através do uso de abordagens como random forest, least absolute shrinkage selection estimator (lasso), krigagem com regressão, krigagem ordinária, ou mesmo a abordagem híbrida recentemente apresentada a literatura (ALTOÉ, 2016; REIS *et al.*, 2018; SCOLFORO *et al.*, 2016).

Conforme verificado por David et al. (2019), por exemplo, a modelagem da distribuição do estoque de carbono em florestas da Amazônia a partir de variáveis de sensoriamento remoto apresentaram resultados precisos acerca da distribuição espacial da variável de interesse. Resultados semelhantes foram reportados por Mascaro et al.

(2014) com o ajuste de random forest em função de variáveis bioclimáticas para estimar estoque de carbono no Peru.

Para as florestas nativas brasileiras, muitas vezes por inexistência de inventários de longo prazo, a distribuição espacial de estoque de carbono arbóreo é exclusivamente modelada em função de variáveis auxiliares relacionadas a clima, sensoriamento remoto, solos, etc (SILVEIRA *et al.*, 2019). Em países com inventários florestais nacionais de longo prazo, no entanto, modelos de crescimento para estoque de carbono estão disponíveis para uso, o que naturalmente permite abordagens que utilizem informações dos inventários passados (NYSTRÖM *et al.*, 2015). Isso permite maiores informações a respeito da variável de interesse na escala espacial. Nyström *et al.* (2015), inclusive, reportaram em florestas da Suécia que a combinação de informações de sensoriamento remoto com o aproveitamento dos inventários passados, possibilita a redução da incerteza associada às estimativas da variável de interesse, além do incremento em precisão.

A garantia de gestão pública, cada vez mais efetiva no cerrado de Minas Gerais, passa pelo conhecimento da distribuição do estoque de carbono. Naturalmente, esse conhecimento requer estimativas precisas da variável de interesse ao nível de fragmento florestal, sendo possível assim definir áreas prioritárias para conservação. Neste contexto, o presente estudo objetivou apresentar uma abordagem de crescimento e predição (CP), contrasta-la à abordagem híbrida (desenvolvimento do artigo 1 desta tese) na quantificação do estoque de carbono arbóreo do cerrado de Minas Gerais.

Espera-se a partir da abordagem de CP um ganho de precisão nas estimativas de estoque de carbono ao nível de fragmento florestal, assim como a redução de incertezas associadas. A abordagem CP permite a correção de duas situações normalmente observadas na modelagem de carbono em florestas nativas brasileiras: (1) reaproveitamento de inventários passados, costumeiramente descartados, a partir de modelos de crescimento e (2) modelagem preditiva baseada em informações de estoque de carbono dos inventários atuais e passados.

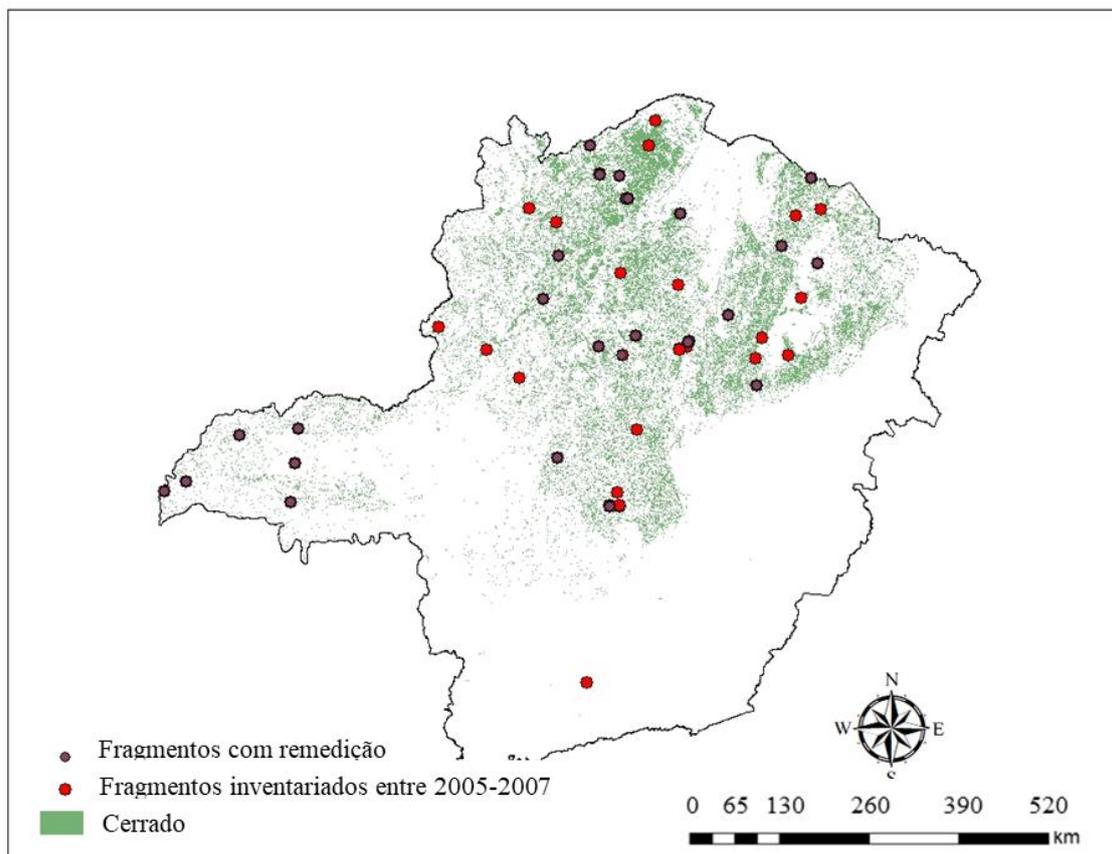
## **2. MATERIAL E MÉTODOS**

### **2.1 Caracterização da área de estudo**

A área de estudo compreende o bioma cerrado em Minas Gerais. O cerrado é o segundo bioma mais devastado no Brasil, sendo estimado que 55% da sua cobertura

vegetal já tenha sido desmatada (MACHADO *et al.*, 2004). Observa-se que os fragmentos de cerrado estão amplamente distribuídos em Minas Gerais (Figura 1).

**Figura 1.** Distribuição dos fragmentos amostrados em Minas Gerais.



## 2.2. Amostragem do estoque de carbono arbóreo do cerrado de Minas Gerais

Neste estudo foram utilizados dados de inventários conduzidos entre 2005 e 2011 totalizando 58 fragmentos de cerrado. Os dados são resultados do Inventário Florestal de Minas Gerais (IFMG), o qual é detalhadamente descrito por Scolforo *et al.* (2015). Resumidamente, o IFMG foi construído da seguinte maneira:

- (1): amostragem em multi estágio: onde o primeiro estágio foi a estratificação das florestas de cerrado por fitofisionomia e subbacia hidrográfica. Já o segundo estágio foi a partir da seleção aleatória de fragmentos a serem inventariados em todos os estratos do estágio 1. Por fim, o estágio 3 correspondeu ao lançamento de parcelas distribuídas de forma sistemática em cada fragmento selecionado para amostragem no estágio 2.

- (2): 58 fragmentos de cerrado foram inicialmente mensurados entre 2005-2007, sendo 26 destes selecionados para novas mensurações entre 2010-2011. Conforme observado na Figura 1, a estratégia de amostragem do IFMG permitiu um conhecimento sem viés do cerrado em Minas Gerais.

