SAMUEL JOSÉ SILVA SOARES DA ROCHA

MÉTODOS DE APRENDIZADO DE MÁQUINA APLICADOS A MODELAGEM DE FLORESTAS INEQUIÂNEAS

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência Florestal, para obtenção do título de *Doctor Scientiae*.

Orientador: Carlos Moreira M. Eleto Torres

Coorientadores: Helio Garcia Leite
Laércio Antônio Gonçalves Jacovine

VIÇOSA - MINAS GERAIS
2021

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T

R672m
2021

Rocha, Samuel José Silva Soares da, 1992-
Métodos de aprendizado de máquina aplicados a modelagem de florestas inequiâneas / Samuel José Silva Soares da Rocha. – Viçosa, MG, 2021.
139 f. : il. (algumas color.) ; 29 cm.

Inclui apêndice.
Orientador: Carlos Moreira Miquelino Eleto Torres.
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Florestas tropicais. 2. Inteligência artificial. 3. Dinâmica florestal - Modelos matemáticos. 4. Recrutamento (População biológica). 5. Florestas - Mortalidade. 6. Florestas - Crescimento. I. Universidade Federal de Viçosa. Departamento de Engenharia Florestal. Programa de Pós-Graduação em Ciência Florestal. II. Título.

CDO adapt. CDD 22. ed. 634.91811

SAMUEL JOSÉ SILVA SOARES DA ROCHA

MÉTODOS DE APRENDIZADO DE MÁQUINA APLICADOS A MODELAGEM DE FLORESTAS INEQUIÂNEAS

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência Florestal, para obtenção do título de *Doctor Scientiae*.

APROVADA: 06 de abril de 2021.

Assentimento:

Samuel José S S da Rocha

Samuel José Silva Soares da Rocha
Autor

Carlos M. M. Eleto Torres

Carlos Moreira Miquelino Eleto Torres
Orientador

AGRADECIMENTOS

A Deus Pai, Filho e Espírito Santo.

Nossa Senhora e todos os anjos e santos por sempre cuidarem de mim.

À minha família, em especial aos meus pais, Waldemiro Soares da Rocha (*in memoriam*) e Maria de Fátima Silva Soares, pelo amor insubstituível. Ao meu irmão, Davi Silva, pelo companheirismo.

À minha namorada Raquel Ker, por toda compreensão e carinho.

Ao professor e amigo Carlos Eleto Torres pelos ensinamentos, amizade, incentivo e por nunca medir esforços em me auxiliar.

Aos professores Laércio Jacovine e Helio Garcia Leite pela orientação e confiança desde o início da graduação.

Aos professores Carlos Pedro Boechat Soares e Liniker Fernandes da Silva pela orientação e pelas contribuições.

Ao pesquisador Leonardo Reis por todo auxílio.

Aos professores Lívia Thaís Moreira de Figueiredo e Renato Vinícius Oliveira Castro pela disponibilidade e atenção.

Aos pesquisadores Bruno Leão Said Schettini, Paulo Henrique Villanova e Indira Bifano Comini por compartilharem suas experiências.

À Universidade Federal de Viçosa (UFV), ao Departamento de Engenharia Florestal (DEF) e ao Programa de Pós-Graduação em Ciência Florestal pela oportunidade e por todo subsídio.

Aos funcionários da UFV e do DEF, em especial: Adão, Alexandre, Antônio, Dilson, Francisco, Elizabeth, Gabriel, Josimar, Marquione, Maria Auxiliadora, Neuza, Paulo, Rômulo e Sebastião.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão da bolsa de estudos.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), pelos financiamentos.

Ao Grupo de Estudos em Economia Ambiental (GEEA). Em especial, aos estagiários e tantos outros que colaboraram.

Ao programa Carbono Zero por todos os ensinamentos e oportunidades.

Aos amigos da Pós-Graduação. Em especial: Aline Farias, Arthur Araújo, Caio Eleto, Eliana Boaventura, Fernanda Lambrecht, Flora Magdaline, Hugo Tharner, Ivaldo Tavares, Lauana Blenda, Maria Paula Rufino, Mindaye Teshome, Otávio Verly, Tamara Braga, Valéria Silva, Vicente Toledo e Vinícius Rodrigues.

Aos amigos da turma de 2010. Em especial: Àguida Beatriz, João Paulo Ladeira, Lucas Arthur, Pedro Mota e Valéria Silva.

Aos amigos do ASC (Adolescentes Seguidores de Cristo), JSC (Jovens Seguidores de Cristo) e ESC (Eternos Seguidores de Cristo) da paróquia Santa Rita de Cássia pela formação humana e cristã.

Aos amigos da Capelania da UFV por todo o suporte e experiências.

A todos aqueles que, direta ou indiretamente, contribuíram. Muito obrigado!

"Ninguém é suficientemente perfeito, que não possa aprender com o outro e, ninguém é totalmente destituído de valores que não possa ensinar algo ao seu irmão."

São Francisco de Assis

RESUMO

ROCHA, Samuel José Silva Soares da, D.Sc., Universidade Federal de Viçosa, abril de 2021. **Métodos de aprendizado de máquina aplicados a modelagem de florestas inequiâneas**. Orientador: Carlos Moreira Miquelino Eleto Torres. Coorientadores: Helio Garcia Leite e Laércio Antônio Gonçalves Jacovine.

As florestas tropicais vêm sofrendo com a fragmentação e o desmatamento. No Brasil, a Mata Atlântica e a floresta Amazônica são exemplos disso. Nesses biomas, o conhecimento e a modelagem dos atributos florestais podem ser ferramentas úteis na compreensão destes biomas e no auxílio para adoção de medidas e práticas de manutenção da biodiversidade. Modelos de aprendizado de máquina podem ser técnicas promissoras nessas tarefas. Assim, o objetivo dessa pesquisa foi avaliar modelos de aprendizado de máquina para estimativas de atributos no bioma da Mata Atlântica e da Floresta Amazônica, no Brasil. Utilizamos dados de sete fragmentos de Mata Atlântica localizados em Minas Gerais, Brasil e uma área sob manejo florestal no sudoeste da Amazônia, no município de Porto Acre, estado do Acre, Brasil. Para atender os objetivos o trabalho foi dividido em quatro artigos. No primeiro artigo, aplicou-se modelos de aprendizado de máquina, Redes Neurais Artificiais (RNA), Máquina de Vetor de Suporte (MVS) e Random Forests (RF), para prever o recrutamento em nível de parcela na Mata Atlântica do Brasil. Atributos florestais, histórico de uso do solo, paisagem, solo e características climáticas foram usados na modelagem. O método Recursive Feature Elimination foi usado para selecionar o melhor subconjunto de variáveis preditoras. Observou-se que clima, paisagem, histórico de uso da terra e atributos da floresta são variáveis importantes para prever o recrutamento de árvores na Mata Atlântica no Brasil. O RF apresentou o melhor desempenho para estimar as taxas de recrutamento, com o maior Coeficiente de Correlação de Pearson ($r_{y\hat{y}}$) e os menores Raiz do Erro Quadrático Médio (REQM) e Erro Médio Absoluto (EMA) para todas as repetições. No segundo artigo, utilizou-se as variáveis e modelos do primeiro artigo para estimar as taxas de mortalidade em nível de parcela. Constatou-se que o clima (precipitação, déficit hídrico climático e temperatura), idade de abandono e área basal são variáveis importantes para predizer a mortalidade de árvores na Mata Atlântica no Brasil. O RF também apresentou o melhor desempenho para estimar as taxas de mortalidade. No terceiro artigo, estimou-se o crescimento líquido com as informações utilizadas nos dois primeiros artigos. Observou-se que as variáveis edáficas, atributos da floresta e climáticas são importantes preditores das taxas de crescimento líquido na Mata Atlântica brasileira. Os métodos de aprendizado de máquina foram eficientes.

O método Random Forests também mostrou superioridade sobre os demais para modelagem de crescimento na Mata Atlântica. No quarto artigo, estimou-se o volume e biomassa de árvores comerciais no sudoeste da Amazônia. Utilizou-se variáveis dendrométricas, climáticas e topográficas. O Algoritmo de Boruta foi aplicado para selecionar o melhor conjunto de variáveis. Máquina de Vetor de Suporte (MVS), Redes Neurais Artificiais (RNA), Random Forests (RF) e Modelo Linear Generalizado (MLG) foram os métodos de aprendizado de máquina avaliados. Em geral, os métodos avaliados mostraram um poder de generalização satisfatório. Os resultados mostraram que as previsões de volume e biomassa de árvores comerciais na floresta amazônica diferiram entre as técnicas ($p < 0{,}05$). As RNAs apresentaram os melhores desempenhos para prever o volume e a biomassa das árvores comerciais, com o maior $r_{y\hat{y}}$ e os menores REQM e EMA. Por fim, constatamos que a utilização de Aprendizado de Máquina é uma abordagem promissora em estimativas de atributos na Mata Atlântica e na Floresta Amazônica. Esses modelos representam uma alternativa para modelagem de florestas tropicais ao redor do mundo, sobretudo as ameaçadas, como as aqui estudadas.

Palavras-chave: Florestas tropicais. Inteligência Artificial. Dinâmica Florestal. Recrutamento. Mortalidade. Crescimento.

ABSTRACT

ROCHA, Samuel José Silva Soares da, D.Sc., Universidade Federal de Viçosa, April, 2021. **Machine learning methods applied to uneven-aged mixed forest modelling**. Adviser: Carlos Moreira Miquelino Eleto Torres. Co-advisers: Helio Garcia Leite and Laércio Antônio Gonçalves Jacovine.

Tropical forests have been suffering from fragmentation and deforestation. In Brazil, the Atlantic Forest and the Amazon rainforest are examples of this. In these biomes, knowledge and modeling of forest attributes can be useful tools in understanding these biomes and in helping to adopt measures and practices for maintaining biodiversity. Machine learning models can be promising techniques in these tasks. Thus, the objective of this research was to evaluate machine learning models for estimating attributes in the Atlantic Forest and Amazon Forest biome, in Brazil. We used data from seven fragments of Atlantic Forest located in Minas Gerais, Brazil and an area under forest management in southwestern Amazonia, in the municipality of Porto Acre, state of Acre, Brazil. To meet the objectives, the work was divided into four scientific papers. In the first, machine learning models, Artificial Neural Networks (ANN), Support Vector Machine (SVM) and Random Forests (RF) were applied to predict recruitment at plot level in the Brazilian Atlantic Forest. Forest attributes, history of land use, landscape, soil and climatic characteristics were used in the modeling. The Recursive Feature Elimination method was used to select the best subset of predictor variables. It was observed that climate, landscape, history of land use and attributes of the forest are important variables to predict the recruitment of trees in the Atlantic Forest in Brazil. The RF presented the best performance to estimate recruitment rates, with the highest Pearson correlation coefficient ($r_{y\hat{y}}$) and the lowest Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) for all repetitions. In the second article, the variables and models of the first article were used to estimate mortality rates at the plot level. It was found that the climate (precipitation, water deficit and temperature), age of abandonment and basal area are important variables to predict the mortality of trees in the Atlantic Forest in Brazil. The RF also performed best to estimate mortality rates. In the third paper, net growth was estimated with the information used in the first two articles. It was observed that the edaphic, forest attributes and climatic variables are important predictors of the net growth rates in the Brazilian Atlantic Forest. Machine learning methods were efficient. The Random Forests method also showed superiority over the others for modeling growth in the Atlantic Forest. In the fourth paper, the volume and biomass of commercial trees in the southwest of the Amazon were estimated. Dendrometric, climatic and topographic variables were used. The Boruta Algorithm was applied to select the best set of variables. Support Vector Machine (SVM), Artificial Neural Networks (ANN), Random Forests (RF) and Generalized Linear Model (GLM) were the machine learning methods evaluated. In general, the evaluated methods showed a satisfactory generalization power. The results showed that the volume and biomass predictions of commercial trees in the Amazon rainforest differed between the techniques ($p < 0.05$). The ANNs showed the best performances to predict the volume and biomass of commercial trees, with the highest $r_{y\hat{y}}$ and the lowest RSME and MAE. Finally, we found that the use of Machine Learning is a promising approach in estimating attributes in the Atlantic Forest and the Amazon Forest. These models represent an alternative for modeling tropical forests around the world, especially threatened, such as those studied here.

Keywords: Tropical forests. Artificial Intelligence. Forest Dynamics. Recruitment. Mortality. Growth.

SUMÁRIO

INTRODUÇÃO GERAL

As florestas tropicais são vitais para a biodiversidade global, pois são meios de subsistência para muitas comunidades locais, além de fornecerem diversos serviços ecossistêmicos (Neef, 2020; Sullivan et al., 2017). No Brasil, a Mata Atlântica e a Floresta Amazônica merecem uma atenção especial. A Mata Atlântica é uma das florestas mais fragmentadas do mundo (Haddad et al., 2015) e a Floresta Amazônica nos últimos anos, vêm sofrendo bastante com o desmatamento (Fearnside, 2021).

Nesse contexto, a compreensão dos processos demográficos (crescimento, ingresso e mortalidade) destas florestas, podem ser úteis na manutenção da biodiversidade e garantia de manejo florestal sustentável. Esses processos e a produtividade primária em regiões tropicais são influenciadas pelas interações entre clima, solo, ações antrópicas e paisagem (Malhi et al., 2015; Wagner et al., 2016).

Por isso, modelos de aprendizagem de máquina que integrem essas variáveis podem servir como base para prognosticar a dinâmica de árvores em florestas tropicais, além de fornecerem subsídios para avaliação dos seus consequentes impactos na dinâmica florestal, tal como nos estoques e perdas de carbono ocasionadas por variações no clima (Allen et al., 2015; Anderegg et al., 2012; Intergovernmental Panel on Climate Change, 2014).

Aprendizado de máquina é uma área de estudo da inteligência artificial em rápido crescimento que deve se tornar mais comum para a modelagem florestal devido ao seu potencial para produzir modelos melhores do que as abordagens tradicionais de modelagem de dados (Gleason and Im, 2012; Jachowski et al., 2013; Zhao et al., 2011). As aplicações desta técnica de inteligência computacional na área florestal têm ganhado alta relevância (Reis et al., 2018, 2016; Rocha et al., 2018; Tavares Júnior et al., 2020). Modelos de aprendizado de máquina já foram testados com eficiência para estimar o crescimento das árvores (Reis et al., 2016), biomassa e carbono (Corona-Núñez et al., 2017; Nandy et al., 2017; Santi et al., 2017), mapear a riqueza e composição de espécies (Foody and Cutler, 2006), prognósticos de diâmetro e altura de árvores (Diamantopoulou et al., 2015; Diamantopoulou and Özçelik, 2012; Vieira et al., 2018); mapeamento da estrutura de floresta tropical (Ingram et al., 2005) e avaliar os parâmetros da qualidade florestal (Zhao et al., 2019). Desse modo, a aplicação desses modelos em estimativas de dinâmica e estoque em florestas tropicias podem representar uma alternativa promissora.

Neste estudo, realizou-se uma comparação de técnicas de aprendizagem de máquina, a saber: Redes Neurais Artificiais, Máquina de Vetores de Suporte, Random Forest e Modelo

Linear Generalizado para prever atributos de florestas no bioma Mata Atlântica e na Amazônia brasileira. Para isso, objetivou-se avaliar modelos de aprendizado de máquina para estimativas de atributos no bioma da Mata Atlântica e da Floresta Amazônica, no Brasil. Para atender a estes objetivos, o trabalho foi dividido em artigos, conforme descrito a seguir.

Artigo 1: Machine learning: Modeling tree recruitment rates in Atlantic Forest, Brazil;

Artigo 2: Machine learning: Modeling tree mortality rates in Atlantic Forest, Brazil;

Artigo 3: Comparison of machine learning methods in net growth estimates in the Atlantic Forest of Brazil;

Artigo 4: Volume and biomass estimates of commercial trees in the Amazon forest using machine learning.

Referências bibliográficas

Allen, C.D., Breshears, D.D., McDowell, N.G., 2015. On underestimation of global vulnerability to tree mortality and forest die-off from hotter drought in the Anthropocene. Ecosphere 6, 1–55. https://doi.org/10.1890/ES15-00203.1

Anderegg, W.R.L., Berry, J.A., Smith, D.D., Sperry, J.S., Anderegg, L.D.L., Field, C.B., 2012. The roles of hydraulic and carbon stress in a widespread climate-induced forest die-off. PNAS 109, 233–237. https://doi.org/10.1073/pnas.1107891109

Corona-Núñez, R.O., Mendoza-Ponce, A., López-Martínez, R., 2017. Model selection changes the spatial heterogeneity and total potential carbon in a tropical dry forest. Forest Ecology and Management 405, 69–80. https://doi.org/10.1016/j.foreco.2017.09.018

Diamantopoulou, M.J., Özçelik, R., 2012. Evaluation of different modeling approaches for total tree-height estimation in Mediterranean Region of Turkey. Forest Systems 21, 383–397. https://doi.org/10.5424/fs/2012213-02338

Diamantopoulou, M.J., Özçelik, R., Crecente-Campo, F., Eler, Ü., 2015. Estimation of Weibull function parameters for modelling tree diameter distribution using least squares and artificial neural networks methods. Biosystems Engineering 133, 33–45. https://doi.org/10.1016/j.biosystemseng.2015.02.013

Fearnside, P.M., 2021. The intrinsic value of Amazon biodiversity. Biodivers Conserv 30, 1199–1202. https://doi.org/10.1007/s10531-021-02133-7

Foody, G.M., Cutler, M.E.J., 2006. Mapping the species richness and composition of tropical forests from remotely sensed data with neural networks. Ecological Modelling, Selected Papers from the Third Conference of the International Society for Ecological

Informatics (ISEI), August 26--30, 2002, Grottaferrata, Rome, Italy 195, 37–42. https://doi.org/10.1016/j.ecolmodel.2005.11.007

Gleason, C.J., Im, J., 2012. Forest biomass estimation from airborne LiDAR data using machine learning approaches. Remote Sensing of Environment 125, 80–91. https://doi.org/10.1016/j.rse.2012.07.006

Haddad, N.M., Brudvig, L.A., Clobert, J., Davies, K.F., Gonzalez, A., Holt, R.D., Lovejoy, T.E., Sexton, J.O., Austin, M.P., Collins, C.D., Cook, W.M., Damschen, E.I., Ewers, R.M., Foster, B.L., Jenkins, C.N., King, A.J., Laurance, W.F., Levey, D.J., Margules, C.R., Melbourne, B.A., Nicholls, A.O., Orrock, J.L., Song, D.-X., Townshend, J.R., 2015. Habitat fragmentation and its lasting impact on Earth's ecosystems. Science Advances 1, e1500052. https://doi.org/10.1126/sciadv.1500052

Ingram, J.C., Whittaker, R.J., Dawson, T.P., 2005. Tree structure and diversity in human-impacted littoral forests, madagascar. Environ Manage 35, 779–798. https://doi.org/10.1007/s00267-004-0079-9

Intergovernmental Panel on Climate Change, 2014. Climate Change 2013 – The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781107415324

Jachowski, N.R.A., Quak, M.S.Y., Friess, D.A., Duangnamon, D., Webb, E.L., Ziegler, A.D., 2013. Mangrove biomass estimation in Southwest Thailand using machine learning. Applied Geography 45, 311–321. https://doi.org/10.1016/j.apgeog.2013.09.024

Malhi, Y., Doughty, C.E., Goldsmith, G.R., Metcalfe, D.B., Girardin, C.A.J., Marthews, T.R., Aguila-Pasquel, J. del, Aragão, L.E.O.C., Araujo-Murakami, A., Brando, P., Costa, A.C.L. da, Silva-Espejo, J.E., Amézquita, F.F., Galbraith, D.R., Quesada, C.A., Rocha, W., Salinas-Revilla, N., Silvério, D., Meir, P., Phillips, O.L., 2015. The linkages between photosynthesis, productivity, growth and biomass in lowland Amazonian forests. Global Change Biology 21, 2283–2295. https://doi.org/10.1111/gcb.12859

Nandy, S., Singh, R., Ghosh, S., Watham, T., Kushwaha, S.P.S., Kumar, A.S., Dadhwal, V.K., 2017. Neural network-based modelling for forest biomass assessment. Carbon Management 8, 305–317. https://doi.org/10.1080/17583004.2017.1357402

Neef, A., 2020. Tropical forests lost to land grabbing. Nature Geoscience 13, 460–461. https://doi.org/10.1038/s41561-020-0604-3

Reis, L.P., de Souza, A.L., dos Reis, P.C.M., Mazzei, L., Soares, C.P.B., Miquelino Eleto Torres, C.M., da Silva, L.F., Ruschel, A.R., Rêgo, L.J.S., Leite, H.G., 2018. Estimation of mortality and survival of individual trees after harvesting wood using artificial neural networks in the amazon rain forest. Ecological Engineering 112, 140–147. https://doi.org/10.1016/j.ecoleng.2017.12.014

Reis, L.P., de Souza, A.L., Mazzei, L., dos Reis, P.C.M., Leite, H.G., Soares, C.P.B., Torres, C.M.M.E., da Silva, L.F., Ruschel, A.R., 2016. Prognosis on the diameter of individual trees on the eastern region of the amazon using artificial neural networks. Forest Ecology and Management 382, 161–167. https://doi.org/10.1016/j.foreco.2016.10.022

Ribeiro, M.C., Metzger, J.P., Martensen, A.C., Ponzoni, F.J., Hirota, M.M., 2009. The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. Biological Conservation, Conservation Issues in the Brazilian Atlantic Forest 142, 1141–1153. https://doi.org/10.1016/j.biocon.2009.02.021

Rocha, S.J.S.S. da, Torres, C.M.M.E., Jacovine, L.A.G., Leite, H.G., Gelcer, E.M., Neves, K.M., Schettini, B.L.S., Villanova, P.H., Silva, L.F. da, Reis, L.P., Zanuncio, J.C., 2018. Artificial neural networks: Modeling tree survival and mortality in the Atlantic Forest biome in Brazil. Science of The Total Environment 645, 655–661. https://doi.org/10.1016/j.scitotenv.2018.07.123

Santi, E., Paloscia, S., Pettinato, S., Fontanelli, G., Mura, M., Zolli, C., Maselli, F., Chiesi, M., Bottai, L., Chirici, G., 2017. The potential of multifrequency SAR images for estimating forest biomass in Mediterranean areas. Remote Sensing of Environment 200, 63–73. https://doi.org/10.1016/j.rse.2017.07.038

Sullivan, M.J.P., Talbot, J., Lewis, S.L., Phillips, O.L., Qie, L., Begne, S.K., Chave, J., Cuni-Sanchez, A., Hubau, W., Lopez-Gonzalez, G., Miles, L., Monteagudo-Mendoza, A., Sonké, B., Sunderland, T., ter Steege, H., White, L.J.T., Affum-Baffoe, K., Aiba, S., de Almeida, E.C., de Oliveira, E.A., Alvarez-Loayza, P., Dávila, E.Á., Andrade, A., Aragão, L.E.O.C., Ashton, P., C, G.A.A., Baker, T.R., Balinga, M., Banin, L.F., Baraloto, C., Bastin, J.-F., Berry, N., Bogaert, J., Bonal, D., Bongers, F., Brienen, R., Camargo, J.L.C., Cerón, C., Moscoso, V.C., Chezeaux, E., Clark, C.J., Pacheco, Á.C., Comiskey, J.A., Valverde, F.C., Coronado, E.N.H., Dargie, G., Davies, S.J., De Canniere, C., K, M.N.D., Doucet, J.-L., Erwin, T.L., Espejo, J.S., Ewango, C.E.N., Fauset, S., Feldpausch, T.R., Herrera, R., Gilpin, M., Gloor, E., Hall, J.S., Harris, D.J., Hart, T.B., Kartawinata, K., Kho, L.K., Kitayama, K., Laurance, S.G.W., Laurance, W.F., Leal, M.E., Lovejoy, T., Lovett, J.C., Lukasu, F.M., Makana, J.-R., Malhi, Y., Maracahipes, L., Marimon, B.S., Junior, B.H.M., Marshall, A.R., Morandi, P.S., Mukendi, J.T., Mukinzi, J., Nilus, R., Vargas, P.N., Camacho, N.C.P., Pardo, G., Peña-

Claros, M., Pétronelli, P., Pickavance, G.C., Poulsen, A.D., Poulsen, J.R., Primack, R.B., Priyadi, H., Quesada, C.A., Reitsma, J., Réjou-Méchain, M., Restrepo, Z., Rutishauser, E., Salim, K.A., Salomão, R.P., Samsoedin, I., Sheil, D., Sierra, R., Silveira, M., Slik, J.W.F., Steel, L., Taedoumg, H., Tan, S., Terborgh, J.W., Thomas, S.C., Toledo, M., Umunay, P.M., Gamarra, L.V., Vieira, I.C.G., Vos, V.A., Wang, O., Willcock, S., Zemagho, L., 2017. Diversity and carbon storage across the tropical forest biome. Scientific Reports 7, 39102. https://doi.org/10.1038/srep39102

Tavares Júnior, I. da S., Torres, C.M.M.E., Leite, H.G., Castro, N.L.M. de, Soares, C.P.B., Castro, R.V.O., Farias, A.A., 2020. Machine learning: Modeling increment in diameter of individual trees on Atlantic Forest fragments. Ecological Indicators 117, 106685. https://doi.org/10.1016/j.ecolind.2020.106685

Vieira, G.C., de Mendonça, A.R., da Silva, G.F., Zanetti, S.S., da Silva, M.M., dos Santos, A.R., 2018. Prognoses of diameter and height of trees of eucalyptus using artificial intelligence. Science of The Total Environment 619–620, 1473–1481. https://doi.org/10.1016/j.scitotenv.2017.11.138

Wagner, F.H., Hérault, B., Bonal, D., Stahl, C., Anderson, L.O., Baker, T.R., Becker, G.S., Beeckman, H., Boanerges Souza, D., Botosso, P.C., Bowman, D.M.J.S., Bräuning, A., Brede, B., Brown, F.I., Camarero, J.J., Camargo, P.B., Cardoso, F.C.G., Carvalho, F.A., Castro, W., Chagas, R.K., Chave, J., Chidumayo, E.N., Clark, D.A., Costa, F.R.C., Couralet, C., Silva Mauricio, P.H. da, Dalitz, H., Castro, V.R. de, Freitas Milani, J.E. de, Oliveira, E.C. de, Souza Arruda, L. de, Devineau, J.-L., Drew, D.M., Dünisch, O., Durigan, G., Elifuraha, E., Fedele, M., Ferreira Fedele, L., Figueiredo Filho, A., Finger, C.A.G., Franco, A.C., Freitas Júnior, J.L., Galvão, F., Gebrekirstos, A., Gliniars, R., Graça, P.M.L. de A., Griffiths, A.D., Grogan, J., Guan, K., Homeier, J., Kanieski, M.R., Kho, L.K., Koenig, J., Kohler, S.V., Krepkowski, J., Lemos-Filho, J.P., Lieberman, D., Lieberman, M.E., Lisi, C.S., Longhi Santos, T., López Ayala, J.L., Maeda, E.E., Malhi, Y., Maria, V.R.B., Marques, M.C.M., Marques, R., Maza Chamba, H., Mbwambo, L., Melgaço, K.L.L., Mendivelso, H.A., Murphy, B.P., O'Brien, J.J., Oberbauer, S.F., Okada, N., Pélissier, R., Prior, L.D., Roig, F.A., Ross, M., Rossatto, D.R., Rossi, V., Rowland, L., Rutishauser, E., Santana, H., Schulze, M., Selhorst, D., Silva, W.R., Silveira, M., Spannl, S., Swaine, M.D., Toledo, J.J., Toledo, M.M., Toledo, M., Toma, T., Tomazello Filho, M., Valdez Hernández, J.I., Verbesselt, J., Vieira, S.A., Vincent, G., Volkmer de Castilho, C., Volland, F., Worbes, M., Zanon, M.L.B., Aragão, L.E.O.C., 2016. Climate seasonality limits leaf carbon assimilation and wood productivity in tropical forests. Biogeosciences 13, 2537–2562. https://doi.org/10.5194/bg-13-2537-2016

Zhao, K., Popescu, S., Meng, X., Pang, Y., Agca, M., 2011. Characterizing forest canopy structure with lidar composite metrics and machine learning. Remote Sensing of Environment 115, 1978–1996. https://doi.org/10.1016/j.rse.2011.04.001

Zhao, Q., Yu, S., Zhao, F., Tian, L., Zhao, Z., 2019. Comparison of machine learning algorithms for forest parameter estimations and application for forest quality assessments. Forest Ecology and Management 434, 224–234. https://doi.org/10.1016/j.foreco.2018.12.019

CAPÍTULO I: Machine learning: Modeling tree recruitment rates in Atlantic Forest, Brazil

Abstract
Tree recruitment models are important in projections future forest growth and production. However, it is often neglected in dynamics models. In this study, we evaluated machine learning techniques to predict recruitment rates. We used Artificial Neural Networks (ANN), Suport Vector Machine (SVM) and Random Forest (RF). We calculated the recruitment rate (R) and relative recruitment rate (Rrel), % basal area (BA) year^{-1} at plot level. Seven Atlantic Forest fragments located in Brazil were studied. We collect information about Forest attributes, land use history, landscape, soil and climatic characteristics. Recursive Feature Elimination was used to select the best subset of predictor variables. We have found that climate, landscape, land use history and forest attributes are important variables to predict tree recruitment in the Atlantic Forest in Brazil. Machine Leaning are efficient methods to estimate recruitment rates at plot-level with these variables. The results show differences in the prediction of recruitment rates in the Atlantic Forest, the models used differed from each other ($p < 0.05$). The Friedman and Nemenyi nonparametric test confirm that the RF model is the best. Our results, they suggest a new alternative for modeling this important component of forest dynamics. And it can help improve estimates of the dynamics, growth, and production of native forests, especially those of the Atlantic Forest.

Keywords: forest dynamics, recruitment, tropical forest, Brazil.

1. Introduction

The Brazilian Atlantic Forest is one of the most endangered biodiversity hostpots in the world (Bellard et al., 2014; Ribeiro et al., 2009; Scarano and Ceotto, 2015). Urbanization, industrialization, and agricultural land use have caused the fragmentation of the biome (Scarano and Ceotto, 2015). The remaining fragments suffer from the erosion of biomass and biodiversity (de Lima et al., 2020). In this context, knowledge and modeling of forest dynamics can help in understanding it and in habitat conservation strategies.

Forest demographic processes include three components: tree mortality, tree growth, and recruitment of new trees (Zhang et al., 2012). Mortality and growth models in the biome have already been performed (Rocha et al., 2018; Tavares Júnior et al., 2020). As for recruitment, there are few studies. Because forest management requires a long-term perspective, exploring patterns of tree recruitment under perspectives of global change in climate, can increase our understanding of the composition, structure, and function of forests around the world (Perea et al., 2020), and is an important step in understanding the biome.

Recruitment models are important tools for predicting forest dynamics, especially for long-term projections of future forest composition (de Avila et al., 2017; Xiang et al., 2016). Disregarding it will provide a biased prediction of future forest growth and productivity (Zhang et al., 2012). The recruitment of trees in secondary threatened forests, such as those of the Atlantic Forest, is important for the design of local conservation and restoration strategies (Safar et al., 2020).

However, as a highly variable, complicated and largely stochastic process, tree recruitment remains difficult to model accurately (Xiang et al., 2016). Several factors affect tree recruitment (Clement et al., 2019) and are often neglected in models of native forest dynamics. In addition, the use of edaphic, climatic, landscape, and especially land use history variables can improve the estimates. In human-modified landscapes, successional pathways are largely defined by land-use history (Jakovac et al., 2021).

Because of this high variability and stochastic process, machine learning models may be a promising approach in predicting this component of the dynamics. Machine learning is a fast-growing area of artificial intelligence that performs well in tropical forest dynamics models (Reis et al., 2018, 2016; Rocha et al., 2018; Tavares Júnior et al., 2020)..

The objective of the present study is to develop models of machine learning, Artificial Neural Networks (ANN), Support Vector Machine (SVM) and *Random Forest* (RF) capable of estimating the recruitment rates in Atlantic Forest fragments, based on forest attributes, land use history, landscape, soil and climatic characteristics. (i) What are the most important variables to estimate recruitment in these forests? (ii) Are the machine learning methods evaluated efficient for estimating the recruitment rates at plot-level? (iii) What is the best method to estimate the recruitment rates at plot-level?

2. Material and methods

2.1. Study sites and plot characteristics

We used data from seven Atlantic Forest fragments located in Minas Gerais, Brazil (Table 1 and Figure 1). The vegetation is classified as semideciduous seasonal forest (IBGE, 2012). Across sites the annual rainfall varies from 701 to 1737 mm. year^{-1}, elevation from 242 to 1169 m above sea level, and slope from 2.97 to 65.81% (Table S1).

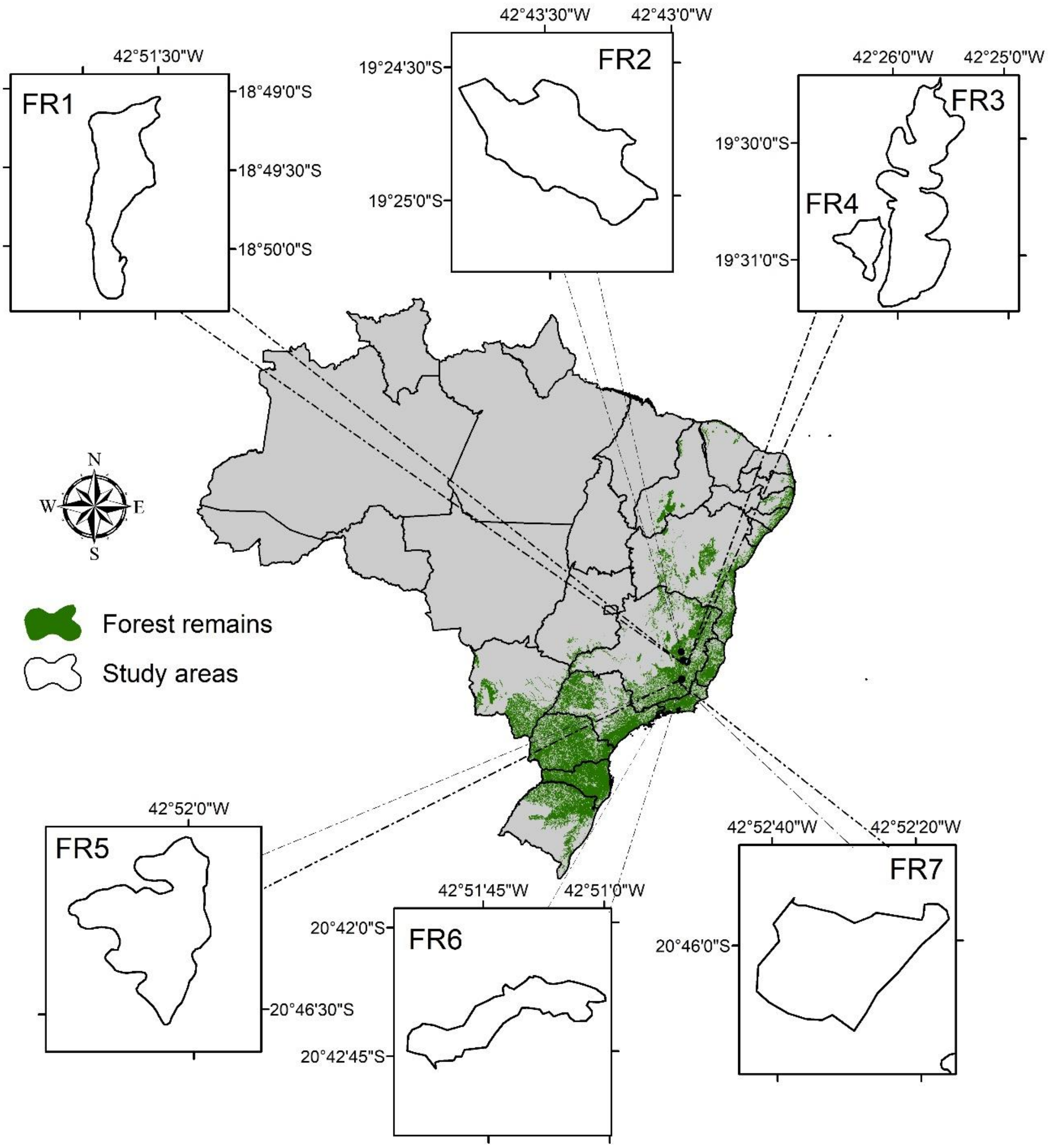

Figure 1. Locations of the seven studied Atlantic Forest fragments (FR) in Minas Gerais, Brazil. FR1: Cachoeira das Pombas; FR2: Mata da Garagem; FR3: Ipaba Mata1; FR4: Ipaba Mata2; FR5: Centev; FR6: São José and FR7: Mata da Silvicultura.

We measured 104 plots located in seven forest fragments over several years (Table 1). In all plots, we measured and identified botanically all stems with diameter at breast height (dbh), 1.3 m ≥ 5 cm.

Table 1. Overview of the seven studied Atlantic Forest fragments (FR) in Minas Gerais, Brazil. Forest fragments location (municipality), size, number and size of plots, and years of forest inventory are also provided

Fragment	Municipality	Forest Size (ha)	Plots	Plots size (ha)	Measurement

FR1	Guanhães	106.0	20	0.05	2002, 2007, 2012, 2017
FR2	Viçosa	21.8	50	0.01	1996, 1998, 2001, 2003, 2013, 2018
FR3	Caratinga	264.0	16	0.05	2002, 2007, 2012, 2017
FR4	Caratinga	37.3	6	0.05	2002, 2007, 2012, 2017
FR5	Viçosa	44.1	20	0.05	2010, 2015
FR6	Coronel Fabriciano	38.4	12	0.05	2002, 2007, 2012, 2017
FR7	Viçosa	17.0	10	0.10	1994, 1997, 2000, 2004, 2008, 2010, 2013, 2016

FR1: Cachoeira das Pombas; FR2: Mata da Garagem; FR3: Ipaba Mata1; FR4: Ipaba Mata2; FR5: Centev; FR6: São José and FR7: Mata da Silvicultura.

We calculated the recruitment rate at plot level (van der Sande et al., 2017). The recruitment rate (% year^{-1}) was calculated as proposed (Shiel and May, 1996) (Equation 1).

$$R=\left(\left((N_0+r)/N_0\right)^{\frac{1}{y}}-1\right)x100 \tag{1}$$

Where:

R - annual recruitment rate, % year $^{-1}$

N_0 - number of individuals in the initial population;

r - number of recruits trees;

y - time, years.

As demographic processes may be influence by the total basal area (BA) of the plot (Breugel et al., 2019; Carreño-Rocabado et al., 2012; Menezes and Melo, 2019; Rozendaal and Chazdon, 2015). We calculated the BA of each individual tree and measurement year to estimate the relative recruitment rate at plot level (van der Sande et al., 2017). Recruitment is the basal area of stems that reached the minimum diameter for inclusion (dbh > 5 cm) in census t_n + 1. Relative recruitment rate (Rrel) was calculated (Lebrija-Trejos et al., 2010; Martínez-Ramos et al., 2018) (Equation 2).

$$Rrel=\left(1-\left((N_0+r)/N_0\right)^{\frac{1}{y}}\right)x100 \tag{2}$$

Where:

Rrel - Relative recruitment rate, % BA year^{-1};

N_0 - Basal area of the stems alive on census in the initial population;

r - Basal area of recruitment (m^2);

y - time, years.

2.2. Forest attributes and Anthropogenic variables

We use forest attributes from Atlantic Forest fragments as predictive variables. The attributes of the forest used were: Basal area of the plot and Number of stems. This information was collected through forest inventories carried out over the years.

The Anthropogenic variables used to estimate tree recruitment rates were: Land use history, forest cover, forest size and edge distance, age of abandonment (that means, time since abandonment previous anthropogenic uses). Based on information collected from aerial photographs (from around 1960, 1980 and 1990), Landsat satellite images from 1985, landowner interviews and land titles, we determined the Land use history and age of abandonment of the areas. The categories of land use used were: deforestation, agricultural production, eucalyptus plantation, and selective logging.