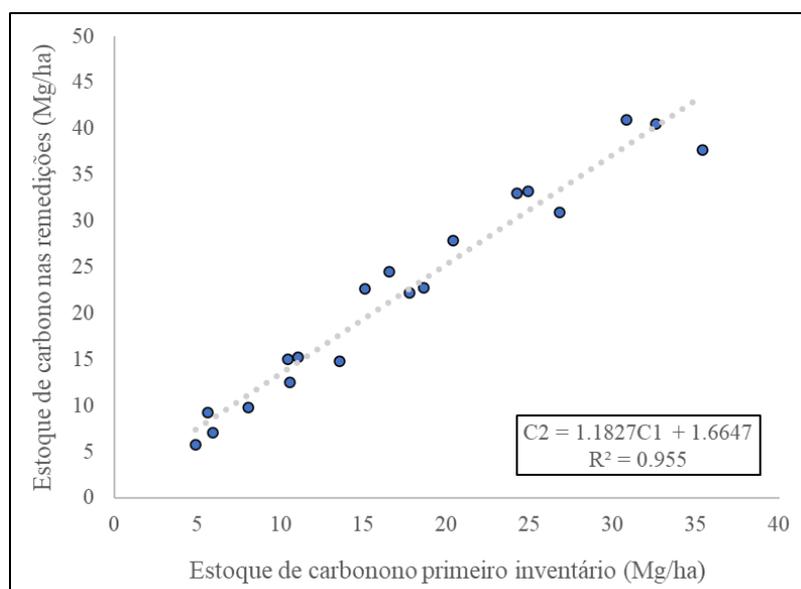
- (3): mais de 500 parcelas de 1000 m<sup>2</sup> foram alocadas nos 58 fragmentos de cerrado. Dentro das parcelas, todas as árvores com diâmetro a 1,30 metros acima do solo (DAP)  $\geq 5$  cm tiveram seus DAPs e alturas (em metros). Estas árvores também foram devidamente plaqueteadas para fins de remedição e identificadas botanicamente.

De maneira geral, todos os fragmentos remediados de cerrado apresentaram incremento de estoque de carbono arbóreo, independente do seu estágio sucessional (Tabela 1 e Figura 2). O crescimento em um intervalo de 4,3 anos (entre 2007 a 2011), foi de 5 MG/ha.

**Tabela 1.** Estoque de carbono arbóreo no cerrado de Minas Gerais.

Fragmentos inventariados	Estoque de carbono (Mg/ha)
Geral -> 2005 – 2007	14,7
26 fragmentos -> 2005 – 2007	15,9
26 fragmentos -> 2010 – 2011	20,9

**Figura 2.** Gráfico 1 x 1 retratando a aderência e a expectativa de ganho proporcional de estoque de carbono arbóreo entre os 26 fragmentos de cerrado que foram inicialmente inventariados entre 2005-2007 e remediados entre 2010-2011.



### 2.3. Variáveis bioclimáticas e de sensoriamento remoto

Dados bioclimáticos relacionados ao regime pluviométrico de Minas Gerais, com resolução de 1 km, foram extraídos do *WorldClim* (HIJMAN et al., 2015) para todos os fragmentos do estado:

- Bio13: precipitação (mm) do mês mais úmido
- Bio14: precipitação (mm) do mês mais seco
- Bio15: coeficiente de variação da precipitação
- Bio16: precipitação (mm) do trimestre mais úmido
- Bio17: precipitação (mm) do trimestre mais seco
- Bio18: precipitação (mm) do trimestre mais quente
- Bio19: precipitação (mm) do trimestre mais frio

Dados de sensoriamento remoto foram extraídos da Landsat TM. Foram adquiridas 35 cenas, sendo estas suficientes para cobrir toda a área de estudo. O download das cenas se deu através do “United States Geological Survey for Earth Observation and Science (USGS\EROS)” com as devidas correções geométricas e de reflectância (SILVEIRA *et al.*, 2019). Informações das imagens Landsat foram extraídas para o ano de 2010 (sempre sem nuvens). De maneira geral, valores de diferentes bandas foram extraídas para os fragmentos de Minas Gerais, entretanto, as variáveis de interesse deste estudo se concentraram nos seguintes índices:

NDVI: Normalized Difference Vegetation Index

SAVI: Soil Adjusted Vegetation Index

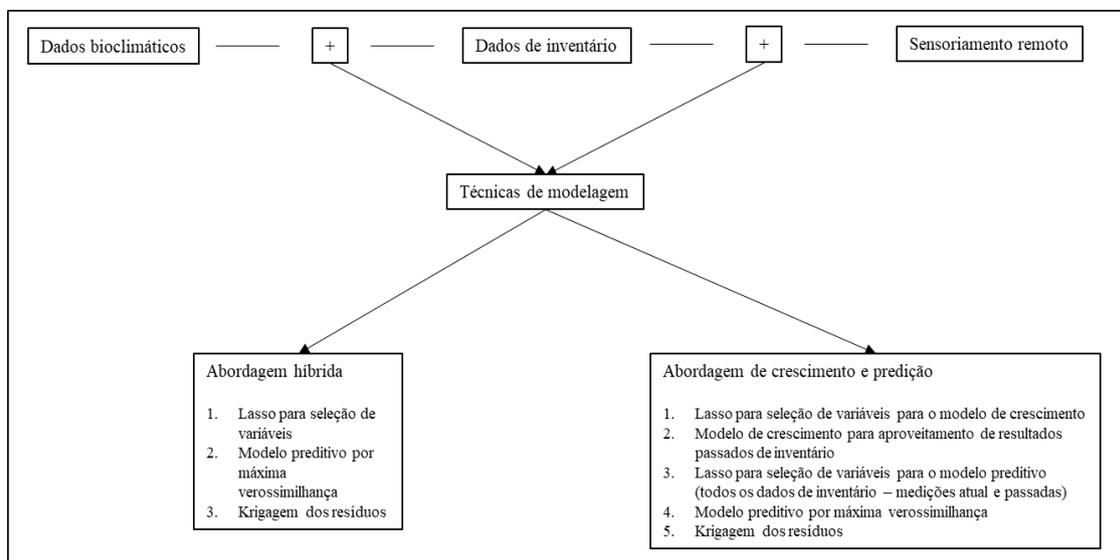
SR: Simple Ratio

### 2.4 Sistema de modelagem

A ideia central deste estudo foi apresentar uma abordagem que permita a combinação de modelagem de crescimento florestal e modelagem geoespacial, de forma que a incerteza nas estimativas de estoque de carbono arbóreo ao nível de fragmento florestal seja minimizada e, desta forma, maior confiabilidade seja gerada para tomada de decisão pública em Minas Gerais. Sendo assim, além do ganho de precisão de estimativas, o estudo também discute o efeito das incertezas associadas as estimativas de estoque de carbono em Minas Gerais.

Abaixo são descritas as duas abordagens: (1) híbrida e (2) crescimento e predição (CP) (Figura 3).

**Figura 3.** Diagrama resumindo o funcionamento das duas abordagens.



Análises estatísticas foram realizadas através do software R (R Core Team, 2016). Para tanto, os seguintes pacotes foram utilizados: glmnet (FRIEDMAN; HASTIE; TIBSHIRANI, 2010), gdata (WARNES *et al.*, 2018), reshape (WICKHAM, 2005), gstat (PEBESMA, 2004), optimx (NASH; VARADHAN, 2011), sp (PEBESMA; BIVAND, 2005), raster (HIJMANS; van ET TEN, 2012) e rgdal (KEITH *et al.*, 2010).

#### 2.4.1. Abordagem híbrida

A abordagem híbrida conforme apresentada no Artigo 1 deste tese, é composta de 3 passos, apresentados de forma resumida a seguir:

##### **Etapa 1: Least Absolute Shrinkage Selection Operator (lasso)**

O lasso (TIBSHIRANI, 1996) é uma técnica de aprendizado de máquina com foco principal em selecionar a partir de um modelo linear as variáveis (bioclimáticas e de sensoriamento remoto) com maior capacidade preditiva, assim como mitigar colinearidade presente entre as variáveis preditoras (1). Sendo assim, o lasso penaliza os coeficientes do modelo em questão, de forma que a soma absoluta destes seja inferior ao parâmetro de penalização  $\lambda$ . Naturalmente, essa penalização gera encolhimento dos

coeficientes de forma geral, zerando inclusive aqueles com menor relevância preditiva (SIMON, FRIEDMAN e HASTIE, 2013).

$$\min_{\beta_0, \beta} \{ \sum_{i=1}^N (C_i - \beta_0 - x_i^T \beta)^2 \} \text{ subject to } \sum_{i=1}^N |\beta_j| \leq t \quad (1)$$

Em que:  $\beta$ 's correspondem ao vetor de coeficientes; min é a função de minimizar;  $y_i$  é o vetor para a variável dependente de  $i = 1 \dots n$ ;  $x_i$  é o vetor com as variáveis independentes de  $i = 1 \dots n$ ;  $t$  é um parâmetro livre que determina o grau de regularização;  $\sum |\beta_i|$  é a restrição que garante com que os coeficientes não ultrapassem o valor de  $t$ .