Forest cover was calculated for each plot using circular buffers with radii of 500, 1000, and 2000 m, for the year 1985, 2002 and 2017, with data from MapsBiomas (MapBiomas Project, 2019) in ArcGIS 10.3.1 (ESRI, 2015). Forest cover area was produced from the pixel-per-pixel (30 x 30 m) classification of Landsat satellite images through the Google Earth Engine platform (MapBiomas Project, 2019). We calculated distance to the nearest edge of the forest (edge distance) with near tool in ArcGIS 10.3.1 1 (ESRI, 2015).

2.3. Climate variables

We obtained annual precipitation, the number of months with less than 100 mm of rainfall, precipitation in the three driest months, and average annual temperature for each fragment from the nearest climatological station (Figure S1). We then estimated the climatic water deficit (CWD) – a water balance between precipitation and evapotranspiration – as a proxy for drought conditions following Lutz et al. (2010) and using the R function CWD and AET (actual evapotranspiration) from Redmond (2019).

We used slope, latitude, aspect, precipitation and temperature of the site for monthly calculations. AET as the evaporative water loss from a site covered by a hypothetical standard crop, given the prevailing water availability (Stephenson, 1998).

CWD reflects drought conditions more accurately than total annual rainfall (Chave et al., 2014), and was consider anthropogenic variable. More negative CWD indicates high water stress conditions and values close to 0 (zero) indicates not water stressed (Poorter et al., 2017). We calculated the average of total annual precipitation and CWD for one, two, three and four years before the measurement year.

2.4.Soil and topography variables

In predicted climate change scenarios of increased frequency of extreme storms, soil and topography may become more useful for improving estimates of tree recruitment and biomass losses over large areas (de Toledo et al., 2012). Thus, edaphic and topographic information was collected.

We collect soil samples for each plot. We obtained information for the depths of 0-20 cm and 20-40 cm. 20-30 samples were collected per plot to obtain a composite sample. Soil pH in H_2O, exchangeable cations (Ca^{2+}, Mg^{2+} and Al^{3+}), total acidity (H^{+}+ Al^{3+}), cation exchange capacity (CEC), base saturation (V), available phosphorus (P), P remaining in solution (P-res) and soil organic matter (SOM) were determined using standard methods (Teixeira et al., 2017). The soil analyses were performed at the laboratory of Soil Fertility at the *Universidade Federal de Viçosa* (UFV), Brazil.

To account for differences in topography, we calculated for each plot the elevation, slope and aspect using Spatial Analyst Tools of surface in ArcGIS 10.3.1 1 (ESRI, 2015). We used the Shuttle Radar Topography Mission (SRTM) and Digital Elevation Models (DEMs) for the analysis.

2.5.Data analyses

2.5.1. Variables selection

We performed tests to determine the variables to be included in the statistical models to assess the effect of anthropogenic and environmental variables (Table S1) on relative recruitment.

Quantitative variables were standardized to accelerate the convergence rate and reduce the iteration process in training (Equation 3). The scale function of R Software was used in this step.

$$Z_i = (x_i - \bar{x})/\sigma \tag{3}$$

where:

Z_i = standardized value of the i-th observation;

x_i = value of the i-th observation;

$\bar{x}$ = average of the observed values;
σ = standard deviation.

First, we exclude highly correlated variables using a correlation coefficient limit of ± 0.9 (Leite et al., 2020; Silva et al., 2016). Subsequently, a method based on Recursive Feature Elimination (RFE) (Gomes et al., 2019), was used, considering only the variables not excluded in the first step to select the best subset of variables. This method is a reverse selection algorithm that calculates the importance of the resource in each iteration, classifying them from most important to least important, removing a user-defined subset at each stage (Kuhn and Johnson, 2013a, 2013b). Although resource collinearity cannot severely affect nonparametric methods, the exclusion of highly correlated methods was important to make RFE iterations more constant, as resources can be interchangeable within models (Leite et al., 2020).

2.5.2. Model selection, evaluation, and inference

The tested models to estimate the recruitment rates were: SVM, ANN and RF. The trained ANN was the multilayer perceptron, also known as the multilayer perceptron (MLP), with a hidden layer. The range of neurons in this layer was defined by the Fletcher-Gloss method (Silva et al., 2010): $2 \times n0.5 + n2 \leq n1 \leq 2 \times n + 1$; where n = number of network inputs; n1 = amount of neurons in the hidden layer; and n2 = number of neurons in the output layer. The activation functions tested were exponential, identity, logistic, and hyperbolic tangent. The training algorithms used were resilient propagation (Rprop) and scaled conjugate gradient (SCG). The initial ANN weights were randomly generated, and the maximum number of iterations was 100 due to the error becoming constant before this number. The ANNs were implemented with the MLP function of the "RSNNS" Package in R (Bergmeir and Benítez, 2012). The function SVM of the "e1071" Package on *R* was used for training SVMs. Thus, four configurations were used in the SVR training, represented by four kernel functions: Linear, Polynomial, Radial basis and Sigmoid.

In the RF training, three essential parameters were configured: the number of random regression trees (ntree, tested 20 to 100 trees); the number of division variables (mtry, used to determine the number of variables available to each node of the tree, with the default number of 1/3 of the independent variables); and the minimum size of nodes (node size, value = 5).

The performance of the models in the estimation was assessed using the *k*–fold cross-validation method, with the data divided into 5 folds (4 for adjustments/training and 1 for validation). At each adjustment/training of the folds the metrics of Root Mean Square Error –

RMSE (Equation 4); Mean Absolute Error – MAE (Equation 5), Pearson correlation coefficient - $r_{y\hat{y}}$ (Equation 6); BIAS (Equation 7) and Relative Bias (%) (Equation 8) were calculated. This process was repeated 50 times, obtaining the average of the metrics for comparison of all models. The data were selected randomly in each of the 50 repetitions, resulting in different data sets, for greater robustness of the evaluation.

$$RMSE = \sqrt{\sum_{r=1}^{R} \frac{\sum_{i=1}^{n}(X_i - \hat{X}_i)^2}{n}} \tag{4}$$

$$MAE = \sum_{r=1}^{R} \frac{\left|\sum_{i=1}^{n}(X_i - \hat{X}_i)\right|}{n} \tag{5}$$

$$r_{y\hat{y}} = \frac{cov\left(X, \hat{X}\right)}{\sqrt{s^2(X) \times s^2\left(\hat{X}\right)}} \tag{6}$$

$$Bias = \frac{\left(\sum_{i}^{n} X_i - \hat{X}_i\right)}{n} \tag{7}$$

$$rBias\ (\%) = \frac{Bias}{\bar{X}} \times 100 \tag{8}$$

Where:

n = number of observations;

X_i = observed variable from the i-th plot;

$\hat{X}_i$ = estimated variable of the i-th plot.

The averages of RMSE, MAE, $r_{y\hat{y}}$ and Bias of each method in each repetition were ranked with weight assignments from 1 to 3, with 1 for the lowest value and 3 for the highest value. With the result of these sums, the values were submitted to the Friedman – Nemenyi test, at the 5% significance level (Equation 9).

The Friedman and Nemenyi nonparametric tests were used to compare ANN, SVR, and RF, based on the cross-validation RMSE, MAE, r and Bias means. The null hypothesis of

Friedman's test is that all algorithms are equivalent. Nemenyi's post hoc test is applied to report significant differences between the techniques if the null hypothesis is rejected. The techniques' performance differs when the mean RMSE by at least one calculated critical difference (CD) differs (Tavares Júnior et al., 2020).

$$CD = q_{\alpha}\sqrt{\frac{k(k+1)}{6N}} \tag{9}$$

Where:

CD = critical difference;

q_{α} = critical value calculated based on Studentized interval statistics divided by √2;

k = number of algorithms being compared;

N = number of data sets.

In total, 280 ANNs, 162 RFs and 8 SVM were trained for all fragments for estimate annual recruitment rate and relative recruitment rate of the Brazilian Atlantic Forest. The methodological flowchart used is presented below (Figure 2).

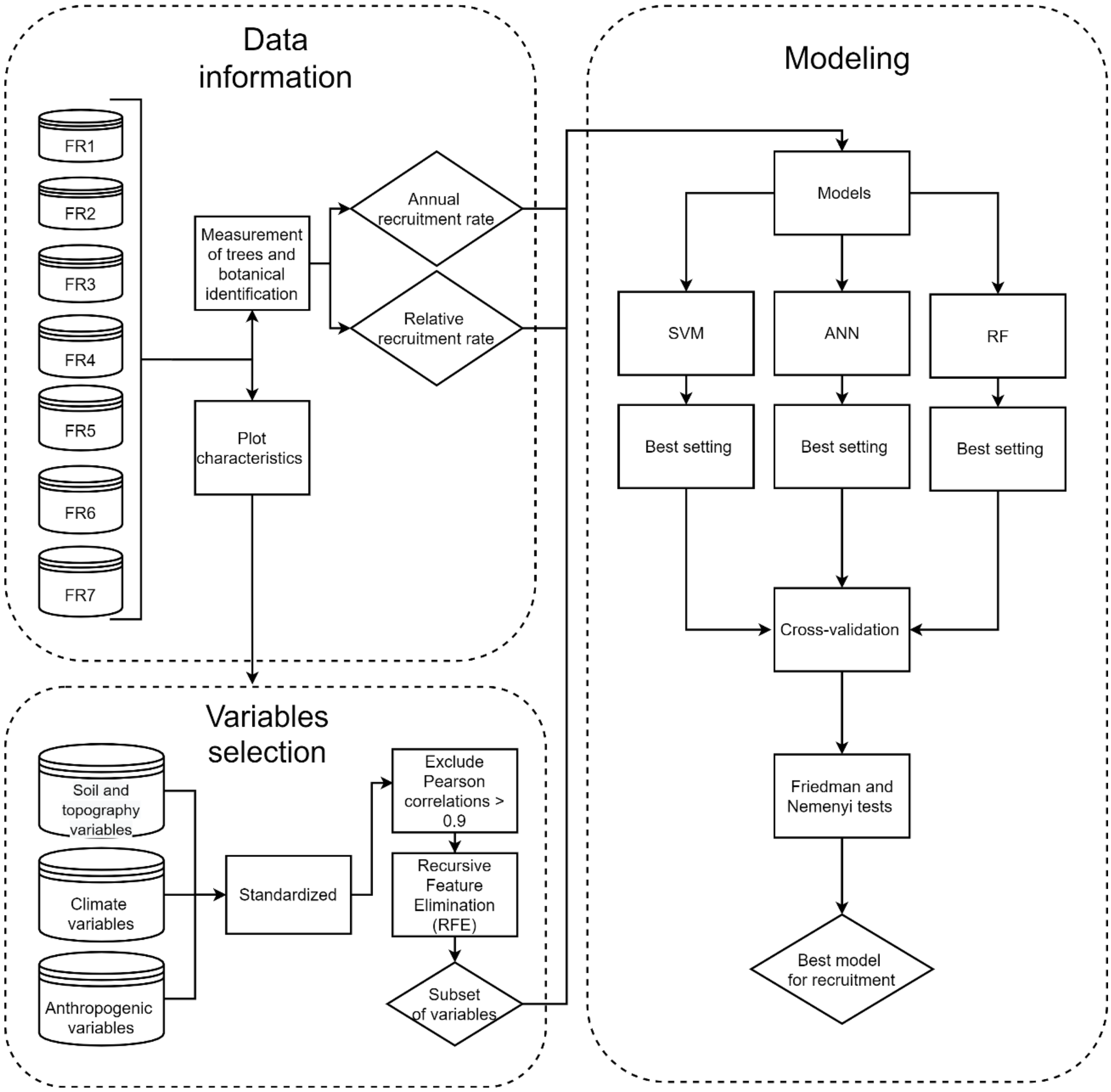

Figure 2. Methodological flowchart for modeling tree recruitment rates in Atlantic Forest in Brazil.

3. Results

We found that CWD, age of abandonment, BA, forest cover, annual and average temperature and precipitation and mean annual precipitation (4 years before measurement) were the most important variables to predict the annual recruitment rate of trees in the Atlantic Forest (Figure 3).

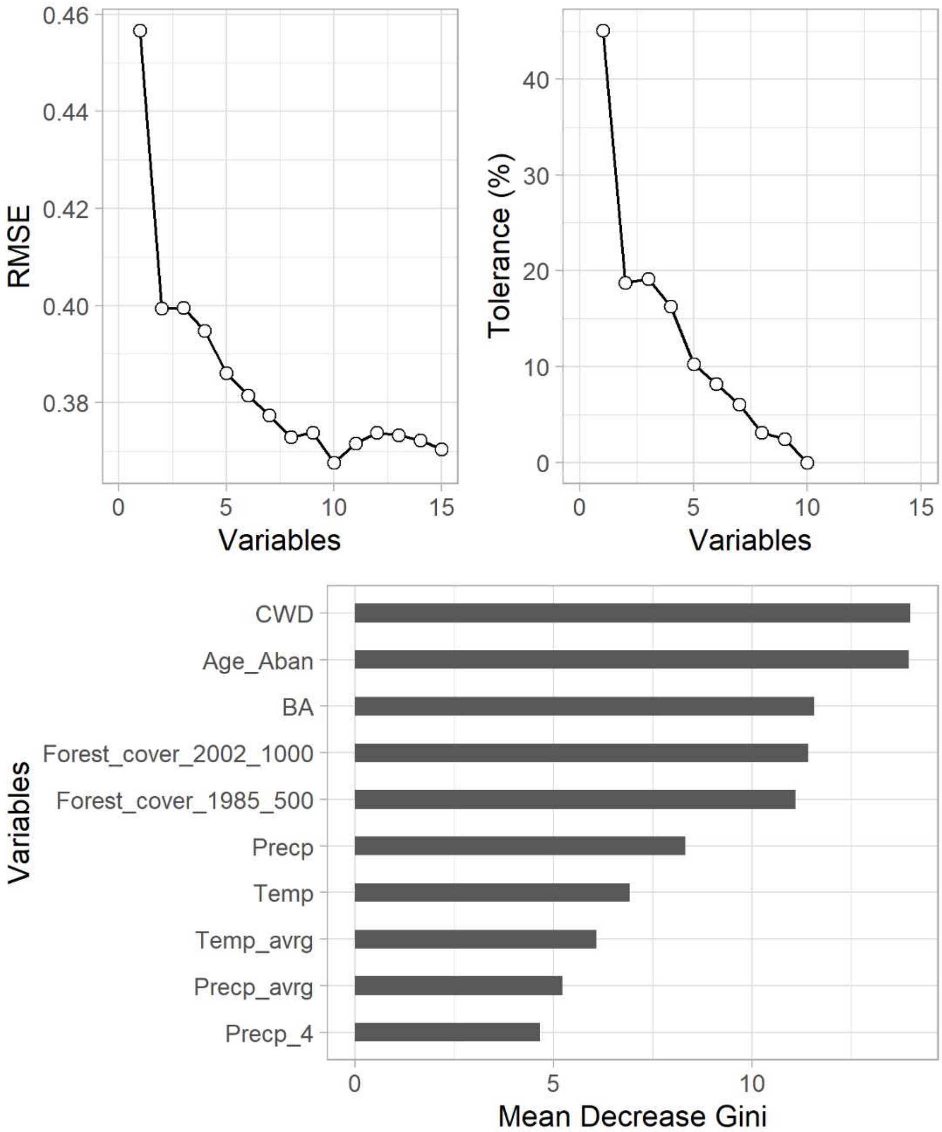

Figure 3. Most important variables for modeling the recruitment rate (R) of studied Atlantic Forest fragments in Minas Gerais, Brazil. CWD: climatic water deficit (from 1989 to 1 year before measurement); Age_Aban: age of Abandonment; BA: basal area; Forest_cover_2002_1000: Forest cover (ha); Year=2002-Buffer=1000m; Forest_cover_1985_500: Forest cover (ha); Year=1985 - Buffer=500m; Precp: Annual precipitation (measurement year); Temp: mean annual temperature 1 year before measurement; Temp_avrg: Average of Mean annual temperature (from 1989 to 1 year before measurement); Precp_avrg: Mean annual precipitation (from 1989 to 1 year before measurement); Precp_4: mean annual precipitation (4 years before measurement).

The variables BA, land use history (eucalyptus and selective logging), forest cover, elevation, age of abandonment, average temperature, annual and average precipitation and mean annual precipitation (1 and 2 years before measurement) most important in predicting the relative recruitment rates in BA (Figure 4).

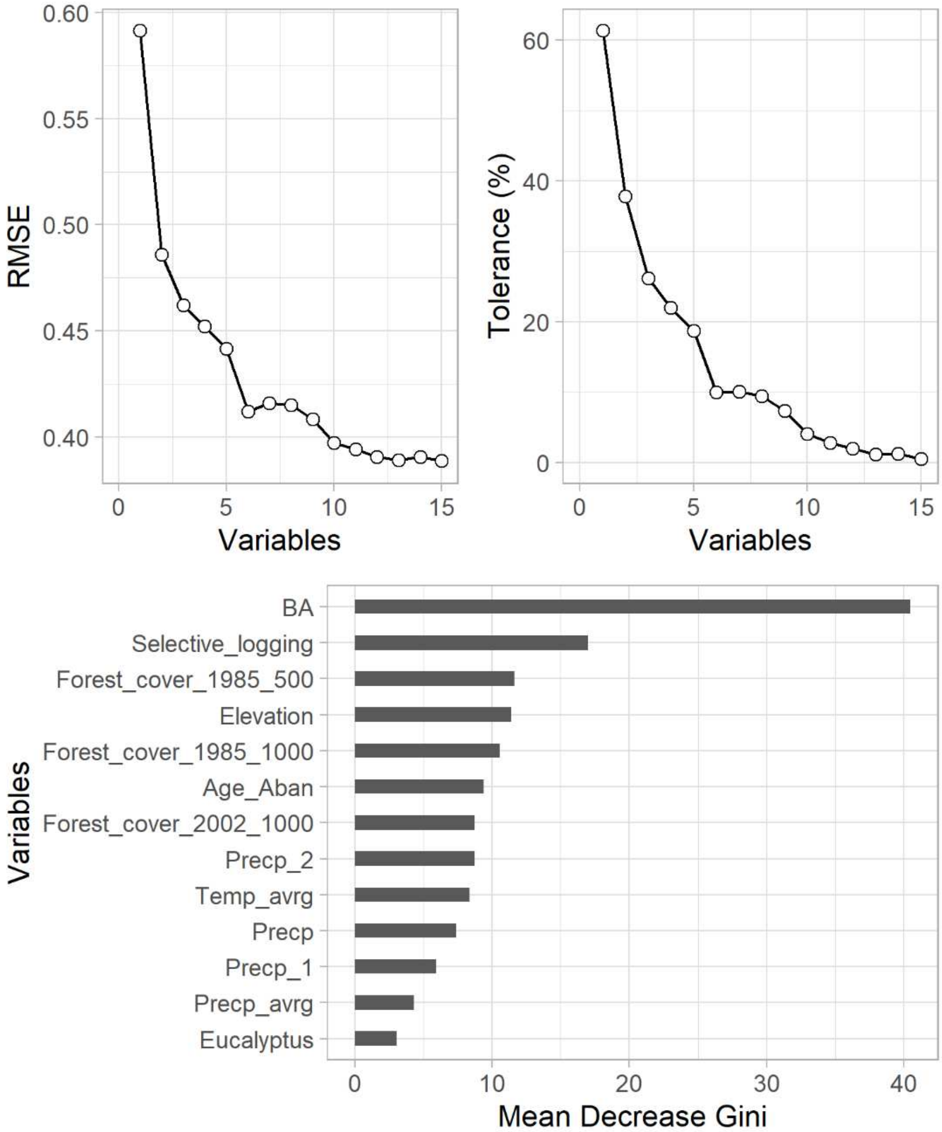

Figure 4. Most important variables for modeling the relative recruitment rate (Rrel) of studied Atlantic Forest fragments in Minas Gerais, Brazil. BA: basal area; Selective_logging: Land Use History - Selective Logging; Forest_cover_1985_500: Forest cover (ha); Year=1985 - Buffer=500m; Elevation: Elevation (m); Forest_cover_1985_1000: Forest cover (ha); Year=1985 - Buffer=1000m; Age_Aban: age of Abandonment; Forest_cover_2002_1000: Forest cover (ha); Year=2002-Buffer=1000m; Precp_2: mean annual precipitation (1 and 2 years before measurement); Temp_avrg: Average of Mean annual temperature (from 1989 to 1 year before measurement); Precp: Annual precipitation (measurement year); Precp_1: Annual precipitation (1 year before measurement); Precp_avrg: Mean annual precipitation (from 1989 to 1 year before measurement); Eucalyptus: Land Use History - Selective Logging

The evaluated models showed a satisfactory generalization power, indicated by similar precision results between the observed and estimated data in the validation for all variables studied (Figure 5 and 6). The trained models showed similar patterns of error distribution, with the largest errors in the plots with the highest rates (Figure 5 and 6).

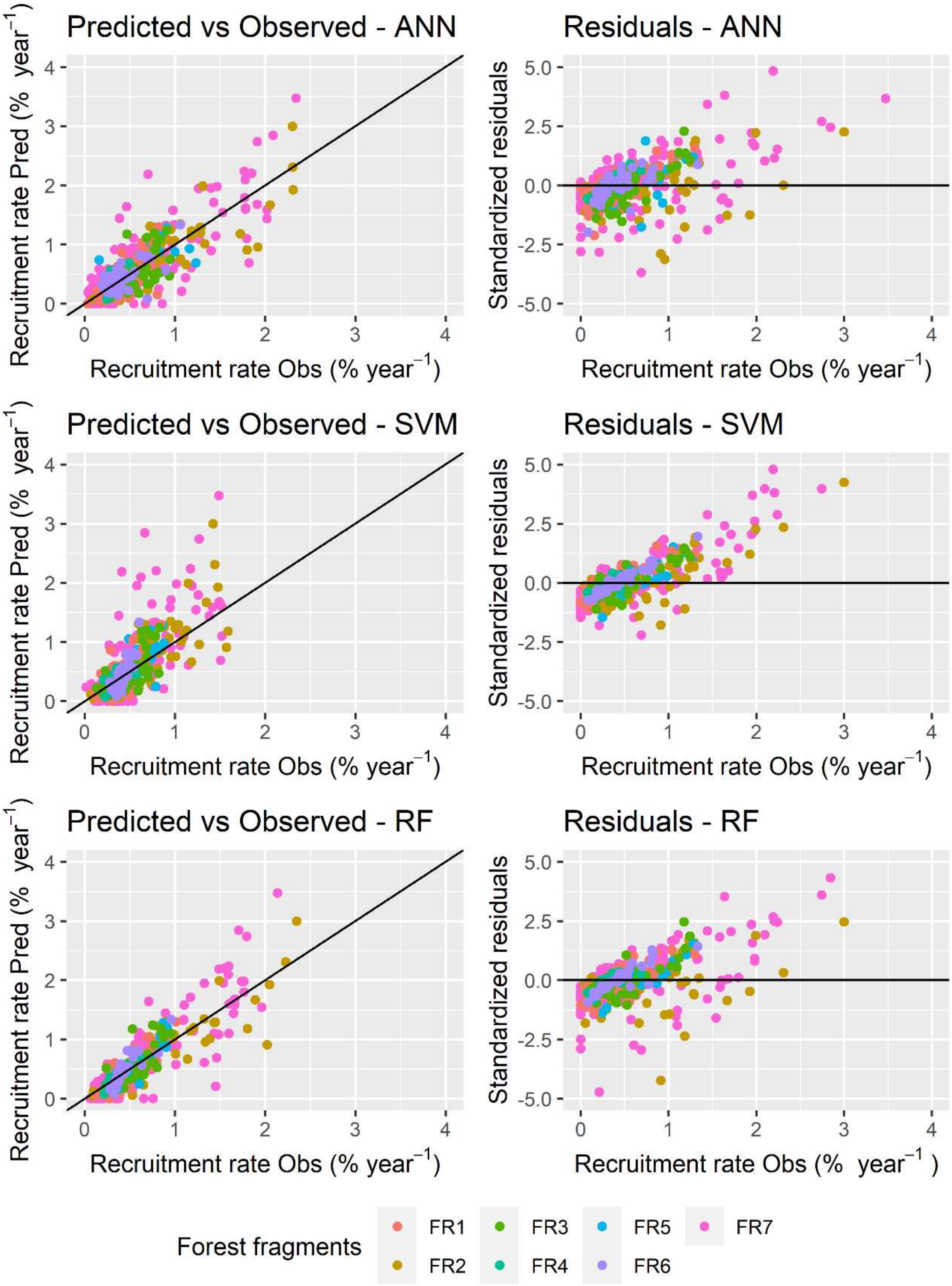

Figure 5. Observed and predicted and residuals values of recruitment rate (R) for the different machine learning models, SVM, ANN and RF tested in Atlantic Forest fragments in Minas Gerais, Brazil. Colors represent the areas. Each small point represents the plots by areas. FR1: Cachoeira das Pombas; FR2: Mata da Garagem; FR3: Ipaba Mata1; FR4: Ipaba Mata2; FR5: Centev; FR6: São José and FR7: Mata da Silvicultura.

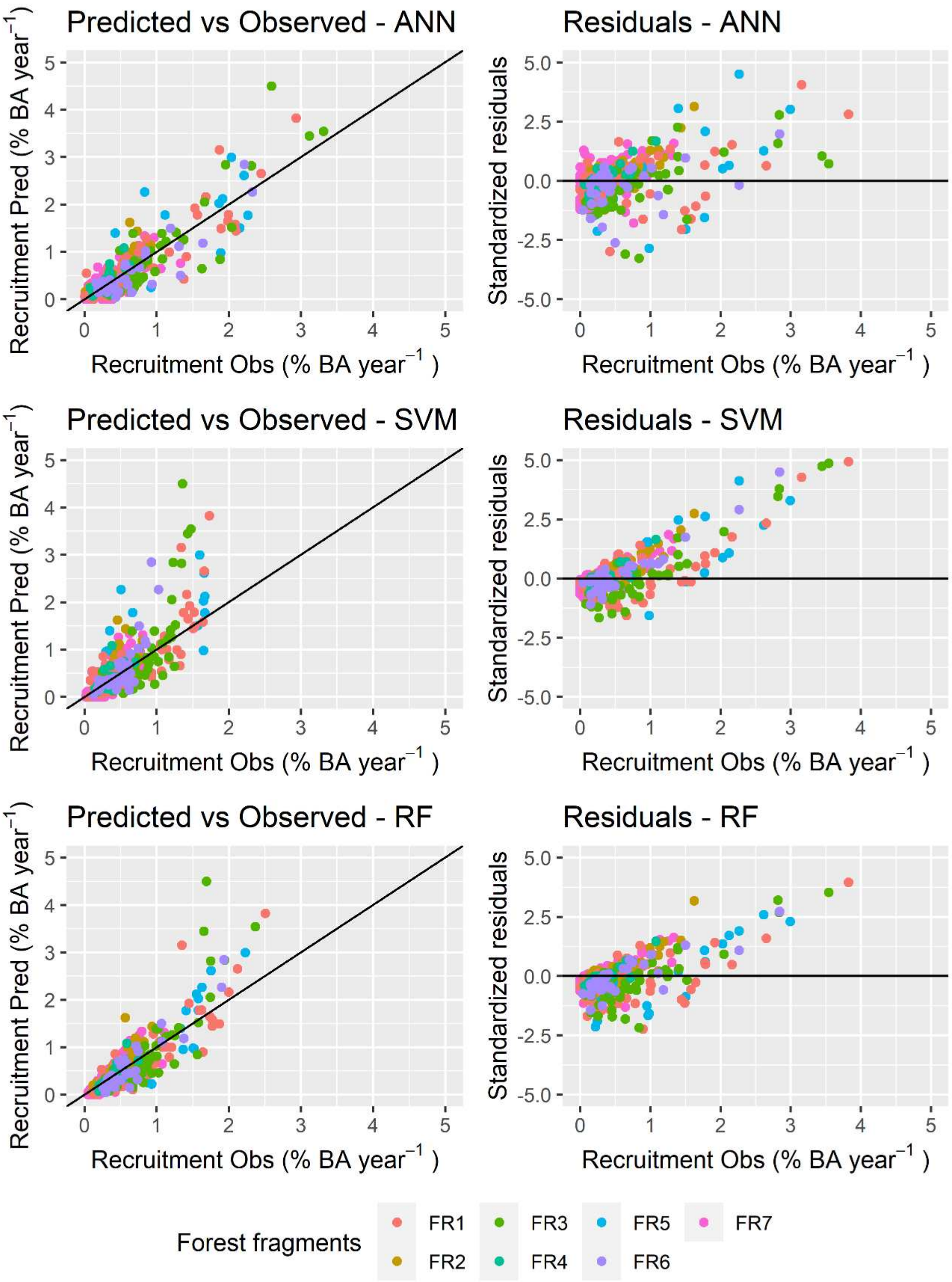

Figure 6. Observed and predicted and residuals values of relative recruitment rate (Rrel) for the different machine learning models, SVM, ANN and RF tested in Atlantic Forest fragments in Minas Gerais, Brazil. Colors represent the areas. Each small point represents the plots by areas. FR1: Cachoeira das Pombas; FR2: Mata da Garagem; FR3: Ipaba Mata1; FR4: Ipaba Mata2; FR5: Centev; FR6: São José and FR7: Mata da Silvicultura.

RF showed the best performance to estimate the recruitment trees, with the highest $r_{y\hat{y}}$ and the lowest RMSE and MAE for all repetitions (Table 3). ANN had the moderate performance to predict R and Rrel. The SVM had the worst performance for predicting the recruitment in Atlantic Forest.

Table 3. Statistics of the machine learning models, SVM, ANN and RF tested in Atlantic Forest fragments in Minas Gerais, Brazil

Type	Input	Output	Technique	Neur./Trees	Algorithm	Function	RMSE	MAE	r	Bias
Train	Climate; Land use history; Forest Attributes	R	RF	62			0.260±0.032	0.178±0.017	0.871±0.044	-0.907±4.796
			ANN	9	Rprop	Tangential	0.306±0.032	0.219±0.021	0.813±0.056	-0.125±5.22
			SVM			Radial	0.371±0.057	0.24±0.026	0.733±0.071	10.013±5.43
Test	Climate; Land use history; Forest Attributes	R	RF	62			0.262±0.008	0.178±0.004	0.872±0.01	-0.683±1.189
			ANN	9	Rprop	Tangential	0.307±0.008	0.219±0.005	0.813±0.056	-0.125±5.22
			SVM			Radial	0.375±0.014	0.24±0.007	0.731±0.017	10.318±1.338
Train	Climate; Landscape; Land use history; Forest Attributes	Rrel	RF	75			0.324±0.078	0.203±0.029	0.882±0.034	-1.345±6.229
			ANN	10	Rprop	Logistic	0.314±0.045	0.221±0.025	0.872±0.037	-1.722±6.238
			SVM			Radial	0.423±0.087	0.245±0.039	0.786±0.046	13.065±6.803
Test	Climate; Landscape; Land use history; Forest Attributes	Rrel	RF	75			0.332±0.021	0.203±0.007	0.876±0.01	-0.86±1.573
			ANN	10	Rprop	Logistic	0.317±0.012	0.221±0.006	0.872±0.037	-1.722±6.238
			SVM			Radial	0.431±0.023	0.245±0.01	0.781±0.011	13.558±1.714

Where: RMSE: Root Mean Square Error; MAE: mean absolute error; SVM: Support Vector Machine; ANN: Artificial Neural Networks; RF: *Random Forest*. Rprop: Resilient backpropagation. Climate: CWD; Precp; Temp; Temp_avrg; Precp_avrg; Precp_4; Precp_2; Landscape: Elevation, Forest Cover; Forest Attributes: BA, Age_Aban.

We observed some stability and little variation in the metrics over the 50 repetitions of the cross-validation The RMSE averages of RF, over the 50 repetitions in the cross-validation, showed the lowest values to estimate the recruitment of trees and the highest values of correlation (Figure 7) for all variables evaluated in the seven fragments of the Atlantic Forest studied.

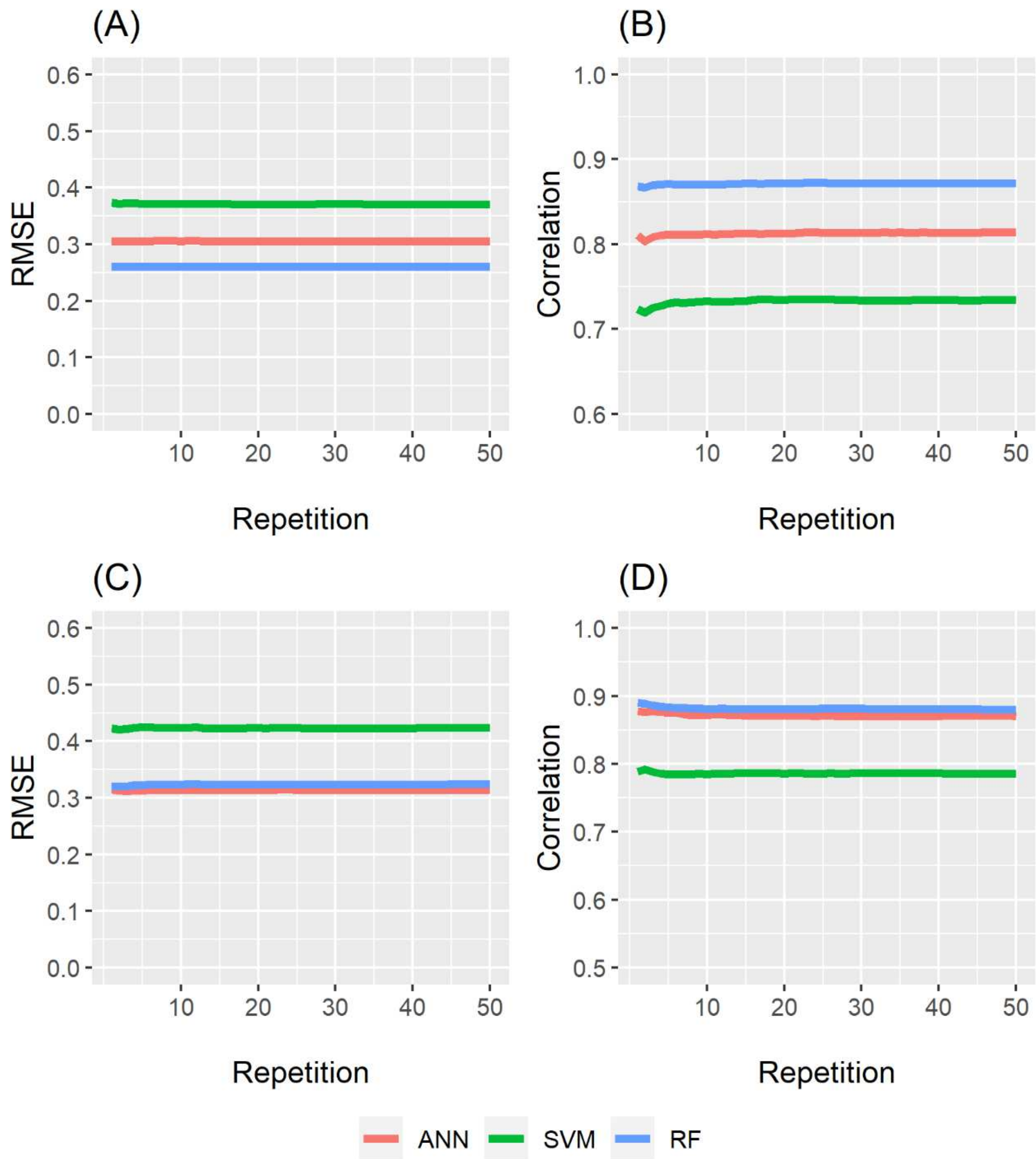

Figure 7. Root Mean Square Error (RMSE) and correlation of the machine learning models SVM, ANN and RF, in the modeling of the recruitment rates (A and B) and relative recruitment rates (C and D) in the in Atlantic Forest fragments in Minas Gerais, Brazil.

The Friedman test with the means of cross-validation RMSE showed that the predictions of recruitment rates in the Atlantic Forest differed between the techniques (p <0.05). Thus, the hypothesis that at least one average of one of the techniques differs from the others was accepted. The Nemenyi test pointed out that the difference between the RF model and the other techniques was greater than the calculated critical difference (CD). The calculated critical difference (CD) of the ANN and SVM it was not significant to estimate recruitment rates in Atlantic Forest (Figure 8).

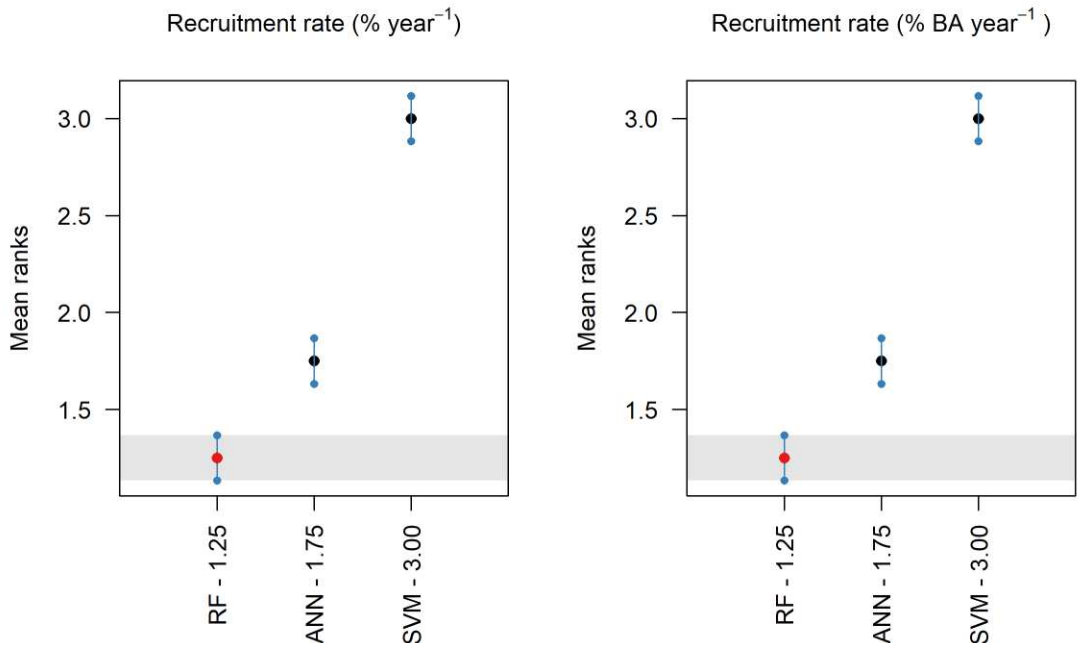

Figure 8. Nemenyi test of the machine learning models SVM, ANN and RF, in the modeling of the recruitment rates in the in Atlantic Forest fragments in Minas Gerais, Brazil.

4. Discussion

Modeling the tree recruitment is an important component for predictions of forest dynamics. However, sometimes, this component is neglected, due to the difficulty of forecasting and data availability. In our study, we found that the use of climate and anthropogenic variables, coupled with machine learning models (SVM, ANN and RF) can represent a promising approach to perform this task.

We identify which climate variables (CWD, precipitation, temperature) are important for predicting annual recruitment rate (% year $^{-1}$) and relative recruitment rate (% BA year^{-1}) in Atlantic Forest fragments. Climatic factors are essential for predicting recruitment, as they regulate forest maintenance (Badano et al., 2015; Massad and Castigo, 2016). The establishment of successional tree species generally depends on specific light, temperature, and moisture conditions that occur in the forest understory (Badano et al., 2015; Benavides et al., 2016). Recruitment biomass growth increases with soil water availability and light availability and decreases under dry conditions (Esquivel-Muelbert et al., 2020; Sande et al., 2017). Any change in these environmental variables can prevent the recruitment of new individuals (Benavides et al., 2016). In addition, the demographic responses of these events can have varying timings (Aleixo et al., 2019), *e.g.:* drought-induced tree death in the Atlantic Forest can be detected 4 years after the drought event (Rocha et al., 2020). Therefore, it is important to consider temperature and precipitation averages from years before the measurement.