### **Etapa 2: modelagem por máxima verossimilhança**

Com as variáveis selecionadas na Etapa 1, a abordagem híbrida parte para a etapa 2, a qual consiste no ajuste de um modelo preditivo por meio da máxima verossimilhança (Equação 2):

$$\lambda(\theta, \sigma | x_i, y_i) = \sum_{i=1}^n \log \left[ \frac{1}{\sqrt{2\pi}\sigma} \exp \left( \frac{-y_i + f(x_i, \theta)}{\sigma^2} \right) \right] \quad (2)$$

Em que:  $\lambda$  é a verossimilhança em logaritmo neperiano;  $f(\cdot)$  é a função do modelo com a expectativa do parâmetro  $\theta$ ;  $y_i$  é o vetor para a variável dependente de  $i = 1 \dots n$ ;  $x_i$  é o vetor com as variáveis independentes de  $i = 1 \dots n$ ;  $\sigma^2$  representa o parâmetro da variância;  $\pi$  é uma constante.

A vantagem do uso da máxima verossimilhança (em relação aos mínimos quadrados ordinários) é que se qualquer pressuposto residual não for atendido (distribuição normal, independência e com variância desconhecida) torna-se possível mitigar tal problema a partir da modelagem explícita da variância/erro do modelo.

Em especial para este estudo, este é um aspecto crítico, uma vez que somente pode-se avaliar efeito das incertezas a respeito das estimativas se a distribuição utilizada para o cálculo da incerteza tiver coerência com padrão de distribuição residual.

### **Etapa 3: krigagem ordinária dos resíduos**

A ideia central desta etapa consiste na modelagem espacial dos resíduos, de forma que parte deixa de ser aleatório. A etapa 3 da abordagem híbrida garante com que a

variância residual seja reduzida, além de ganhos de precisão das estimativas, em especial para regiões de limite do banco de dados.

A modelagem espacial dos resíduos é realizada através de um modelo exponencial (determinado via análise variográfica) e os parâmetros do modelo ajustado são então utilizados para interpolação via krigagem ordinária (JOURNAL; HUIJBREGTS, 1978).

$$\hat{Z}_{X0} = \sum_{i=1}^n \lambda_i Z_{xi} \quad (3)$$

Em que:  $\hat{Z}_{X0}$  é a estimativa do valor na posição  $X0$ ;  $n$  é o número de pontos amostrais vizinhos utilizados para predição dos pontos não amostrados  $\hat{Z}_{X0}$ ;  $\lambda_i$  é o  $i$ -th peso associado para cada  $i$ -th observação da variável de interesse na posição  $X$ ; e  $Z_{xi}$ , é o valor observado.

#### **2.4.2. Abordagem de crescimento e predição (CP)**

A abordagem CP apresenta propriedades semelhantes à abordagem híbrida, como utilização de lasso para seleção de variáveis, e a krigagem ordinária dos resíduos. Essa abordagem, entretanto, introduz a modelagem de crescimento florestal, a qual permite a utilização de resultados de inventários passados. As seguintes etapas foram consideradas:

##### **Etapa 1: lasso**

Lasso é utilizado para seleção das variáveis que melhor expressem o crescimento de estoque de carbono arbóreo de cerrado em Minas Gerais. Nesta fase, apenas as variáveis bioclimáticas foram consideradas.

##### **Etapa 2: modelagem de crescimento por meio da máxima verossimilhança**

Um modelo de crescimento (4) é utilizado para estimar ao longo do tempo o quanto um determinado fragmento florestal cresceu. Além disso, porém, é preciso considerar que também existe um efeito crescente de erro associado a tal estimativa de crescimento. Sendo assim, nessa etapa utilizou-se a técnica da máxima verossimilhança, onde um modelo explícito para o erro (5) foi simultaneamente ajustado ao modelo de crescimento:

$$\ln(C_2 - C_1) = \beta_0 + \beta_1 \times t + \beta_2 \text{Variável bioclimática} \quad (4)$$

$$\text{Erro} = \beta_3 \times \ln(C_2 - C_1)^{\beta_4} \quad (5)$$

Em que:  $C_2$  é o estoque de carbono arboreo remedido para um determinado fragmento de cerrado;  $C_1$  é o estoque de carbono arboreo inventariado pela primeira vez para um determinado fragmento de cerrado;  $t$  é o intervalo de tempo entre remedições; *Variável bioclimática* é a variável selecionada pelo Lasso na Etapa 1;  $\beta'_s$  são os coeficientes a serem ajustados.

O modelo foi ajustado de forma a garantir que o crescimento fosse expresso em função do tempo entre medições, assim como de uma variável que pudesse ser uma proxy de sítio. A modelagem explícita do erro permitiu que a incerteza associada a projeção de resultados de inventários passados para o tempo presente (no caso 2010) fosse contemplada nos demais passos da análise.

Por fim, vale comentar que para atualizar resultados de inventários de 2005-2007 para o tempo presente (2010, últimos inventários), basta rearranjar o modelo (4) já parametrizado na sua forma de projeção (assim como modelo de Erro).

### **Etapa 3: lasso**

Lasso foi utilizado conforme descrito na abordagem híbrida, ou seja, para seleção de variáveis para modelagem preditiva de estoque de carbono arboreo de cerrado em Minas Gerais. Nesta etapa, as variáveis bioclimáticas e de sensoriamento remoto foram consideradas.

### **Etapa 4: modelagem preditiva por meio da máxima verossimilhança**

Modelagem preditiva foi utilizada, conforme descrito na abordagem híbrida. Porém, enquanto a abordagem híbrida se baseia apenas nos 26 fragmentos inventariados em 2010, a etapa 4 também utiliza os resultados das projeções (e dos respectivos erros associados) dos inventários dos outros 32 fragmentos inventariados entre 2005-2007. Isso quer dizer que essa modelagem se beneficia de toda informação atual de carbono do cerrado de Minas Gerais, assim como de toda informação passada já apresentada.

## Etapa 5: krigagem ordinária do erro

Na etapa 5, a krigagem ordinária do erro foi realizada conforme descrito na abordagem híbrida. A diferença é que aqui os mapas que contabilizam o erro do modelo preditivo e o erro do modelo de projeção são somados.

### 2.4.3. Precisão das técnicas de modelagem

Dada a quantidade de fragmentos utilizados para modelagem do estoque de carbono arbóreo decidiu-se pela utilização de *bootstrapping* não paramétrico (reamostragem com repetição) para simulação de 1.000 bancos de dados para fins de validação das diferentes abordagens.

A validação foi realizada em duas etapas, sendo a primeira via a simulação dos 1.000 bancos de dados através da utilização dos 26 fragmentos inventariadas entre 2010-2011. Sendo assim, as estimativas geradas pelas duas abordagens puderam ser contrastadas quanto as suas precisões (erro médio (T), erro absoluto (MAE) e erro quadrático médio (RMSE)).

Em um segundo momento foi avaliado via análise gráfica 1x1, a aderência entre as estimativas geradas por cada abordagem nos demais 32 fragmentos inventariados entre 2005-2007. Desta forma, nesse segundo momento avaliou-se o quanto a modelagem via abordagem híbrida e abordagem de CP tem convergência quanto as suas estimativas geradas.

$$T \text{ (Mg/ha)} = \frac{\sum_{i=1}^n (obs-est)}{n} \quad (6)$$

$$MAE \text{ (Mg/ha)} = \frac{\sum_{i=1}^n (|obs-est|)}{n} \quad (7)$$

$$RMSE \text{ (Mg/ha)} = \sqrt{\frac{\sum_{i=1}^n (obs-est)^2}{n}} \quad (8)$$

Em que:  $i$  é o número total de fragmentos simulados; obs: valores observados de estoque de carbono arbóreo (Mg/ha); valores estimados de estoque de carbono arbóreo (Mg/ha).

## **2.5 Incerteza das estimativas geradas pelas duas abordagens**

Além da precisão das estimativas, também é importante a avaliação da incerteza global do estoque de carbono arbóreo no cerrado de Minas Gerais. Somente assim torna-se possível corrigir a rede amostral (inventário de novos fragmentos) para que continuamente a informação reportada nos mapas gerais de estoque de carbono arbóreo de cerrado de Minas Gerais seja a mais confiável, o que permitirá ações públicas específicas nas diferentes regiões de Minas Gerais.

Partiu-se da premissa que as duas abordagens avaliadas neste estudo geram resíduos que não ferem os pressupostos de normalidade. Sendo assim, foi possível calcular a diferença tanto de estimativa de estoque de carbono gerado entre os dois métodos, como a diferença do erro esperado (incerteza).

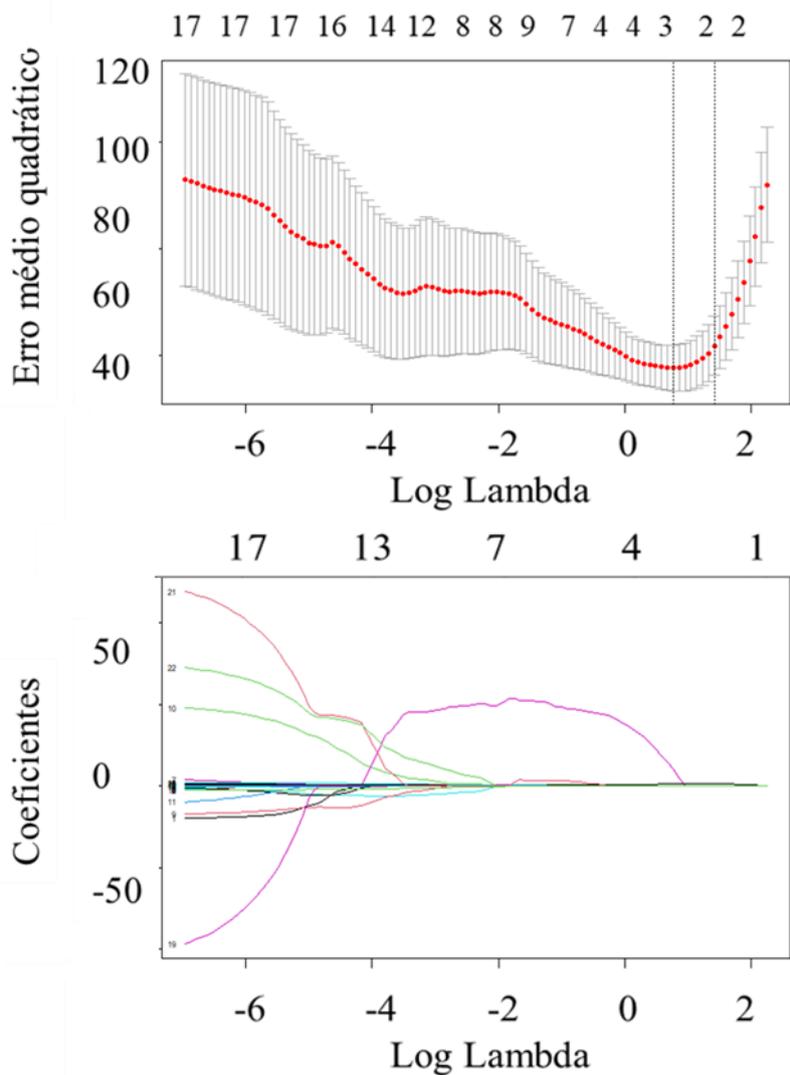
## **3. RESULTADOS**

### **3.1 Avaliação estatística das diferentes abordagens**

#### **3.1.1 Abordagem híbrida**

As variáveis selecionadas por lasso, utilizando  $\lambda$  a um desvio, foram SR e Bio18, sendo que estas apresentam correlação positiva com a variável resposta (Bio18 x C = 0,70 e SR x C = 0,85). A figura 4 apresenta a localização e o comportamento dos coeficientes em relação à penalização de  $\lambda$  ( $\lambda = 2,3$ ).

**Figura 4.** Comportamento do erro quadrático médio - RMSE (Mg/ha) e dos coeficientes com  $\lambda$ .



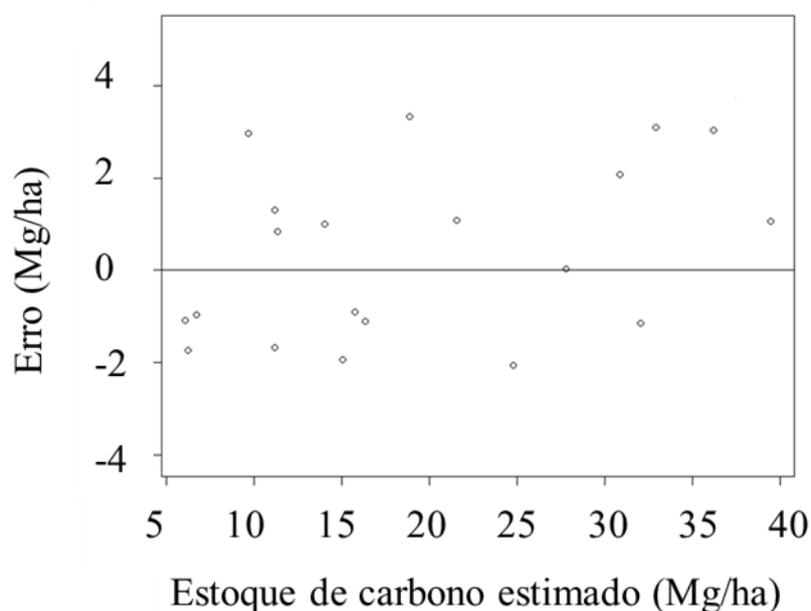
A etapa 2 da abordagem híbrida consiste no ajuste do modelo preditivo por meio da máxima verossimilhança:

$$\hat{C} = -10,10 + 0,02795Bio18 + 2,27995SR \quad (9)$$

Em que:  $\hat{C}$  = estoque de carbono predito (Mg/ha); demais variáveis foram previamente descritas.

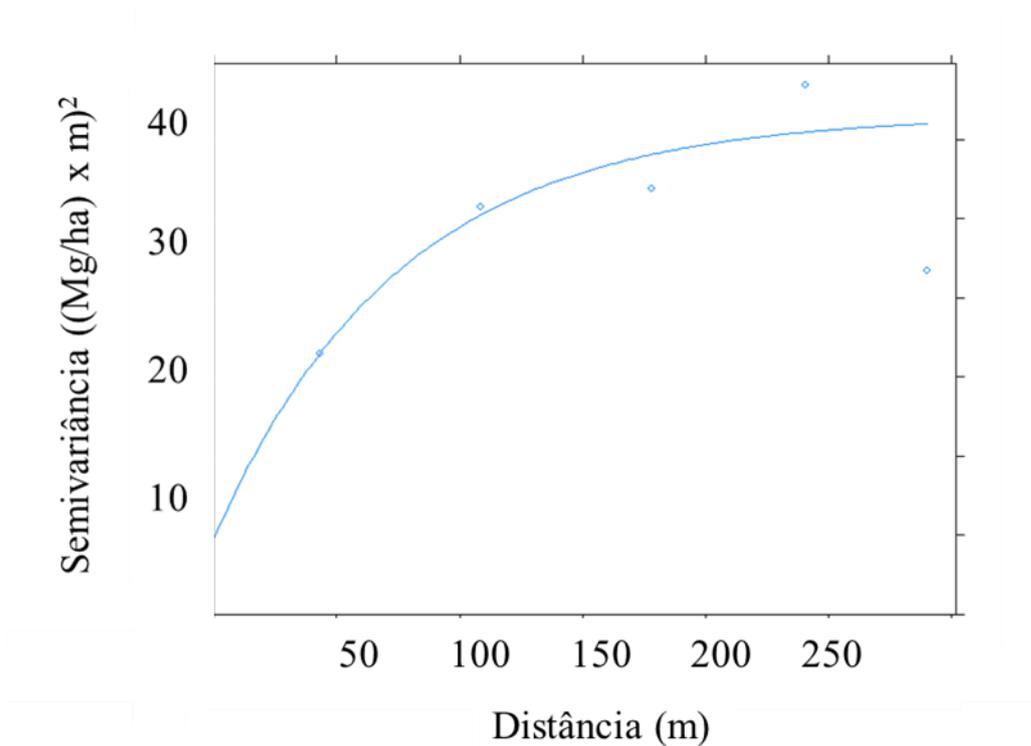
O modelo preditivo de estoque de carbono arbóreo no cerrado de Minas Gerais apresentou coeficientes biologicamente consistentes e significativos para um nível de significância de 0,05. Adicionalmente, as estatísticas de ajuste foram precisas ( $R^2 = 79\%$  e Erro padrão residual = 5,5 Mg/ha), e a distribuição residual não apresentou tendência (Figura 5).