Anthropogenic variables (land use history and forest cover), age of Abandonment and BA are also important predictors of recruitment. The dynamics of the highly degraded Brazilian Atlantic Forest is mostly driven by its anthropogenic context (Souza et al., 2021). While climatic conditions generate variation among regions, land use history plays a central role in driving alternative successional pathways in human-modified landscapes, (Jakovac et al., 2021), especially in secondary Atlantic Forest forests (Sansevero et al., 2017), such as those studied here.

Disturbances alter the course of forest dynamics and ecosystem services in the Atlantic Forest (Souza et al., 2021). Small variations in land use history, like selective logging of trees, can interfere with biotic interactions (Arroyo-Rodríguez et al., 2017), affecting regeneration rates, vegetation structure, and species composition (Jakovac et al., 2021). Human interference in forest fragments, such as selective logging, promotes the opening of clearings and increase in BA per plot, resulting in competition for resources and this that can inhibit regeneration (Liebsch et al., 2021). In addition, the forest cover around the plots may be a source of seed trees. Distance to seed source habitat alters tree recruitment patterns (Muñiz-Castro et al., 2006; Toledo-Aceves et al., 2021), with compromised seed dispersal and greater seed predation in fragments, altered recruitment patterns in fragments are expected.

We also observed an influence of landscape (altitude) on tree recruitment patterns in the Atlantic Forest. Trees need to have adequate soil and moisture conditions to regenerate (Clement et al., 2019). These conditions can vary spatially with latitude and altitude (Rogers and Mittanck, 2014). Topography restricts the local nutrients and hydraulic conditions within which trees grow (Jucker et al., 2018). Therefore, altimetric variables can affect resource availability and consequently tree recruitment.

Using these variables, we found that machine learning methods were efficient and are important tools for modeling growth in forest fragments in the Brazilian Atlantic Forest. They can help in understanding the biome and in developing management strategies aimed at recovering biodiversity and reducing the deleterious effects of fragmentation. The *Random Forest* method showed superiority over the others for modeling growth in the Atlantic Forest.

The observed metrics and graphs and residuals corroborated this statement. This method produces the most accurate and stable predictions (Sun et al., 2019), being increasingly used in ecological studies because it is suitable for the analysis of large complex data sets (Reise et al., 2019). As a non-parametric method, it benefits from its ability to take into account data variability and non-linear relationships Alternatively, parametric models are simpler and more widely known, and easier to share and explain (Leite et al., 2020).

Finally, we observed that the use of machine learning (ANN, SVM, and RF) can be a promising way to accurately indicate tree recruitment in tropical forests, especially in the Brazilian Atlantic Forest. In addition, the use of climate, landscape, and land use history variables should be taken into consideration in models to predict this component of forest dynamics. Land use history defines successional pathways through impacts on various processes that define the availability of species for succession (Jakovac et al., 2021). Thus, studies like this should be encouraged and can help in better understanding tree recruitment and assist in conservation practices in forests around the world, especially those threatened by human pressure and fragmentation.

5. Conclusion

We found out that climate, landscape, land use history and forest attributes are important variables to predict tree recruitment in the Atlantic Forest in Brazil. Machine Leaning are efficient methods to estimate recruitment rates at plot-level with these variables. *Random Forest* is more efficient in estimating. Our findings support a new approach for modeling tree recruitment in tropical forests around the world, especially in forest fragments of the Brazilian Atlantic Forest. This approach may represent improvement of future estimates in forest dynamics models.

6. References

Aleixo, I., Norris, D., Hemerik, L., Barbosa, A., Prata, E., Costa, F., Poorter, L., 2019. Amazonian rainforest tree mortality driven by climate and functional traits. Nature Climate Change 9, 384–388. https://doi.org/10.1038/s41558-019-0458-0

Arroyo-Rodríguez, V., Melo, F.P.L., Martínez-Ramos, M., Bongers, F., Chazdon, R.L., Meave, J.A., Norden, N., Santos, B.A., Leal, I.R., Tabarelli, M., 2017. Multiple successional pathways in human-modified tropical landscapes: new insights from forest succession, forest fragmentation and landscape ecology research. Biological Reviews 92, 326–340. https://doi.org/10.1111/brv.12231

Badano, E.I., Samour-Nieva, O.R., Flores, J., Douterlungne, D., 2015. Microclimate and seeding predation as drivers of tree recruitment in human-disturbed oak forests. Forest Ecology and Management, Special Section: The characteristics, impacts and management of forest fire in China 356, 93–100. https://doi.org/10.1016/j.foreco.2015.07.031

Bellard, C., Leclerc, C., Leroy, B., Bakkenes, M., Veloz, S., Thuiller, W., Courchamp, F., 2014. Vulnerability of biodiversity hotspots to global change. Global Ecology and Biogeography 23, 1376–1386. https://doi.org/10.1111/geb.12228

Benavides, R., Escudero, A., Coll, L., Ferrandis, P., Ogaya, R., Gouriveau, F., Peñuelas, J., Valladares, F., 2016. Recruitment patterns of four tree species along elevation gradients in Mediterranean mountains: Not only climate matters. Forest Ecology and Management, Special Section: Forest Management for Climate Change 360, 287–296. https://doi.org/10.1016/j.foreco.2015.10.043

Bergmeir, C., Benítez, J.M., 2012. Neural Networks in R Using the Stuttgart Neural Network Simulator: RSNNS. Journal of Statistical Software 46, 1–26. https://doi.org/10.18637/jss.v046.i07

Breugel, M. van, Craven, D., Lai, H.R., Baillon, M., Turner, B.L., Hall, J.S., 2019. Soil nutrients and dispersal limitation shape compositional variation in secondary tropical forests across multiple scales. Journal of Ecology 107, 566–581. https://doi.org/10.1111/1365-2745.13126

Carreño-Rocabado, G., Peña-Claros, M., Bongers, F., Alarcón, A., Licona, J.-C., Poorter, L., 2012. Effects of disturbance intensity on species and functional diversity in a tropical forest. Journal of Ecology 100, 1453–1463. https://doi.org/10.1111/j.1365-2745.2012.02015.x

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M.S., Delitti, W.B.C., Duque, A., Eid, T., Fearnside, P.M., Goodman, R.C., Henry, M., Martínez-Yrízar, A., Mugasha, W.A., Muller-Landau, H.C., Mencuccini, M., Nelson, B.W., Ngomanda, A., Nogueira, E.M., Ortiz-Malavassi, E., Pélissier, R., Ploton, P., Ryan, C.M., Saldarriaga, J.G., Vieilledent, G., 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. Global Change Biology 20, 3177–3190. https://doi.org/10.1111/gcb.12629

Clement, M.J., Harding, L.E., Lucas, R.W., Rubin, E.S., 2019. The relative importance of biotic and abiotic factors influencing aspen recruitment in Arizona. Forest Ecology and Management 441, 32–41. https://doi.org/10.1016/j.foreco.2019.03.026

de Avila, A.L., Schwartz, G., Ruschel, A.R., Lopes, J. do C., Silva, J.N.M., Carvalho, J.O.P. de, Dormann, C.F., Mazzei, L., Soares, M.H.M., Bauhus, J., 2017. Recruitment, growth and recovery of commercial tree species over 30years following logging and thinning in a tropical rain forest. Forest Ecology and Management 385, 225–235. https://doi.org/10.1016/j.foreco.2016.11.039

de Lima, R.A.F., Oliveira, A.A., Pitta, G.R., de Gasper, A.L., Vibrans, A.C., Chave, J., ter Steege, H., Prado, P.I., 2020. The erosion of biodiversity and biomass in the Atlantic Forest biodiversity hotspot. Nature Communications 11, 6347. https://doi.org/10.1038/s41467-020-20217-w

de Toledo, J.J., Magnusson, W.E., Castilho, C.V., Nascimento, H.E.M., 2012. Tree mode of death in Central Amazonia: Effects of soil and topography on tree mortality associated with storm disturbances. Forest Ecology and Management 263, 253–261. https://doi.org/10.1016/j.foreco.2011.09.017

Esquivel-Muelbert, A., Phillips, O.L., Brienen, R.J.W., Fauset, S., Sullivan, M.J.P., Baker, T.R., Chao, K.-J., Feldpausch, T.R., Gloor, E., Higuchi, N., Houwing-Duistermaat, J., Lloyd, J., Liu, H., Malhi, Y., Marimon, B., Marimon Junior, B.H., Monteagudo-Mendoza, A., Poorter, L., Silveira, M., Torre, E.V., Dávila, E.A., del Aguila Pasquel, J., Almeida, E., Loayza, P.A., Andrade, A., Aragão, L.E.O.C., Araujo-Murakami, A., Arets, E., Arroyo, L., C, G.A.A., Baisie, M., Baraloto, C., Camargo, P.B., Barroso, J., Blanc, L., Bonal, D., Bongers, F., Boot, R., Brown, F., Burban, B., Camargo, J.L., Castro, W., Moscoso, V.C., Chave, J., Comiskey, J., Valverde, F.C., da Costa, A.L., Cardozo, N.D., Di Fiore, A., Dourdain, A., Erwin, T., Llampazo, G.F., Vieira, I.C.G., Herrera, R., Honorio Coronado, E., Huamantupa-Chuquimaco, I., Jimenez-Rojas, E., Killeen, T., Laurance, S., Laurance, W., Levesley, A., Lewis, S.L., Ladvocat, K.L.L.M., Lopez-Gonzalez, G., Lovejoy, T., Meir, P., Mendoza, C., Morandi, P., Neill, D., Nogueira Lima, A.J., Vargas, P.N., de Oliveira, E.A., Camacho, N.P., Pardo, G., Peacock, J., Peña-Claros, M., Peñuela-Mora, M.C., Pickavance, G., Pipoly, J., Pitman, N., Prieto, A., Pugh, T.A.M., Quesada, C., Ramirez-Angulo, H., de Almeida Reis, S.M., Rejou-Machain, M., Correa, Z.R., Bayona, L.R., Rudas, A., Salomão, R., Serrano, J., Espejo, J.S., Silva, N., Singh, J., Stahl, C., Stropp, J., Swamy, V., Talbot, J., ter Steege, H., Terborgh, J., Thomas, R., Toledo, M., Torres-Lezama, A., Gamarra, L.V., van der

Heijden, G., van der Meer, P., van der Hout, P., Martinez, R.V., Vieira, S.A., Cayo, J.V., Vos, V., Zagt, R., Zuidema, P., Galbraith, D., 2020. Tree mode of death and mortality risk factors across Amazon forests. Nature Communications 11, 5515. https://doi.org/10.1038/s41467-020-18996-3

ESRI, 2015. ArcGIS Desktop: Release 10.3.1. Environmental Systems Research Institute, Redlands, CA.

Gomes, L.C., Faria, R.M., de Souza, E., Veloso, G.V., Schaefer, C.E.G.R., Filho, E.I.F., 2019. Modelling and mapping soil organic carbon stocks in Brazil. Geoderma 340, 337–350. https://doi.org/10.1016/j.geoderma.2019.01.007

IBGE, 2012. Manual Técnico da Vegetação Brasileira, 2ª. ed. IBGE, Rio de Janeiro.

Jakovac, C.C., Junqueira, A.B., Crouzeilles, R., Peña-Claros, M., Mesquita, R.C.G., Bongers, F., 2021. The role of land-use history in driving successional pathways and its implications for the restoration of tropical forests. Biological Reviews n/a. https://doi.org/10.1111/brv.12694

Jucker, T., Bongalov, B., Burslem, D.F.R.P., Nilus, R., Dalponte, M., Lewis, S.L., Phillips, O.L., Qie, L., Coomes, D.A., 2018. Topography shapes the structure, composition and function of tropical forest landscapes. Ecology Letters 21, 989–1000. https://doi.org/10.1111/ele.12964

Kuhn, M., Johnson, K., 2013a. Applied Predictive Modeling; Springer: New York, NY, USA,.

Kuhn, M., Johnson, K., 2013b. An Introduction to Feature Selection, in: Kuhn, M., Johnson, K. (Eds.), Applied Predictive Modeling. Springer, New York, NY, pp. 487–519. https://doi.org/10.1007/978-1-4614-6849-3_19

Lebrija-Trejos, E., Meave, J.A., Poorter, L., Pérez-García, E.A., Bongers, F., 2010. Pathways, mechanisms and predictability of vegetation change during tropical dry forest succession. Perspectives in Plant Ecology, Evolution and Systematics 12, 267–275. https://doi.org/10.1016/j.ppees.2010.09.002

Leite, R.V., Amaral, C.H. do, Pires, R. de P., Silva, C.A., Soares, C.P.B., Macedo, R.P., Silva, A.A.L. da, Broadbent, E.N., Mohan, M., Leite, H.G., 2020. Estimating Stem Volume in Eucalyptus Plantations Using Airborne LiDAR: A Comparison of Area- and Individual Tree-Based Approaches. Remote Sensing 12, 1513. https://doi.org/10.3390/rs12091513

Liebsch, D., Velazco, S.J.E., Mikich, S.B., Marques, M.C.M., Galvão, F., 2021. Effects of selective logging, fragmentation, and dominance of bamboos on the structure and diversity of Araucaria Forest fragments. Forest Ecology and Management 487, 118961. https://doi.org/10.1016/j.foreco.2021.118961

Lutz, J.A., Wagtendonk, J.W. van, Franklin, J.F., 2010. Climatic water deficit, tree species ranges, and climate change in Yosemite National Park. Journal of Biogeography 37, 936–950. https://doi.org/10.1111/j.1365-2699.2009.02268.x

MapBiomas Project, 2019. Collection V3.1 of Brazilian Land Cover & Use Map Series [WWW Document]. URL http://mapbiomas.org (accessed 8.15.19).

Martínez-Ramos, M., Balvanera, P., Arreola Villa, F., Mora, F., Maass, J.M., Maza-Villalobos Méndez, S., 2018. Effects of long-term inter-annual rainfall variation on the dynamics of regenerative communities during the old-field succession of a neotropical dry forest. Forest Ecology and Management, Resilience of tropical dry forests to extreme disturbance events 426, 91–100. https://doi.org/10.1016/j.foreco.2018.04.048

Massad, T.J., Castigo, T., 2016. Investigating possible effects of climate change on tree recruitment: Responses of abundant species to water stress in Gorongosa National Park. South African Journal of Botany 106, 96–100. https://doi.org/10.1016/j.sajb.2016.06.002

Menezes, T.G.C., Melo, F.P.L., 2019. Assembly patterns of tree seedling communities in a human-dominated Tropical landscape. Austral Ecology 44, 1204–1212. https://doi.org/10.1111/aec.12798

Muñiz-Castro, M.A., Williams-Linera, G., Benayas, J.M.R., 2006. Distance effect from cloud forest fragments on plant community structure in abandoned pastures in Veracruz, Mexico. Journal of Tropical Ecology 22, 431–440. https://doi.org/10.1017/S0266467406003221

Perea, R., López-Sánchez, A., Pallarés, J., Gordaliza, G.G., González-Doncel, I., Gil, L., Rodríguez-Calcerrada, J., 2020. Tree recruitment in a drought- and herbivory-stressed oak-beech forest: Implications for future species coexistence. Forest Ecology and Management 477, 118489. https://doi.org/10.1016/j.foreco.2020.118489

Poorter, L., Sande, M.T. van der, Arets, E.J.M.M., Ascarrunz, N., Enquist, B.J., Finegan, B., Licona, J.C., Martínez-Ramos, M., Mazzei, L., Meave, J.A., Muñoz, R., Nytch, C.J., Oliveira, A.A. de, Pérez-García, E.A., Prado-Junior, J., Rodríguez-Velázques, J., Ruschel, A.R., Salgado-Negret, B., Schiavini, I., Swenson, N.G., Tenorio, E.A., Thompson, J., Toledo, M., Uriarte, M., Hout, P. van der, Zimmerman, J.K., Peña-Claros, M., 2017. Biodiversity and climate determine the functioning of Neotropical forests. Global Ecology and Biogeography 26, 1423–1434. https://doi.org/10.1111/geb.12668

Redmond, M.D., 2019. CWD and AET function.

Reis, L.P., de Souza, A.L., dos Reis, P.C.M., Mazzei, L., Soares, C.P.B., Miquelino Eleto Torres, C.M., da Silva, L.F., Ruschel, A.R., Rêgo, L.J.S., Leite, H.G., 2018. Estimation of mortality and survival of individual trees after harvesting wood using artificial neural networks in the amazon rain forest. Ecological Engineering 112, 140–147. https://doi.org/10.1016/j.ecoleng.2017.12.014

Reis, L.P., de Souza, A.L., Mazzei, L., dos Reis, P.C.M., Leite, H.G., Soares, C.P.B., Torres, C.M.M.E., da Silva, L.F., Ruschel, A.R., 2016. Prognosis on the diameter of individual trees on the eastern region of the amazon using artificial neural networks. Forest Ecology and Management 382, 161–167. https://doi.org/10.1016/j.foreco.2016.10.022

Reise, J., Kukulka, F., Flade, M., Winter, S., 2019. Characterising the richness and diversity of forest bird species using National Forest Inventory data in Germany. Forest Ecology and Management 432, 799–811. https://doi.org/10.1016/j.foreco.2018.10.012

Ribeiro, M.C., Metzger, J.P., Martensen, A.C., Ponzoni, F.J., Hirota, M.M., 2009. The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. Biological Conservation, Conservation Issues in the Brazilian Atlantic Forest 142, 1141–1153. https://doi.org/10.1016/j.biocon.2009.02.021

Rocha, S.J.S.S. da, Torres, C.M.M.E., Jacovine, L.A.G., Leite, H.G., Gelcer, E.M., Neves, K.M., Schettini, B.L.S., Villanova, P.H., Silva, L.F. da, Reis, L.P., Zanuncio, J.C., 2018. Artificial neural networks: Modeling tree survival and mortality in the Atlantic Forest biome in Brazil. Science of The Total Environment 645, 655–661. https://doi.org/10.1016/j.scitotenv.2018.07.123

Rocha, S.J.S.S. da, Torres, C.M.M.E., Villanova, P.H., Schettini, B.L.S., Jacovine, L.A.G., Leite, H.G., Gelcer, E.M., Reis, L.P., Neves, K.M., Comini, I.B., Silva, L.F. da, 2020. Drought effects on carbon dynamics of trees in a secondary Atlantic Forest. Forest Ecology and Management 465, 118097. https://doi.org/10.1016/j.foreco.2020.118097

Rogers, P.C., Mittanck, C.M., 2014. Herbivory strains resilience in drought-prone aspen landscapes of the western United States. Journal of Vegetation Science 25, 457–469. https://doi.org/10.1111/jvs.12099

Rozendaal, D.M.A., Chazdon, R.L., 2015. Demographic drivers of tree biomass change during secondary succession in northeastern Costa Rica. Ecological Applications 25, 506–516. https://doi.org/10.1890/14-0054.1

Safar, N.V.H., Magnago, L.F.S., Schaefer, C.E.G.R., 2020. Resilience of lowland Atlantic forests in a highly fragmented landscape: Insights on the temporal scale of landscape restoration. Forest Ecology and Management 470–471, 118183. https://doi.org/10.1016/j.foreco.2020.118183

Sande, M.T. van der, Peña-Claros, M., Ascarrunz, N., Arets, E.J.M.M., Licona, J.C., Toledo, M., Poorter, L., 2017. Abiotic and biotic drivers of biomass change in a Neotropical forest. Journal of Ecology 105, 1223–1234. https://doi.org/10.1111/1365-2745.12756

Sansevero, J.B.B., Prieto, P.V., Sánchez-Tapia, A., Braga, J.M.A., Rodrigues, P.J.F.P., 2017. Past land-use and ecological resilience in a lowland Brazilian Atlantic Forest: implications for passive restoration. New Forests 48, 573–586. https://doi.org/10.1007/s11056-017-9586-4

Scarano, F.R., Ceotto, P., 2015. Brazilian Atlantic forest: impact, vulnerability, and adaptation to climate change. Biodivers Conserv 24, 2319–2331. https://doi.org/10.1007/s10531-015-0972-y

Shiel, D., May, R.M., 1996. Mortality and recruitment rate evaluations in heterogeneous tropical forests. Journal of Ecology 84, 91–100.

Silva, C.A., Klauberg, C., Hudak, A.T., Vierling, L.A., Liesenberg, V., Carvalho, S.P.C. e, Rodriguez, L.C.E., 2016. A principal component approach for predicting the stem volume in Eucalyptus plantations in Brazil using airborne LiDAR data. Forestry: An International Journal of Forest Research 89, 422–433. https://doi.org/10.1093/forestry/cpw016

Silva, I.N., Spatti, D.H., Flauzin, R.A., 2010. Redes neurais artificiais: para engenharia e ciências aplicadas. São Paulo: Artliber, 396 p.

Souza, C.R., Maia, V.A., Aguiar-Campos, N. de, Santos, A.B.M., Rodrigues, A.F., Farrapo, C.L., Gianasi, F.M., Paula, G.G.P. de, Fagundes, N.C.A., Silva, W.B., Santos, R.M., 2021. Long-term ecological trends of small secondary forests of the atlantic forest hotspot: A 30-year study case. Forest Ecology and Management 489, 119043. https://doi.org/10.1016/j.foreco.2021.119043

Stephenson, N., 1998. Actual evapotranspiration and deficit: biologically meaningful correlates of vegetation distribution across spatial scales. Journal of Biogeography 25, 855–870. https://doi.org/10.1046/j.1365-2699.1998.00233.x

Sun, T., Chen, F., Zhong, L., Liu, W., Wang, Y., 2019. GIS-based mineral prospectivity mapping using machine learning methods: A case study from Tongling ore district, eastern China. Ore Geology Reviews 109, 26–49. https://doi.org/10.1016/j.oregeorev.2019.04.003

Tavares Júnior, I. da S., Torres, C.M.M.E., Leite, H.G., Castro, N.L.M. de, Soares, C.P.B., Castro, R.V.O., Farias, A.A., 2020. Machine learning: Modeling increment in diameter of individual trees on Atlantic Forest fragments. Ecological Indicators 117, 106685. https://doi.org/10.1016/j.ecolind.2020.106685

Teixeira, P.C., Donagemma, G.K., Fontana, A., Teixeira, W.G. (Eds.), 2017. Manual de métodos de análise de solo, 3ª. ed. Embrapa, Brasília, Brazil.

Toledo-Aceves, T., Trujillo-Miranda, A.L., López-Barrera, F., 2021. Tree regeneration in active and passive cloud forest restoration: Functional groups and timber species. Forest Ecology and Management 489, 119050. https://doi.org/10.1016/j.foreco.2021.119050

van der Sande, M.T., Peña-Claros, M., Ascarrunz, N., Arets, E.J.M.M., Licona, J.C., Toledo, M., Poorter, L., 2017. Abiotic and biotic drivers of biomass change in a Neotropical forest. Journal of Ecology 105, 1223–1234. https://doi.org/10.1111/1365-2745.12756

Xiang, W., Lei, X., Zhang, X., 2016. Modelling tree recruitment in relation to climate and competition in semi-natural Larix-Picea-Abies forests in northeast China. Forest Ecology and Management 382, 100–109. https://doi.org/10.1016/j.foreco.2016.09.050

Zhang, X., Lei, Y., Cai, D., Liu, F., 2012. Predicting tree recruitment with negative binomial mixture models. Forest Ecology and Management 270, 209–215. https://doi.org/10.1016/j.foreco.2012.01.028

CAPÍTULO II: Machine learning: Modeling tree mortality rates in Atlantic Forest, Brazil

Abstract
Mortality models are essential to understand the dynamics of tropical forests. But modeling mortality in natural forests is not an easy task. Machine learning methods can help to solve this. In this study, we applied machine learning models, Artificial Neural Networks (ANN), Suport Vector Machine (SVM) and Random Forest (RF), to predict mortality at plot-level in the Atlantic Forest in Brazil. We calculated the mortality rate (M) and relative mortality rate (Mrel), % basal area (BA) year^{-1} at plot level. We used data from seven Atlantic Forest fragments located in Minas Gerais, Brazil. Forest attributes, land use history, landscape, soil and climatic characteristics were used in the modeling. Recursive Feature Elimination was used to select the best subset of predictor variables. We found out that climate (precipitation, climatic water deficit and temperature), anthropogenic variables (age of abandonment) and forest attributes (BA) are important variables to predict tree mortality in the Atlantic Forest in Brazil. RF showed the best performance to estimate the mortality rates, with the highest $r_{y\hat{y}}$ and the lowest RMSE and MAE for all repetitions. ANN had the moderate performance to predict M and Mrel. The SVM had the worst performance for predicting the mortality in Atlantic Forest. The Friedman and Nemenyi nonparametric test confirm that the RF model is a powerful machine learning algorithm for predictions. Our results, they suggest a new alternative for modeling this important component of forest dynamics.
Keywords: forest dynamics, mortality, tropical forest, Brazil.

1. Introduction

Tropical forests deforestation has contributed to the formation of fragmented landscapes, composed mainly by a secondary forests matrix, grassland and agriculture (Santo-Silva et al., 2016; Sousa et al., 2017). These changes in land use have caused many pressures on ecosystems (Diniz et al., 2021; Melo et al., 2013). They affect not only the quantity of native vegetation, but also the spatial configuration and quality of the remaining forest through habitat fragmentation (Diniz et al., 2021; dos Santos et al., 2020; Haddad et al., 2015). This results in isolation of plant and/or animal populations, which increases the risks of inbreeding, genetic drift, and extinction (Dixo et al., 2009; dos Santos et al., 2020).

In Brazil, the Atlantic Forest biome has been suffering from this process since 1500 (Colombo and Joly, 2010). The Atlantic Forest is a global biodiversity hotspot (de Lima et al., 2020; Macedo et al., 2021; Myers et al., 2000; Ribeiro et al., 2009). The biome has one of the highest species richness and endemism rates on the planet. The fragmentation of the biome has meant that a large proportion of its vast biodiversity is threatened with extinction. (Ribeiro et al., 2009).

In this context, knowledge and prediction of tree population dynamics can help mitigate Atlantic Forest degradation and preserve its remaining fragments. Forest dynamics are driven by the balance between four forest demographic changes: growth, recruitment, mortality, and forest composition (Pretzsch, 2009; Vanclay, 1994). These are important variables to consider in landscape planning and management for biodiversity conservation.

Among these variables, tree mortality is a critical ecological phenomenon that shapes the dynamics, structure, and composition of the forest ecosystem, and its effects are of global relevance because of its relationship to forest conditions and environmental change (Salas-Eljatib and Weiskittel, 2020; Synek et al., 2020). In addition, it plays a key role in the carbon storage capacity of forests. The carbon sink capacity of tropical forests is substantially affected by tree mortality (Esquivel-Muelbert et al., 2020).

However, modeling tree mortality is challenging, especially in natural forests (Reis et al., 2018; Rocha et al., 2018; Ruiz-Benito et al., 2013; Vanoni et al., 2019). Mortality of individuals is highly uncertain and difficult to model (Salas-Eljatib and Weiskittel, 2020). Tree mortality can be affected by a variety of environmental, physiological, pathological, and entomological factors, as well as random events that are difficult to predict (Hallinger et al., 2016; Hülsmann et al., 2016), making modeling difficult. Besides that, tree mortality remains poorly evaluated at the stand scale, particularly quantitatively, due to the lack of adequate tree mortality demographic data (Zhu et al., 2019).

Modeling mortality requires choosing an appropriate approach to estimating the model parameters (Tavares Júnior et al., 2020). The two main methods used are regression models and machine learning techniques (Breiman, 2001), most commonly using artificial neural networks (ANN), followed by support vector regression (SVR) and random forests (RF) (Jachowski et al., 2013). The applications of this computational intelligence technique in the forestry area have gained high relevance (Bayat et al., 2019; Hamidi et al., 2021; Reis et al., 2018, 2016; Rocha et al., 2018; Tavares Júnior et al., 2020). These techniques have been used to improve local, regional and global estimates (Vahedi, 2016, Silva et al., 2019).

The objective of the present study is to develop models of machine learning (Artificial Neural Networks - ANN, Support Vector Machine - SVM and Random Forest - RF) capable of estimating the mortality rates in Atlantic Forest fragments, based on forest attributes, land use history, landscape, soil and climatic characteristics. (i) What are the most important variables to estimate mortality in these forests? (ii) Are the machine learning methods evaluated efficient for estimating the mortality rates at plot-level? (iii) What is the best method to estimate the mortality rates at plot-level?

2. Material and methods

2.1. Study sites and plot characteristics

We used data from seven Atlantic forest fragments located in Minas Gerais, Brazil (Table 1 and Figure 1). The vegetation is classified as semideciduous seasonal forest (IBGE, 2012). Across sites the annual rainfall varies from 701 to 1737 mm. year^{-1}, elevation from 242 to 1169 m above sea level, and slope from 2.97 to 65.81% (Table S1).

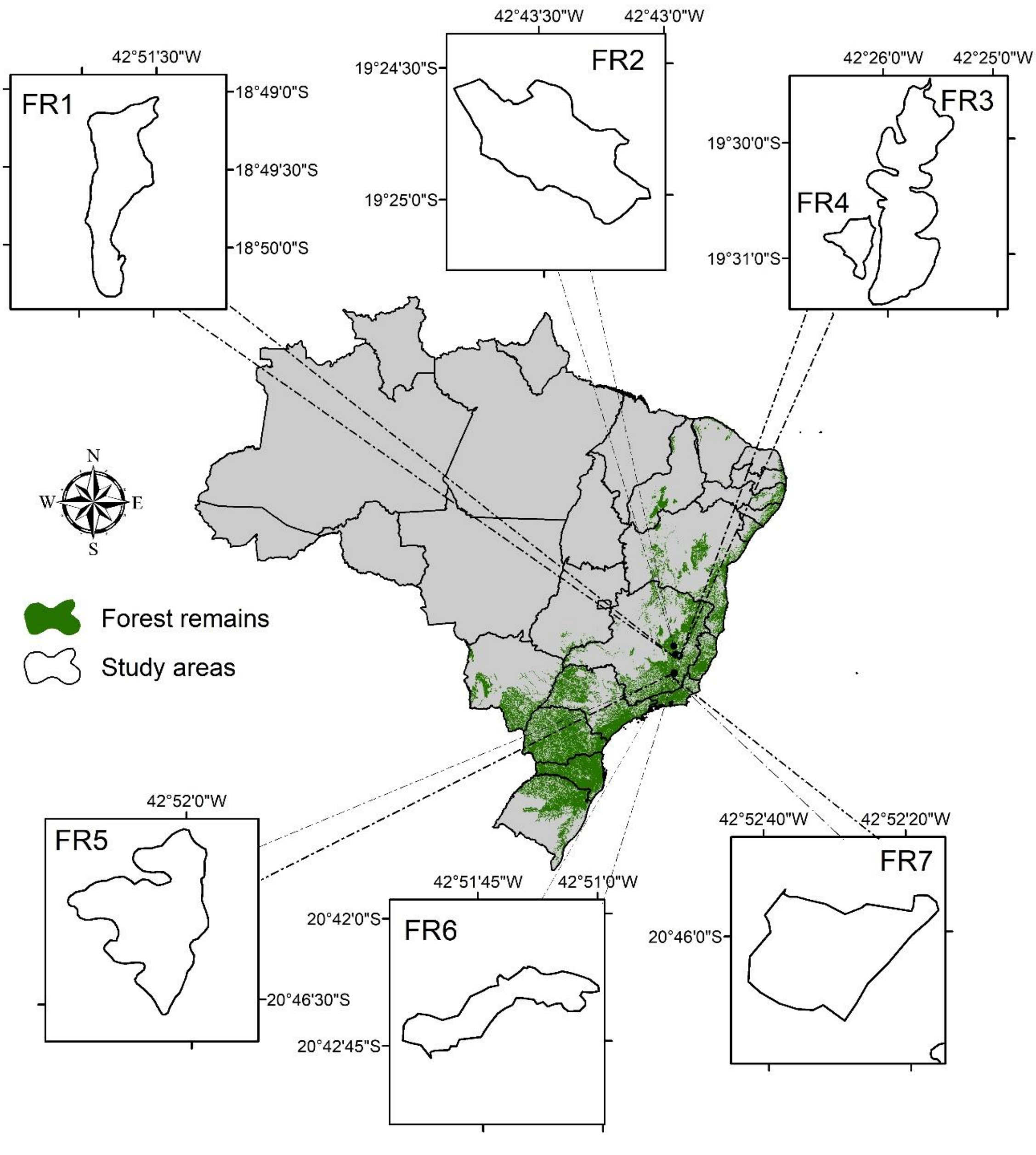

Figure 1. Locations of the seven studied Atlantic Forest fragments (FR) in Minas Gerais, Brazil. FR1: Cachoeira das Pombas; FR2: Mata da Garagem; FR3: Ipaba Mata1; FR4: Ipaba Mata2; FR5: Centev; FR6: São José and FR7: Mata da Silvicultura.

We measured 104 plots located in seven forest fragments over several years (Table 1). In all plots, we measured and identified botanically all stems with diameter at breast height (dbh), 1.3 m ≥ 5 cm.

Table 1. Overview of the seven studied Atlantic Forest fragments (FR) in Minas Gerais, Brazil. Forest fragments location (municipality), size, number and size of plots, and years of forest inventory are also provided

Fragment	Municipality	Forest Size (ha)	Plots	Plots size (ha)	Measurement
FR1	Guanhães	106.0	20	0.05	2002, 2007, 2012, 2017
FR2	Viçosa	21.8	50	0.01	1996, 1998, 2001, 2003, 2013, 2018
FR3	Caratinga	264.0	16	0.05	2002, 2007, 2012, 2017
FR4	Caratinga	37.3	6	0.05	2002, 2007, 2012, 2017
FR5	Viçosa	44.1	20	0.05	2010, 2015
FR6	Coronel Fabriciano	38.4	12	0.05	2002, 2007, 2012, 2017
FR7	Viçosa	17.0	10	0.10	1994, 1997, 2000, 2004, 2008, 2010, 2013, 2016

FR1: Cachoeira das Pombas; FR2: Mata da Garagem; FR3: Ipaba Mata1; FR4: Ipaba Mata2; FR5: Centev; FR6: São José and FR7: Mata da Silvicultura.

We calculated the mortality rate at plot level (van der Sande et al., 2017). The mortality rate (% $year^{-1}$) was calculated as proposed by (Sheil et al., 1995) (Equation 1).

$$M=\left(1-\left((N_0-m)/N_0\right)^{\frac{1}{y}}\right)x100 \quad (1)$$

Where:

M - annual mortality rate, % $year^{-1}$;

N_0 - number of individuals in the initial population;

m - number of dead trees;

y - time, years.

As demographic processes may be influence by the total basal area of the plot (Breugel et al., 2019; Carreño-Rocabado et al., 2012; Menezes and Melo, 2019; Rozendaal and Chazdon, 2015).We calculated the basal area (BA) of each individual tree and measurement year to estimate the rate of mortality at plot level (van der Sande et al., 2017). Mortality is the basal area of the stems alive on census tn, where t indicates the census in time n, but dead on the census tn + 1. Relative mortality rate was calculated (Lebrija-Trejos et al., 2010; Martínez-Ramos et al., 2018) (Equation 2).

$$Mrel = \left(1\text{-}\left((N_0\text{-}m)/N_0\right)^{\frac{1}{y}}\right) x100 \tag{2}$$

Where:

Mrel - Relative mortality rate, % BA year^{-1};

N_0 - Basal area of the stems alive on census in the initial population;

m - Basal area of mortality (m^2);

y - time, years.

2.2. Forest attributes and Anthropogenic variables

We use forest attributes from Atlantic Forest fragments as predictive variables. The attributes of the forest used were: Basal area of the plot and Number of stems. This information was collected through forest inventories carried out over the years.

The Anthropogenic variables used to estimate tree mortality rates were: Land use history, forest cover, forest size and edge distance, age of abandonment (that means, time since abandonment previous anthropogenic uses). Based on information collected from aerial photographs (from around 1960, 1980 and 1990), Landsat satellite images from 1985, landowner interviews and land titles, we determined the Land use history and age of abandonment of the areas. The categories of land use used were: deforestation, agricultural production, eucalyptus plantation, and selective logging.

Forest cover was calculated for each plot using circular buffers with radii of 500, 1000, and 2000 m, for the year 1985, 2002 and 2017, with data from MapsBiomas (MapBiomas Project, 2019) in ArcGIS 10.3.1 (ESRI, 2015). Forest cover area was produced from the pixel-per-pixel (30 x 30 m) classification of Landsat satellite images through the Google Earth Engine platform (MapBiomas Project, 2019). We calculated distance to the nearest edge of the forest (edge distance) with near tool in ArcGIS 10.3.1 1 (ESRI, 2015).

2.3.Climate variables

We obtained annual precipitation, the number of months with less than 100 mm of rainfall, precipitation in the three driest months, and average annual temperature for each fragment from the nearest climatological station (Figure S1). We then estimated the climatic water deficit (CWD) – a water balance between precipitation and evapotranspiration – as a proxy for drought conditions following Lutz et al. (2010) and using the R function CWD and AET (actual evapotranspiration) from Redmond (2019).

We used slope, latitude, aspect, precipitation and temperature of the site for monthly calculations. AET as the evaporative water loss from a site covered by a hypothetical standard crop, given the prevailing water availability (Stephenson, 1998).

CWD reflects drought conditions more accurately than total annual rainfall (Chave et al., 2014), and was consider anthropogenic variable. More negative CWD indicates high water stress conditions and values close to 0 (zero) indicates not water stressed (Poorter et al., 2017). We calculated the average of total annual precipitation and CWD for one, two, three and four years before the measurement year.

2.4.Soil and topography variables

Soil and topography can be useful to improve estimates of tree mortality and biomass losses over large areas (de Toledo et al., 2012). Thus, edaphic and topographic information was collected.

We collect soil samples for each plot. We obtained information for the depths of 0-20 cm and 20-40 cm. 20-30 samples were collected per plot to obtain a composite sample. Soil pH in H_2O, exchangeable cations (Ca^{2+}, Mg^{2+} and Al^{3+}), total acidity (H^{+}+ Al^{3+}), cation exchange capacity (CEC), base saturation (V), available phosphorus (P), P remaining in solution (P-res) and soil organic matter (SOM) were determined using standard methods (Teixeira et al., 2017). The soil analyses were performed at the laboratory of Soil Fertility at the Universidade Federal de Viçosa (UFV), Brazil.

To account for differences in topography, we calculated for each plot the elevation, slope and aspect using Spatial Analyst Tools of surface in ArcGIS 10.3.1 1 (ESRI, 2015). We used the Shuttle Radar Topography Mission (SRTM) and Digital Elevation Models (DEMs) for the analysis.

2.5.Data analyses

2.5.1. Variables selection

We performed tests to determine the variables to be included in the statistical models to assess the effect of anthropogenic and environmental variables (Table S1) on relative mortality.

Quantitative variables were standardized to accelerate the convergence rate and reduce the iteration process in training (Equation 3). The scale function of R Software was used in this step.

$$Z_i = (x_i - \bar{x})/\sigma \quad (3)$$

where:
Z_i = standardized value of the i-th observation;
x_i = value of the i-th observation;
$\bar{x}$ = average of the observed values;
σ = standard deviation.

First, we exclude highly correlated variables using a correlation coefficient limit of ± 0.9 (Leite et al., 2020; Silva et al., 2016). Subsequently, a method based on Recursive Feature Elimination (RFE) (Gomes et al., 2019), was used, considering only the variables not excluded in the first step to select the best subset of variables. This method is a reverse selection algorithm that calculates the importance of the resource in each iteration, classifying them from most important to least important, removing a user-defined subset at each stage (Kuhn and Johnson, 2013a, 2013b). Although resource collinearity cannot severely affect nonparametric methods, the exclusion of highly correlated methods was important to make RFE iterations more constant, as resources can be interchangeable within models (Leite et al., 2020).

2.5.2. Model selection, evaluation, and inference

The tested models to estimate the mortality rates were: Support Vector Machines (SVM), Artificial Neural Networks (ANN) and Random Forests (RF).

The trained ANN was the multilayer perceptron, also known as the multilayer perceptron (MLP), with a hidden layer. The range of neurons in this layer was defined by the Fletcher-Gloss method (Silva et al., 2010): 2 × n0.5 + n2 ≤ n1 ≤ 2 × n + 1; where n = number of network inputs; n1 = amount of neurons in the hidden layer; and n2 = number of neurons in the output layer. The activation functions tested were exponential, identity, logistic, and hyperbolic tangent. The training algorithms used were resilient propagation (Rprop) and scaled conjugate gradient (SCG). The initial ANN weights were randomly generated, and the maximum number of iterations was 100 due to the error becoming constant before this number. The ANNs were implemented with the MLP function of the "RSNNS" Package in R (Bergmeir and Benítez, 2012). The function SVM of the "e1071" Package on *R* was used for training SVMs. Thus, four configurations were used in the SVR training, represented by four kernel functions: Linear, Polynomial, Radial basis and Sigmoid.

In the RF training, three essential parameters were configured: the number of random regression trees (ntree, tested 20 to 100 trees); the number of division variables (mtry, used to determine the number of variables available to each node of the tree, with the default number of 1/3 of the independent variables); and the minimum size of nodes (node size, value = 5).

The performance of the models in the estimation was assessed using the *k*–fold cross-validation method, with the data divided into 5 folds (4 for adjustments/training and 1 for validation). At each adjustment/training of the folds the metrics of Root Mean Square Error – RMSE (Equation 4); Mean Absolute Error – MAE (Equation 5), Pearson correlation coefficient - $r_{y\hat{y}}$ (Equation 6); BIAS (Equation 7) and Relative Bias (%) (Equation 8) were calculated. This process was repeated 50 times, obtaining the average of the metrics for comparison of all models. The data were selected randomly in each of the 50 repetitions, resulting in different data sets, for greater robustness of the evaluation.

$$RMSE = \sqrt{\sum_{r=1}^{R} \frac{\sum_{i=1}^{n}(X_i - \hat{X}_i)^2}{n}} \tag{4}$$

$$MAE = \sum_{r=1}^{R} \frac{\left|\sum_{i=1}^{n}(X_i - \hat{X}_i)\right|}{n} \tag{5}$$

$$r_{y\hat{y}} = \frac{cov(X, \hat{X})}{\sqrt{s^2(X) \times s^2(\hat{X})}} \tag{6}$$

$$Bias = \frac{\left(\sum_i^n X_i - \hat{X}_i\right)}{n} \tag{7}$$

$$rBias\,(\%) = \frac{Bias}{\bar{X}} \times 100 \tag{8}$$

Where:

n = number of observations;

X_i = observed variable from the i-th plot;

$\hat{X}_i$ = estimated variable of the i-th plot.

The averages of RMSE, EAM, $r_{y\hat{y}}$ and Bias of each method in each repetition were ranked with weight assignments from 1 to 3, with 1 for the lowest value and 3 for the highest value. With the result of these sums, the values were submitted to the Friedman – Nemenyi test, at the 5% significance level (Equation 9).

The Friedman and Nemenyi nonparametric tests were used to compare ANN, SVR, and RF, based on the cross-validation RMSE, EAM, r and Bias means. The null hypothesis of Friedman's test is that all algorithms are equivalent. Nemenyi's post hoc test is applied to report significant differences between the techniques if the null hypothesis is rejected. The techniques' performance differs when the mean RMSE by at least one calculated critical difference (CD) differs (Tavares Júnior et al., 2020).

$$CD = q_{\alpha}\sqrt{\frac{k(k+1)}{6N}} \tag{9}$$

Where:

CD = critical difference;

q_{α} = critical value calculated based on Studentized interval statistics divided by √2;

k = number of algorithms being compared;

N = number of data sets.

In total, 120 ANNs, 161 RFs and 8 SVM were trained for all fragments for estimate annual mortality rate and relative mortality rate of the Brazilian Atlantic Forest. The methodological flowchart used is presented below (Figure 2).

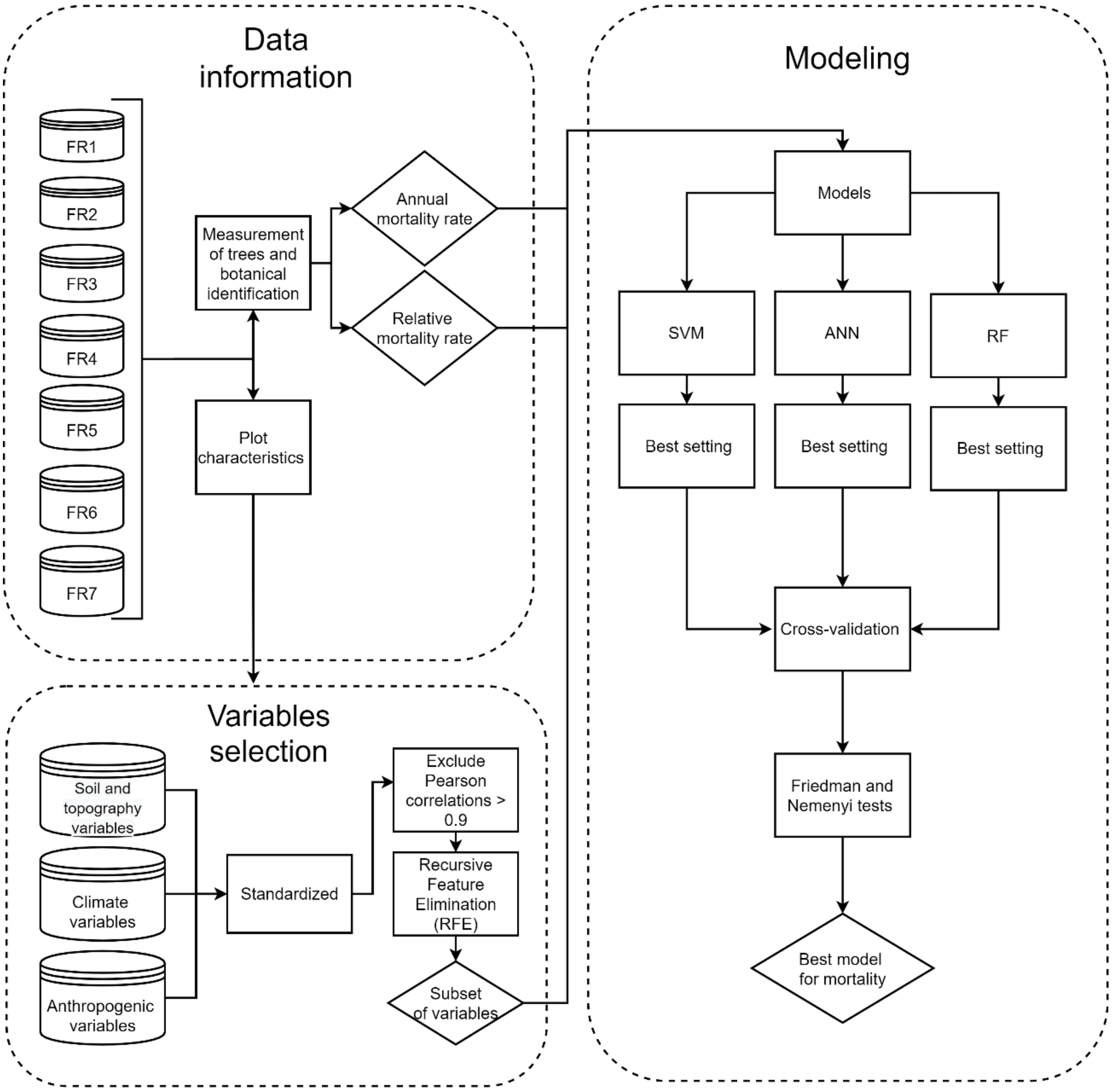

Figure 2. Methodological flowchart for modeling tree mortality rates in Atlantic forest in Brazil.

3. Results

The applied variable selection procedure allowed the choice of the best model based on the ideal subset of variables. We found that BA, age of abandonment, CWD, mean annual precipitation (2 and 4 years before measurement), mean annual temperature 1 year before measurement and total precipitation of the three driest months were the most important variables to predict the annual mortality rate of trees in the Atlantic Forest (Figure 3).

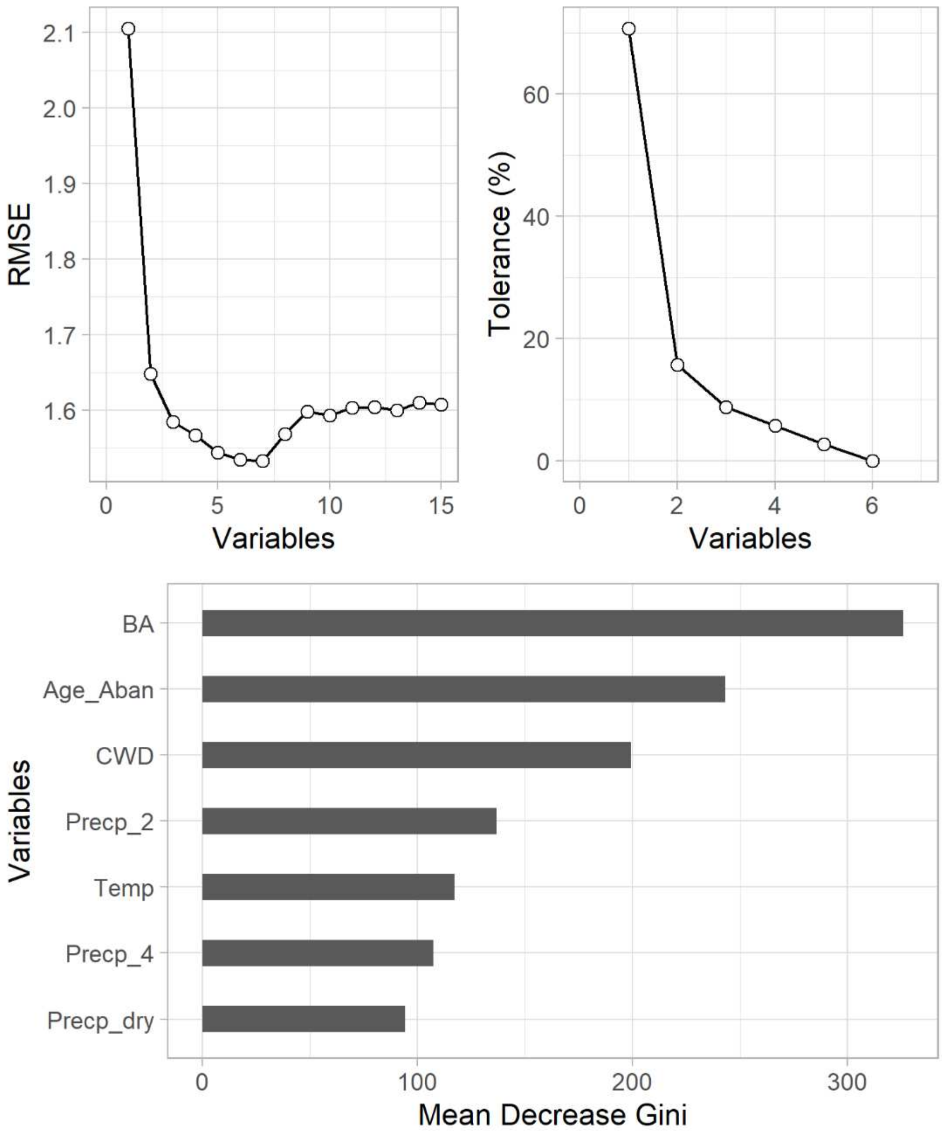

Figure 3. Most important variables for modeling the mortality rate (M) of studied Atlantic Forest fragments in Minas Gerais, Brazil. BA: basal area; Age_Aban: age of Abandonment; CWD: climatic water deficit (from 1989 to 1 year before measurement); Precp_2: mean annual precipitation (1 and 2 years before measurement); Temp: mean annual temperature 1 year before measurement; Precp_4: mean annual precipitation (1, 2, 3 and 4 years before measurement); Precp_dry: total precipitation of the three driest months.

The variables BA - plots, mean annual temperature 1 year before measurement and total precipitation of the three driest months were the most important in predicting the relative mortality rates in BA (Figure 4).

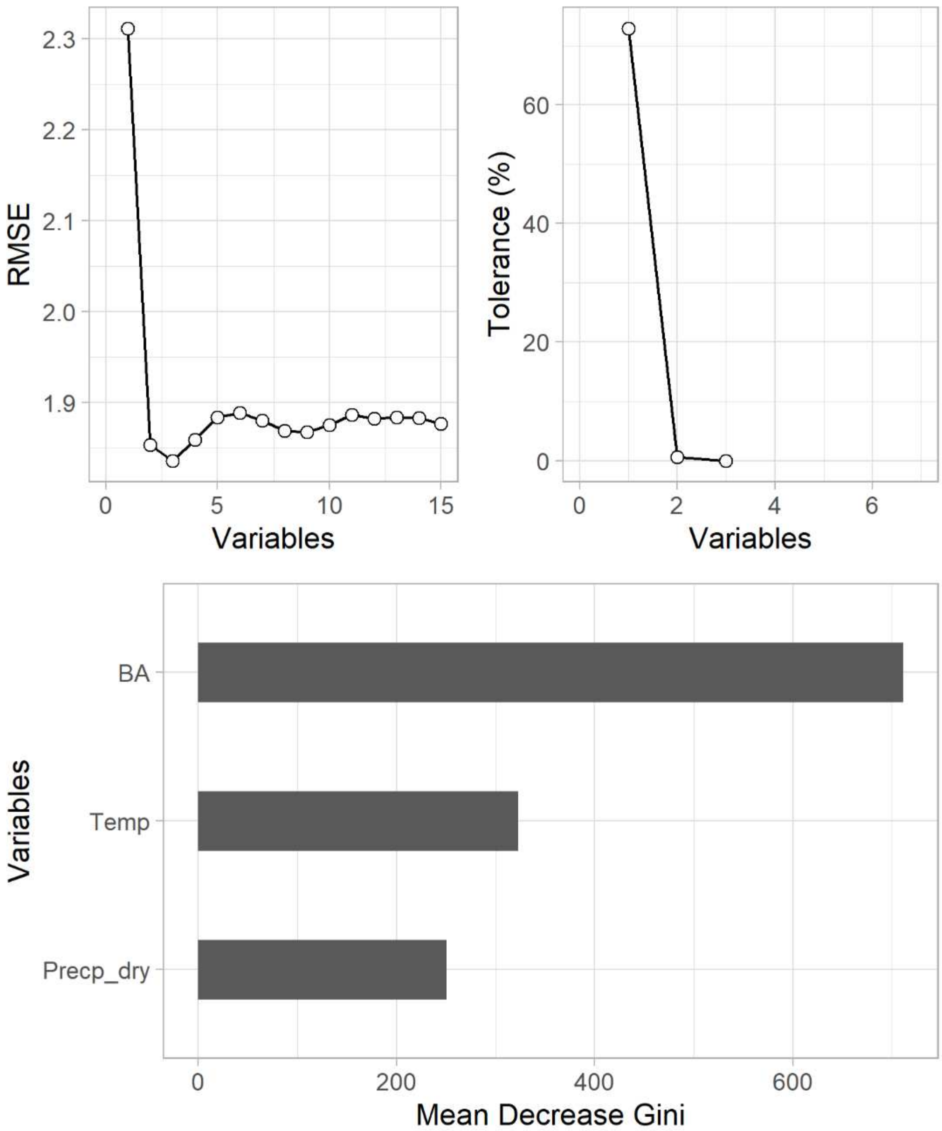

Figure 4. Most important variables for modeling the relative mortality rate (Mrel) of studied Atlantic Forest fragments in Minas Gerais, Brazil. BA: basal area; Temp: mean annual temperature 1 year before measurement and Precp_dry: total precipitation of the three driest months.

In general, the evaluated models showed a satisfactory generalization power, indicated by similar precision results between the observed and estimated data in the validation for all variables studied (Figure 5 and 6).

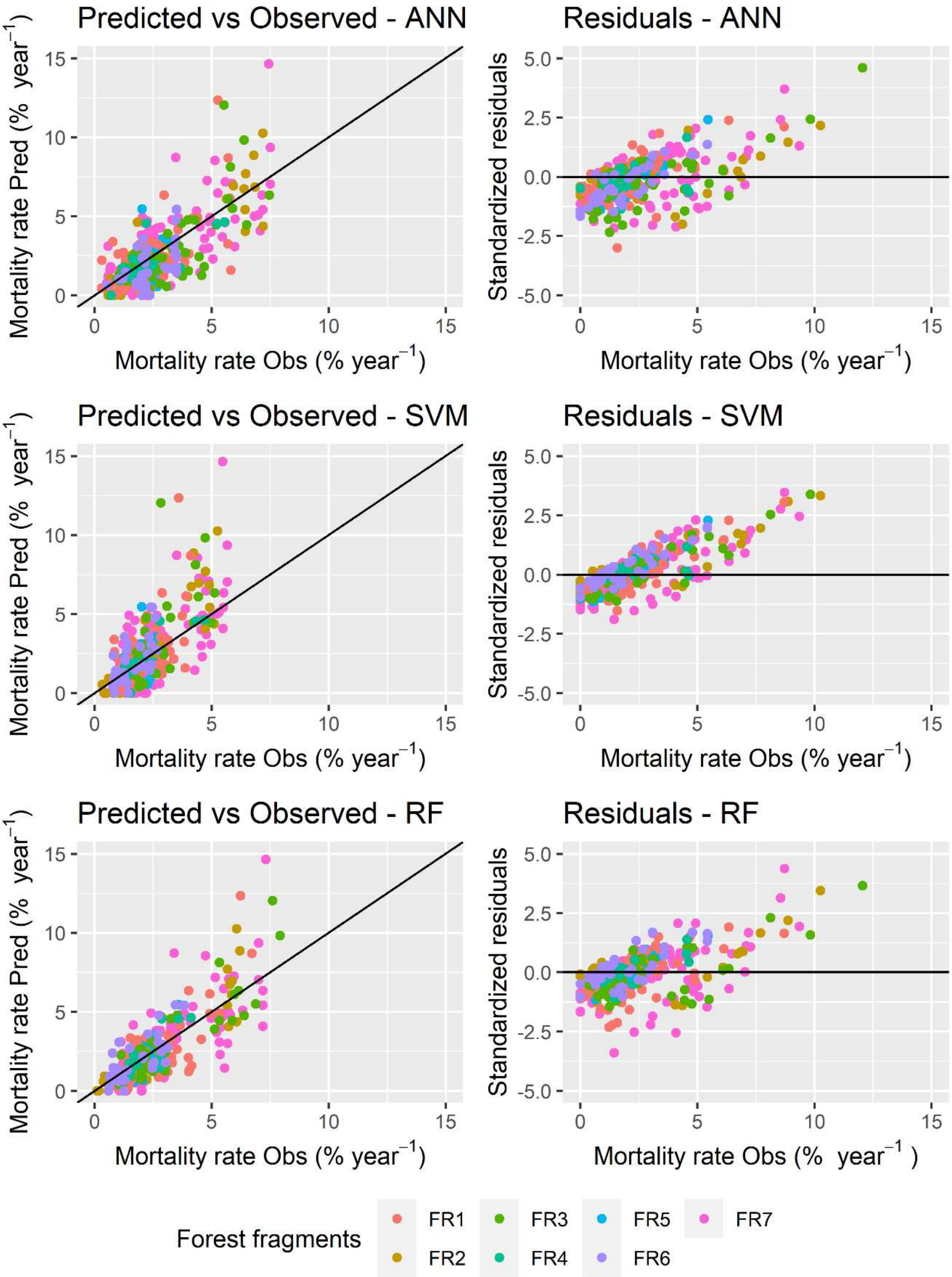

Figure 5. Observed and predicted and residuals values of mortality rate (M) for the different machine learning models, SVM, ANN and RF tested in Atlantic Forest fragments in Minas Gerais, Brazil. Colors represent the areas. Each small point represents the plots by areas. FR1: Cachoeira das Pombas; FR2: Mata da Garagem; FR3: Ipaba Mata1; FR4: Ipaba Mata2; FR5: Centev; FR6: São José and FR7: Mata da Silvicultura.

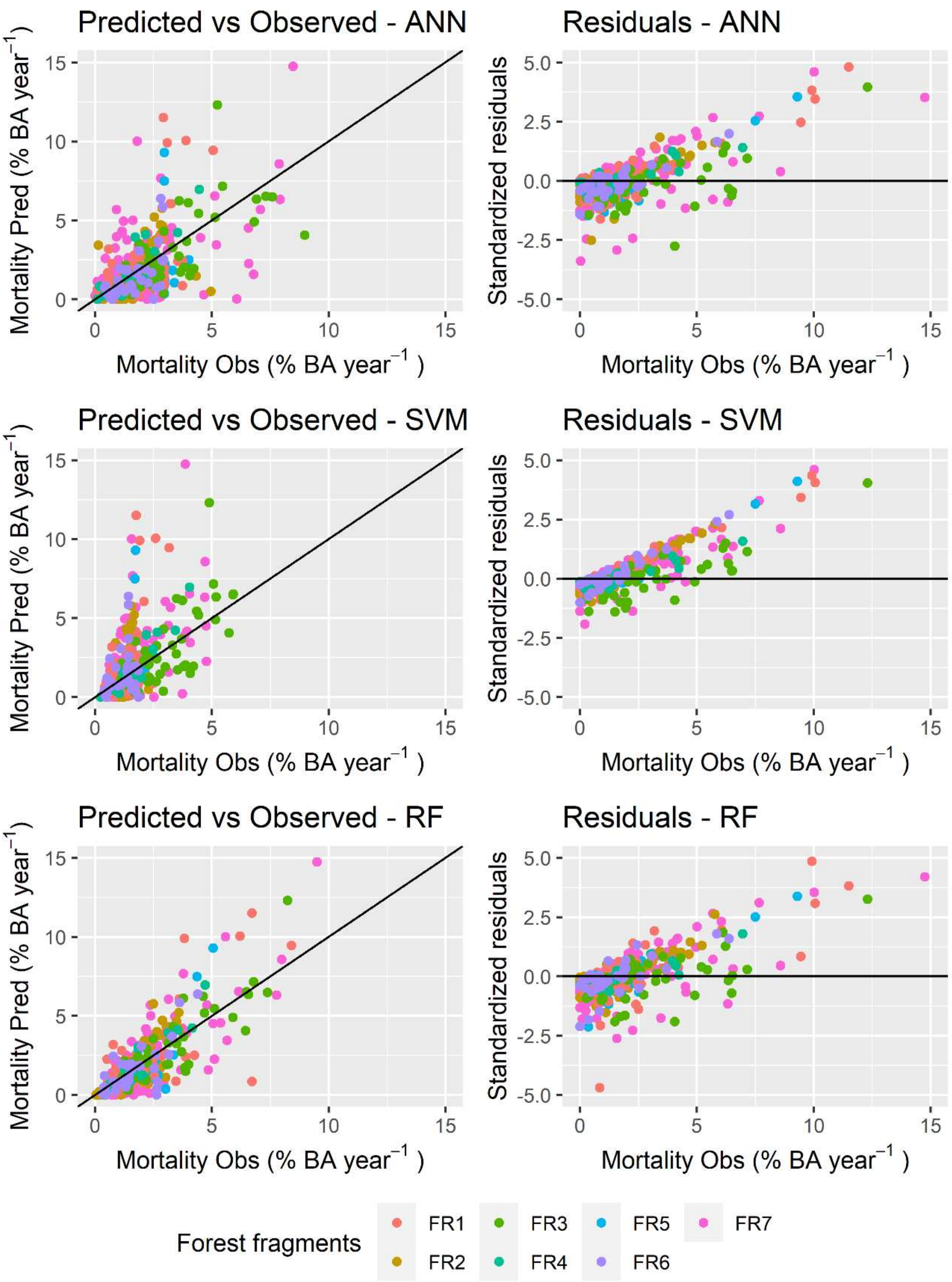

Figure 6. Observed and predicted and residuals values of relative mortality rate (Mrel) for the different machine learning models, SVM, ANN and RF tested in Atlantic Forest fragments in Minas Gerais, Brazil. Colors represent the areas. Each small point represents the plots by areas. FR1: Cachoeira das Pombas; FR2: Mata da Garagem; FR3: Ipaba Mata1; FR4: Ipaba Mata2; FR5: Centev; FR6: São José and FR7: Mata da Silvicultura.

RF showed the best performance to estimate the mortality trees, with the highest $r_{y\hat{y}}$ and the lowest RMSE and MAE for all repetitions (Table 3). ANN had the moderate performance to predict M and Mrel. The SVM had the worst performance for predicting the mortality in Atlantic Forest.

Table 3. Statistics of the machine learning models, SVM, ANN and RF tested in Atlantic Forest fragments in Minas Gerais, Brazil

Type	Input	Output	Technique	Neur./Trees	Algorithm	Function	RMSE	EAM	r	Bias
Train	BA, Age Aban, CWD, Precp2, Temp, Precp4, Precp_dry	M	RF	46			1.201±0.163	0.844±0.083	0.82±0.042	-1.804±5.643
			ANN	9	SCG	Exponential	1.416±0.165	1.042±0.095	0.738±0.052	-7.541±6.421
			SVM			Radial	1.508±0.248	0.988±0.114	0.722±0.06	10.675±5.692
Test	BA, Age Aban, CWD, Precp2, Temp, Precp4, Precp_dry	M	RF	46			1.211±0.042	0.844±0.021	0.819±0.011	-1.467±1.405
			ANN	9	SCG	Exponential	1.424±0.041	1.042±0.024	0.738±0.052	-7.541±6.421
			SVM			Radial	1.527±0.062	0.988±0.028	0.721±0.015	11.058±1.433
Train	BA, Temp, Precp_dry	Mrel	RF	69			1.238±0.158	0.84±0.089	0.829±0.066	-1.562±6.334
			ANN	6	SCG	Exponential	1.768±0.216	1.213±0.123	0.598±0.11	-3.601±9.557
			SVM			Radial	1.848±0.309	1.138±0.141	0.58±0.103	18.392±7.117
Test	BA, Temp, Precp_dry	Mrel	RF	69			1.248±0.04	0.84±0.022	0.837±0.016	-1.109±1.55
			ANN	6	SCG	Exponential	1.78±0.054	1.213±0.031	0.598±0.11	-3.601±9.557
			SVM			Radial	1.872±0.076	1.138±0.035	0.584±0.025	18.946±1.726

Where: RMSE: Root Mean Square Error; MAE: mean absolute error; SVM: Support Vector Machine; ANN: Artificial Neural Networks; RF: *Random Forest*. SCG: Scaled Conjugate Gradient. BA: Basal area; Age_Aban: Age of Abandonment; CWD: Climatic water deficit (from 1989 to 1 year before measurement); Precp_2: Mean annual precipitation (1 and 2 years before measurement); Temp: Mean annual temperature 1 year before measurement; Precp_4: Mean annual precipitation (1, 2, 3 and 4 years before measurement); Precp_dry: Total precipitation of the three driest months.

The means of RMSE and Correlation varied over the repetitions for each technique. The RMSE averages of RF, over the 50 repetitions in the cross-validation, showed the lowest values to estimate the mortality of trees and the highest values of Correlation (Figure 7) for all variables evaluated in the present study.

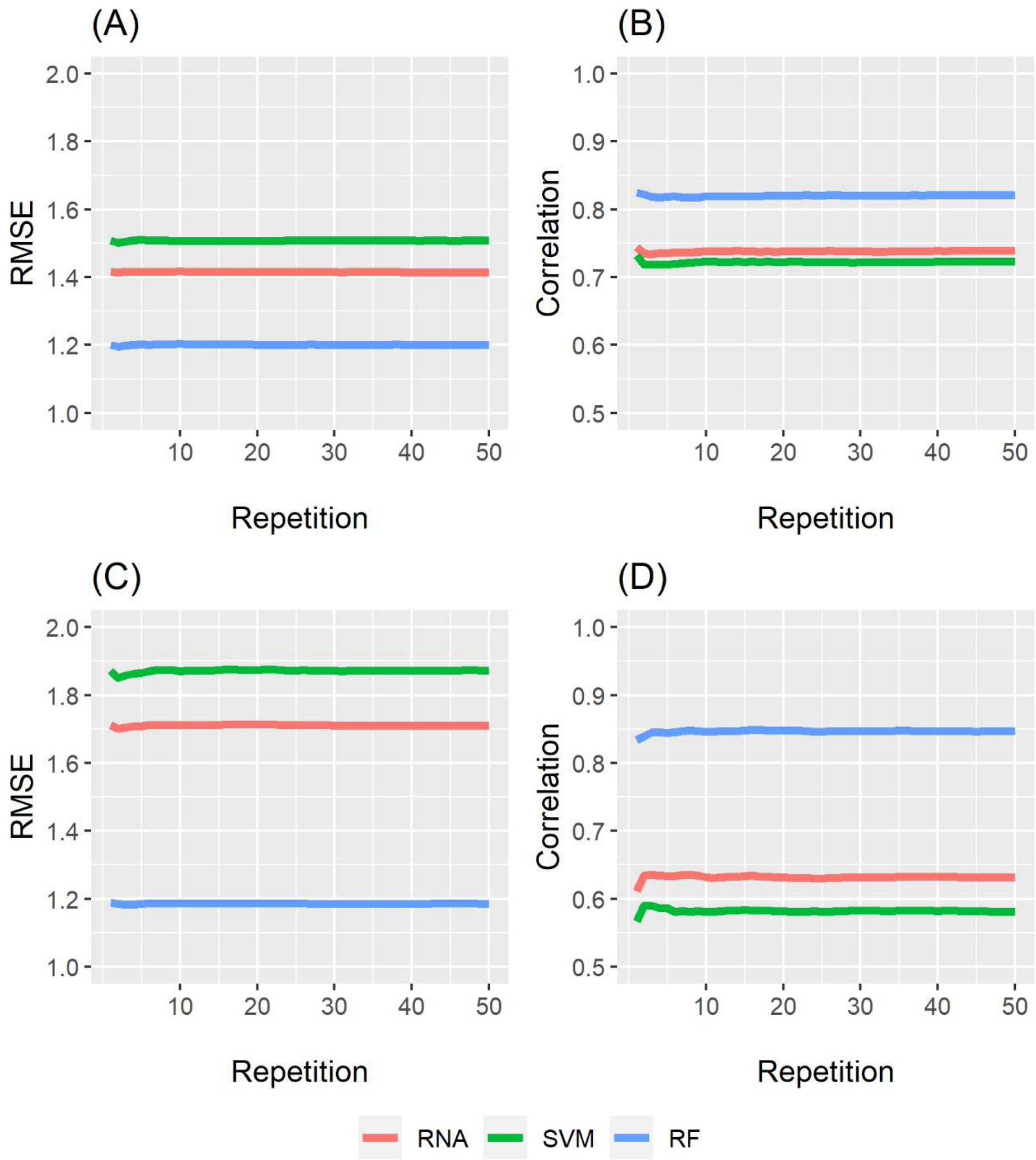

Figure 7. Root Mean Square Error (RMSE) and Correlation of the machine learning models SVM, ANN and RF, in the modeling of the mortality rates (A and B) and relative mortality rates (C and D) in the in Atlantic Forest fragments in Minas Gerais, Brazil.

The Nemenyi test pointed out that the difference between the RF model and the other techniques was greater than the calculated critical difference (CD). The calculated critical difference (CD) of the ANN and SVM it was not significant to estimate mortality rates in Atlantic Forest (Figure 8).

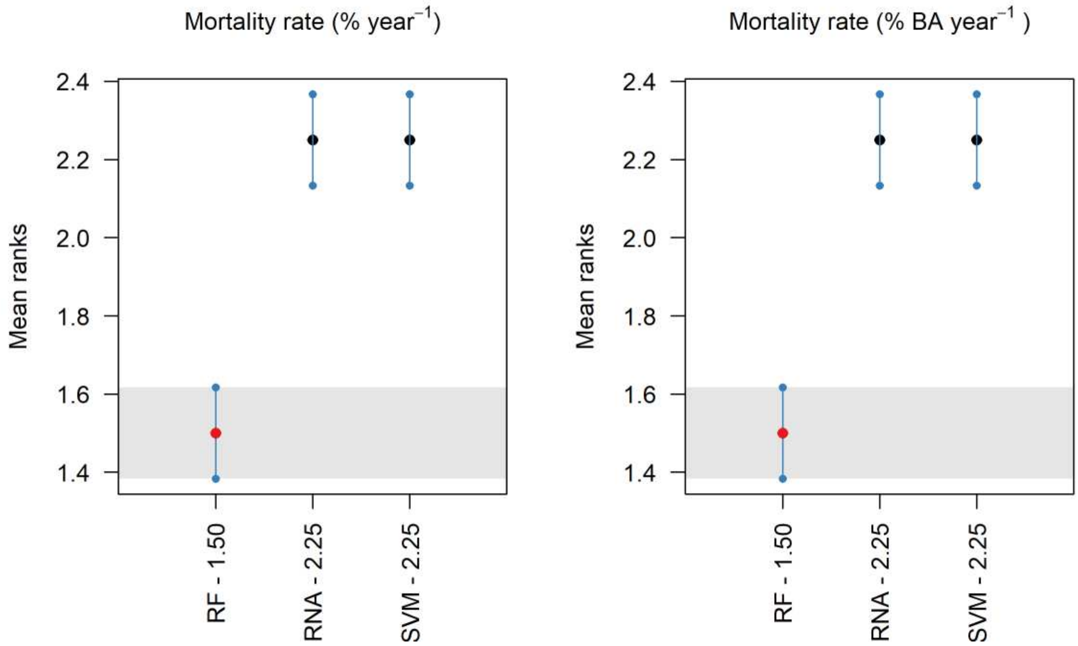

Figure 8. Nemenyi test of the machine learning models SVM, ANN and RF, in the modeling of the mortality rates in the in Atlantic Forest fragments in Minas Gerais, Brazil.

4. Discussion

Mortality models are recognized as key components for projecting forest ecosystem dynamics, structure, and composition (Salas-Eljatib and Weiskittel, 2020). However, in many ecosystems, obtaining an accurate estimate of mortality remains a challenge due to the interaction of several factors and the scarcity of data (Bayat et al., 2019). In this study, we obtained models for three of the most commonly used machine learning techniques (ANN, SVR e RF). The study has shown that modeling mortality at the stand level can be approached using climate and forest attributes.

Our results suggest that the survival is a function of Climate variables, BA and Age of Abandonment. Forest dynamics is the process of recruitment, growth, death, and renewal of the constituent tree species of the forest community. These processes are driven by natural and anthropogenic disturbances (McDowell et al., 2020).

The basal area is important to model mortality, as it is associated with competition. Annual tree mortality rates vary with forest composition and tree size structure (Sheil et al., 1995; Vanoni et al., 2019). Age of Abandonment reflect the stage of forest succession and the probability of survival, as they integrate many environmental influences (Dobbertin, 2005; Esquivel-Muelbert et al., 2020; Vanoni et al., 2019). Allied to these variables, the climate plays a fundamental role. The mortality of trees in tropical forests can be affected by the interaction of climate and the ecological characteristics of the trees (Phillips et al., 2010). For example, mortality rates tend to be higher for fast-growing early successional and softwood species because they have more acquisitive and less secure lifestyles, shorter life expectancy, and their wood is physically less protected against wind and pathogens (Aleixo et al., 2019).

Climate variables, like the ones observed, (CWD, Mean annual precipitation (2 and 4 years before measurement), Mean annual temperature 1 year before measurement and Total precipitation of the three driest months) are related to tree death (Aleixo et al., 2019; Meir et al., 2015; Seidl et al., 2017), especially in the Brazilian Atlantic Forest (Rocha et al., 2020). Mortality is affected by competition for resources (light, water, nutrient). Under conditions of resource scarcity, suppressed trees die by carbon deficit, hydraulic failure, or biotic attack as a result of reduced light, water, and nutrients due to increased competition (McDowell et al., 2018). Studies suggest that the timing of responses to weather events are not immediate, and that this can occur up to 6 years after the event (Seidl et al., 2017).

Our results confirm that the Random Forests model is a powerful machine learning algorithm for predictions. The prediction statistics (Table 4) show no overfitting, as the R^2 of the training and validation sets are similar. O RMSE e o MAE corroborate this assertion, showing small differences in training and validation. This method produces the most accurate and stable predictions (Sun et al., 2019), being increasingly used in ecological studies because it is suitable for the analysis of large complex data sets (Reise et al., 2019).

Observed, predicted and residuals plots suggest a difficulty in modeling in areas with higher mortality rates. This may be related to the type of mortality. One of the problems in modeling mortality is that several random factors can cause the death of trees (Reis et al., 2018). Tree mortality is a complex process that results from an interaction between regular mortality (due to competition and senescence) and catastrophic mortality due to extreme weather events and/or insect outbreaks (which are often induced by other disturbances) (Csilléry et al., 2013; Hawkes, 2000). Therefore, in these areas, high rates may be related to catastrophic events that are difficult to predict.

The inference and robustness of mortality predictions depend heavily on the modeling strategy (Salas-Eljatib and Weiskittel, 2020). The different configurations of the tested models showed that ANN requires a larger number of parameter settings. This is a disadvantage of this technique (Dernoncourt and Lee, 2016). SVR and RF, on the other hand, are easier to use, because few hyperparameters need to be set by the user (Ao et al., 2019; Tavares Júnior et al., 2020).

Finally, we found that the use of machine learning models can generate promising results in mortality estimates in natural forests. The future development of forest ecosystems depends critically on tree mortality (Hülsmann et al., 2017) Our study, presents a new way to do this, especially in forest fragments of the Atlantic Forest. It is worth noting that new approaches such as including other variables and using other models (e.g.: Cubist, Regression Trees Models, etc.) can improve the estimates. Studies of this kind should be encouraged and can help in better understanding tree mortality and assist in conservation practices in forests around the world, especially those threatened by human pressure and fragmentation.

5. Conclusion

We found out that Climate variables (Precipitation, Climatic water deficit and Temperature), Age of Abandonment and Basal area are important variables to predict tree mortality in the Atlantic Forest in Brazil. Machine Leaning are efficient methods to estimate mortality rates at plot-level with these variables. Random Forest is more efficient in estimating. Our findings support a new approach for modeling tree mortality in tropical forests around the world, especially in forest fragments of the Brazilian Atlantic Forest.