**Figura 5.** Comportamento residual do modelo 9.



A etapa 3, por fim, consistiu na modelagem espacial dos resíduos gerados a partir da Etapa 2. A Figura 4 apresenta o semivariograma residual com a curva ajustada através do modelo exponencial. O semivariograma apresenta grau de dependência espacial superior a 75%, com alcance de 225 metros.

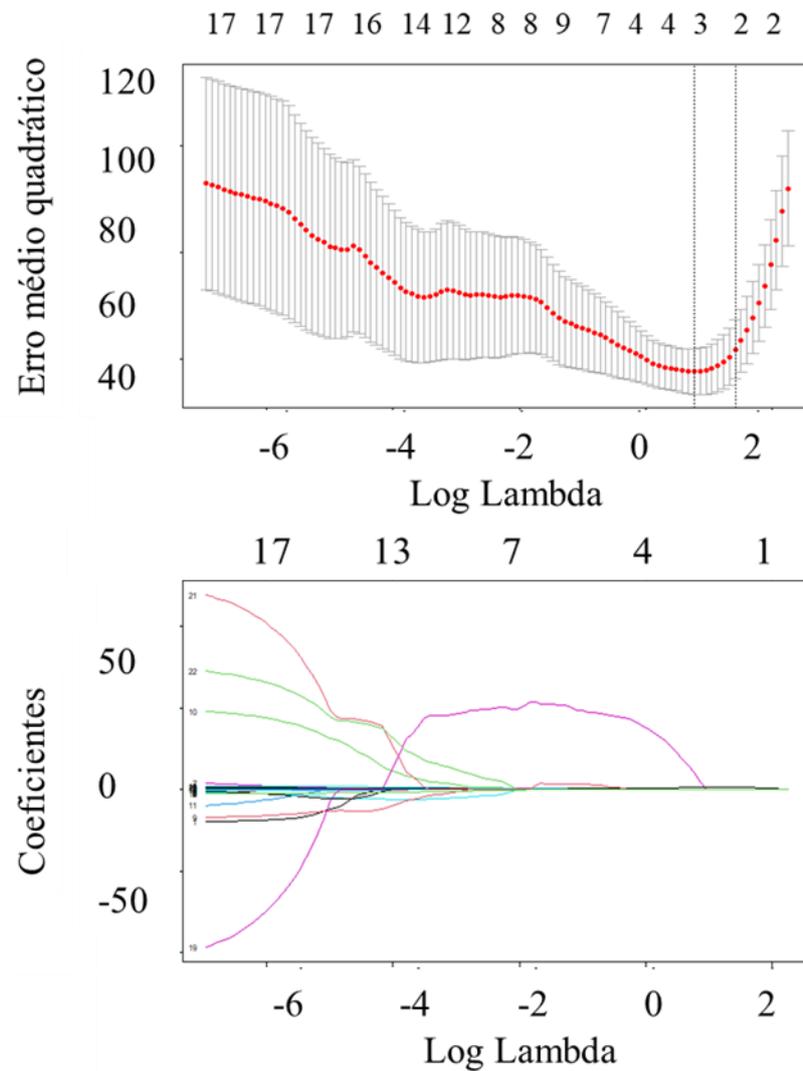
**Figura 4.** Semivariograma ajustado para os resíduos do estoque de carbono arbóreo (Mg/ha) do cerrado de Minas Gerais.



### 3.1.2 Abordagem de CP

As variáveis selecionadas como as mais relevantes para o modelo de crescimento foram o intervalo entre medições ( $t$ ) e Bio18 (correlação entre Bio18  $\times \ln(C_2 - C_1) = 0,54$ ; correlação entre  $t \times \ln(C_2 - C_1) = 0,80$ ). A figura 5 apresenta a relação entre o RMSE  $\times \lambda$ , além da relação entre o comportamento dos coeficientes  $\times \lambda$ .

**Figura 5.** Comportamento do erro quadrático médio - RMSE (Mg/ha) e dos coeficientes com  $\lambda$  ( $\lambda = 0,22$ ).



A etapa 2 consistiu no ajuste simultâneo do modelo de crescimento, assim como a modelagem explícita da variância (erro):

$$\ln(\widehat{C_2} - C_1) = 0,1912 \times t + 0,001159 \text{Bio18} \quad (10)$$

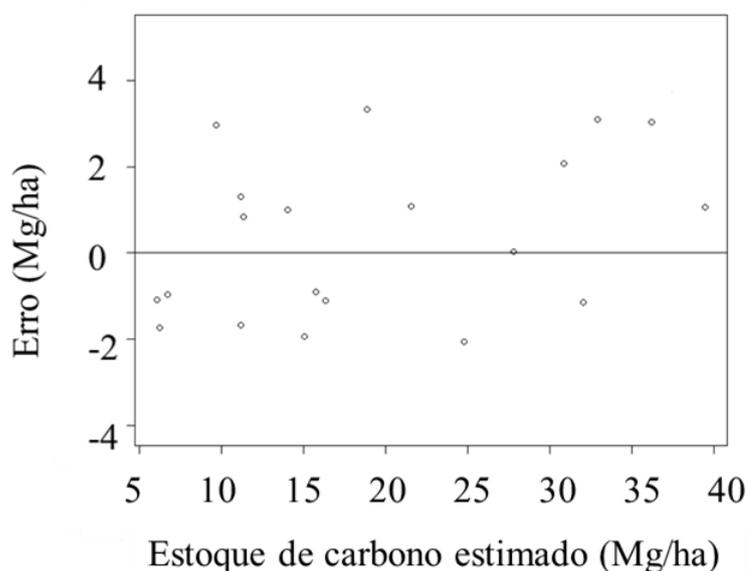
$$\widehat{\text{Erro}} = 1,0978 \times \ln(\widehat{C_2} - C_1)^{-1,4716} \quad (11)$$

Em que: todas as variáveis foram descritas anteriormente.

O modelo de crescimento apresentou coeficientes significativos (ao nível de 0,05 de significância), além de sinais biologicamente coerentes para os coeficientes ajustados. De forma geral, o modelo ajustado apresentou estatísticas de precisão adequadas ( $R^2 = 75\%$  e Erro padrão residual = 0,9 Mg/ha).

A tendência residual também atende aos pressupostos de normalidade residual (Figura 6). Por fim, o modelo de crescimento foi rearranjado para projetar e atualizar o inventário dos 32 fragmentos de cerrado inventariados apenas entre os anos de 2005-2007, além de projetar as incertezas associadas a essas estimativas.