6. References

Aleixo, I., Norris, D., Hemerik, L., Barbosa, A., Prata, E., Costa, F., Poorter, L., 2019. Amazonian rainforest tree mortality driven by climate and functional traits. Nature Climate Change 9, 384–388. https://doi.org/10.1038/s41558-019-0458-0

Ao, Y., Li, H., Zhu, L., Ali, S., Yang, Z., 2019. The linear random forest algorithm and its advantages in machine learning assisted logging regression modeling. Journal of Petroleum Science and Engineering 174, 776–789. https://doi.org/10.1016/j.petrol.2018.11.067

Bayat, M., Ghorbanpour, M., Zare, R., Jaafari, A., Thai Pham, B., 2019. Application of artificial neural networks for predicting tree survival and mortality in the Hyrcanian forest of Iran. Computers and Electronics in Agriculture 164, 104929. https://doi.org/10.1016/j.compag.2019.104929

Bergmeir, C., Benítez, J.M., 2012. Neural Networks in R Using the Stuttgart Neural Network Simulator: RSNNS. Journal of Statistical Software 46, 1–26. https://doi.org/10.18637/jss.v046.i07

Breugel, M. van, Craven, D., Lai, H.R., Baillon, M., Turner, B.L., Hall, J.S., 2019. Soil nutrients and dispersal limitation shape compositional variation in secondary tropical forests across multiple scales. Journal of Ecology 107, 566–581. https://doi.org/10.1111/1365-2745.13126

Carreño-Rocabado, G., Peña-Claros, M., Bongers, F., Alarcón, A., Licona, J.-C., Poorter, L., 2012. Effects of disturbance intensity on species and functional diversity in a tropical forest. Journal of Ecology 100, 1453–1463. https://doi.org/10.1111/j.1365-2745.2012.02015.x

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M.S., Delitti, W.B.C., Duque, A., Eid, T., Fearnside, P.M., Goodman, R.C., Henry, M., Martínez-Yrízar, A., Mugasha, W.A., Muller-Landau, H.C., Mencuccini, M., Nelson, B.W., Ngomanda, A., Nogueira, E.M., Ortiz-Malavassi, E., Pélissier, R., Ploton, P., Ryan, C.M., Saldarriaga, J.G., Vieilledent, G., 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. Global Change Biology 20, 3177–3190. https://doi.org/10.1111/gcb.12629

Colombo, A.F., Joly, C.A., 2010. Brazilian Atlantic Forest lato sensu: the most ancient Brazilian forest, and a biodiversity hotspot, is highly threatened by climate change. Brazilian Journal of Biology 70, 697–708. https://doi.org/10.1590/S1519-69842010000400002

Csilléry, K., Seignobosc, M., Lafond, V., Kunstler, G., Courbaud, B., 2013. Estimating long-term tree mortality rate time series by combining data from periodic inventories and harvest reports in a Bayesian state-space model. Forest Ecology and Management 292, 64–74. https://doi.org/10.1016/j.foreco.2012.12.022

de Lima, R.A.F., Souza, V.C., de Siqueira, M.F., ter Steege, H., 2020. Defining endemism levels for biodiversity conservation: Tree species in the Atlantic Forest hotspot. Biological Conservation 252, 108825. https://doi.org/10.1016/j.biocon.2020.108825

de Toledo, J.J., Magnusson, W.E., Castilho, C.V., Nascimento, H.E.M., 2012. Tree mode of death in Central Amazonia: Effects of soil and topography on tree mortality associated with storm disturbances. Forest Ecology and Management 263, 253–261. https://doi.org/10.1016/j.foreco.2011.09.017

Dernoncourt, F., Lee, J.Y., 2016. Optimizing neural network hyperparameters with Gaussian processes for dialog act classification, in: 2016 IEEE Spoken Language Technology Workshop (SLT). Presented at the 2016 IEEE Spoken Language Technology Workshop (SLT), pp. 406–413. https://doi.org/10.1109/SLT.2016.7846296

Diniz, M.F., Coelho, M.T.P., de Sousa, F.G., Hasui, É., Loyola, R., 2021. The underestimated role of small fragments for carnivore dispersal in the Atlantic Forest. Perspectives in Ecology and Conservation. https://doi.org/10.1016/j.pecon.2020.12.001

Dixo, M., Metzger, J.P., Morgante, J.S., Zamudio, K.R., 2009. Habitat fragmentation reduces genetic diversity and connectivity among toad populations in the Brazilian Atlantic Coastal Forest. Biological Conservation 142, 1560–1569. https://doi.org/10.1016/j.biocon.2008.11.016

Dobbertin, M., 2005. Tree growth as indicator of tree vitality and of tree reaction to environmental stress: a review. Eur J Forest Res 124, 319–333. https://doi.org/10.1007/s10342-005-0085-3

dos Santos, A.R., Araújo, E.F., Barros, Q.S., Fernandes, M.M., de Moura Fernandes, M.R., Moreira, T.R., de Souza, K.B., da Silva, E.F., Silva, J.P.M., Santos, J.S., Billo, D., Silva, R.F., Nascimento, G.S.P., da Silva Gandine, S.M., Pinheiro, A.A., Ribeiro, W.R., Gonçalves, M.S., da Silva, S.F., Senhorelo, A.P., Heitor, F.D., Berude, L.C., de AlmeidaTelles, L.A., 2020. Fuzzy concept applied in determining potential forest fragments for deployment of a network of ecological corridors in the Brazilian Atlantic Forest. Ecological Indicators 115, 106423. https://doi.org/10.1016/j.ecolind.2020.106423

Esquivel-Muelbert, A., Phillips, O.L., Brienen, R.J.W., Fauset, S., Sullivan, M.J.P., Baker, T.R., Chao, K.-J., Feldpausch, T.R., Gloor, E., Higuchi, N., Houwing-Duistermaat, J., Lloyd, J., Liu, H., Malhi, Y., Marimon, B., Marimon Junior, B.H., Monteagudo-Mendoza, A., Poorter, L., Silveira, M., Torre, E.V., Dávila, E.A., del Aguila Pasquel, J., Almeida, E., Loayza, P.A., Andrade, A., Aragão, L.E.O.C., Araujo-Murakami, A., Arets, E., Arroyo, L., C, G.A.A., Baisie, M., Baraloto, C., Camargo, P.B., Barroso, J., Blanc, L., Bonal, D., Bongers, F., Boot, R., Brown, F., Burban, B., Camargo, J.L., Castro, W., Moscoso, V.C., Chave, J., Comiskey, J., Valverde, F.C., da Costa, A.L., Cardozo, N.D., Di Fiore, A., Dourdain, A., Erwin, T., Llampazo, G.F., Vieira, I.C.G., Herrera, R., Honorio Coronado, E., Huamantupa-Chuquimaco, I., Jimenez-Rojas, E., Killeen, T., Laurance, S., Laurance, W., Levesley, A., Lewis, S.L., Ladvocat, K.L.L.M., Lopez-Gonzalez, G., Lovejoy, T., Meir, P., Mendoza, C., Morandi, P., Neill, D., Nogueira Lima, A.J., Vargas, P.N., de Oliveira, E.A., Camacho, N.P., Pardo, G., Peacock, J., Peña-Claros, M., Peñuela-Mora, M.C., Pickavance, G., Pipoly, J., Pitman, N., Prieto, A., Pugh, T.A.M., Quesada, C., Ramirez-Angulo, H., de Almeida Reis, S.M., Rejou-Machain, M., Correa, Z.R., Bayona, L.R., Rudas, A., Salomão, R., Serrano, J., Espejo, J.S., Silva, N., Singh, J., Stahl, C., Stropp, J., Swamy, V., Talbot, J., ter Steege, H., Terborgh, J., Thomas, R., Toledo, M., Torres-Lezama, A., Gamarra, L.V., van der Heijden, G., van der Meer, P., van der Hout, P., Martinez, R.V., Vieira, S.A., Cayo, J.V., Vos, V., Zagt, R., Zuidema, P., Galbraith, D., 2020. Tree mode of death and mortality risk factors across Amazon forests. Nature Communications 11, 5515. https://doi.org/10.1038/s41467-020-18996-3

ESRI, 2015. ArcGIS Desktop: Release 10.3.1. Environmental Systems Research Institute, Redlands, CA.

Gomes, L.C., Faria, R.M., de Souza, E., Veloso, G.V., Schaefer, C.E.G.R., Filho, E.I.F., 2019. Modelling and mapping soil organic carbon stocks in Brazil. Geoderma 340, 337–350. https://doi.org/10.1016/j.geoderma.2019.01.007

Haddad, N.M., Brudvig, L.A., Clobert, J., Davies, K.F., Gonzalez, A., Holt, R.D., Lovejoy, T.E., Sexton, J.O., Austin, M.P., Collins, C.D., Cook, W.M., Damschen, E.I., Ewers, R.M., Foster, B.L., Jenkins, C.N., King, A.J., Laurance, W.F., Levey, D.J., Margules, C.R., Melbourne, B.A., Nicholls, A.O., Orrock, J.L., Song, D.-X., Townshend, J.R., 2015. Habitat fragmentation and its lasting impact on Earth's ecosystems. Science Advances 1, e1500052. https://doi.org/10.1126/sciadv.1500052

Hallinger, M., Johansson, V., Schmalholz, M., Sjöberg, S., Ranius, T., 2016. Factors driving tree mortality in retained forest fragments. Forest Ecology and Management 368, 163–172. https://doi.org/10.1016/j.foreco.2016.03.023

Hamidi, S.K., Zenner, E.K., Bayat, M., Fallah, A., 2021. Analysis of plot-level volume increment models developed from machine learning methods applied to an uneven-aged mixed forest. Annals of Forest Science 78, 4. https://doi.org/10.1007/s13595-020-01011-6

Hawkes, C., 2000. Woody plant mortality algorithms: description, problems and progress. Ecological Modelling 126, 225–248. https://doi.org/10.1016/S0304-3800(00)00267-2

Hülsmann, L., Bugmann, H., Brang, P., 2017. How to predict tree death from inventory data — lessons from a systematic assessment of European tree mortality models. Canadian Journal of Forest Research. https://doi.org/10.1139/cjfr-2016-0224

Hülsmann, L., Bugmann, H.K.M., Commarmot, B., Meyer, P., Zimmermann, S., Brang, P., 2016. Does one model fit all? Patterns of beech mortality in natural forests of three European regions. Ecological Applications 26, 2465–2479. https://doi.org/10.1002/eap.1388

IBGE, 2012. Manual Técnico da Vegetação Brasileira, 2ª. ed. IBGE, Rio de Janeiro.

Kuhn, M., Johnson, K., 2013a. Applied Predictive Modeling; Springer: New York, NY, USA,.

Kuhn, M., Johnson, K., 2013b. An Introduction to Feature Selection, in: Kuhn, M., Johnson, K. (Eds.), Applied Predictive Modeling. Springer, New York, NY, pp. 487–519. https://doi.org/10.1007/978-1-4614-6849-3_19

Lebrija-Trejos, E., Meave, J.A., Poorter, L., Pérez-García, E.A., Bongers, F., 2010. Pathways, mechanisms and predictability of vegetation change during tropical dry forest succession. Perspectives in Plant Ecology, Evolution and Systematics 12, 267–275. https://doi.org/10.1016/j.ppees.2010.09.002

Leite, R.V., Amaral, C.H. do, Pires, R. de P., Silva, C.A., Soares, C.P.B., Macedo, R.P., Silva, A.A.L. da, Broadbent, E.N., Mohan, M., Leite, H.G., 2020. Estimating Stem Volume in Eucalyptus Plantations Using Airborne LiDAR: A Comparison of Area- and Individual Tree-Based Approaches. Remote Sensing 12, 1513. https://doi.org/10.3390/rs12091513

Lutz, J.A., Wagtendonk, J.W. van, Franklin, J.F., 2010. Climatic water deficit, tree species ranges, and climate change in Yosemite National Park. Journal of Biogeography 37, 936–950. https://doi.org/10.1111/j.1365-2699.2009.02268.x

Macedo, T.M., Costa, W.S. da, Brandes, A.F. das N., Valladares, F., Barros, C.F., 2021. Diversity of growth responses to recent droughts reveals the capacity of Atlantic Forest trees to cope well with current climatic variability. Forest Ecology and Management 480, 118656. https://doi.org/10.1016/j.foreco.2020.118656

MapBiomas Project, 2019. Collection V3.1 of Brazilian Land Cover & Use Map Series [WWW Document]. URL http://mapbiomas.org (accessed 8.15.19).

Martínez-Ramos, M., Balvanera, P., Arreola Villa, F., Mora, F., Maass, J.M., Maza-Villalobos Méndez, S., 2018. Effects of long-term inter-annual rainfall variation on the dynamics of regenerative communities during the old-field succession of a neotropical dry forest. Forest Ecology and Management, Resilience of tropical dry forests to extreme disturbance events 426, 91–100. https://doi.org/10.1016/j.foreco.2018.04.048

McDowell, N., Allen, C.D., Anderson-Teixeira, K., Brando, P., Brienen, R., Chambers, J., Christoffersen, B., Davies, S., Doughty, C., Duque, A., Espirito-Santo, F., Fisher, R., Fontes, C.G., Galbraith, D., Goodsman, D., Grossiord, C., Hartmann, H., Holm, J., Johnson, D.J., Kassim, A.R., Keller, M., Koven, C., Kueppers, L., Kumagai, T., Malhi, Y., McMahon, S.M., Mencuccini, M., Meir, P., Moorcroft, P., Muller-Landau, H.C., Phillips, O.L., Powell, T., Sierra, C.A., Sperry, J., Warren, J., Xu, C., Xu, X., 2018. Drivers and mechanisms of tree mortality in moist tropical forests. New Phytologist 219, 851–869. https://doi.org/10.1111/nph.15027

McDowell, N.G., Allen, C.D., Anderson-Teixeira, K., Aukema, B.H., Bond-Lamberty, B., Chini, L., Clark, J.S., Dietze, M., Grossiord, C., Hanbury-Brown, A., Hurtt, G.C., Jackson, R.B., Johnson, D.J., Kueppers, L., Lichstein, J.W., Ogle, K., Poulter, B., Pugh, T.A.M., Seidl, R., Turner, M.G., Uriarte, M., Walker, A.P., Xu, C., 2020. Pervasive shifts in forest dynamics in a changing world. Science 368. https://doi.org/10.1126/science.aaz9463

Meir, P., Mencuccini, M., Dewar, R.C., 2015. Drought-related tree mortality: addressing the gaps in understanding and prediction. New Phytologist 207, 28–33. https://doi.org/10.1111/nph.13382

Melo, F.P.L., Pinto, S.R.R., Brancalion, P.H.S., Castro, P.S., Rodrigues, R.R., Aronson, J., Tabarelli, M., 2013. Priority setting for scaling-up tropical forest restoration projects: Early lessons from the Atlantic Forest Restoration Pact. Environmental Science & Policy 33, 395–404. https://doi.org/10.1016/j.envsci.2013.07.013

Menezes, T.G.C., Melo, F.P.L., 2019. Assembly patterns of tree seedling communities in a human-dominated Tropical landscape. Austral Ecology 44, 1204–1212. https://doi.org/10.1111/aec.12798

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Kent, J., 2000. Biodiversity hotspots for conservation priorities. Nature 403, 853–858. https://doi.org/10.1038/35002501

Phillips, O.L., Heijden, G. van der, Lewis, S.L., López-González, G., Aragão, L.E.O.C., Lloyd, J., Malhi, Y., Monteagudo, A., Almeida, S., Dávila, E.A., Amaral, I., Andelman, S., Andrade, A., Arroyo, L., Aymard, G., Baker, T.R., Blanc, L., Bonal, D., Oliveira, Á.C.A. de, Chao, K.-J., Cardozo, N.D., Costa, L. da, Feldpausch, T.R., Fisher, J.B., Fyllas, N.M., Freitas, M.A., Galbraith, D., Gloor, E., Higuchi, N., Honorio, E., Jiménez, E., Keeling, H., Killeen, T.J., Lovett, J.C., Meir, P., Mendoza, C., Morel, A., Vargas, P.N., Patiño, S., Peh, K.S.-H., Cruz, A.P., Prieto, A., Quesada, C.A., Ramírez, F., Ramírez, H., Rudas, A., Salamão, R., Schwarz, M., Silva, J., Silveira, M., Slik, J.W.F., Sonké, B., Thomas, A.S., Stropp, J., Taplin, J.R.D., Vásquez, R., Vilanova, E., 2010. Drought–mortality relationships for tropical forests. New Phytologist 187, 631–646. https://doi.org/10.1111/j.1469-8137.2010.03359.x

Poorter, L., Sande, M.T. van der, Arets, E.J.M.M., Ascarrunz, N., Enquist, B.J., Finegan, B., Licona, J.C., Martínez-Ramos, M., Mazzei, L., Meave, J.A., Muñoz, R., Nytch, C.J., Oliveira, A.A. de, Pérez-García, E.A., Prado-Junior, J., Rodríguez-Velázques, J., Ruschel, A.R., Salgado-Negret, B., Schiavini, I., Swenson, N.G., Tenorio, E.A., Thompson, J., Toledo, M., Uriarte, M., Hout, P. van der, Zimmerman, J.K., Peña-Claros, M., 2017. Biodiversity and climate determine the functioning of Neotropical forests. Global Ecology and Biogeography 26, 1423–1434. https://doi.org/10.1111/geb.12668

Pretzsch, H., 2009. Forest Dynamics, Growth, and Yield. pp. 1–39. https://doi.org/10.1007/978-3-540-88307-4_1

Redmond, M.D., 2019. CWD and AET function.

Reis, L.P., de Souza, A.L., dos Reis, P.C.M., Mazzei, L., Soares, C.P.B., Miquelino Eleto Torres, C.M., da Silva, L.F., Ruschel, A.R., Rêgo, L.J.S., Leite, H.G., 2018. Estimation of mortality and survival of individual trees after harvesting wood using artificial neural networks in the amazon rain forest. Ecological Engineering 112, 140–147. https://doi.org/10.1016/j.ecoleng.2017.12.014

Reis, L.P., de Souza, A.L., Mazzei, L., dos Reis, P.C.M., Leite, H.G., Soares, C.P.B., Torres, C.M.M.E., da Silva, L.F., Ruschel, A.R., 2016. Prognosis on the diameter of individual trees on the eastern region of the amazon using artificial neural networks. Forest Ecology and Management 382, 161–167. https://doi.org/10.1016/j.foreco.2016.10.022

Reise, J., Kukulka, F., Flade, M., Winter, S., 2019. Characterising the richness and diversity of forest bird species using National Forest Inventory data in Germany. Forest Ecology and Management 432, 799–811. https://doi.org/10.1016/j.foreco.2018.10.012

Ribeiro, M.C., Metzger, J.P., Martensen, A.C., Ponzoni, F.J., Hirota, M.M., 2009. The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. Biological Conservation, Conservation Issues in the Brazilian Atlantic Forest 142, 1141–1153. https://doi.org/10.1016/j.biocon.2009.02.021

Rocha, S.J.S.S. da, Torres, C.M.M.E., Jacovine, L.A.G., Leite, H.G., Gelcer, E.M., Neves, K.M., Schettini, B.L.S., Villanova, P.H., Silva, L.F. da, Reis, L.P., Zanuncio, J.C., 2018. Artificial neural networks: Modeling tree survival and mortality in the Atlantic Forest biome in Brazil. Science of The Total Environment 645, 655–661. https://doi.org/10.1016/j.scitotenv.2018.07.123

Rocha, S.J.S.S. da, Torres, C.M.M.E., Villanova, P.H., Schettini, B.L.S., Jacovine, L.A.G., Leite, H.G., Gelcer, E.M., Reis, L.P., Neves, K.M., Comini, I.B., Silva, L.F. da, 2020. Drought effects on carbon dynamics of trees in a secondary Atlantic Forest. Forest Ecology and Management 465, 118097. https://doi.org/10.1016/j.foreco.2020.118097

Rozendaal, D.M.A., Chazdon, R.L., 2015. Demographic drivers of tree biomass change during secondary succession in northeastern Costa Rica. Ecological Applications 25, 506–516. https://doi.org/10.1890/14-0054.1

Ruiz-Benito, P., Lines, E.R., Gómez-Aparicio, L., Zavala, M.A., Coomes, D.A., 2013. Patterns and Drivers of Tree Mortality in Iberian Forests: Climatic Effects Are Modified by Competition. PLOS ONE 8, e56843. https://doi.org/10.1371/journal.pone.0056843

Salas-Eljatib, C., Weiskittel, A.R., 2020. On studying the patterns of individual-based tree mortality in natural forests: A modelling analysis. Forest Ecology and Management 475, 118369. https://doi.org/10.1016/j.foreco.2020.118369

Santo-Silva, E.E., Almeida, W.R., Tabarelli, M., Peres, C.A., 2016. Habitat fragmentation and the future structure of tree assemblages in a fragmented Atlantic forest landscape. Plant Ecol 217, 1129–1140. https://doi.org/10.1007/s11258-016-0638-1

Seidl, R., Thom, D., Kautz, M., Martin-Benito, D., Peltoniemi, M., Vacchiano, G., Wild, J., Ascoli, D., Petr, M., Honkaniemi, J., Lexer, M.J., Trotsiuk, V., Mairota, P., Svoboda, M., Fabrika, M., Nagel, T.A., Reyer, C.P.O., 2017. Forest disturbances under climate change. Nature Climate Change 7, 395–402. https://doi.org/10.1038/nclimate3303

Sheil, D., Burslem, D.F.R.P., Alder, D., 1995. The Interpretation and Misinterpretation of Mortality Rate Measures. Journal of Ecology 83, 331–333. https://doi.org/10.2307/2261571

Silva, C.A., Klauberg, C., Hudak, A.T., Vierling, L.A., Liesenberg, V., Carvalho, S.P.C. e, Rodriguez, L.C.E., 2016. A principal component approach for predicting the stem volume in Eucalyptus plantations in Brazil using airborne LiDAR data. Forestry: An International Journal of Forest Research 89, 422–433. https://doi.org/10.1093/forestry/cpw016

Silva, I.N., Spatti, D.H., Flauzin, R.A., 2010. Redes neurais artificiais: para engenharia e ciências aplicadas. São Paulo: Artliber, 396 p.

Sousa, T.R., Costa, F.R.C., Bentos, T.V., Leal Filho, N., Mesquita, R.C.G., Ribeiro, I.O., 2017. The effect of forest fragmentation on the soil seed bank of Central Amazonia. Forest Ecology and Management 393, 105–112. https://doi.org/10.1016/j.foreco.2017.03.020

Stephenson, N., 1998. Actual evapotranspiration and deficit: biologically meaningful correlates of vegetation distribution across spatial scales. Journal of Biogeography 25, 855–870. https://doi.org/10.1046/j.1365-2699.1998.00233.x

Sun, T., Chen, F., Zhong, L., Liu, W., Wang, Y., 2019. GIS-based mineral prospectivity mapping using machine learning methods: A case study from Tongling ore district, eastern China. Ore Geology Reviews 109, 26–49. https://doi.org/10.1016/j.oregeorev.2019.04.003

Synek, M., Janda, P., Mikoláš, M., Nagel, T.A., Schurman, J.S., Pettit, J.L., Trotsiuk, V., Morrissey, R.C., Bače, R., Čada, V., Brang, P., Bugmann, H., Begovič, K., Chaskovskyy, O., Dušátko, M., Frankovič, M., Kameniar, O., Kníř, T., Kozák, D., Langbehn, T., Málek, J., Rodrigo, R., Saulnier, M., Teodosiu, M., Vostarek, O., Svoboda, M., 2020. Contrasting patterns of natural mortality in primary Picea forests of the Carpathian Mountains. Forest Ecology and Management 457, 117734. https://doi.org/10.1016/j.foreco.2019.117734

Tavares Júnior, I. da S., Torres, C.M.M.E., Leite, H.G., Castro, N.L.M. de, Soares, C.P.B., Castro, R.V.O., Farias, A.A., 2020. Machine learning: Modeling increment in diameter of individual trees on Atlantic Forest fragments. Ecological Indicators 117, 106685. https://doi.org/10.1016/j.ecolind.2020.106685

Teixeira, P.C., Donagemma, G.K., Fontana, A., Teixeira, W.G. (Eds.), 2017. Manual de métodos de análise de solo, 3ª. ed. Embrapa, Brasília, Brazil.

van der Sande, M.T., Peña-Claros, M., Ascarrunz, N., Arets, E.J.M.M., Licona, J.C., Toledo, M., Poorter, L., 2017. Abiotic and biotic drivers of biomass change in a Neotropical forest. Journal of Ecology 105, 1223–1234. https://doi.org/10.1111/1365-2745.12756

Vanclay, J., 1994. Modelling Forest Growth and Yield: Applications to Mixed Tropical Forests.

Vanoni, M., Cailleret, M., Hülsmann, L., Bugmann, H., Bigler, C., 2019. How do tree mortality models from combined tree-ring and inventory data affect projections of forest succession? Forest Ecology and Management 433, 606–617. https://doi.org/10.1016/j.foreco.2018.11.042

CAPÍTULO III: Comparison of machine learning methods in net growth estimates in the Atlantic Forest of Brazil

Abstract
Tree growth models are an important and essential part of modeling forest dynamics and valuable tools for management planning and biodiversity conservation strategies. We applied machine learning models, Artificial Neural Networks (ANN), Suport Vector Machine (SVM) and Random Forest (RF), to predict tree growth at plot-level in the Atlantic Forest of Brazil. Forest attributes, land use history, landscape, soil and climatic characteristics were used in the modeling. Recursive Feature Elimination was used to select the best subset of predictor variables. We found that edaphic, forest attributes and climatic variables are important in shaping growth in the Brazilian Atlantic Forest. Soil acidity was the most important characteristic. The machine learning methods were efficient. The Random Forest method showed superiority over the others for modeling growth in the Atlantic Forest. The Nemenyi test pointed out that the difference between the RF model and the other techniques was greater than the calculated critical difference (CD). Machine learning can be an important tool for modeling growth in forest fragments in the Brazilian Atlantic Forest. They can help in understanding the biome and in developing management strategies aimed at recovering biodiversity and reducing the deleterious effects of fragmentation.

Keywords: forest dynamics, net growth, tropical forest, Brazil.

1. Introduction

Biodiversity hotspots are among habitats most threatened by climate change around the world, and the Brazilian Atlantic Forest is a case in point (Scarano and Ceotto, 2015). The Atlantic Forest is classified as one of the 3 biodiversity hotspots most vulnerable to climate change (Bellard et al., 2014). The biome has a rich biodiversity (de Lima et al., 2015; Joly et al., 2014; Myers et al., 2000) and provides various ecosystem services, such as water supply, climate balance, food production, timber assortment and medicines, which contribute to human well-being (Bullock et al., 2011; Melo et al., 2013). Despite providing several environmental services to society, this is the Brazilian biome that has suffered most from fragmentation (Lewis et al., 2015; Magnago et al., 2014).

The biome is located in areas of intense urbanization, industrialization and agricultural activities (Scarano and Ceotto, 2015). The Atlantic Forest is the most threatened biome in Brazil. The area of the biome has reduced to about 12,4% of its original cover (SOS Mata Atlântica, 2019). Fragmentation directly impacts biodiversity (Haddad et al., 2015), plant and animal structure and composition (Câmara et al., 2017), microclimate (Schmidt et al., 2017), seed dispersal (Emer et al., 2018) and population dynamics (Arroyo-Rodríguez et al., 2017).

Monitoring forest dynamics in these fragmented areas is important to determine how tropical forests respond to land use and cover changes and global climate change (Bustamante et al., 2016). Forest dynamics arise from the interaction of environmental factors and disturbances with the demographic processes of recruitment, growth, and mortality, subsequently driving biomass and species composition (McDowell et al., 2020; Xu et al., 2016).

Net growth is the result of the three components of forest dynamics. Tree growth models are basic and essential components of forest dynamics modeling and valuable tools for forest management planning at any level (Uzoh and Oliver, 2008). Developing and validating these models can provide a better understanding of the tree growth and increment causes and mechanisms and predict the condition of plants at future times (Huy et al., 2021; Ma and Lei, 2015).

There is a large amount of literature on growth modeling forest plantation, however, many of the modeling approaches for such forests are not applicable to stands with multiple tree species and age groups (Vanclay, 1994). Mixed tropical forests present a special challenge because of the diversity of species and a wide range of sizes and ages (Huy et al., 2021).

Due to this complexity, machine learning models may represent a promising approach to this task Machine learning is a rapidly growing area of study that should become more common for modeling forest dynamics because of its potential to produce better models than traditional data modeling approaches (Gleason and Im, 2012; Jachowski et al., 2013; Reis et al., 2018, 2016; Rocha et al., 2018; Tavares Júnior et al., 2020; Zhao et al., 2011).

The objective of the present study was to develop models of machine learning capable of estimating the net growth in Atlantic Forest fragments, based on Forest Attributes, Land use history, Landscape, Soil and Climatic characteristics. (i) What are the most important variables to estimate net growth in these forests? (ii) Are the machine learning methods evaluated efficient for estimating the net growth at plot-level? (iii) What is the best method to estimate the net growth at plot-level?

2. Material and methods

2.1. Study sites and plot characteristics

We used data from seven Atlantic Forest fragments located in Minas Gerais, Brazil (Table 1 and Figure 1). The vegetation is classified as semideciduous seasonal forest (IBGE, 2012). Across sites the annual rainfall varies from 701 to 1737 mm. year^{-1}, elevation from 242 to 1169 m above sea level, and slope from 2.97 to 65.81% (Table S1).

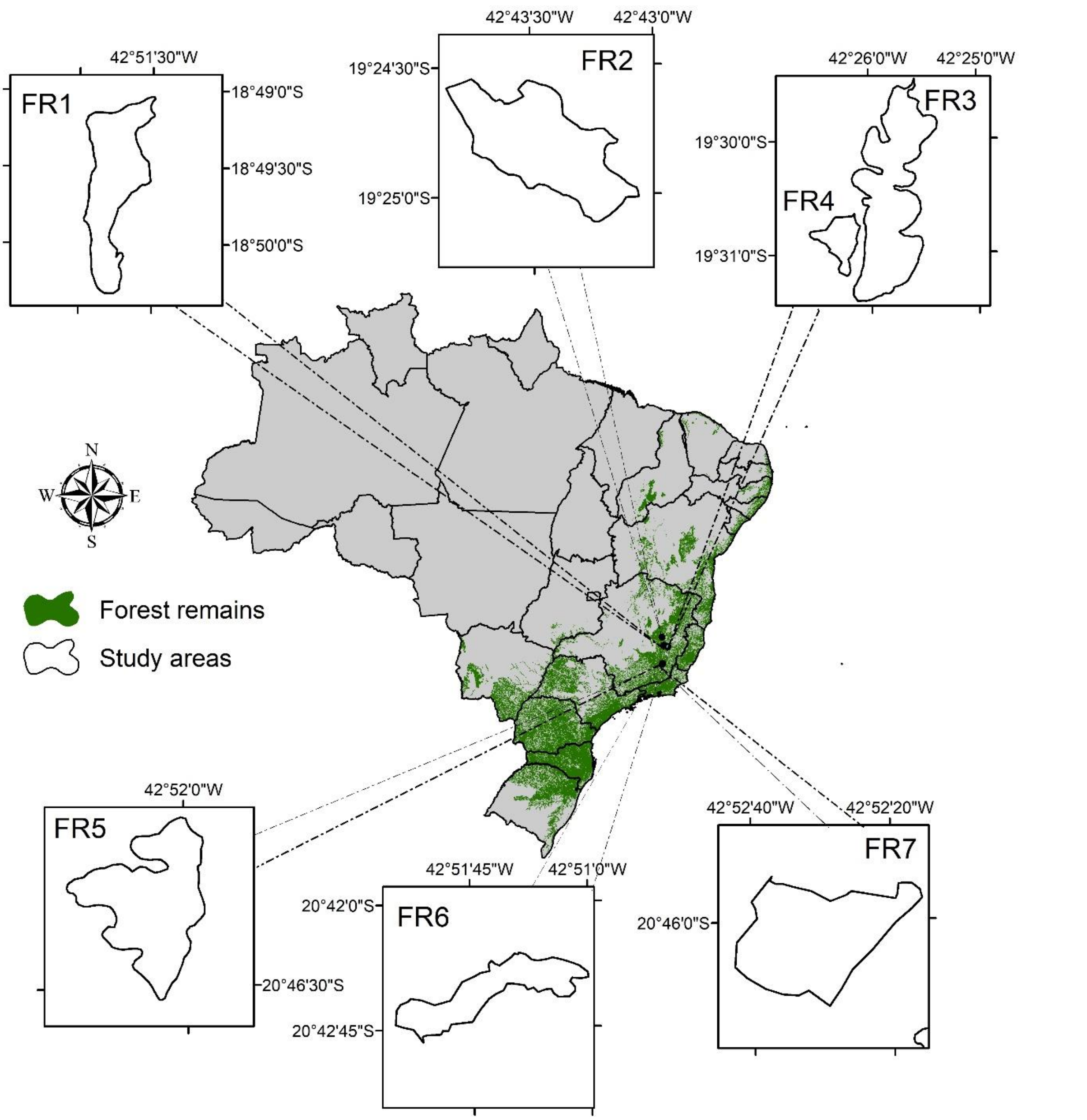

Figure 1. Locations of the seven studied Atlantic Forest fragments (FR) in Minas Gerais, Brazil. FR1: Cachoeira das Pombas; FR2: Mata da Garagem; FR3: Ipaba Mata1; FR4: Ipaba Mata2; FR5: Centev; FR6: São José and FR7: Mata da Silvicultura.

We measured 104 plots located in seven forest fragments over several years (Table 1). In all plots, we measured and identified botanically all stems with diameter at breast height (dbh), 1.3 m ≥ 5 cm.

Table 1. Overview of the seven studied Atlantic Forest fragments in Minas Gerais, Brazil. Forest fragments location (municipality), size, number and size of plots, and years of forest inventory are also provided

Fragment	Municipality	Forest Size (ha)	Plots	Plots Size (ha)	Measurement

FR1	Guanhães	106.0	20	0.05	2002, 2007, 2012, 2017
FR2	Viçosa	21.8	50	0.01	1996, 1998, 2001, 2003, 2013, 2018
FR3	Caratinga	264.0	16	0.05	2002, 2007, 2012, 2017
FR4	Caratinga	37.3	6	0.05	2002, 2007, 2012, 2017
FR5	Viçosa	44.1	20	0.05	2010, 2015
FR6	Coronel Fabriciano	38.4	12	0.05	2002, 2007, 2012, 2017
FR7	Viçosa	17.0	10	0.10	1994, 1997, 2000, 2004, 2008, 2010, 2013, 2016

FR1: Cachoeira das Pombas; FR2: Mata da Garagem; FR3: Ipaba Mata1; FR4: Ipaba Mata2; FR5: Centev; FR6: São José and FR7: Mata da Silvicultura.

As demographic processes may be influence by the total basal area of the plot (Carreño-Rocabado et al., 2012; Rozendaal and Chazdon, 2015). We calculated the basal area (BA) of each individual tree and measurement year to estimate the rate of mortality, recruitment and survivor growth at plot level (van der Sande et al., 2017).

Mortality is the BA of the stems alive on census t_n, where t indicates the census in time n, but dead on the census $t_n + 1$. Recruitment is the BA of stems that reached the minimum diameter for inclusion (dbh $\geq$ 5 cm) in census $t_n + 1$. Survivor growth was considered as the basal area increment due to the growth of surviving stems during a measurement period. Then, net growth was calculated summing recruitment and growth, and subtracting mortality.

2.2. Forest attributes and Anthropogenic variables

We use forest attributes from Atlantic Forest fragments as predictive variables. The attributes of the forest used were: Basal area of the plot and Number of stems. This information was collected through forest inventories carried out over the years.

The Anthropogenic variables used to estimate tree growth rate were: Land use history, forest cover, forest size and edge distance, age of abandonment (that means, time since abandonment previous anthropogenic uses). Based on information collected from aerial photographs (from around 1960, 1980 and 1990), Landsat satellite images from 1985, landowner interviews and land titles, we determined the Land use history and age of abandonment of the areas. The categories of land use used were: deforestation, agricultural production, eucalyptus plantation, and selective logging.

Forest cover was calculated for each plot using circular buffers with radii of 500, 1000, and 2000 m, for the year 1985, 2002 and 2017, with data from MapsBiomas (MapBiomas

Project, 2019) in ArcGIS 10.3.1 (ESRI, 2015). Forest cover area was produced from the pixel-per-pixel (30 x 30 m) classification of Landsat satellite images through the Google Earth Engine platform (MapBiomas Project, 2019). We calculated distance to the nearest edge of the forest (Edge distance) with near tool in ArcGIS 10.3.1 1 (ESRI, 2015).

2.3. Climate variables

We obtained annual precipitation, the number of months with less than 100 mm of rainfall, precipitation in the three driest months, and average annual temperature for each fragment from the nearest climatological station (Figure S1). We then estimated the climatic water deficit CWD (a water balance between precipitation and evapotranspiration) as a proxy for drought conditions following Lutz et al. (2010) and using the R function CWD and AET (actual evapotranspiration) from Redmond (2019).

We used slope, latitude, aspect, precipitation and temperature of the site for monthly calculations. AET as the evaporative water loss from a site covered by a hypothetical standard crop, given the prevailing water availability (Stephenson, 1998).

CWD reflects drought conditions more accurately than total annual rainfall (Chave et al., 2014), and was consider anthropogenic variable. More negative CWD indicates high water stress conditions and values close to 0 (zero) indicates not water stressed (Poorter et al., 2017). We calculated the average of total annual precipitation and CWD for one, two, three and four years before the measurement year.

2.4. Soil and topography variables

We collect soil samples for each plot. We obtained information for the depths of 0-20 cm and 20-40 cm. 20-30 samples were collected per plot to obtain a composite sample. Soil pH in H_2O, exchangeable cations (Ca^{2+}, Mg^{2+} and Al^{3+}), total acidity (H^{+}+ Al^{3+}), cation exchange capacity (CEC), base saturation (V), available phosphorus (P), P remaining in solution (P-res) and soil organic matter (SOM) were determined using standard methods (Teixeira et al., 2017). The soil analyses were performed at the laboratory of Soil Fertility at the Universidade Federal de Viçosa (UFV), Brazil.

To account for differences in topography, we calculated for each plot the elevation, slope and aspect using Spatial Analyst Tools of surface in ArcGIS 10.3.1 1 (ESRI, 2015). We used the Shuttle Radar Topography Mission (SRTM) and Digital Elevation Models (DEMs) for the analysis.

2.5.Data analyses

2.5.1. Variables selection

We performed tests to determine the variables to be included in the statistical models to assess the effect of anthropogenic and environmental variables (Table S1) on growth. Quantitative variables were standardized to accelerate the convergence rate and reduce the iteration process in training (Equation 1). The scale function of R Software was used in this step.

$$Z_i = (x_i - \bar{x})/\sigma \tag{1}$$

where:

Z_i = standardized value of the i-th observation;

x_i = value of the i-th observation;

$\bar{x}$ = average of the observed values;

σ = standard deviation.

First, we exclude highly correlated variables using a correlation coefficient limit of ± 0.9 (Leite et al., 2020; Silva et al., 2016). Subsequently, a method based on Recursive Feature Elimination (RFE) (Gomes et al., 2019), was used, considering only the variables not excluded in the first step to select the best subset of variables. This method is a reverse selection algorithm that calculates the importance of the resource in each iteration, classifying them from most important to least important, removing a user-defined subset at each stage (Kuhn and Johnson, 2013a, 2013b). Although resource collinearity cannot severely affect nonparametric methods, the exclusion of highly correlated methods was important to make RFE iterations more constant, as resources can be interchangeable within models (Leite et al., 2020).

2.5.2. Model selection, evaluation, and inference

The tested models to estimate the growth rates were: Support Vector Machines (SVM), Artificial Neural Networks (ANN) and *Random Forests* (RF).

The trained ANN was the multilayer perceptron, also known as the multilayer perceptron (MLP), with a hidden layer. The range of neurons in this layer was defined by the Fletcher-Gloss method (Silva et al., 2010): 2 × n0.5 + n2 ≤ n1 ≤ 2 × n + 1; where n = number of network inputs; n1 = amount of neurons in the hidden layer; and n2 = number of neurons in the output layer. The activation functions tested were exponential, identity, logistic, and hyperbolic tangent. The training algorithms used were resilient propagation (Rprop) and scaled conjugate gradient (SCG). The initial ANN weights were randomly generated, and the maximum number of iterations was 100 due to the error becoming constant before this number. The ANNs were implemented with the MLP function of the "RSNNS" Package in R (Bergmeir and Benítez, 2012). The function svm of the "e1071" Package on *R* was used for training SVMs. Thus, four configurations were used in the SVR training, represented by four kernel functions: Linear, Polynomial, Radial basis and Sigmoid.

In the RF training, three essential parameters were configured: the number of random regression trees (ntree, tested 20 to 100 trees); the number of division variables (mtry, used to determine the number of variables available to each node of the tree, with the default number of 1/3 of the independent variables); and the minimum size of nodes (node size, value = 5).

The performance of the models in the estimation was assessed using the *k*–fold cross-validation method, with the data divided into 5 folds (4 for adjustments/training and 1 for validation). At each adjustment/training of the folds the metrics of Root Mean Square Error – RMSE (Equation 2); Mean Absolute Error – MAE (Equation 3), Pearson correlation coefficient - $r_{y\hat{y}}$ (Equation 4); BIAS (Equation 5) and Relative Bias (%) (Equation 6) were calculated. This process was repeated 50 times, obtaining the average of the metrics for comparison of all models. The data were selected randomly in each of the 50 repetitions, resulting in different data sets, for greater robustness of the evaluation.

$$RMSE = \sqrt{\sum_{r=1}^{R} \frac{\sum_{i=1}^{n}(X_i - \hat{X}_i)^2}{n}} \quad (2)$$

$$MAE = \sum_{r=1}^{R} \frac{\left|\sum_{i=1}^{n}(X_i - \hat{X}_i)\right|}{n} \quad (3)$$

$$r_{y\hat{y}} = \frac{cov(X, X)}{\sqrt{s^2(X) \times s^2(\hat{X})}} \quad (4)$$

$$Bias = \frac{\left(\sum_i^n X_i - \hat{X}_i\right)}{n} \quad (5)$$

$$rBias\,(\%) = \frac{Bias}{\bar{X}} \times 100 \quad (6)$$

Where:
n = number of observations;
X_i = observed variable from the i-th plot;
$\hat{X}_i$ = estimated variable of the i-th plot.

The averages of RMSE, EAM, $r_{y\hat{y}}$ and Bias of each method in each repetition were ranked with weight assignments from 1 to 3, with 1 for the lowest value and 3 for the highest value. With the result of these sums, the values were submitted to the Friedman – Nemenyi test, at the 5% significance level (Equation 7).

The Friedman and Nemenyi nonparametric tests were used to compare ANN, SVR, and RF, based on the cross-validation RMSE, EAM, r and Bias means. The null hypothesis of Friedman's test is that all algorithms are equivalent. Nemenyi's post hoc test is applied to report significant differences between the techniques if the null hypothesis is rejected. The techniques' performance differs when the mean RMSE by at least one calculated critical difference (CD) differs (Tavares Júnior et al., 2020).

$$CD = q_{\alpha}\sqrt{\frac{k(k+1)}{6N}} \quad (7)$$

Where:
CD = critical difference;
q_{α} = critical value calculated based on Studentized interval statistics divided by √2;
k = number of algorithms being compared;
N = number of data sets.

In total, 64 ANNs, 80 RFs and 4 SVM were trained for all fragments for estimate annual growth rate of the Brazilian Atlantic Forest. The methodological flowchart used is presented below (Figure 2).

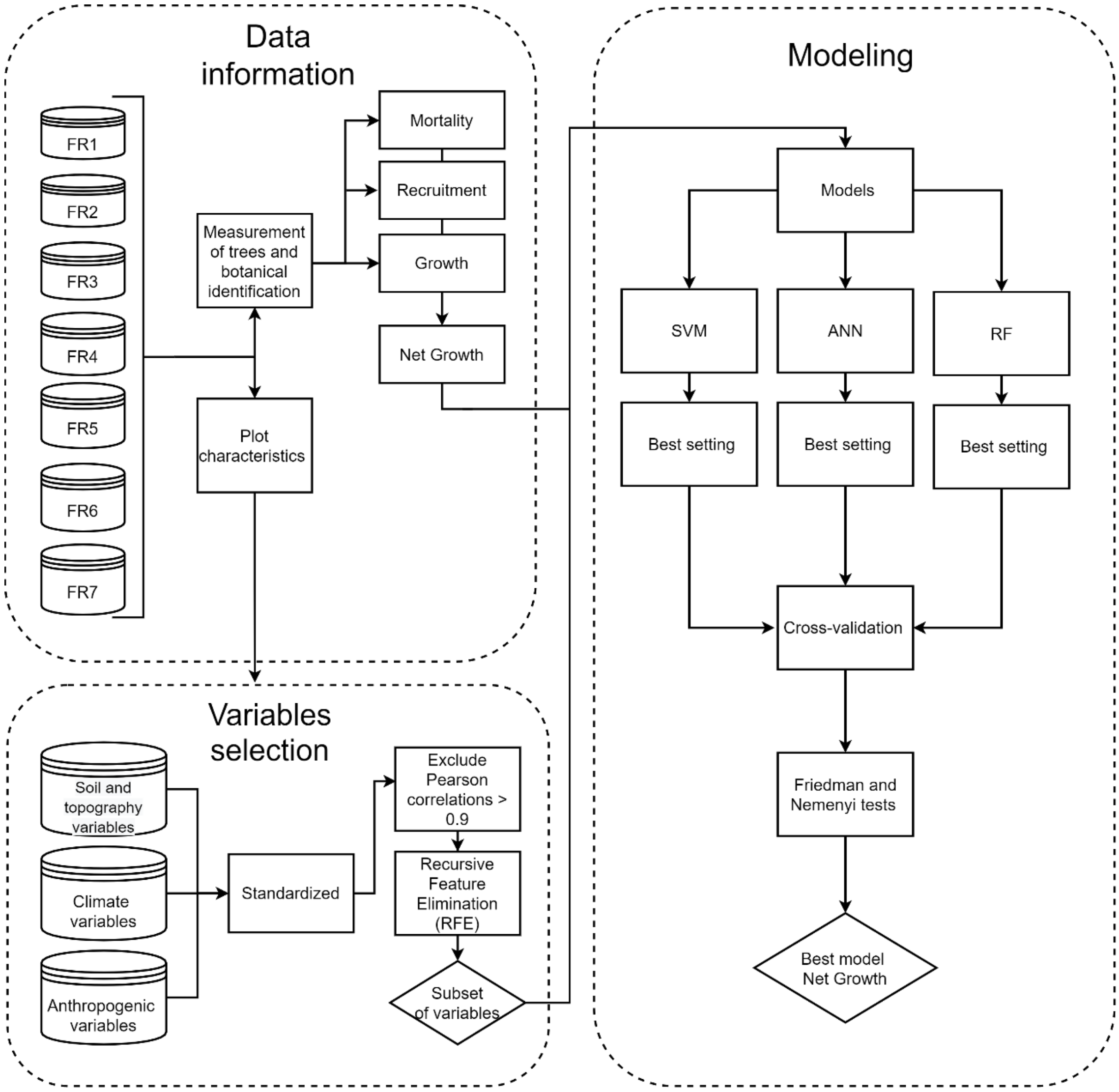

Figure 2. Methodological flowchart for modeling tree growth in Atlantic Forest in Brazil.

3. Results

We found that pH H_2O 20cm - 40 cm, BA, Age of Abandonment, annual precipitation, mean annual precipitation, and total precipitation of the three driest months were the most important variables to predict the net growth of trees in the Atlantic Forest (Figure 3).

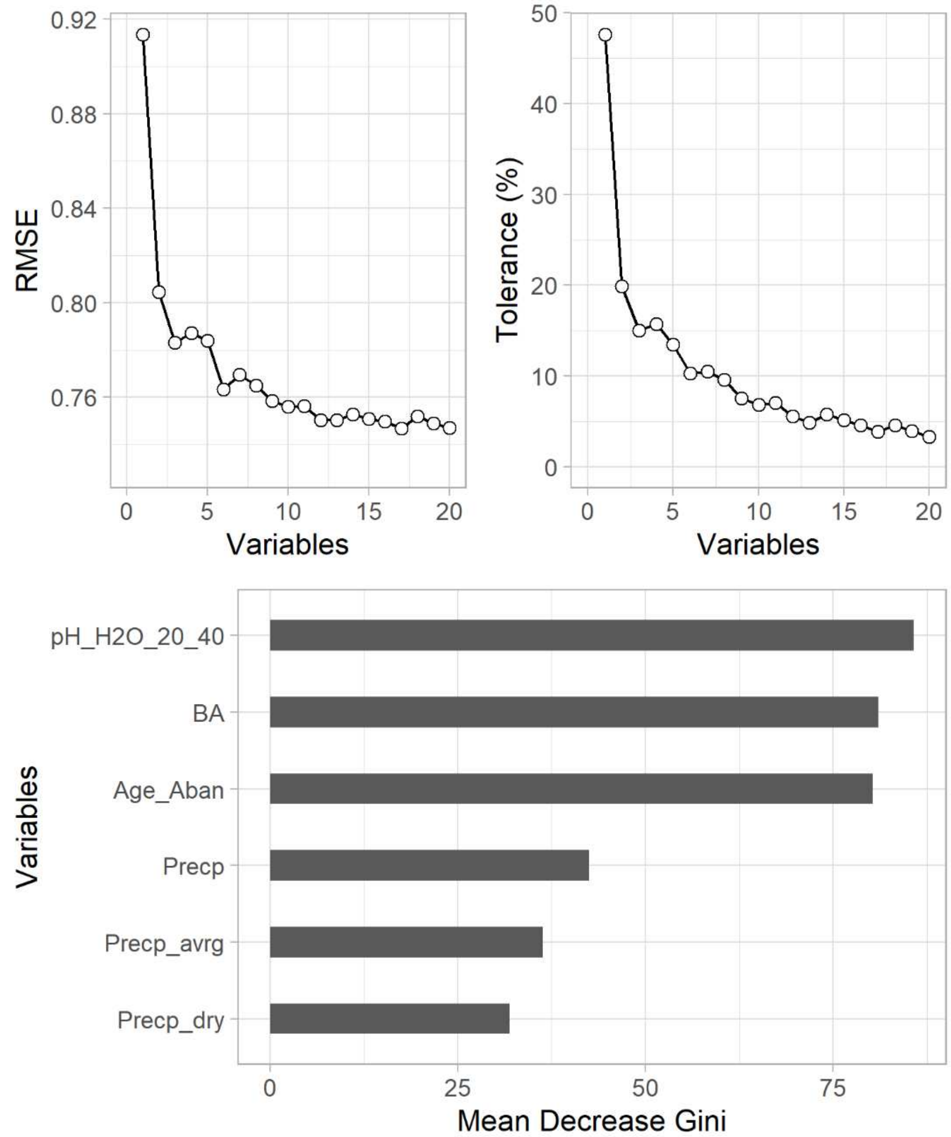

Figure 3. Most important variables for modeling the net growth of studied Atlantic Forest fragments in Minas Gerais, Brazil. pH_H20_20_40: pH H_2O 0cm - 40 cm; BA: Basal area; Age_Aban: Age of Abandonment; Precp: Annual precipitation; Precp_avrg: Mean annual precipitation; Precp_dry: Total precipitation of the three driest months.

In general, the evaluated models showed a satisfactory generalization power, indicated by similar precision results between the observed and estimated data in the validation for all variables studied (Figure 4).

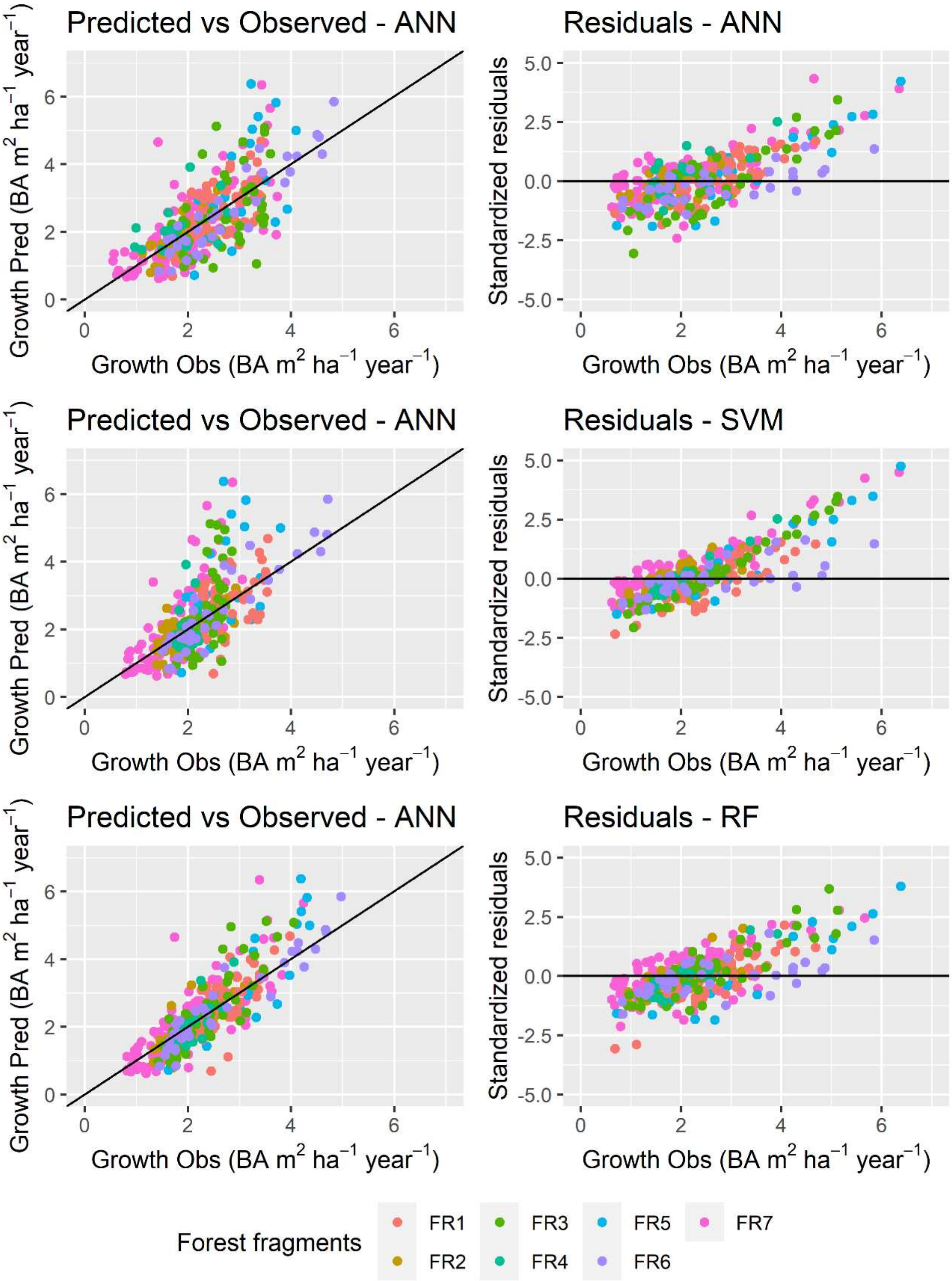

Figure 4. Observed and predicted and residuals values of M for the different machine learning models, SVM, ANN and RF tested in Atlantic Forest fragments in Minas Gerais, Brazil. Colors represent the areas. Each small point represents the plots by areas. FR1: Cachoeira das Pombas; FR2: Mata da Garagem; FR3: Ipaba Mata1; FR4: Ipaba Mata2; FR5: Centev; FR6: São José and FR7: Mata da Silvicultura.

RF showed the best performance to estimate net growth, with the highest $r_{y\hat{y}}$ and the lowest RMSE and MAE for all repetitions (Table 3). ANN had the moderate performance to predict the growth. The SVM had the worst performance for predicting the net growth in Atlantic Forest.

Table 3. Statistics of the machine learning models, SVM, ANN and RF tested in Atlantic Forest fragments in Minas Gerais, Brazil

Tipo	Input	Output	Technique	Neur./Trees	Algorithm	Function	RMSE	EAM	r	Bias
Train	pH_H20_20_40, BA, Age_Aban, Precp, Precp_avrg and Precp_dry	Net growth	RF	37			0.571±0.073	0.409±0.042	0.854±0.039	0.95±2.847
			ANN	7	Rprop	Tangencial	0.742±0.079	0.548±0.049	0.717±0.058	0.001±3.493
			SVM			Radial	0.775±0.106	0.518±0.061	0.699±0.059	4.99±3.635
Test	pH_H20_20_40, BA, Age_Aban, Precp, Precp_avrg and Precp_dry	Net growth	RF	37			0.575±0.019	0.409±0.011	0.855±0.009	1.06±0.717
			ANN	7	Rprop	Tangencial	0.746±0.02	0.548±0.012	0.717±0.058	0.001±3.493
			SVM			Radial	0.782±0.027	0.518±0.015	0.698±0.015	5.139±0.915

Where: RMSE: Root Mean Square Error; MAE: mean absolute error; SVM: Support Vector Machine; ANN: Artificial Neural Networks; RF: *Random Forest*. Rprop: Resilient backpropagation. pH_H20_20_40: pH H_2O 0cm - 40 cm; BA: Basal area; Age_Aban: Age of Abandonment; ; Precp: Annual precipitation; Precp_avrg: Mean annual precipitation; Precp_dry: Total precipitation of the three driest months.

The means of RMSE, MAE, Correlation and Relative Bias varied over the repetitions for each technique. RF showed the lowest values of RMSE (Figure 5-A) and MAE (Figure 5-B) and highest values of correlation (Figure 5-C) to estimate net growth of trees in the cross-validation. The RNA showed the lowest values of relative Bias (Figure 5-D).

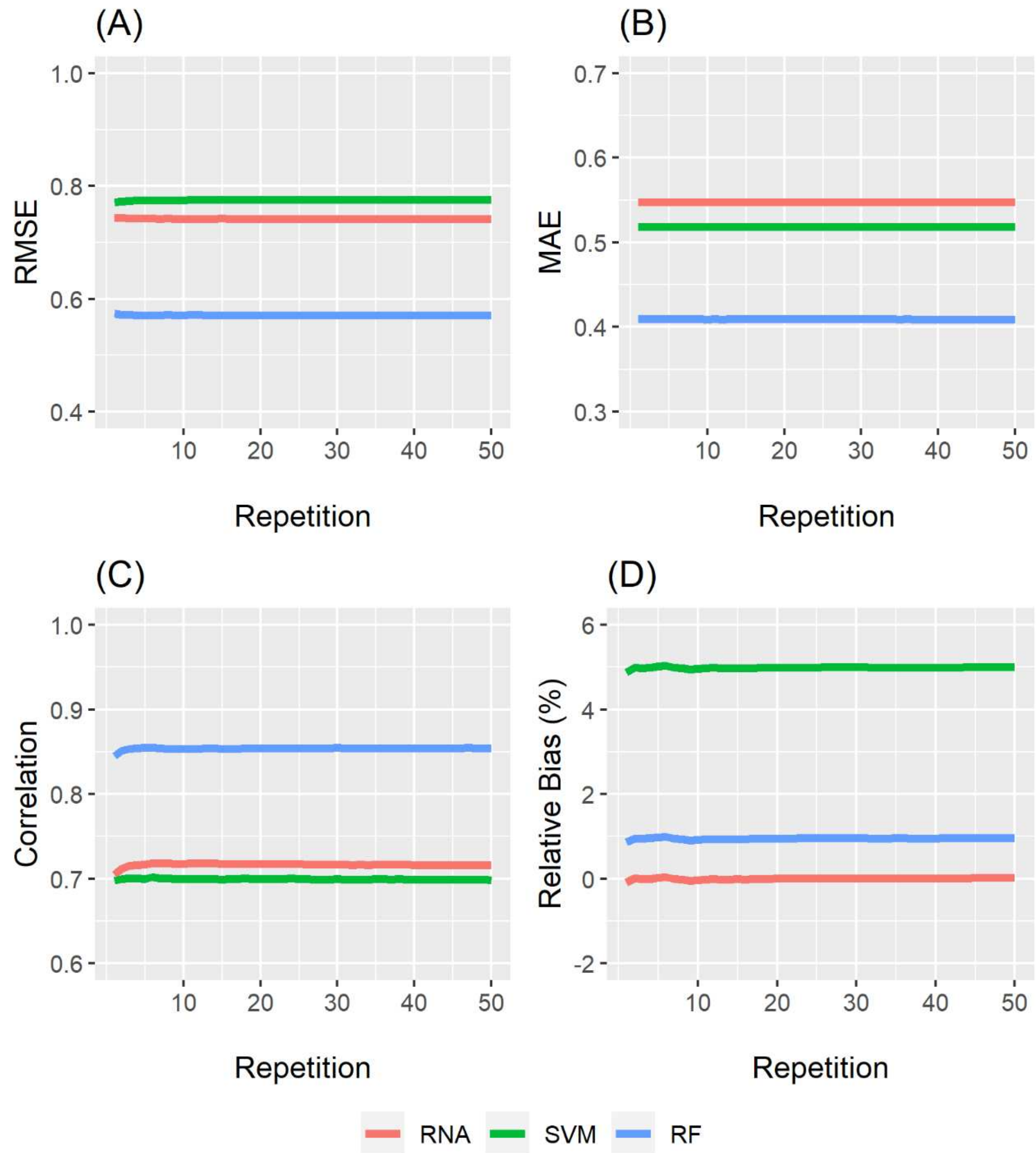

Figure 5. Root Mean Square Error – RMSE (A), Mean Absolute Error – MAE (B), Correlation (C) and Relative Bias (D) of the machine learning models SVM, ANN and RF, in the modeling of the net growth in the in Atlantic Forest fragments in Minas Gerais, Brazil.

The Friedman test with the means of cross-validation RMSE showed that the predictions of net growth in the Atlantic Forest differed between the techniques ($p < 0.05$). The Nemenyi test pointed out that the difference between the RF model and the other techniques was greater than the calculated critical difference (CD). The calculated critical difference (CD) of the ANN and SVM it was not significant to estimate net growth in Atlantic Forest (Figure 6).

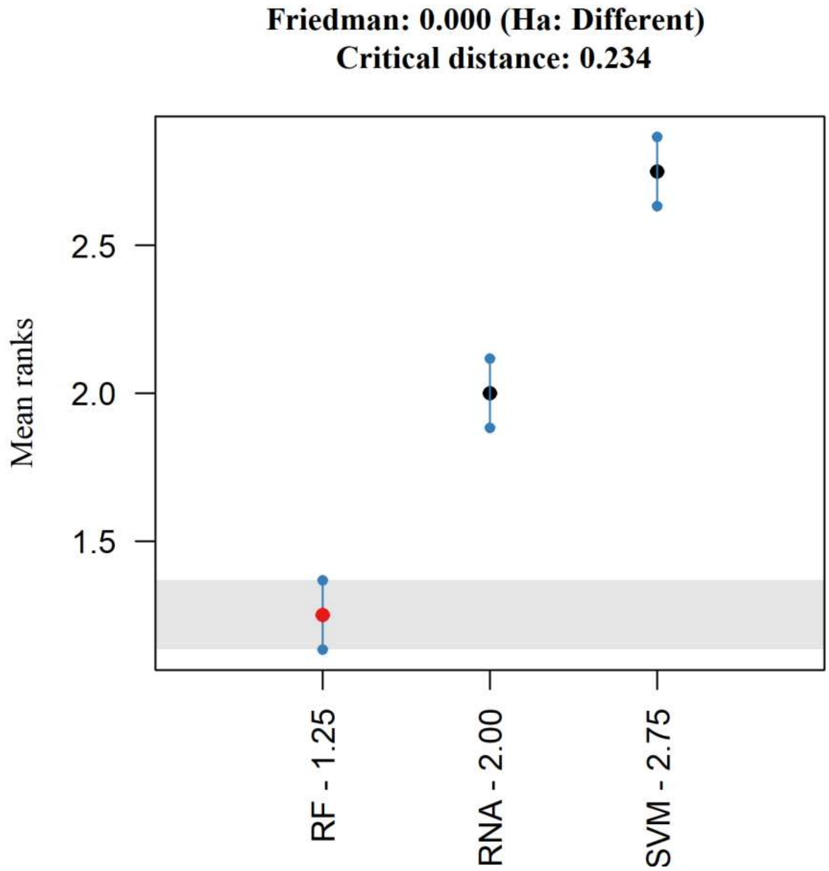

Figure 6. Nemenyi test of the machine learning models SVM, ANN and RF, in the modeling of the net growth in the in Atlantic Forest fragments in Minas Gerais, Brazil.

4. Discussion

Ongoing changes in environmental factors and disturbance regimes are consistently increasing mortality and forcing forests to have younger, smaller stands, reducing growth and carbon storage potential (McDowell et al., 2020). Understanding tropical forest dynamics and planning for their sustainable management require efficient, yet accurate, predictions of the joint dynamics of hundreds of tree species (Rüger et al., 2020). Growth models are essential for this. Using machine learning models, we showed that net growth in the basal area can be accurately predicted, using variable edaphic, anthropogenic and climatic variables.

We found that edaphic, forest attributes and climatic variables are important in shaping tree growth in the Brazilian Atlantic Forest. The different configurations of a forest occur because of natural and/or anthropic changes that take place in the structure and composition of the vegetation (Bezerra et al., 2021). The growth of individual trees is the result of the combined effects of several factors, such as age, tree size, microenvironment, genetic traits, and competitive status (Kunstler et al., 2012; Pretzsch, 2009; Weiskittel et al., 2011). Therefore, variables related to soils, climate, and effects are key factors in species growth and distribution (Martini et al., 2020).

In our study, soil acidity was the most important characteristic for shaping growth. It can have deleterious effects on forest ecosystems (Šantrůčková et al., 2019), by negatively affecting plant growth. Soil acidification impairs the long-term functioning of forest ecosystems by altering the availability of critical macro- and micronutrients in the soil (Desie et al., 2020; Schaberg et al., 2001).

The variables BA and age at abandonment were also important. The structural attributes and functional composition of the stand determine the growth of the trees and consequently the aboveground biomass (Manuel Villa et al., 2020). A study in the Atlantic Forest has already revealed that variables related to average tree size, i.e. basal area, are important for models in the biome (David et al., 2017). In general, tree size variables are essential components of growth modeling, because they express the competition between individuals. Basal area (BA) is an effective measure, since it incorporates the number of trees in a stand and their diameters, to express competition and aid in growth estimation (Weiskittel et al., 2011). The time of abandonment, on the other hand, influences the diversity of plants (Mangueira et al., 2021) and occurs due to changes in the forest succession stages and, consequently, affects tree growth.

Finally, we identify that precipitation is also associated with tree growth in the Atlantic Forest. Studies in the biome (David et al., 2017; Rocha et al., 2020) corroborated this assertion Potential increases in tree mortality associated with climate-induced physiological stress and interactions with other climate-mediated processes such as insect outbreaks and forest fires affect forest growth (Allen et al., 2010).

We found that machine learning methods were efficient and are important tools for modeling growth in forest fragments in the Brazilian Atlantic Forest. They can help in understanding the biome and in developing management strategies aimed at recovering biodiversity and reducing the deleterious effects of fragmentation

The Random Forest method showed superiority over the others for modeling growth in the Atlantic Forest. The metrics and the graphs of observed and estimates residuals used corroborate this statement. Random Forest is being increasingly used in ecological studies because it is suitable for the analysis of large complex data sets (Reise et al., 2019). This method has shown excellent results in forestry estimates (Freeman et al., 2015; Mascaro et al., 2014). As a non-parametric method, it benefits from its ability to take into account data variability and non-linear relationships Alternatively, parametric models are simpler and more widely known, and easier to share and explain (Freeman et al., 2015; Leite et al., 2020).

Machine learning models can generate promising results for tree growth estimates in tropical forests. Our study presents a new way to do this, especially in forest fragments of the Atlantic Forest. It is worth noting that new approaches, such as the inclusion of other variables and the use of other models (e.g., Cubist, Regression Trees Models, etc.) can improve the estimates. Studies of this kind should be encouraged and can help to better understand tree dynamics and assist in conservation practices in forests around the world, especially those threatened by human pressure and fragmentation.

5. Conclusion

Soil, forest attributes and climatic variables are important for modeling growth in the Brazilian Atlantic Forest. With the use of these variables machine learning models (RF, ANN, SVM) are promising in estimating the net growth in basal area in this biome. RF is the best algorithm to perform this task, as observed in our study area. Our results represent a new approach to accurately predict tree growth in Atlantic Forest fragments and can assist in biodiversity management and conservation strategies.

6. Referencias

Allen, C.D., Macalady, A.K., Chenchouni, H., Bachelet, D., McDowell, N., Vennetier, M., Kitzberger, T., Rigling, A., Breshears, D.D., Hogg, E.H. (Ted), Gonzalez, P., Fensham, R., Zhang, Z., Castro, J., Demidova, N., Lim, J.-H., Allard, G., Running, S.W., Semerci, A., Cobb, N., 2010. A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. Forest Ecology and Management, Adaptation of Forests and Forest Management to Changing Climate 259, 660–684. https://doi.org/10.1016/j.foreco.2009.09.001

Arroyo-Rodríguez, V., Melo, F.P.L., Martínez-Ramos, M., Bongers, F., Chazdon, R.L., Meave, J.A., Norden, N., Santos, B.A., Leal, I.R., Tabarelli, M., 2017. Multiple successional pathways in human-modified tropical landscapes: new insights from forest succession, forest fragmentation and landscape ecology research. Biological Reviews 92, 326–340. https://doi.org/10.1111/brv.12231

Bellard, C., Leclerc, C., Leroy, B., Bakkenes, M., Veloz, S., Thuiller, W., Courchamp, F., 2014. Vulnerability of biodiversity hotspots to global change. Global Ecology and Biogeography 23, 1376–1386. https://doi.org/10.1111/geb.12228

Bergmeir, C., Benítez, J.M., 2012. Neural Networks in R Using the Stuttgart Neural Network Simulator: RSNNS. Journal of Statistical Software 46, 1–26. https://doi.org/10.18637/jss.v046.i07

Bezerra, T.G., Ruschel, A.R., Emmert, F., Nascimento, R.G.M., 2021. Changes caused by forest logging in structure and floristic diversity of natural regeneration: Relationship between climate variables and forest dynamics in the eastern Amazon. Forest Ecology and Management 482, 118862. https://doi.org/10.1016/j.foreco.2020.118862

Bullock, J.M., Aronson, J., Newton, A.C., Pywell, R.F., Rey-Benayas, J.M., 2011. Restoration of ecosystem services and biodiversity: conflicts and opportunities. Trends in Ecology & Evolution 26, 541–549. https://doi.org/10.1016/j.tree.2011.06.011

Bustamante, M.M.C., Roitman, I., Aide, T.M., Alencar, A., Anderson, L.O., Aragão, L., Asner, G.P., Barlow, J., Berenguer, E., Chambers, J., Costa, M.H., Fanin, T., Ferreira, L.G., Ferreira, J., Keller, M., Magnusson, W.E., Morales-Barquero, L., Morton, D., Ometto, J.P.H.B., Palace, M., Peres, C.A., Silvério, D., Trumbore, S., Vieira, I.C.G., 2016. Toward an integrated monitoring framework to assess the effects of tropical forest degradation and recovery on carbon stocks and biodiversity. Glob Chang Biol 22, 92–109. https://doi.org/10.1111/gcb.13087

Câmara, T., Almeida, W.R., Tabarelli, M., Andersen, A.N., Leal, I.R., 2017. Habitat fragmentation, EFN-bearing trees and ant communities: Ecological cascades in Atlantic Forest of northeastern Brazil. Austral Ecology 42, 31–39. https://doi.org/10.1111/aec.12393

Carreño-Rocabado, G., Peña-Claros, M., Bongers, F., Alarcón, A., Licona, J.-C., Poorter, L., 2012. Effects of disturbance intensity on species and functional diversity in a tropical forest. Journal of Ecology 100, 1453–1463. https://doi.org/10.1111/j.1365-2745.2012.02015.x

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M.S., Delitti, W.B.C., Duque, A., Eid, T., Fearnside, P.M., Goodman, R.C., Henry, M., Martínez-Yrízar, A., Mugasha, W.A., Muller-Landau, H.C., Mencuccini, M., Nelson, B.W., Ngomanda, A., Nogueira, E.M., Ortiz-Malavassi, E., Pélissier, R., Ploton, P., Ryan, C.M., Saldarriaga, J.G., Vieilledent, G., 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. Global Change Biology 20, 3177–3190. https://doi.org/10.1111/gcb.12629

David, H.C., Araújo, E.J.G. de, Morais, V.A., Scolforo, J.R.S., Marques, J.M., Péllico Netto, S., MacFarlane, D.W., 2017. Carbon stock classification for tropical forests in Brazil: Understanding the effect of stand and climate variables. Forest Ecology and Management 404, 241–250. https://doi.org/10.1016/j.foreco.2017.08.044

de Lima, R.A.F., Mori, D.P., Pitta, G., Melito, M.O., Bello, C., Magnago, L.F., Zwiener, V.P., Saraiva, D.D., Marques, M.C.M., de Oliveira, A.A., Prado, P.I., 2015. How much do we know about the endangered Atlantic Forest? Reviewing nearly 70 years of information on tree community surveys. Biodivers Conserv 24, 2135–2148. https://doi.org/10.1007/s10531-015-0953-1

Desie, E., Vancampenhout, K., van den Berg, L., Nyssen, B., Weijters, M., den Ouden, J., Muys, B., 2020. Litter share and clay content determine soil restoration effects of rich litter tree species in forests on acidified sandy soils. Forest Ecology and Management 474, 118377. https://doi.org/10.1016/j.foreco.2020.118377

Emer, C., Galetti, M., Pizo, M.A., Guimarães, P.R., Moraes, S., Piratelli, A., Jordano, P., 2018. Seed-dispersal interactions in fragmented landscapes – a metanetwork approach. Ecology Letters 21, 484–493. https://doi.org/10.1111/ele.12909

ESRI, 2015. ArcGIS Desktop: Release 10.3.1. Environmental Systems Research Institute, Redlands, CA.

Freeman, E.A., Moisen, G.G., Coulston, J.W., Wilson, B.T., 2015. Random forests and stochastic gradient boosting for predicting tree canopy cover: comparing tuning processes and model performance1. Canadian Journal of Forest Research. https://doi.org/10.1139/cjfr-2014-0562

Gleason, C.J., Im, J., 2012. Forest biomass estimation from airborne LiDAR data using machine learning approaches. Remote Sensing of Environment 125, 80–91. https://doi.org/10.1016/j.rse.2012.07.006

Gomes, L.C., Faria, R.M., de Souza, E., Veloso, G.V., Schaefer, C.E.G.R., Filho, E.I.F., 2019. Modelling and mapping soil organic carbon stocks in Brazil. Geoderma 340, 337–350. https://doi.org/10.1016/j.geoderma.2019.01.007

Haddad, N.M., Brudvig, L.A., Clobert, J., Davies, K.F., Gonzalez, A., Holt, R.D., Lovejoy, T.E., Sexton, J.O., Austin, M.P., Collins, C.D., Cook, W.M., Damschen, E.I., Ewers, R.M., Foster, B.L., Jenkins, C.N., King, A.J., Laurance, W.F., Levey, D.J., Margules, C.R., Melbourne, B.A., Nicholls, A.O., Orrock, J.L., Song, D.-X., Townshend, J.R., 2015. Habitat fragmentation and its lasting impact on Earth's ecosystems. Science Advances 1, e1500052. https://doi.org/10.1126/sciadv.1500052

Huy, B., Canh Nam, L., Poudel, K.P., Temesgen, H., 2021. Individual tree diameter growth modeling system for Dalat pine (Pinus dalatensis Ferré) of the upland mixed tropical forests. Forest Ecology and Management 480, 118612. https://doi.org/10.1016/j.foreco.2020.118612

IBGE, 2012. Manual Técnico da Vegetação Brasileira, 2ª. ed. IBGE, Rio de Janeiro.

Jachowski, N.R.A., Quak, M.S.Y., Friess, D.A., Duangnamon, D., Webb, E.L., Ziegler, A.D., 2013. Mangrove biomass estimation in Southwest Thailand using machine learning. Applied Geography 45, 311–321. https://doi.org/10.1016/j.apgeog.2013.09.024

Joly, C.A., Metzger, J.P., Tabarelli, M., 2014. Experiences from the Brazilian Atlantic Forest: ecological findings and conservation initiatives. New Phytologist 204, 459–473. https://doi.org/10.1111/nph.12989

Kuhn, M., Johnson, K., 2013a. Applied Predictive Modeling; Springer: New York, NY, USA,.

Kuhn, M., Johnson, K., 2013b. An Introduction to Feature Selection, in: Kuhn, M., Johnson, K. (Eds.), Applied Predictive Modeling. Springer, New York, NY, pp. 487–519. https://doi.org/10.1007/978-1-4614-6849-3_19

Kunstler, G., Lavergne, S., Courbaud, B., Thuiller, W., Vieilledent, G., Zimmermann, N.E., Kattge, J., Coomes, D.A., 2012. Competitive interactions between forest trees are driven by species' trait hierarchy, not phylogenetic or functional similarity: implications for forest community assembly. Ecology Letters 15, 831–840. https://doi.org/10.1111/j.1461-0248.2012.01803.x

Leite, R.V., Amaral, C.H. do, Pires, R. de P., Silva, C.A., Soares, C.P.B., Macedo, R.P., Silva, A.A.L. da, Broadbent, E.N., Mohan, M., Leite, H.G., 2020. Estimating Stem Volume in Eucalyptus Plantations Using Airborne LiDAR: A Comparison of Area- and Individual Tree-Based Approaches. Remote Sensing 12, 1513. https://doi.org/10.3390/rs12091513

Lewis, S.L., Edwards, D.P., Galbraith, D., 2015. Increasing human dominance of tropical forests. Science 349, 827–832. https://doi.org/10.1126/science.aaa9932

Lutz, J.A., Wagtendonk, J.W. van, Franklin, J.F., 2010. Climatic water deficit, tree species ranges, and climate change in Yosemite National Park. Journal of Biogeography 37, 936–950. https://doi.org/10.1111/j.1365-2699.2009.02268.x

Ma, W., Lei, X., 2015. Nonlinear Simultaneous Equations for Individual-Tree Diameter Growth and Mortality Model of Natural Mongolian Oak Forests in Northeast China. Forests 6, 2261–2280. https://doi.org/10.3390/f6062261

Magnago, L.F.S., Edwards, D.P., Edwards, F.A., Magrach, A., Martins, S.V., Laurance, W.F., 2014. Functional attributes change but functional richness is unchanged after fragmentation of Brazilian Atlantic forests. Journal of Ecology 102, 475–485. https://doi.org/10.1111/1365-2745.12206

Mangueira, J.R.S.A., Vieira, L.T.A., Azevedo, T.N., Sabino, A.P.S., Ferraz, K.M.P.M.B., Ferraz, S.F.B., Rother, D.C., Rodrigues, R.R., 2021. Plant diversity conservation in highly deforested landscapes of the Brazilian Atlantic Forest. Perspectives in Ecology and Conservation 19, 69–80. https://doi.org/10.1016/j.pecon.2020.12.003

Manuel Villa, P., Ali, A., Venâncio Martins, S., Nolasco de Oliveira Neto, S., Cristina Rodrigues, A., Teshome, M., Alvim Carvalho, F., Heringer, G., Gastauer, M., 2020. Stand structural attributes and functional trait composition overrule the effects of functional divergence on aboveground biomass during Amazon forest succession. Forest Ecology and Management 477, 118481. https://doi.org/10.1016/j.foreco.2020.118481

MapBiomas Project, 2019. Collection V3.1 of Brazilian Land Cover & Use Map Series [WWW Document]. URL http://mapbiomas.org (accessed 8.15.19).