**Figura 6.** Comportamento residual do modelo 10.



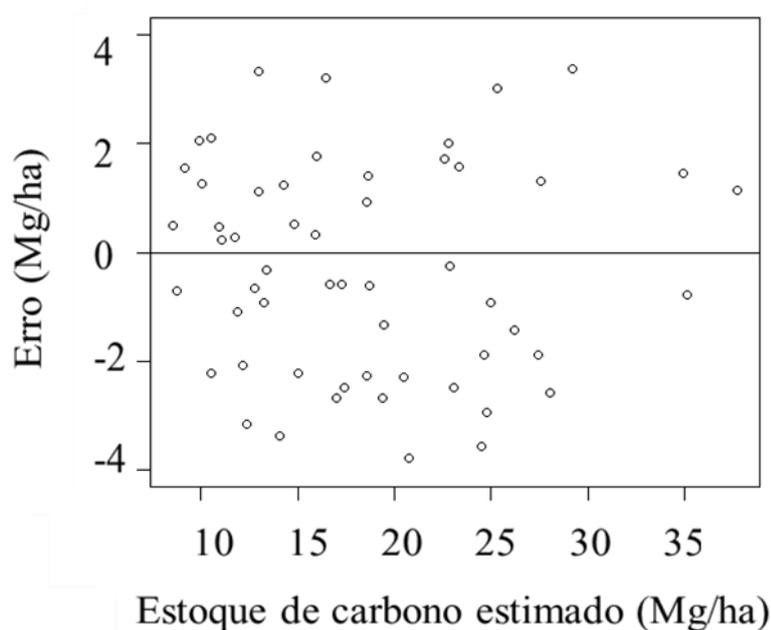
Para a seleção de variáveis por lasso (Etapa 3), repetiu-se a análise feita na abordagem híbrida e as variáveis utilizadas na modelagem preditiva foram Bio18 e SR. O modelo preditivo foi ajustado então com base em todos os fragmentos já inventariados, tanto os 26 remeidos em 2010, como os outros 32 fragmentos que tiveram seus estoques de carbono atualizados (Etapa 4). Finalmente, o modelo preditivo para estoque de carbono arbóreo de cerrado de Minas Gerais é apresentado abaixo:

$$\hat{C} = -3,77 + 0,01506Bio18 + 2,1351SR \quad (12)$$

Em que:  $\hat{C}$  = estoque de carbono predito (Mg/ha); demais variáveis foram previamente descritas.

Assim como o modelo de crescimento da Etapa 2, o modelo preditivo da Etapa 4 apresentou coeficientes consistentes e significativos ao nível de 0,05. Finalmente, este apresentou estatísticas de ajuste precisas ( $R^2 = 70\%$  e Erro padrão residual = 5,6 Mg/ha), além da distribuição residual não apresentar tendência (Figura 7).

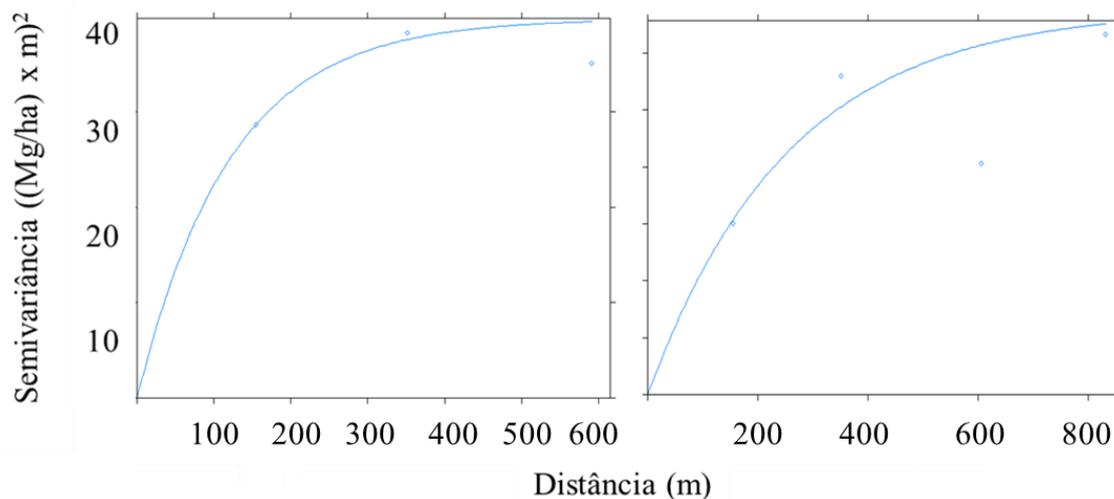
**Figura 7.** Comportamento residual do modelo 12.



Por fim, a modelagem espacial dos erros gerados a partir do modelo de crescimento (contabilizado na modelagem explícita do erro), assim como a modelagem espacial dos erros resultantes do modelo preditivo da Etapa 4 são apresentadas na Figura 8 (Etapa 5). Os semivariogramas tiveram as curvas ajustadas através de modelos exponenciais.

Para ambos os semivariogramas, o grau de dependência espacial é superior a 80%, com alcance superior a 500 metros. Vale ressaltar nessa etapa, que diferentemente da abordagem híbrida, não há somente o erro associado ao modelo preditivo, mas também é necessário contabilizar o erro/incerteza associada às projeções pela atualização dos inventários passados.

**Figura 8.** Semivariogramas ajustados para os erros do modelo de projeção (a) e do modelo preditivo (b).



### 3.2 Comparação entre os diferentes métodos

A tabela 2 expressa como ambas as técnicas apresentaram boa precisão de estimativa. Notadamente não há viés de estimativa, assim como o MAE e RMSE também são baixos. Sendo assim, fica evidente que ambas as técnicas estimam estoque de carbono com boa precisão, apesar da clara vantagem para a técnica de CP.

**Tabela 2.** Precisão das abordagens híbrida e de crescimento e predição (CP) na estimativa de estoque de carbono arbóreo dos fragmentos inventariados em 2010 no cerrado de Minas Gerais.

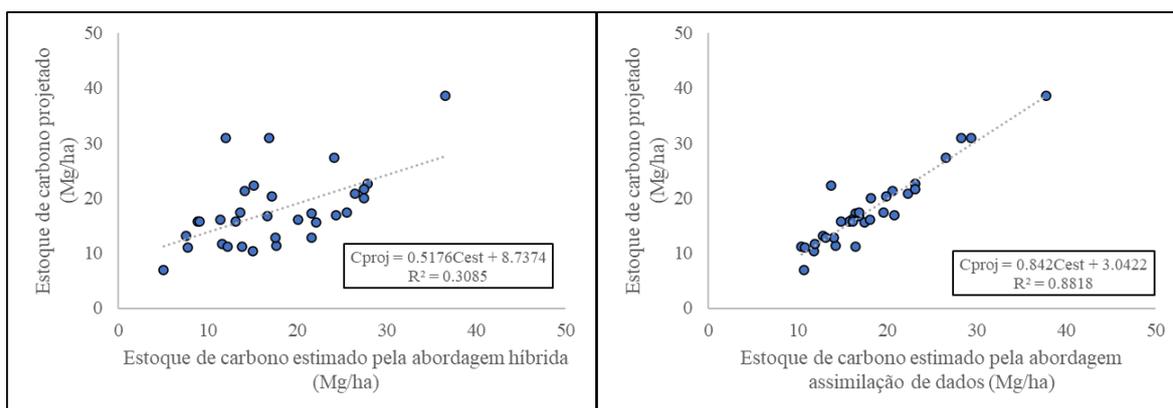
Estatística	Híbrida	Crescimento e predição
T (Mg/ha)	-0,33	-0,36
MAE (Mg/ha)	1,50	1,20
RMSE (Mg/ha)	1,95	1,59

### 3.2.1 Avaliação das abordagens aplicadas aos resultados de inventários atualizados

Na Figura 9 é possível observar como os estoques de carbono projetados pelo modelo de crescimento x estoques de carbono estimados pela abordagem de CP tem alta aderência (conforme esperado). Chama a atenção, porém, como a correlação entre projeções x estimativas a partir da abordagem híbrida apresentam baixa aderência.