Martini, F., Xia, S.-W., Zou, C., Goodale, U.M., 2020. Seedling growth and survival responses to multiple soil properties in subtropical forests of south China. Forest Ecology and Management 474, 118382. https://doi.org/10.1016/j.foreco.2020.118382

Mascaro, J., Asner, G.P., Knapp, D.E., Kennedy-Bowdoin, T., Martin, R.E., Anderson, C., Higgins, M., Chadwick, K.D., 2014. A Tale of Two "Forests": Random Forest Machine Learning Aids Tropical Forest Carbon Mapping. PLOS ONE 9, e85993. https://doi.org/10.1371/journal.pone.0085993

McDowell, N.G., Allen, C.D., Anderson-Teixeira, K., Aukema, B.H., Bond-Lamberty, B., Chini, L., Clark, J.S., Dietze, M., Grossiord, C., Hanbury-Brown, A., Hurtt, G.C., Jackson, R.B., Johnson, D.J., Kueppers, L., Lichstein, J.W., Ogle, K., Poulter, B., Pugh, T.A.M., Seidl, R., Turner, M.G., Uriarte, M., Walker, A.P., Xu, C., 2020. Pervasive shifts in forest dynamics in a changing world. Science 368. https://doi.org/10.1126/science.aaz9463

Melo, F.P.L., Pinto, S.R.R., Brancalion, P.H.S., Castro, P.S., Rodrigues, R.R., Aronson, J., Tabarelli, M., 2013. Priority setting for scaling-up tropical forest restoration projects: Early lessons from the Atlantic Forest Restoration Pact. Environmental Science & Policy 33, 395–404. https://doi.org/10.1016/j.envsci.2013.07.013

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Kent, J., 2000. Biodiversity hotspots for conservation priorities. Nature 403, 853–858. https://doi.org/10.1038/35002501

Poorter, L., Sande, M.T. van der, Arets, E.J.M.M., Ascarrunz, N., Enquist, B.J., Finegan, B., Licona, J.C., Martínez-Ramos, M., Mazzei, L., Meave, J.A., Muñoz, R., Nytch, C.J., Oliveira, A.A. de, Pérez-García, E.A., Prado-Junior, J., Rodríguez-Velázques, J., Ruschel, A.R., Salgado-Negret, B., Schiavini, I., Swenson, N.G., Tenorio, E.A., Thompson, J., Toledo, M., Uriarte, M., Hout, P. van der, Zimmerman, J.K., Peña-Claros, M., 2017. Biodiversity and climate determine the functioning of Neotropical forests. Global Ecology and Biogeography 26, 1423–1434. https://doi.org/10.1111/geb.12668

Pretzsch, H., 2009. Forest Dynamics, Growth and Yield: From Measurement to Model. Springer-Verlag, Berlin Heidelberg. https://doi.org/10.1007/978-3-540-88307-4

Redmond, M.D., 2019. CWD and AET function.

Reis, L.P., de Souza, A.L., dos Reis, P.C.M., Mazzei, L., Soares, C.P.B., Miquelino Eleto Torres, C.M., da Silva, L.F., Ruschel, A.R., Rêgo, L.J.S., Leite, H.G., 2018. Estimation of mortality and survival of individual trees after harvesting wood using artificial neural networks in the amazon rain forest. Ecological Engineering 112, 140–147. https://doi.org/10.1016/j.ecoleng.2017.12.014

Reis, L.P., de Souza, A.L., Mazzei, L., dos Reis, P.C.M., Leite, H.G., Soares, C.P.B., Torres, C.M.M.E., da Silva, L.F., Ruschel, A.R., 2016. Prognosis on the diameter of individual trees on the eastern region of the amazon using artificial neural networks. Forest Ecology and Management 382, 161–167. https://doi.org/10.1016/j.foreco.2016.10.022

Reise, J., Kukulka, F., Flade, M., Winter, S., 2019. Characterising the richness and diversity of forest bird species using National Forest Inventory data in Germany. Forest Ecology and Management 432, 799–811. https://doi.org/10.1016/j.foreco.2018.10.012

Rocha, S.J.S.S. da, Torres, C.M.M.E., Jacovine, L.A.G., Leite, H.G., Gelcer, E.M., Neves, K.M., Schettini, B.L.S., Villanova, P.H., Silva, L.F. da, Reis, L.P., Zanuncio, J.C., 2018. Artificial neural networks: Modeling tree survival and mortality in the Atlantic Forest biome in Brazil. Science of The Total Environment 645, 655–661. https://doi.org/10.1016/j.scitotenv.2018.07.123

Rocha, S.J.S.S. da, Torres, C.M.M.E., Villanova, P.H., Schettini, B.L.S., Jacovine, L.A.G., Leite, H.G., Gelcer, E.M., Reis, L.P., Neves, K.M., Comini, I.B., Silva, L.F. da, 2020. Drought effects on carbon dynamics of trees in a secondary Atlantic Forest. Forest Ecology and Management 465, 118097. https://doi.org/10.1016/j.foreco.2020.118097

Rozendaal, D.M.A., Chazdon, R.L., 2015. Demographic drivers of tree biomass change during secondary succession in northeastern Costa Rica. Ecological Applications 25, 506–516. https://doi.org/10.1890/14-0054.1

Rüger, N., Condit, R., Dent, D.H., DeWalt, S.J., Hubbell, S.P., Lichstein, J.W., Lopez, O.R., Wirth, C., Farrior, C.E., 2020. Demographic trade-offs predict tropical forest dynamics. Science 368, 165–168. https://doi.org/10.1126/science.aaz4797

Šantrůčková, H., Cienciala, E., Kaňa, J., Kopáček, J., 2019. The chemical composition of forest soils and their degree of acidity in Central Europe. Science of The Total Environment 687, 96–103. https://doi.org/10.1016/j.scitotenv.2019.06.078

Scarano, F.R., Ceotto, P., 2015. Brazilian Atlantic forest: impact, vulnerability, and adaptation to climate change. Biodivers Conserv 24, 2319–2331. https://doi.org/10.1007/s10531-015-0972-y

Schaberg, P.G., DeHayes, D.H., Hawley, G.J., 2001. Anthropogenic Calcium Depletion: A Unique Threat to Forest Ecosystem Health? Ecosystem Health 7, 214–228. https://doi.org/10.1046/j.1526-0992.2001.01046.x

Schmidt, M., Jochheim, H., Kersebaum, K.-C., Lischeid, G., Nendel, C., 2017. Gradients of microclimate, carbon and nitrogen in transition zones of fragmented landscapes – a review. Agricultural and Forest Meteorology 232, 659–671. https://doi.org/10.1016/j.agrformet.2016.10.022

Silva, C.A., Klauberg, C., Hudak, A.T., Vierling, L.A., Liesenberg, V., Carvalho, S.P.C. e, Rodriguez, L.C.E., 2016. A principal component approach for predicting the stem volume in Eucalyptus plantations in Brazil using airborne LiDAR data. Forestry: An International Journal of Forest Research 89, 422–433. https://doi.org/10.1093/forestry/cpw016

Silva, I.N., Spatti, D.H., Flauzin, R.A., 2010. Redes neurais artificiais: para engenharia e ciências aplicadas. São Paulo: Artliber, 396 p.

SOS Mata Atlântica, 2019. Relatório Anual 2019.

Stephenson, N., 1998. Actual evapotranspiration and deficit: biologically meaningful correlates of vegetation distribution across spatial scales. Journal of Biogeography 25, 855–870. https://doi.org/10.1046/j.1365-2699.1998.00233.x

Tavares Júnior, I. da S., Torres, C.M.M.E., Leite, H.G., Castro, N.L.M. de, Soares, C.P.B., Castro, R.V.O., Farias, A.A., 2020. Machine learning: Modeling increment in diameter of individual trees on Atlantic Forest fragments. Ecological Indicators 117, 106685. https://doi.org/10.1016/j.ecolind.2020.106685

Teixeira, P.C., Donagemma, G.K., Fontana, A., Teixeira, W.G. (Eds.), 2017. Manual de métodos de análise de solo, 3ª. ed. Embrapa, Brasília, Brazil.

Uzoh, F.C.C., Oliver, W.W., 2008. Individual tree diameter increment model for managed even-aged stands of ponderosa pine throughout the western United States using a multilevel linear mixed effects model. Forest Ecol. Manage. 256 (3), 438–445. 256, 438–445.

van der Sande, M.T., Peña-Claros, M., Ascarrunz, N., Arets, E.J.M.M., Licona, J.C., Toledo, M., Poorter, L., 2017. Abiotic and biotic drivers of biomass change in a Neotropical forest. Journal of Ecology 105, 1223–1234. https://doi.org/10.1111/1365-2745.12756

Vanclay, J., 1994. Modelling Forest Growth and Yield: Applications to Mixed Tropical Forests.

Weiskittel, A.R., Hann, D.W., Kershaw, J.A., Vanclay, J.K., 2011. Forest Growth and Yield Modeling, 1ª edição. ed. Wiley.

Xu, X., Medvigy, D., Powers, J.S., Becknell, J.M., Guan, K., 2016. Diversity in plant hydraulic traits explains seasonal and inter-annual variations of vegetation dynamics in seasonally dry tropical forests. New Phytologist 212, 80–95. https://doi.org/10.1111/nph.14009

Zhao, K., Popescu, S., Meng, X., Pang, Y., Agca, M., 2011. Characterizing forest canopy structure with lidar composite metrics and machine learning. Remote Sensing of Environment 115, 1978–1996. https://doi.org/10.1016/j.rse.2011.04.001

CAPITULO IV: Volume and biomass estimates of commercial trees in the Amazon forest using machine learning

Abstract
The accuracy of the volume and above-ground biomass estimation of exploitable trees by the practice of selective logging is essential for the elaboration of a sustainable management plan. The objective of the study is using machine learning models capable of estimating volume and biomass in commercial trees in Southwestern Amazon. The study was carried out in the Southwestern Amazon, in the municipality of Porto Acre, Acre state, Brazil. Determining volume and biomass of sample trees was performed using dendrometric, climatic and topographic variables. The Boruta Algorithm was applied to select the best set of variables. Support Vector Machines (SVM), Artificial Neural Networks (ANN), Random Forests (RF) and Generalized Linear Model (GLM) were the machine learning methods evaluated. In general, the evaluated methods showed a satisfactory generalization power. The results showed that the volume and biomass predictions of commercial trees in the Amazon rainforest differed between the techniques ($p < 0.05$). ANNs showed the best performances to predict the volume and biomass of commercial trees, with the highest $r_{y\hat{y}}$ and the lowest RSME and MAE. In this way, machine learning methods such as SVM, ANN, RF and GLM are useful and efficient tools for estimating volume and biomass of commercial trees in the Amazon rainforest. These methods can be useful tools to improve the accuracy of estimates in forest management plans.

Keywords: allometry; dense rainforest; models.

1. Introduction

Amazon rainforest presents one of the greatest biodiversity on the planet (Andrade et al., 2019; Luize et al., 2018; Steege et al., 2016), covering more than a third of all the diversity of neotropical plants (Antonelli and Sanmartín, 2011) and housing between 6,700-16,000 species of trees (Steege et al., 2013; Cardoso et al., 2017). In addition, the Amazon have an important role in the carbon cycle, storing around 150–200 Pg of carbon in biomass and living things (Brienen et al., 2015).

The Brazilian Amazon rainforest is the largest remnant of the forest (Hansen et al., 2013). In this biome, selective logging, which includes cutting of individuals from timber tree species, is one of the most common land uses (Gaui et al., 2019). This practice, when associated with correct planning and the use of low impact logging techniques, reduces the damage caused by logging to the ecosystem and contributes to the conservation of global biodiversity (Peña-Claros et al., 2008; Andrade et al., 2019; Chaudhary et al., 2016).

The exact quantification of biomass stocks and the above-ground volume of exploitable trees by the practice of selective logging is essential for the elaboration of an effectively sustainable management plan. The sustainability of forest management is associated with the continuous production of wood in the future in similar quantity and quality (Fortini, 2019; Putz et al., 2008). In addition, quantifying these components of forest production is important for understanding the role of these ecosystems in the global carbon cycle (Goodman et al., 2014a) and for the successful implementation of climate change mitigation policies (Chave et al., 2014).

Researches on allometric equations in the Amazon region have already been carried out (Alvarez et al., 2012; Chambers et al., 2001; Chave et al., 2005, 2014; Cummings et al., 2002; Goodman et al., 2013, 2014a; Keller et al., 2001; Lima et al., 2012; Nelson et al., 1999; Nogueira et al., 2008; Vieilledent et al., 2012). However, studies involving machine learning models are rare in the region, which are more common for even-aged forests (*e.g.* Domingues et al., 2020; Souza et al., 2019). Machine learning models are a rapidly growing area of study and have been increasingly used for modeling in tropical forests because they can generate more accurate estimates than traditional data modeling approaches (Gleason and Im, 2012; Jachowski et al., 2013; Zhao et al., 2011).

This superiority in the generation of estimates is associated with a lower number of assumptions about data and processes (Diamantopoulou and Milios, 2010), which allows the generation of better results in prediction, in view of the complex relationships of forest dynamics. In view of the above, this study aims to develop models of machine learning capable of estimating the volume and biomass in commercial trees in Southwestern Amazon, based on dendrometric, climatic and topographic characteristics. The research questions of this study are: (i) Are the machine learning methods evaluated efficient for estimating the volume and biomass of commercial trees?; (ii) What is the best method to estimate the volume of commercial trees?; and (iii) What is the best method to estimate the biomass of commercial trees?

2. Material and Methods

2.1 Characterization of the study area

The study was carried out at Antimary Farm I and II, located in the Southwestern Amazon, in the municipality of Porto Acre, Acre, Brazil (Figure 1). The area under sustainable management comprises 1,253.02 ha. The region's vegetation is classified as a “terra firme” - forest with solid ground- and wetland rainforest (Alencar et al., 1979). The climate of the region is of the Am type, according to the Köppen classification (Alvares et al., 2013). The study area presents two types of soil, Red Argisol and Dystrophic Red Yellow Latosol (Acre, 2010). Topography is predominantly flat, with a slope of around 5%. Altimetry varies between 220 to 300 m above the sea level.

A census was conducted in the exploitable area in May 2015 and the Sustainable Forest Management Plan (SFMP) was approved in 2016 by the Acre Environment Institute (*Instituto de Meio Ambiente do Acre*, IMAC).

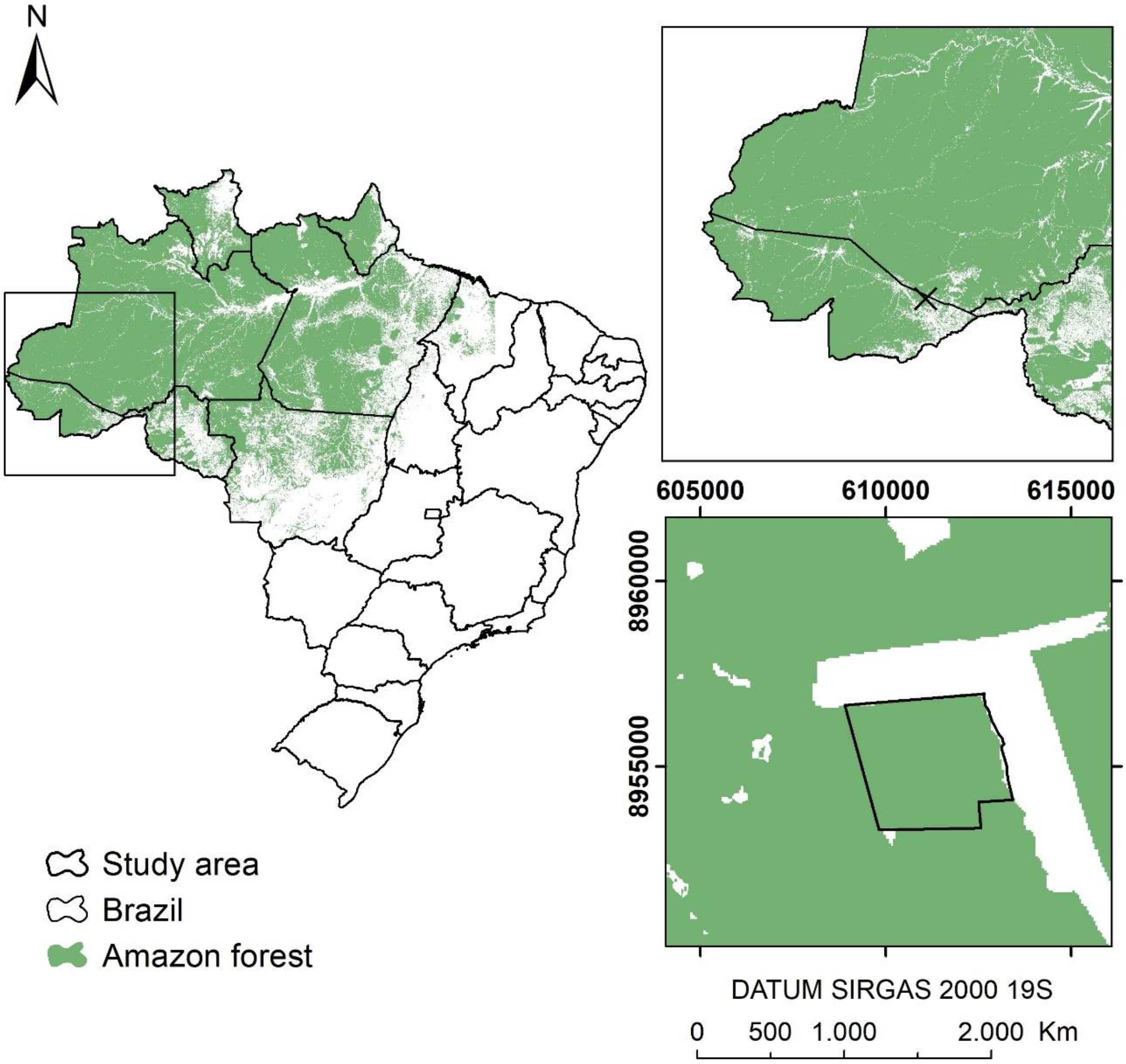

Figure 1 – Location of the study area in the Southwestern Amazon, in the municipality of Porto Acre, Acre, Brazil.

2.2 Determination of volume and biomass stocks

Sample trees were selected based on density and basal area, obtained from information from the census provided by the company responsible for management, in which all trees of commercial interest with a diameter at breast height (DBH; 1.30 m) ≥ 50 cm were measured. Eighteen species of highest importance value were selected, distributed in 214 individuals.

The values of the volume and biomass stocks in the study area were obtained by Romero (2018). The volume of the selected individuals was determined by strict cubing using the method of Smalian (Husch et al., 2003). Wood discs from the base of the logs were collected to determine the basic wood density according to the ABNT standard (2003). Biomass was calculated by multiplying the volume and the basic density of the wood (Chave et al., 2005). The average basic wood density of trees analyzed was 0.59 g cm^{-3} (Table 1).

Table 1 – Number of trees and basic wood density of commercial tree species present in Southwestern Amazon, in the municipality of Porto Acre, Acre, Brazil

SN	F	N	Bd
Albizia niopoides (Spruce ex Benth.) Burkart	Fabaceae Lindl.	7	0.64 ± 0.03
Apuleia leiocarpa (Vogel) J.F.Macbr.	Fabaceae Lindl.	13	0.77 ± 0.03
Astronium lecointei Ducke	Anacardiaceae R.Br.	6	0.82 ± 0.05
Barnebydendron riedelii (Tul.) J.H.Kirkbr.	Fabaceae Lindl.	5	0.57 ± 0.03
Buchenavia tetraphylla (Aubl.) R.A.Howard	Combretaceae R.Br.	9	0.69 ± 0.04
Castilla ulei Warb.	Moraceae Gaudich.	37	0.41 ± 0.04
Cedrela odorata L.	Meliaceae A.Juss.	8	0.43 ± 0.04
Ceiba pentandra (L.) Gaertn.	Malvaceae Juss.	4	0.29 ± 0.03
Ceiba samauma (Mart.) K.Schum.	Malvaceae Juss.	22	0.51 ± 0.05
Copaifera multijuga Hayne	Fabaceae Lindl.	6	0.52 ± 0.05
Dipteryx odorata (Aubl.) Willd.	Fabaceae Lindl.	11	0.80 ± 0.04
Eschweilera bracteosa (Poepp. ex O.Berg) Miers	Lecythidaceae A.Rich.	15	0.65 ± 0.05
Eschweilera grandiflora (Aubl.) Sandwith	Lecythidaceae A.Rich.	13	0.73 ± 0.03
Handroanthus serratifolius (Vahl) S.Grose	Bignoniaceae Juss.	8	0.82 ± 0.04
Hura crepitans L.	Euphorbiaceae Juss.	6	0.36 ± 0.06
Hymenaea courbaril L.	Fabaceae Lindl.	8	0.76 ± 0.04
Parkia paraensis Ducke	Fabaceae Lindl.	20	0.46 ± 0.06
Schizolobium parahyba var. *amazonicum* (Huber ex Ducke) Barneby	Fabaceae Lindl.	16	0.48 ± 0.08
$\bar{X}$ ± CI		11.89 ± 8.12	0.59 ± 0.17

Where: SN = Scientific name; F = family; N = number of individuals; Bd = Basic wood density, in g cm^{-3}; $\bar{X}$ = Mean; CI = confidence interval.

2.3 Predictor variables for modeling

For modeling, dendrometric, qualitative, climatic and topographic variables (predictor variables) were used to estimate volume and biomass stocks (response variables). Dendrometric variables used were: DBH, commercial height (Ch) and basic wood density (Bd) (Romero, 2018). Qualitative variables were the species and the family of the individuals (Romero, 2018). Bioclimatic variables used (Bio 1-19) are derived from the monthly values of temperature and precipitation and were obtained from the WorldClim - Global Climate Data database (Fick and Hijmans, 2017), with a spatial resolution of approximately 1 km^2. The Bio 5 climatic variable was not used because it did not show variability in the study area. The topographic variable used was altitude (Table 2).

Table 2 – Predictor variables used in the modeling of volume and biomass in the Southwestern Amazon, in the municipality of Porto Acre, Acre, Brazil

Variable	Min	1° Quartil	Median	Mean	3° Quartil	Max	SD
DBH	50.38	64.78	75.44	79.60	89.52	149.92	20.14
Ch	7.30	11.71	14.20	14.82	17.87	25.40	4.03
Bd	0.29	0.43	14.20	0.57	0.73	0.82	0.15
Bio1	24.86	24.88	24.88	24.88	24.89	24.92	0.01
Bio2	11.45	11.51	11.52	11.51	11.52	11.53	0.01
Bio3	81.21	81.62	81.68	81.65	81.68	82.02	0.09
Bio4	82.28	83.38	83.42	83.49	83.63	84.39	0.25
Bio6	17.30	17.30	17.30	17.30	17.30	17.40	0.01
Bio7	14.00	14.10	14.10	14.10	14.10	14.10	0.01
Bio8	25.32	25.32	25.33	25.33	25.35	25.38	0.02
Bio9	23.63	23.67	23.67	23.67	23.68	23.73	0.02
Bio10	25.55	25.57	25.57	25.57	25.58	25.60	0.01
Bio11	23.63	23.67	23.67	23.67	23.68	23.73	0.02
Bio12	1830.00	1834.00	1836.00	1836.31	1839.00	1853.00	3.98
Bio13	250.00	251.00	252.00	251.61	252.00	254.00	1.03
Bio14	40.00	40.00	41.00	40.70	41.00	41.00	0.46
Bio15	51.47	51.58	51.77	51.74	51.77	52.05	0.17
Bio16	735.00	736.00	738.00	737.61	738.00	744.00	2.09
Bio17	154.00	155.00	155.00	155.29	156.00	158.00	0.91
Bio18	568.00	570.00	570.00	570.27	571.00	576.00	1.69
Bio19	154.00	155.00	155.00	156.82	156.00	198.00	8.01
Alt	151.05	164.99	173.16	175.08	183.08	248.87	13.75

Where: Min = minimum value; Max = maximum value; SD = standard deviation; DBH = diameter at breast height, in cm; Ch = commercial height, in m; Db = basic wood density; in g cm^{-3}; Bio1 = annual mean temperature (°C); Bio2 = mean diurnal range (°C); Bio3 = Isothermality (%); Bio4 = temperature seasonality; Bio5 = max temperature of warmest month (°C); Bio6 = min temperature of coldest month (°C); Bio7 = temperature annual range (°C); Bio8 = mean temperature of wettest quarter (°C); Bio9 = mean temperature of driest quarter (°C); Bio10 = mean temperature of warmest quarter (°C); Bio11 = mean temperature of coldest quarter (°C); Bio12 = annual precipitation (mm); Bio13 = precipitation of wettest month (mm); Bio14 = precipitation of driest month (mm); Bio15 = precipitation seasonality (Coefficient of Variation) (mm); Bio16 = precipitation of wettest quarter (mm); Bio17 = precipitation of driest quarter (mm); Bio18 = precipitation of warmest quarter (mm); Bio19 = precipitation of coldest quarter (mm); Alt = altitude.

The Boruta Algorithm (Kursa and Rudnicki, 2010) was applied to select the best set of predictor variables to estimate volume and biomass. This algorithm iteratively removes the resources that are proven by a statistical test to be less relevant than random probes (Kursa and Rudnicki, 2010). The Boruta *R* Software Package was used.

Quantitative variables were standardized to accelerate the convergence rate and reduce the iteration process in training:

$$Z_i = (x_i - \bar{x})/\sigma$$

where:

Z_i = standardized value of the *i*-th observation;

x_i = value of the *i*-th observation;

$\bar{x}$ = average of the observed values;

σ = standard deviation.

The scale function of *R* Software was used in this step.

2.4 Model evaluation

The tested models to estimate the volume and biomass were: Support Vector Machines (SVM), Artificial Neural Networks (ANN), *Random Forests* (RF) and Generalized Linear Model (GLM).

The trained ANNs were of the multilayer *perceptron* type, also known as *multilayer perceptron* (MLP). The typical MLP architecture consists of an input layer containing the predictor variables, one or more hidden layers and an output layer containing the predicted variable. The activation function used was logistics. The training algorithms used were resilient propagation. The ANNs were implemented with the MLP function of the "RSNNS" Package in *R*.

The function svm of the "e1071" Package on *R* was used for training SVMs. The Kernel function was of the linear type. The randomForest function of the package of the same name in

R was used for RF training. The glm function and link function of identity type and Gaussian family were used for GLM.

The performance of the models in the estimation of volume and biomass was assessed using the *k*–fold cross-validation method, with the data divided into 5 folds (4 for adjustments/training and 1 for validation). At each adjustment/training of the 5 folds the metrics of Root Mean Square Error – RMSE (Eq. 1) and mean absolute error – MAE (Eq. 2) were calculated. This process was repeated 50 times, obtaining the average of the metrics for comparison of all models. The data were selected randomly in each of the 50 repetitions, resulting in different data sets, for greater robustness of the evaluation.

$$RMSE = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\frac{\sum_{i=1}^{n}(X_i - \hat{X}_i)^2}{n}} \quad (1)$$

$$MAE = \frac{1}{R}\sum_{r=1}^{R}\frac{\left|\sum_{i=1}^{n}(X_i - \hat{X}_i)\right|}{n} \quad (2)$$

Where:

R = number of repetitions (50);

n = number of observations;

H_i = observed variable from the *i*-th tree, in m;

$\hat{H}_i$ = estimated variable of the *i*-th tree, in m.

The averages of RMSE and MAE of each method in each repetition were ranked with weight assignments from 1 to 4, with 1 for the lowest value and 4 for the highest value. The weight p_{ij} assigned to the model m_j for the mean of RMSE was added to the weight p_{ij} assigned to the same model m_j for the mean of MAE, with i = 1, 2, ..., 50. With the result of these sums, the values were submitted to the Friedman – Nemenyi test, at the 5% significance level (Eq. 3).

$$CD = q_\alpha\sqrt{\frac{k(k+1)}{6N}} \quad (3)$$

Where:

CD = critical difference;

q_α = critical value calculated based on Studentized interval statistics (Harter, 1960) divided by √2;

k = number of algorithms being compared;

N = number of data sets.

3. Results

3.1. Selection of variables

The applied variable selection procedure allowed the choice of the best model based on the ideal subset of variables. For that, the variables selected to estimate the volume were DAP, Ch, species and family (Figure 2). In addition to the predictor variables mentioned for the volume, the wood basic density variable was added, for the biomass prediction.

The bioclimatic variables and the topographic variable altitude were not considered significant for the modeling of volume and biomass based on the Boruta algorithm (Figure 2).

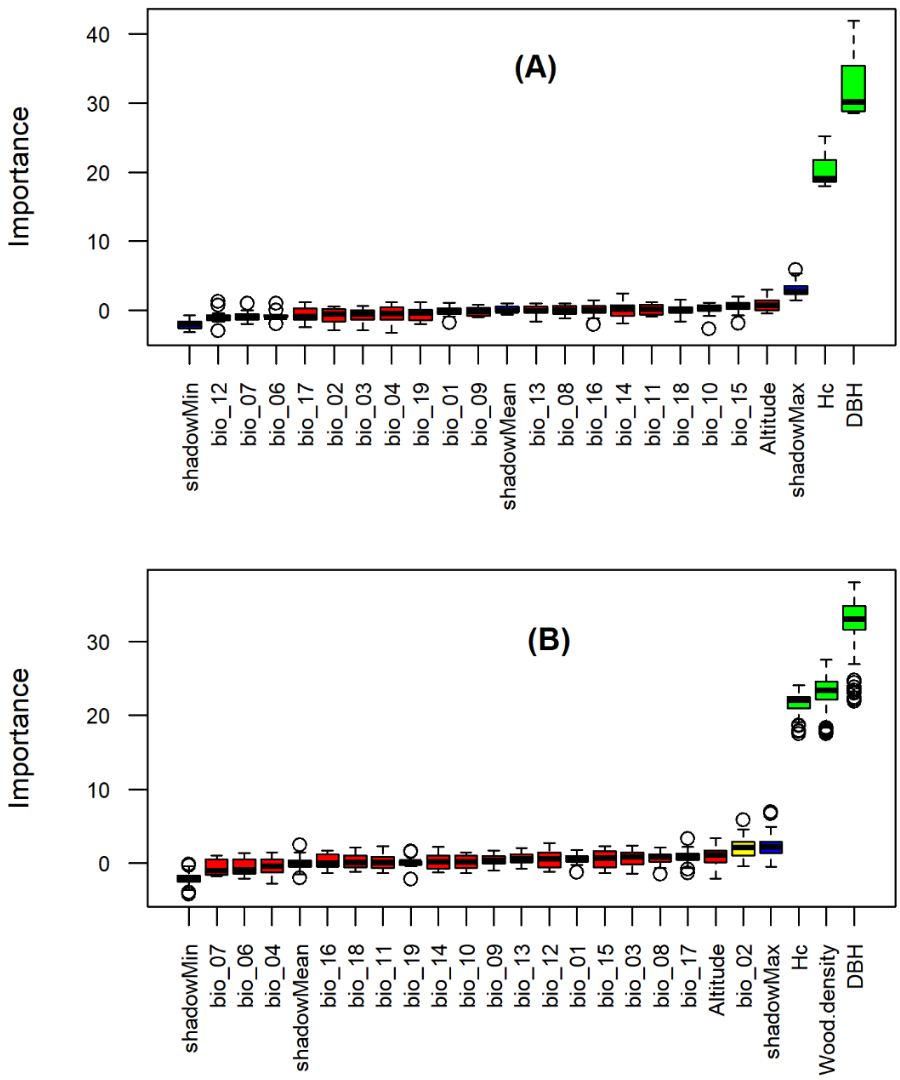

Figure 2 – Most important variables for modeling the volume (A) and biomass (B) of commercial trees in Southwestern Amazon, in the municipality of Porto Acre, Acre, Brazil.

3.2 Model performance

In general, the evaluated models showed a satisfactory generalization power, indicated by similar precision results between the observed and estimated data in the validation for all variables studied (Figure 3). Pearson's correlation coefficient ($r_{y\hat{y}}$) between the estimated and observed volume and biomass data was greater than 0.85 in all applied machine learning models.

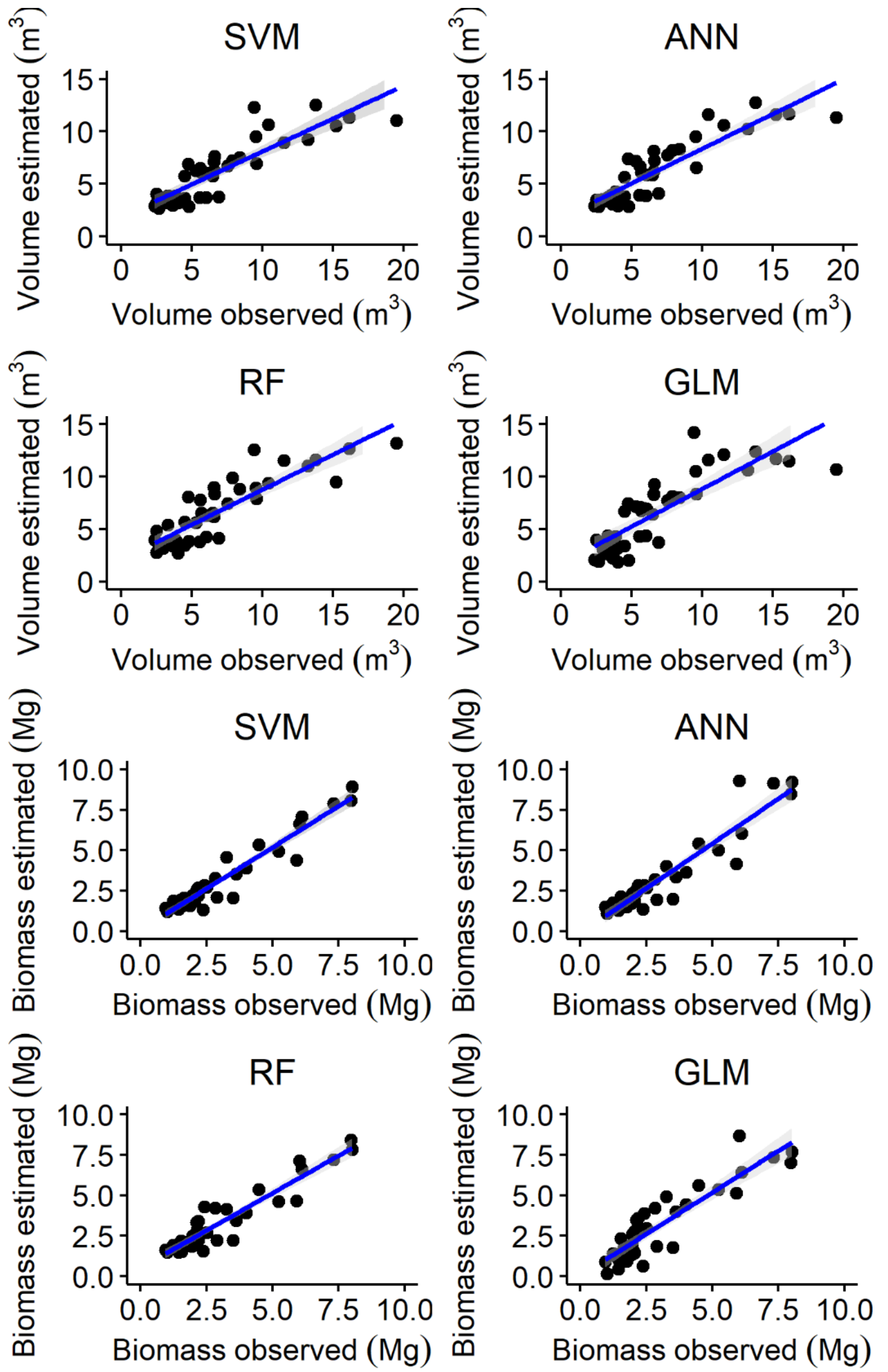

Figure 3 – Observed and estimated values of volume and biomass by the different machine learning models, SVM, ANN, RF and GLM tested.

GLM showed the best performance to estimate the volume of commercial trees, with the highest $r_{y\hat{y}}$ and the lowest RSME and MAE for all repetitions (Table 3). RF had a $r_{y\hat{y}}$ close to the GLM model and the second best performance to predict volume. The ANN algorithm showed moderate performance, and the SVM had the worst performance for predicting the volume of commercial trees in the Amazon.

ANN showed the best performance with the highest $r_{y\hat{y}}$ and the lowest RMSE and MAE for all repetitions, to predict biomass. RF also had the second best performance for predicting biomass. SVM had the worst performance for the prediction of commercial tree biomass in Southwestern Amazon (Table 3).

Table 3 – Statistics of the tested machine learning models (SVM, ANN, RF and GLM) for modeling volume and biomass of commercial trees in the Southwestern Amazon, in the municipality of Porto Acre, Acre, Brazil

Variable	Model	RMSE	MAE	$r_{y\hat{y}}$
Volume	SVM	1.93 ± 0.54	1.19 ± 0.23	0.89 ± 0.04
	ANN	1.67 ± 0.36	1.13 ± 0.19	0.91 ± 0.04
	RF	1.82 ± 0.41	1.24 ± 0.20	0.90 ± 0.04
	GLM	1.82 ± 0.33	1.30 ± 0.18	0.89± 0.04
Biomass	SVM	1.15 ± 0.33	0.67 ± 0.15	0.92 ± 0.03
	ANN	1.10 ± 0.27	0.69 ± 0.13	0.92 ± 0.03
	RF	1.19 ± 0.31	0.76 ± 0.14	0.91 ± 0.03
	GLM	1.35 ± 0.27	0.97 ± 0.12	0.88 ± 0.04

Where: RMSE: Root Mean Square Error; MAE: mean absolute error; SVM: Support Vector Machine; ANN: Artificial Neural Networks; RF: *Random Forest*; GLM: Generalized Linear Model.

The means of RMSE and MAE varied over the repetitions for each technique. The RMSE and MAE averages of ANN and GLM showed the lowest values to estimate the volume of trees (Figure 3). ANN and RF showed the lowest RMSE and MAE over the 50 repetitions in the cross-validation, to estimate the biomass (Figure 4). SVM showed greater instability in the values of RMSE and MAE of the cross-validation and the highest values of RMSE and MAE for all variables evaluated in the present study.

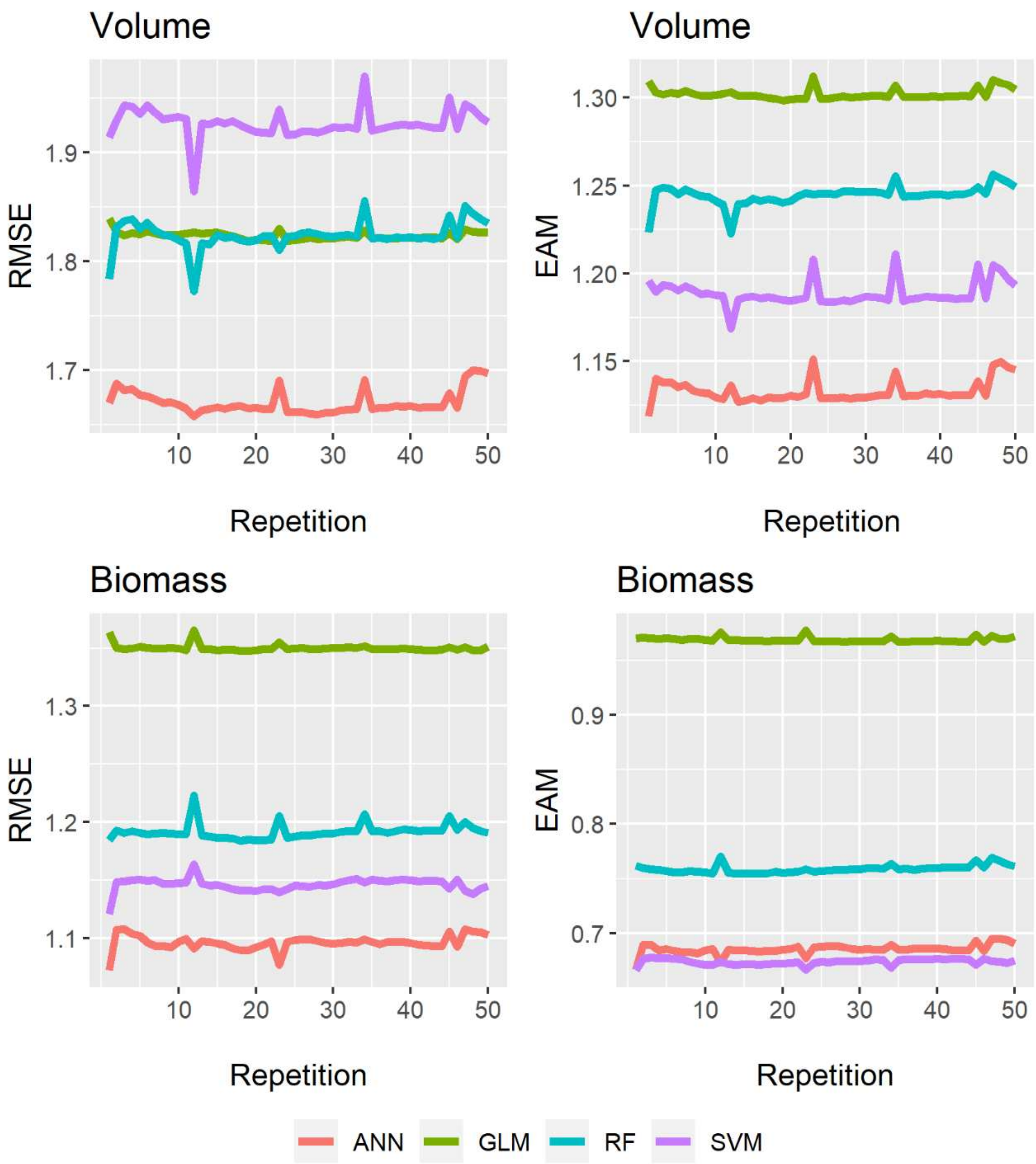

Figure 4 – Root Mean Square Error (RMSE) of the machine learning models SVM, ANN, RF and GLM, in the modeling of the volume and biomass of commercial trees in the Southwestern Amazon, in the municipality of Porto Acre, Acre, Brazil.