Esse fato liga o sinal de alerta que por mais que a abordagem híbrida expresse a tendência da distribuição do estoque de carbono no gradiente climático/sensoriamento remoto de forma precisa, quando se busca aumento de precisão de estimativas ao nível de fragmento florestal tal abordagem tende a não atender.

**Figura 9.** Aderência entre as projeções de estoque de carbono dos inventários dos 32 fragmentos realizados entre 2005-2007 x estimativas da abordagem híbrida (a) e crescimento e predição (b).



De fato quando comparados os valores dos resultados dos inventários dos 32 fragmentos entre 2005-2007 x estimativas de carbono para o presente pela abordagem híbrida, percebe-se que para 9 fragmentos (28% do total) as estimativas desta abordagem foram inferiores ao reportado nos inventários. Isso explicita a incapacidade da abordagem na geração de estimativas precisas ao nível de fragmento florestal.

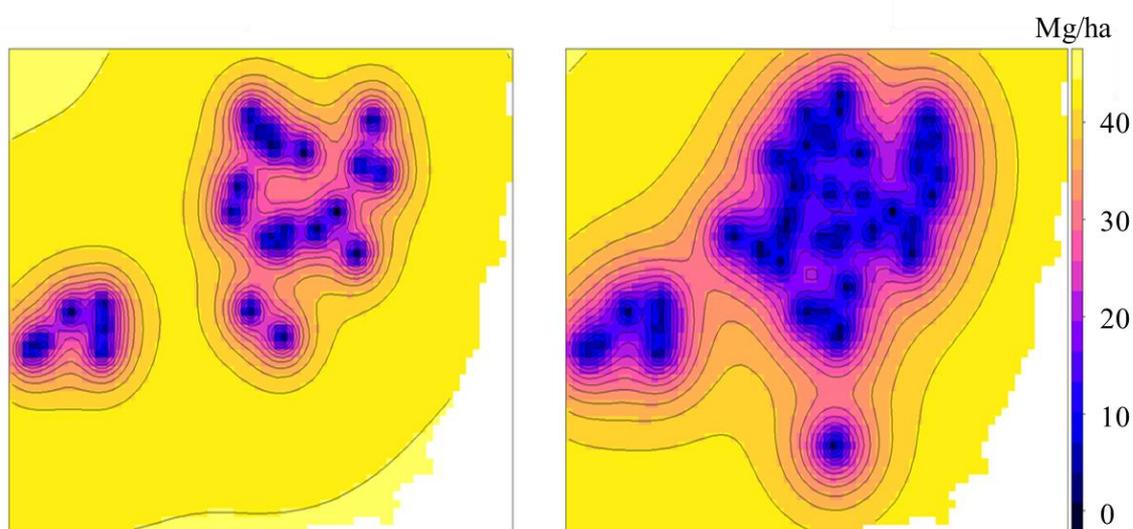
### 3.3 Incertezas associadas às estimativas geradas pelas diferentes abordagens

O tópico anterior demonstrou como a abordagem híbrida expressa bem, em média, a distribuição do estoque de carbono arbóreo do cerrado de Minas Gerais. Entretanto,

quando se requer precisão ao nível de fragmento florestal, essa abordagem não atende as expectativas, ao contrário da abordagem de crescimento e predição.

Naturalmente quando avaliada a distribuição dos erros/incertezas de estimativas via as duas abordagens, percebe-se como a abordagem de CP concentra valores entre 0 a 23 Mg/ha, enquanto a abordagem híbrida apresenta diversas regiões onde as incertezas das estimativas ultrapassam 35 Mg/ha (Figura 10).

**Figura 10.** Erro associado as estimativas das abordagens híbrida (a) e crescimento e predição (b) para o cerrado em Minas Gerais.



Para exemplificar e dar a clareza do efeito da incerteza, a abordagem híbrida tem uma incerteza associada as estimativas dos 32 fragmentos inventariados entre 2005-2007 na ordem de 33 Mg/ha, enquanto a abordagem de CP apresenta incerteza média de 18 Mg/ha (redução de 45%).

A abordagem de CP expressa a importância da modelagem espacial, correção residual por krigagem, assim como o uso de lasso para a seleção de variáveis. Essa técnica, porém, indica que todos resultados passados de inventário devem ser utilizados para melhoria das estimativas globais de carbono, o que necessariamente reflete numa possibilidade de gestão pública localizada.

#### 4. DISCUSSÃO

Este estudo objetivou apresentar uma abordagem de crescimento e predição para quantificação de estoque de carbono arbóreo no cerrado de Minas Gerais. A ideia central do estudo foi de apresentar uma nova abordagem que possibilite o uso de múltiplas informações, como de inventários passados, clima, sensoriamento remoto, etc. Além disso, é importante ressaltar que a combinação de múltiplas informações auxiliares em um processo que combina modelagem de crescimento e modelagem geoespacial, resulta em mapas de estoque de carbono com estimativas substancialmente mais precisas e com menores erros associados, quando comparadas às demais abordagens presentes na literatura para esse bioma brasileiro.

De maneira geral, a quantificação da distribuição do estoque de carbono no cerrado é realizada através de modelagem em função de variáveis geográficas (SCOLFORO *et al.*, 2015; SCOLFORO *et al.*, 2016), ou variáveis de sensoriamento (SILVEIRA *et al.*, 2019). Silveira *et al.* (2019), por exemplo, utilizaram variáveis bioclimáticas e de sensoriamento remoto para estimativa do estoque de carbono e reportaram estimativas precisas quanto a tendência espacial desta no cerrado de Minas Gerais. David *et al.* (2019) reportaram como a aplicação de krigagem com regressão permite estimativas mais adequadas de estoque de carbono, quando combinadas as variáveis bioclimáticas e de sensoriamento remoto.

Estudos conduzidos na Suécia, no entanto, refletem sobre como a presença de modelos de crescimento permitem a utilização de resultados de inventários passados, os quais naturalmente trazem ganho de precisão da estimativa de forma global. Cabe ressaltar também como a combinação de informação de inventários atuais, passados, além das variáveis bioclimáticas e de sensoriamento remoto permitem ganho global de estimativas, mas também redução de incerteza quanto as estimativas ao nível de fragmento florestal (NYSTRÖM *et al.*, 2015).

Existe uma ampla disponibilidade de informações bioclimáticas e mesmo de sensoriamento remoto. As informações de sensoriamento remoto, em especial, ganham cada vez maiores granularidades. Como exemplo, a missão GEDI (POTAPOV *et al.*, 2021) que vem permitindo a extração de informações LiDAR para diferentes fragmentos, o que naturalmente permite um acréscimo de precisão local das estimativas de estoque de

carbono. O que falta no entanto, são modelos de crescimento florestal para aproveitamento dos inventários passados na estimativa atual de carbono.

Este estudo propõe a utilização de uma modelagem de crescimento parcimoniosa e que controle o intervalo de tempo para minimizar incertezas associadas a atualização dos inventários passados para o momento dos últimos inventários.

Por fim, fica nítido como a combinação de modelos de crescimento para atualização de inventários passados, além de modelagem preditiva utilizando variáveis bioclimáticas e de sensoriamento remoto, e finalmente krigagem dos resíduos agregam em estimativas de carbono mais precisas e com menor grau de incerteza. Isso, no entanto, reforça a necessidade de não desativar a rede de parcelas permanentes do cerrado de Minas Gerais, assim como é importante a cada 5 anos existirem novas medições de parcelas temporárias. Essa seria a combinação ótima quanto ao aproveitamento de inventários passados, demais informações auxiliares, assim como aumento de informação para locais onde incerteza momentânea ainda é alta. Adicionalmente, o amplo aproveitamento de todas as informações auxiliares disponíveis tendem a trazer bons resultados, como exemplo, a missão GEDI, a qual deve ser aproveitada para refinamento ainda maior dos resultados já entregues neste estudo.

## **5. CONCLUSÃO**

Uma abordagem baseada na combinação de modelos de crescimento e geoespacial foi apresentada neste estudo. Os resultados reportados reforçam como de maneira geral a abordagem aumenta a precisão das estimativas e reduz as incertezas associadas as mesmas.

Uma das principais vantagens da nova abordagem passa pela possibilidade de uso de diferentes variáveis auxiliares, como variáveis bioclimáticas, sensores orbitais e inventários passados. Entretanto, é fundamental que parcelas temporárias e permanentes sejam inventariadas a cada 5 anos de maneira que a informação do estoque de carbono seja sempre a mais atualizada possível, além de possibilitar a geração de modelos de crescimento cada vez mais confiáveis.

## REFERÊNCIAS

- ALTOÉ, T. F. **Modelagem e espacialização do estoque de carbono em fragmentos nativos de Minas Gerais em função de fatores edafoclimáticos**. 2017. 130 p. Tese. (Doutorado em Engenharia Florestal) – Universidade Federal de Lavras, Lavras, 2017.
- BRASIL. Ministério do Meio Ambiente, dos Recursos Hídricos e da Amazônia Legal. **Ações prioritárias para a conservação da biodiversidade do Cerrado e Pantanal**. Brasília, DF: Ventura Comunicações e Cultura, 1999. 24 p.
- CARVALHO, M.C. *et al.* Algoritmos de aprendizagem de máquina na modelagem da distribuição potencial de habitats de espécies arbóreas. **Nativa**, Sinop, v. 7, n. 5, p. 600-606, 2019. DOI: <http://dx.doi.org/10.31413/nativa.v7i5.7214>.
- DAVID, H.C., *et al.* Exploring coarse-to fine-scale approaches for mapping and estimating forest volume from Brazilian National Forest Inventory data. **Forestry**, Londres, v. 92, n. 5, p. 577–590, oct., 2019. DOI: 10.1093/forestry/cpz030
- EITEN, G. Vegetação do cerrado. In: PINTO, M. N. **Cerrado: caracterização, ocupação e perspectivas**. Brasília, DF: UnB, 1994.p. 11-73.
- FRIEDMAN J, HASTIE T, TIBSHIRANI R. Regularization Paths for Generalized Linear Models via Coordinate Descent. **Journal of Statistical Software**, [S.I.], v. 33, n. 1, p. 1–22, 2010.
- HIJMANS R J.; CAMERON S. E.; PARRA J.L.; JONES P.G.; JARVIS A. Very high-resolution interpolated climate surfaces for global land areas. **International Journal Climatology**, Chichester. v. 25, n. 15, p. 1965-1978, nov. 2005. DOI: 10.1002/joc.1276/full.
- HIJMANS, R. J.; VAN ETTEN; J. **raster: Geographic analysis and modeling with raster data**. 2012. R package version 2.0-12. URL <http://CRAN.R-project.org/package=raster>. (acesso em 10 de agosto de 2021).
- JOURNEL, A. G.; HUIJBREGTS, C. J. **Mining geostatistics**. London: Ed. Academic, 1978. 600 p.

KEITT, T.H.; BIVAND, R.; PEBESMA, E.; ROWLINGSON, B. **rgdal: bindings for the Geospatial Data Abstraction Library**. 2010. R package version 0.6-27. URL: <http://CRAN.R-project.org/package=rgdal>. (acesso em 10 de agosto de 2021).

KLINK, C. A.; MACHADO, R. B. A conservação do cerrado brasileiro. **Mega diversidade**, Belo Horizonte, v. 1, n. 1, p. 147-155, jul. 2005.

MASCARO, J. *et al.* A Tale of Two “Forests”: Random Forest Machine Learning Aids Tropical Forest Carbon Mapping. **PLOS ONE**, San Francisco, v. 9, n. 1, p. e85993, jan., 2014. DOI: 10.1371/journal.pone.0085993.

MENDONÇA, M. P.; LINS, L. V. **Lista vermelha das espécies ameaçadas de extinção da flora de Minas Gerais**. Belo Horizonte: Fundação Biodiversitas; Fundação Zoológica Botânica de Belo Horizonte, 2000. 160 p.

NASH, J.C.; VARADHAN, R., Unifying Optimization Algorithms to Aid Software System Users: optimx for R. **Journal of Statistical Software**, Innsbruck, v. 43, n.9, p. 1–14, 2011. Disponível em <http://www.jstatsoft.org/v43/i09/>. (acesso em 10 de agosto de 2021).

NYSTRÖM, M. *et al.*, Data Assimilation in Forest Inventory: First Empirical Results. **Forests**, Basel, v. 6, p. 4540–4557, dec., 2015. DOI :10.3390/f6124384.

OLIVEIRA, M.C. *et al.* "Avaliação de diferentes níveis de intervenção na florística, diversidade e similaridade de uma área de Cerrado stricto sensu." **Cerne**, Lavras 12.4 (2006): 342-349.

PEBESMA E. J. Multivariable geostatistics in S: the gstat package. **Computers & Geosciences**, Vandœuvre-les-Nancy, v. 30, n. 7, p. 683-691, aug., 2004.

PEBESMA, E.; BIVAND, R. S. S classes and methods for spatial data: the sp package. **R news**, v. 5, n. 2, p. 9-13, 2005.

POTAPOV, P. *et al.* Mapping global forest canopy height through integration of GEDI and Landsat data. **Remote Sensing of Environment**, Toronto, v. 253, feb., p. 112165, oct., 2021. DOI: 10.1016/j.rse.2020.112165.

R CORE TEAM. R: **A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. 2019. Disponível em <https://www.R-project.org/>. (acesso em 10 de agosto de 2021).

REIS, A.A., FRANKLIN, S.E., MELLO, J.M., ACERBI JR, F.W. Volume estimation in a Eucalyptus plantation using multi-source remote sensing and digital terrain data: a case study in Minas Gerais State, Brazil. **International Journal of Remote Sensing**, Londres v. 40, n. 7, p. 1-2018. DOI: 10.1080/01431161.2018.1530808.

SCOLFORO, H. F.; SCOLFORO, J. R. S.; DE MELLO, J. M.; DE MELLO, C. R.; MORAIS, V. A. Spatial interpolators for improving the mapping of carbon stock of the arboreal vegetation in Brazilian biomes of Atlantic forest and Savanna. **Forest Ecology and Management**, Amsterdam, v. 376, p. 24-35, sep., 2016. DOI: 10.1016/j.foreco.2016.05.047

SCOLFORO, H. F.; SCOLFORO, J. R. S.; MELLO, C. R.; MELLO, J. M.; FERRAZ FILHO, A. C. 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. e0128781, 2015. DOI: 10.1371/journal.pone.0128781

SCOLFORO, J. R.; OLIVEIRA, A. D.; ACERBI JÚNIOR, F. W. A. (Eds.). **Inventário florestal de Minas Gerais: equações de volume, peso de matéria seca e carbono para diferentes Fisionomias da flora nativa**. Lavras: Ed. UFLA, 2008. 216 p.

SILVEIRA, E. *et al.* Pre-stratified modelling plus residuals kriging reduces the uncertainty of aboveground biomass estimation and spatial distribution in heterogeneous savannas and forest environments. **International Journal of Applied Earth Observation and Geoinformation**, v. 78, p. 175-188, aug., 2019c. DOI: 10.1016/j.foreco.2019.05.016.

SIMON, N.; FRIEDMAN, J.; HASTIE, T. A blockwise descent algorithm for group-penalized multiresponse and multinomial regression. **Journal of Statistical Software**, [Linz], v. 20, n. 2, p. 1-15, nov. 2013.

TIBSHIRANI, R., Regression Shrinkage and Selection via the Lasso. **Journal of the Royal Society. Series B (Methodological)**, [S.I.], v. 58, n. 1, p. 267-288. 1996.

WARNES, G.R. *et al.* **gdata: Various R programming tools for data manipulation.** 2014. R package version 2.18.0. URL <https://CRAN.R-project.org/package=gdata>. (acesso em 10 de agosto de 2021).

WICKHAM, H., **reshape: Flexibly reshape data.** 2005. R package version 0.71.2. URL: <http://had.co.nz/reshape/>. (acesso em 10 de agosto de 2021).