The Friedman test with the means of cross-validation RMSE showed that the predictions of volume and biomass of commercial trees in the Amazon differed between the techniques ($p <0.05$). Thus, the hypothesis that at least one average of one of the techniques differs from the others was accepted. The Nemenyi test pointed out that the difference between the GLM model and the other techniques was greater than the calculated critical difference (CD) to estimate the volume of the trees. The calculated critical difference (CD) of the ANN was greater than the other machine learning techniques evaluated to estimate biomass (Figure 5).

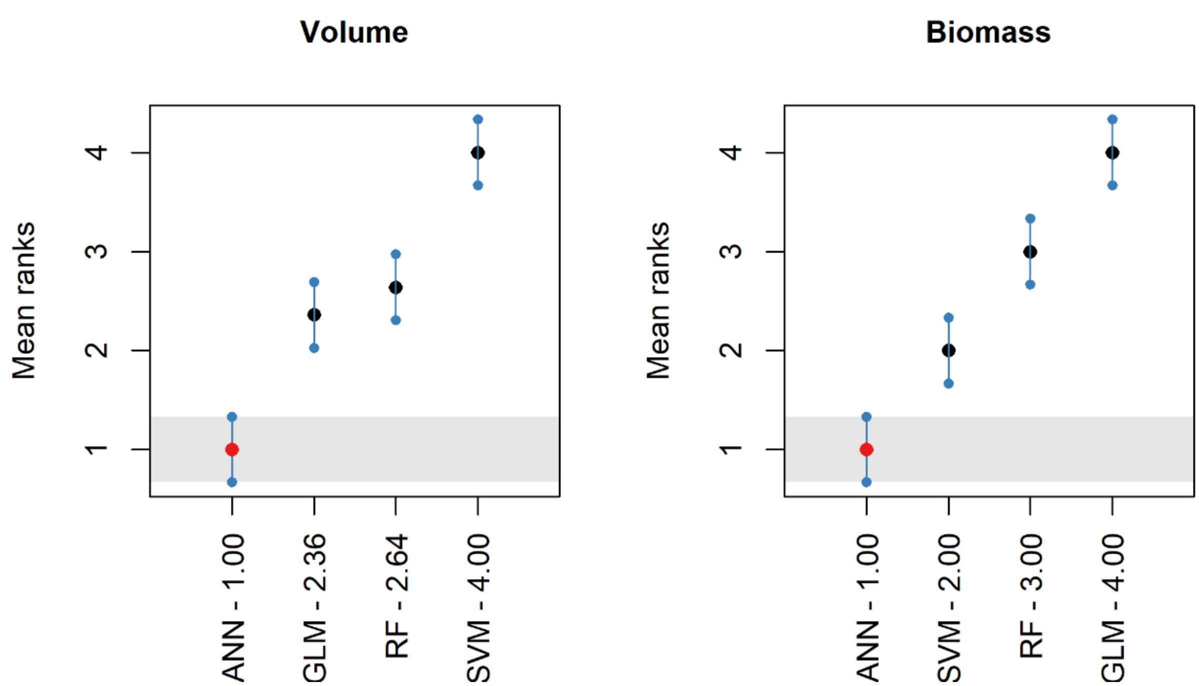

Figure 5 – Nemenyi test in the cross-validation of the estimates of volume and biomass of commercial trees in the Southwestern Amazon, in the municipality of Porto Acre, Acre, Brazil.

4. Discussion

4.1 Selection of variables

The Boruta variable selection method is a preferable algorithm among the variable selection methods because it has a high computational efficiency for working on data sets with many predictors (Speiser et al., 2019). In this study, the method pointed out that the diameter and height variables (total and commercial) are fundamental to explain the allometric attributes of the trees.

The stem diameter is a good predictor (Chambers et al., 2001; Kuyah et al., 2012; Yuen et al., 2016) and this is an important advantage for practical use. However, the integration of tree height significantly reduces uncertainties (Chave et al., 2014; Goodman et al., 2014b; Rutishauser et al., 2013).

Besides that, the inclusion of the wood basic density variable is an important predictor for biomass. This variable, with the diameter of the trunk the height of the tree and the type of forest (dry, moist or wet) are the most important predictors of the biomass (Chave et al., 2005). The inclusion and combination of both provide better quality in adjustment and estimates (Chave et al., 2005; Goodman et al., 2014b).

Our analysis also showed that the bioclimatic and topographic variables were not significant for estimating volume and biomass in commercial trees in the Amazon. This occurs due to the low variability of this information in the area, in view of the uniform distribution of these characteristics in the study region. However, considering the role of climate in predicting forest attributes can provide more accurate estimates (Chave et al., 2014; Feldpausch et al., 2011), since diameter-height relationships in trees depend on a series of physiological and environmental factors (Lines et al., 2012; Marshall et al., 2012). Maximum and minimum temperature, precipitation seasonality and degree of solar radiation have strong correlations with biomass (Banin et al., 2012; Taylor et al., 2019).

4.2 Model performance

In general, the machine learning models with the significant predictor variables pointed out by the Boruta algorithm accurately estimated the production attributes of evaluated commercial trees in Southwestern Amazon. A major advantage of using machine learning methods over traditional models is its applicability to any number of variables (Loh, 2011). This method is a very valuable procedure for working with data sets in large-scale databases (Abdel-Rahman et al., 2014; Belgiu and Drăguţ, 2016) because it can manipulate continuous, categorical and binary data (Ali et al., 2012) and is able to adapt to complex and non-linear relationships between variables, in addition to dealing with interaction effects between them (Reise et al., 2019). These models can encompass several types of information and it is possible to work with a single model for different situations.

The Friedman test confirm ($p < 0.05$) that the ANN model was the best to estimate volume and biomass of commercial trees in Southwestern Amazon. The RMSE and MAE corroborate this statement, showing small differences in training and validation. ANN is considered an important non-parametric algorithm for estimating biophysical parameters of the forest (Nandy et al., 2017). Neural networks can implicitly detect any complex nonlinear relationships between independent and dependent variables (Lazri and Ameur, 2018). In contrast to conventional parametric approaches, ANN does not require any assumptions about the statistical distribution of the data.

RF presented intermediaries results for estimating volume and biomass variables. This method produces the most accurate and stable predictions (Sun et al., 2019). This algorithm has been widely applied in ecological studies, as it can work with complex data analyzes (Reise et al., 2019). In addition, RF has been considered one of the best methods of classification and regression due to the high precision for estimation results, high calculation speed, robustness and the ability to predict important variables (Breiman, 2001). Decision tree-based algorithms are easy to apply, since fewer parameters need to be estimated. Therefore, they have a high degree of automation (Herrera et al., 2010; Rodriguez-Galiano et al., 2015).

SVM presented less precision in the volume estimates compared to the RMSE and MAE values. SVMs have the inconvenience of a delicate and computationally expensive hyperparameter adjustment. In addition, results for SVM compared to other methods showed average accuracy. ANNs and RF generally produce better results than SVMs for regression tasks. The simple statistical procedures and the set methods were very competitive for classification (Meyer et al., 2003).

GLM showed less precision in biomass estimates. This may be related to the link function used. GLMs have the characteristic of being able to choose the residual distribution family, important in the case of non-parametric models, such as those that follow Poisson distribution and negative binomial errors (Lopatin et al., 2016). The choice of one function over another may explain the low performance.

As forest volume and biomass are important for forest management, global change monitoring and modeling of forest productivity, there is a need for reliable methods of assessing and monitoring forest production (Nandy et al., 2017). The results presented here suggest a new alternative to predict these forest attributes.

It is worth mentioning that new approaches, such as the inclusion of climatic and topographic variables with greater variability, in different study areas can improve the estimates. In addition, the evaluation of the application of other variables can be used in view of the ability to work with high dimension data from machine learning models.

More generally, “best estimates”, even in models with all possible variables, should not be considered entirely accurate or baselines against which all other estimates are compared (Goodman et al., 2014b). Thus, new studies may be the subject of future research to improve results. In addition, the applicability of other algorithms (*e.g.,* Cubist, Regression Trees Models, etc.) can be tested and produce better estimates with transparency and computational efficiency (Corona-Núñez et al., 2017).

5. Conclusion

The tested machine learning methods (SVM, ANN, RF and GLM) are useful and efficient tools for estimating volume and biomass of commercial trees in the Southwestern Amazon. This study represents a new approach to estimate these attributes linked to forest production. ANN is the most suitable for estimating volume and biomass of commercial trees.

This study also confirm that the variables diameter, height and basic wood density are important variables for the prediction of volume and tree biomass.

6. Acknowledgments

This research was supported logistically by Empresa Fox Laminados Ltda. We thank the company’s owner, Mr. Antônio Aparecido Barlati and his team for their help in the planning, harvesting and sawmill stages of the study. We thank the Laboratory of Wood Anatomy of the Federal University of Acre (headed by Dr. Moisés Silveira Lobão) and the Laboratory of Wood Panels and Energy of the Federal University of Viçosa (headed by Professor Angelica de Cássia Oliveira Carneiro). This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPqand Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) – CAPES/PrInt [88881.311787/2018-01 and 88887.319055/2019-00]

7. References

ABNT, 2003. ABNT NBR 11941 – Madeira - determinação da densidade básica. ABNT – Associação Brasileira de Normas Técnicas, São Paulo.

Acre, 2010. Zoneamento Ecológico-Econômico do Estado do Acre, Fase II (Escala 1:250.000): Documento Síntese, second ed. SEMA, Rio Branco. http://www.amazonia.cnptia.embrapa.br/publicacoes_estados/Acre/Fase%202/Documento_Sintese.pdf (accessed 09 April 2020)

Abdel-Rahman, E.M., Mutanga, O., Adam, E., Ismail, R., 2014. Detecting *Sirex noctilio* grey-attacked and lightning-struck pine trees using airborne hyperspectral data, random forest and support vector machines classifiers. ISPRS Journal of. Photogrammetry Remote Sensing 88, 48–59. https://doi.org/10.1016/j.isprsjprs.2013.11.013

Alencar, J. da C., Almeida, R.A. de, Fernandes, N.P., Alencar, J. da C., Almeida, R.A. de, Fernandes, N.P., 1979. Fenologia de espécies florestais em floresta tropical úmida de terra firme na Amazônia Central. Acta Amazonica 9, 163–199. https://doi.org/10.1590/1809-43921979091163

Ali, J., Khan, R., Ahmad, N., Maqsood, I., 2012. Random Forests and Decision Trees. International Journal of Computer Science Issues 9, 272–278.

Alvares, C.A., Stape, J.L., Sentelhas, P.C., de Moraes Gonçalves, J.L., Sparovek, G., 2013. Köppen's climate classification map for Brazil. Meteorologische Zeitschrift 22, 711–728. https://doi.org/10.1127/0941-2948/2013/0507

Alvarez, E., Duque, A., Saldarriaga, J., Cabrera, K., de las Salas, G., del Valle, I., Lema, A., Moreno, F., Orrego, S., Rodríguez, L., 2012. Tree above-ground biomass allometries for carbon stocks estimation in the natural forests of Colombia. Forest Ecology and Management 267, 297–308. https://doi.org/10.1016/j.foreco.2011.12.013

Andrade, V.H.F., Machado, S. do A., Figueiredo Filho, A., Botosso, P.C., Miranda, B.P., Schöngart, J., 2019. Growth models for two commercial tree species in upland forests of the Southern Brazilian Amazon. Forest Ecology and Management 438, 215–223. https://doi.org/10.1016/j.foreco.2019.02.030

Antonelli, A., Sanmartín, I., 2011. Why are there so many plant species in the Neotropics? Taxon 60, 403–414. https://doi.org/10.1002/tax.602010

Banin, L., Feldpausch, T.R., Phillips, O.L., Baker, T.R., Lloyd, J., Affum-Baffoe, K., Arets, E.J.M.M., Berry, N.J., Bradford, M., Brienen, R.J.W., Davies, S., Drescher, M., Higuchi, N., Hilbert, D.W., Hladik, A., Iida, Y., Salim, K.A., Kassim, A.R., King, D.A., Lopez-Gonzalez, G., Metcalfe, D., Nilus, R., Peh, K.S.-H., Reitsma, J.M., Sonké, B., Taedoumg, H., Tan, S., White, L., Wöll, H., Lewis, S.L., 2012. What controls tropical forest architecture? Testing environmental, structural and floristic drivers. Global Ecology and Biogeography 21, 1179–1190. https://doi.org/10.1111/j.1466-8238.2012.00778.x

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry Remote Sensing 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Fortini, L.B., 2019. Integrated models show a transient opportunity for sustainable management by tropical forest dwellers. Forest Ecology and. Management 438, 233–242. https://doi.org/10.1016/j.foreco.2019.02.022

Breiman, L., 2001. Random forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

Brienen, R.J.W., Phillips, O.L., Feldpausch, T.R., Gloor, E., Baker, T.R., Lloyd, J., Lopez-Gonzalez, G., Monteagudo-Mendoza, A., Malhi, Y., Lewis, S.L., Vásquez Martinez, R., Alexiades, M., Álvarez Dávila, E., Alvarez-Loayza, P., Andrade, A., Aragão, L.E.O.C., Araujo-Murakami, A., Arets, E.J.M.M., Arroyo, L., C, G.A.A., Bánki, O.S., Baraloto, C., Barroso, J., Bonal, D., Boot, R.G.A., Camargo, J.L.C., Castilho, C.V., Chama, V., Chao, K.J., Chave, J., Comiskey, J.A., Cornejo Valverde, F., da Costa, L., de Oliveira, E.A., Di Fiore, A., Erwin, T.L., Fauset, S., Forsthofer, M., Galbraith, D.R., Grahame, E.S., Groot, N., Hérault, B., Higuchi, N., Honorio Coronado, E.N., Keeling, H., Killeen, T.J., Laurance, W.F., Laurance, S., Licona, J., Magnussen, W.E., Marimon, B.S., Marimon-Junior, B.H., Mendoza, C., Neill, D.A., Nogueira, E.M., Núñez, P., Pallqui Camacho, N.C., Parada, A., Pardo-Molina, G., Peacock, J., Peña-Claros, M., Pickavance, G.C., Pitman, N.C.A., Poorter, L., Prieto, A., Quesada, C.A., Ramírez, F., Ramírez-Angulo, H., Restrepo, Z., Roopsind, A., Rudas, A., Salomão, R.P., Schwarz, M., Silva, N., Silva-Espejo, J.E., Silveira, M., Stropp, J., Talbot, J., ter Steege, H., Teran-Aguilar, J., Terborgh, J., Thomas-

Caesar, R., Toledo, M., Torello-Raventos, M., Umetsu, R.K., van der Heijden, G.M.F., van der Hout, P., Guimarães Vieira, I.C., Vieira, S.A., Vilanova, E., Vos, V.A., Zagt, R.J., 2015. Long-term decline of the Amazon carbon sink. Nature 519, 344–348. https://doi.org/10.1038/nature14283

Cardoso, D., Särkinen, T., Alexander, S., Amorim, A.M., Bittrich, V., Celis, M., Daly, D.C., Fiaschi, P., Funk, V.A., Giacomin, L.L., Goldenberg, R., Heiden, G., Iganci, J., Kelloff, C.L., Knapp, S., Lima, H.C., Machado, A.F.P., Santos, R.M., Mello-Silva, R., Michelangeli, F.A., Mitchell, J., Moonlight, P., Moraes, P.L.R., Mori, S.A., Nunes, T.S., Pennington, T.D., Pirani, J.R., Prance, G.T., Queiroz, L.P., Rapini, A., Riina, R., Rincon, C.A.V., Roque, N., Shimizu, G., Sobral, M., Stehmann, J.R., Stevens, W.D., Taylor, C.M., Trovó, M., van den Berg, C., van der Werff, H., Viana, P.L., Zartman, C.E., Forzza, R.C., 2017. Amazon plant diversity revealed by a taxonomically verified species list. Proceedings of the National Academy of Sciences of the United States of America 114, 10695–10700. https://doi.org/10.1073/pnas.1706756114

Chambers, J.Q., Santos, J. dos, Ribeiro, R.J., Higuchi, N., 2001. Tree damage, allometric relationships, and above-ground net primary production in central Amazon forest. Forest Ecology and Management 152, 73–84. https://doi.org/10.1016/S0378-1127(00)00591-0

Chaudhary, A., Burivalova, Z., Koh, L.P., Hellweg, S., 2016. Impact of Forest Management on Species Richness: Global Meta-Analysis and Economic Trade-Offs. Scientific Reports 6, 23954. https://doi.org/10.1038/srep23954

Chave, J., Andalo, C., Brown, S., Cairns, M.A., Chambers, J.Q., Eamus, D., Fölster, H., Fromard, F., Higuchi, N., Kira, T., Lescure, J.-P., Nelson, B.W., Ogawa, H., Puig, H., Riéra, B., Yamakura, T., 2005. Tree allometry and improved estimation of carbon stocks and balance in tropical forests. Oecologia 145, 87–99. https://doi.org/10.1007/s00442-005-0100-x

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M.S., Delitti, W.B.C., Duque, A., Eid, T., Fearnside, P.M., Goodman, R.C., Henry, M., Martínez-Yrízar, A., Mugasha, W.A., Muller-Landau, H.C., Mencuccini, M., Nelson, B.W., Ngomanda, A., Nogueira, E.M., Ortiz-Malavassi, E., Pélissier, R., Ploton, P., Ryan, C.M., Saldarriaga, J.G., Vieilledent, G., 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. Global Change Biology 20, 3177–3190. https://doi.org/10.1111/gcb.12629

Corona-Núñez, R.O., Mendoza-Ponce, A., López-Martínez, R., 2017. Model selection changes the spatial heterogeneity and total potential carbon in a tropical dry forest. Forest Ecology and Management 405, 69–80. https://doi.org/10.1016/j.foreco.2017.09.018

Cummings, D.L., Boone Kauffman, J., Perry, D.A., Flint Hughes, R., 2002. Aboveground biomass and structure of rainforests in the southwestern Brazilian Amazon. Forest Ecology and Management 163, 293–307. https://doi.org/10.1016/S0378-1127(01)00587-4

Diamantopoulou, M.J., Milios, E., 2010. Modelling total volume of dominant pine trees in reforestations via multivariate analysis and artificial neural network models. Biosystems Engineering 105, 306–315. https://doi.org/10.1016/j.biosystemseng.2009.11.010

Domingues, G.F., Soares, V.P., Leite, H.G., Ferraz, A.S., Ribeiro, C.A.A.S., Lorenzon, A.S., Marcatti, G.E., Teixeira, T.R., de Castro, N.L.M., Mota, P.H.S., de Souza, G.S.A., de Menezes, S.J.M. da C., dos Santos, A.R., do Amaral, C.H., 2020. Artificial neural networks on integrated multispectral and SAR data for high-performance prediction of eucalyptus biomass. Computer and Electronics in Agriculture 168, 105089. https://doi.org/10.1016/j.compag.2019.105089

Feldpausch, T.R., Banin, L., Phillips, O.L., Baker, T.R., Lewis, S.L., Quesada, C.A., Affum-Baffoe, K., Arets, E.J.M.M., Berry, N.J., Bird, M., Brondizio, E.S., Camargo, P. de, Chave, J., Djagbletey, G., Domingues, T.F., Drescher, M., Fearnside, P.M., França, M.B., Fyllas, N.M., Lopez-Gonzalez, G., Hladik, A., Higuchi, N., Hunter, M.O., Iida, Y., Salim, K.A., Kassim, A.R., Keller, M., Kemp, J., King, D.A., Lovett, J.C., Marimon, B.S., Marimon-Junior, B.H., Lenza, E., Marshall, A.R., Metcalfe, D.J., Mitchard, E.T.A., Moran, E.F., Nelson, B.W., Nilus, R., Nogueira, E.M., Palace, M., Patiño, S., Peh, K.S.-H., Raventos, M.T., Reitsma, J.M., Saiz, G., Schrodt, F., Sonké, B., Taedoumg, H.E., Tan, S., White, L., Wöll, H., Lloyd, J., 2011. Height-diameter allometry of tropical forest trees. Biogeosciences 8, 1081–1106. https://doi.org/10.5194/bg-8-1081-2011

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37, 4302–4315. https://doi.org/10.1002/joc.5086

Gaui, T.D., Costa, F.R.C., Coelho de Souza, F., Amaral, M.R.M., de Carvalho, D.C., Reis, F.Q., Higuchi, N., 2019. Long-term effect of selective logging on floristic composition: A 25 year experiment in the Brazilian Amazon. Forest Ecology and Management 440, 258–266. https://doi.org/10.1016/j.foreco.2019.02.033

Gleason, C.J., Im, J., 2012. Forest biomass estimation from airborne LiDAR data using machine learning approaches. Remote Sensing of Environment 125, 80–91. https://doi.org/10.1016/j.rse.2012.07.006

Goodman, R.C., Phillips, O.L., Baker, T.R., 2014a. The importance of crown dimensions to improve tropical tree biomass estimates. Ecological Applications 24, 680–698. https://doi.org/10.1890/13-0070.1

Goodman, R.C., Phillips, O.L., Baker, T.R., 2014b. The importance of crown dimensions to improve tropical tree biomass estimates. Ecological Applications 24, 680–698. https://doi.org/10.1890/13-0070.1

Goodman, R.C., Phillips, O.L., del Castillo Torres, D., Freitas, L., Cortese, S.T., Monteagudo, A., Baker, T.R., 2013. Amazon palm biomass and allometry. Forest Ecology and Management 310, 994–1004. https://doi.org/10.1016/j.foreco.2013.09.045

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. Science 342, 850–853. https://doi.org/10.1126/science.1244693

Herrera, M., Torgo, L., Izquierdo, J., Pérez-García, R., 2010. Predictive models for forecasting hourly urban water demand. Journal of Hydrology 387, 141–150.

Husch B, Beers TW, John AK. Forest Mensuration. New Jersey: John Wiley & Sons; 2003.

Jachowski, N.R.A., Quak, M.S.Y., Friess, D.A., Duangnamon, D., Webb, E.L., Ziegler, A.D., 2013. Mangrove biomass estimation in Southwest Thailand using machine learning. Applied Geography 45, 311–321. https://doi.org/10.1016/j.apgeog.2013.09.024

Keller, M., Palace, M., Hurtt, G., 2001. Biomass estimation in the Tapajos National Forest, Brazil: Examination of sampling and allometric uncertainties. Forest Ecology and Management, New Directions in Tropical Forest Research 154, 371–382. https://doi.org/10.1016/S0378-1127(01)00509-6

Kursa, M.B., Rudnicki, W.R., 2010. Feature Selection with the Boruta Package. Journal of Statistical Software 36, 1–13. https://doi.org/10.18637/jss.v036.i11

Kuyah, S., Dietz, J., Muthuri, C., Jamnadass, R., Mwangi, P., Coe, R., Neufeldt, H., 2012. Allometric equations for estimating biomass in agricultural landscapes: I. Aboveground biomass. Agriculture, Ecosystems & Environment 158, 216–224. https://doi.org/10.1016/j.agee.2012.05.011

Lazri, M., Ameur, S., 2018. Combination of support vector machine, artificial neural network and random forest for improving the classification of convective and stratiform rain using spectral features of SEVIRI data. Atmospheric Research 203, 118–129. https://doi.org/10.1016/j.atmosres.2017.12.006

Lima, A.J.N., Suwa, R., Ribeiro, G.H.P.M., Kajimoto, T., dos Santos, J., da Silva, R.P., de Souza, C.A.S., de Barros, P.C., Noguchi, H., Ishizuka, M., Higuchi, N., 2012. Allometric models for estimating above- and below-ground biomass in Amazonian forests at São Gabriel da Cachoeira in the upper Rio Negro, Brazil. Forest Ecology and Management 277, 163–172. https://doi.org/10.1016/j.foreco.2012.04.028

Lines, E.R., Zavala, M.A., Purves, D.W., Coomes, D.A., 2012. Predictable changes in aboveground allometry of trees along gradients of temperature, aridity and competition. Global Ecology and Biogeography 21, 1017–1028. https://doi.org/10.1111/j.1466-8238.2011.00746.x

Loh, W-Y., 2011. Classification and regression trees. Wires Data Mining and Knowledge Discovery 1, 14–23. https://doi.org/10.1002/widm.8

Lopatin, J., Dolos, K., Hernández, H.J., Galleguillos, M., Fassnacht, F.E., 2016. Comparing Generalized Linear Models and random forest to model vascular plant species richness using LiDAR data in a natural forest in central Chile. Remote Sensing of Environment 173, 200–210. https://doi.org/10.1016/j.rse.2015.11.029

Luize, B.G., Magalhães, J.L.L., Queiroz, H., Lopes, M.A., Venticinque, E.M., Novo, E.M.L. de M., Silva, T.S.F., 2018. The tree species pool of Amazonian wetland forests: Which species can assemble in periodically waterlogged habitats? PLoS ONE 13, e0198130. https://doi.org/10.1371/journal.pone.0198130

Marshall, A.R., Willcock, S., Platts, P.J., Lovett, J.C., Balmford, A., Burgess, N.D., Latham, J.E., Munishi, P.K.T., Salter, R., Shirima, D.D., Lewis, S.L., 2012. Measuring and modelling above-ground carbon and tree allometry along a tropical elevation gradient. Biological Conservation, REDD+ and conservation 154, 20–33. https://doi.org/10.1016/j.biocon.2012.03.017

Meyer, D., Leisch, F., Hornik, K., 2003. The support vector machine under test. Neurocomputing, Support Vector Machines 55, 169–186. https://doi.org/10.1016/S0925-2312(03)00431-4

Nandy, S., Singh, R., Ghosh, S., Watham, T., Kushwaha, S.P.S., Kumar, A.S., Dadhwal, V.K., 2017. Neural network-based modelling for forest biomass assessment. Carbon Management 8, 305–317. https://doi.org/10.1080/17583004.2017.1357402

Nelson, B.W., Mesquita, R., Pereira, J.L.G., Garcia Aquino de Souza, S., Teixeira Batista, G., Bovino Couto, L., 1999. Allometric regressions for improved estimate of secondary forest biomass in the central Amazon. Forest Ecology and Management 117, 149–167. https://doi.org/10.1016/S0378-1127(98)00475-7

Nogueira, E.M., Fearnside, P.M., Nelson, B.W., Barbosa, R.I., Keizer, E.W.H., 2008. Estimates of forest biomass in the Brazilian Amazon: New allometric equations and adjustments to biomass from wood-volume inventories. Forest Ecology and Management 256, 1853–1867. https://doi.org/10.1016/j.foreco.2008.07.022

Peña-Claros, M., Fredericksen, T.S., Alarcón, A., Blate, G.M., Choque, U., Leaño, C., Licona, J.C., Mostacedo, B., Pariona, W., Villegas, Z., Putz, F.E., 2008. Beyond reduced-impact logging:

Silvicultural treatments to increase growth rates of tropical trees. Forest Ecology and Management, Moving beyond reduced impact-logging towards a more holistic management of tropical forests 256, 1458–1467. https://doi.org/10.1016/j.foreco.2007.11.013

Putz, F.E., Sist, P., Fredericksen, T., Dykstra, D., 2008. Reduced-impact logging: Challenges and opportunities. Forest Ecology and Management, Moving beyond reduced impact-logging towards a more holistic management of tropical forests 256, 1427–1433. https://doi.org/10.1016/j.foreco.2008.03.036

Reise, J., Kukulka, F., Flade, M., Winter, S., 2019. Characterising the richness and diversity of forest bird species using National Forest Inventory data in Germany. Forest Ecology and Management 432, 799–811. https://doi.org/10.1016/j.foreco.2018.10.012

Rodriguez-Galiano, V.F., Sanchez-Castillo, M., Chica-Olmo, M., Chica-Rivas, M., 2015. Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geology Reviews, 71. https://doi.org/10.1016/j.oregeorev.2015.01.001

Romero, F.M.B., 2018. Contribuição do manejo sustentável em Floresta do bioma amazônico para minimização de gases de efeito estufa. Ph.D. Thesis, Universidade Federal de Viçosa, Viçosa, Brazil.

Rutishauser, E., Noor'an, F., Laumonier, Y., Halperin, J., Rufi'ie, Hergoualc'h, K., Verchot, L., 2013. Generic allometric models including height best estimate forest biomass and carbon stocks in Indonesia. Forest Ecology and Management 307, 219–225. https://doi.org/10.1016/j.foreco.2013.07.013

Souza, G.S.A. de, Soares, V.P., Leite, H.G., Gleriani, J.M., do Amaral, C.H., Ferraz, A.S., Silveira, M.V. de F., Santos, F.C. dos, Velloso, S.G.S., Domingues, G.F., Silva, S., 2019. Multi-sensor prediction of *Eucalyptus* stand volume: a support vector approach. ISPRS Journal of Photogrammetry and Remote Sensing 156, 135–146. https://doi.org/10.1016/j.isprsjprs.2019.08.002

Speiser, J.L., Miller, M.E., Tooze, J., Ip, E., 2019. A comparison of random forest variable selection methods for classification prediction modeling. Expert Systems with Applications 134, 93–101. https://doi.org/10.1016/j.eswa.2019.05.028

Steege, H. ter, Pitman, N.C.A., Sabatier, D., Baraloto, C., Salomão, R.P., Guevara, J.E., Phillips, O.L., Castilho, C.V., Magnusson, W.E., Molino, J.-F., Monteagudo, A., Vargas, P.N., Montero, J.C., Feldpausch, T.R., Coronado, E.N.H., Killeen, T.J., Mostacedo, B., Vasquez, R., Assis, R.L., Terborgh, J., Wittmann, F., Andrade, A., Laurance, W.F., Laurance, S.G.W., Marimon, B.S., Marimon, B.-H., Vieira, I.C.G., Amaral, I.L., Brienen, R., Castellanos, H., López, D.C., Duivenvoorden, J.F., Mogollón, H.F., Matos, F.D. de A., Dávila, N., García-Villacorta, R., Diaz, P.R.S., Costa, F., Emilio, T., Levis, C., Schietti, J., Souza, P., Alonso, A., Dallmeier, F., Montoya, A.J.D., Piedade, M.T.F., Araujo-Murakami, A., Arroyo, L., Gribel, R., Fine, P.V.A., Peres, C.A., Toledo, M., C, G.A.A., Baker, T.R., Cerón, C., Engel, J., Henkel, T.W., Maas, P., Petronelli, P., Stropp, J., Zartman, C.E., Daly, D., Neill, D., Silveira, M., Paredes, M.R., Chave, J., Filho, D. de A.L., Jørgensen, P.M., Fuentes, A., Schöngart, J., Valverde, F.C., Fiore, A.D., Jimenez, E.M., Mora, M.C.P., Phillips, J.F., Rivas, G., Andel, T.R. van, Hildebrand, P. von, Hoffman, B., Zent, E.L., Malhi, Y., Prieto, A., Rudas, A., Ruschell, A.R., Silva, N., Vos, V., Zent, S., Oliveira, A.A., Schutz, A.C., Gonzales, T., Nascimento, M.T., Ramirez-Angulo, H., Sierra, R., Tirado, M., Medina, M.N.U., Heijden, G. van der, Vela, C.I.A., Torre, E.V., Vriesendorp, C., Wang, O., Young, K.R., Baider, C., Balslev, H., Ferreira, C., Mesones, I., Torres-Lezama, A., Giraldo, L.E.U., Zagt, R., Alexiades, M.N., Hernandez, L., Huamantupa-Chuquimaco, I., Milliken, W., Cuenca, W.P., Pauletto, D., Sandoval, E.V., Gamarra, L.V., Dexter, K.G., Feeley, K., Lopez-Gonzalez, G., Silman, M.R., 2013. Hyperdominance in the Amazonian Tree Flora. Science 342, 1243092. https://doi.org/10.1126/science.1243092

Steege, H. ter, Vaessen, R.W., Cárdenas-López, D., Sabatier, D., Antonelli, A., Oliveira, S.M. de, Pitman, N.C.A., Jørgensen, P.M., Salomão, R.P., 2016. The discovery of the Amazonian tree flora with an updated checklist of all known tree taxa. Scientific Reports 6, 1–15. https://doi.org/10.1038/srep29549

Sun, T., Chen, F., Zhong, L., Liu, W., Wang, Y., 2019. GIS-based mineral prospectivity mapping using machine learning methods: A case study from Tongling ore district, eastern China. Ore Geology Reviews 109, 26–49. https://doi.org/10.1016/j.oregeorev.2019.04.003

Taylor, P.G., Cleveland, C.C., Soper, F., Wieder, W.R., Dobrowski, S.Z., Doughty, C.E., Townsend, A.R., 2019. Greater stem growth, woody allocation, and aboveground biomass in Paleotropical forests than in Neotropical forests. Ecology 100, e02589. https://doi.org/10.1002/ecy.2589

Vieilledent, G., Vaudry, R., Andriamanohisoa, S.F.D., Rakotonarivo, O.S., Randrianasolo, H.Z., Razafindrabe, H.N., Rakotoarivony, C.B., Ebeling, J., Rasamoelina, M., 2012. A universal approach to estimate biomass and carbon stock in tropical forests using generic allometric models. Ecological Applications 22, 572–583. https://doi.org/10.1890/11-0039.1

Yuen, J.Q., Fung, T., Ziegler, A.D., 2016. Review of allometric equations for major land covers in SE Asia: Uncertainty and implications for above- and below-ground carbon estimates. Forest Ecology and Management, Forest Management for Climate Change 360, 323–340. https://doi.org/10.1016/j.foreco.2015.09.016

Zhao, K., Popescu, S., Meng, X., Pang, Y., Agca, M., 2011. Characterizing forest canopy structure with lidar composite metrics and machine learning. Remote Sensing of Environment 115, 1978–1996. https://doi.org/10.1016/j.rse.2011.04.001

CONCLUSÕES GERAIS

Em nosso estudo, observou-se que as técnicas de aprendizado de máquina estudadas representam um caminho promissora para modelagem dos processos demográficos (recrutamento, crescimento e mortalidade) na Mata Atlântica no Brasil **(Capítulos I, II e III)** e em estimativas de estoque e volume na região Amazônica Brasileira (**Capítulo IV**).

Verificou-se que variáveis climáticas, edáficas, topográficas, atributos da floresta e antropogênicas são importantes preditores dos processos demográficos. O entendimento dos efeitos dessas variáveis pode auxiliar em descobertas do comportamento das florestas, sobretudo as ameaças pelo desmatamento e fragmentação.

Nossos resultados, contribuem para o entendimento dos processos da dinâmica florestal em biomas brasileiros e podem fornecer subsídios para o entendimento, manejo e manutenção da biodiversidade em florestas tropicais.

APÊNDICE A – ARTIGOS I, II e III.

Table S1. Overview, climate, anthropogenic variables, soil and topography variables of the seven studied Atlantic Forest fragments in Minas Gerais, Brazil

Variables	Number	Description
1.Soil and topography variables	1.1	pH H_2O 0cm - 20 cm and 0cm - 40 cm
	1.2	P (cmolc/dm³) 0cm - 20 cm and 0cm - 40 cm
	1.3	K (cmolc/dm³) 0cm - 20 cm and 0cm - 40 cm
	1.4	Ca^{2+} (cmolc/dm³) 0cm - 20 cm and 0cm - 40 cm
	1.5	Mg^{2+} (cmolc/dm³) 0cm - 20 cm and 0cm - 40 cm
	1.6	Al^{3+} Alumínio trocável (cmolc/dm³) 0cm - 20 cm and 0cm - 40 cm
	1.7	H + Al (cmolc/dm³) 0cm - 20 cm and 0cm - 40 cm
	1.8	SB (cmolc/dm³) 0cm - 20 cm and 0cm - 40 cm
	1.9	t (cmolc/dm³) 0cm - 20 cm and 0cm - 40 cm
	1.10	T (cmolc/dm³) 0cm - 20 cm and 0cm - 40 cm
	1.11	V (%) 0cm - 20 cm and 0cm - 40 cm
	1.12	m (%) 0cm - 20 cm and 0cm - 40 cm
	1.13	Organic matter(dag/kg) 0cm - 20 cm and 0cm - 40 cm
	1.14	P-Rem (mg/L) 0cm - 20 cm and 0cm - 40 cm
	1.15	pH H_2O 0cm - 20 cm and 0cm - 40 cm
	1.16	Elevation (m)
	1.17	Declivity (%)
	1.18	Slope Angle
2.Climate variables	2.1	Mean annual temperature 1 year before measurement
	2.2	Average of Mean annual temperature
	2.3	Annual precipitation (measurement year)
	2.4	Annual precipitation (1 year before measurement)
	2.5	Mean annual precipitation (1 and 2 years before measurement)
	2.6	Mean annual precipitation (1, 2 and 3 years before measurement)
	2.7	Mean annual precipitation (1, 2, 3 and 4 years before measurement)
	2.8	Mean annual precipitation
	2.9	Total precipitation of the three driest months

	2.10	Number of dry months < 100 mm rainfall
	2.11	Climatic water deficit (from 1989 to 1 year before measurement)
	2.12	Mean climatic water deficit (1, 2, 3 and 4 years before measurement)
	2.13	Mean climatic water deficit (1, 2 and 3 years before measurement)
	2.14	Mean climatic water deficit (1 and 2 years before measurement)
	2.15	Climatic water deficit (1 year before measurement)
	2.16	Climatic water deficit (measurement year)
3.Anthropogenic variables	3.1	Forest cover (ha); Year=1985; 2002;2017 - Buffer=500m
	3.2	Forest cover (ha); Year=1985; 2002; 2017 - Buffer=1000m
	3.3	Forest cover (ha); Year=1985; 2002; 2017 - Buffer=1000m
	3.4	Land Use History
	3.5	Edge distance
	3.6	Age of Abandonment
	3.7	Forest size
4. Forest attributes	4.1	Basal area - plots
	4.2	Number of stems

Land Use History (agricultural production; deforestation; eucalyptus; selective logging); Mean annual temperature 1 year before measurement: average of Mean annual temperature (from 1989 to 1 year before measurement); Mean annual precipitation (from 1989 to 1 year before measurement); BS = Base Sum = $Ca^{2+} + Mg^{2+} + K$; t - Effective Cation Exchange Capacity = $Ca^{2+} + Mg^{2+} + K + AL^{3+}$; T - Cation Exchange Capacity = $Ca^{2+} + Mg^{2+} + K + H^{+} + AL^{3+}$; V= Base Saturation Index = 100*SB/T; m= Aluminum Saturation Index = 100* AL^{3+}/(t)..

Table S2. Overview of seven Atlantic Forest fragments located in Minas Gerais, Brazil included in the study. Mean and Standard Deviation (in brackets) of main characteristics of each area. CWD = climatic water deficit

Description	Atlantic Forest fragments						
	FR1	FR2	FR3	FR4	FR5	FR6	FR7
Forest Attributes							
Tree Species	31 (±7)	33 (±4)	25 (±8)	37 (±9)	24 (±7)	40 (±14)	30 (±5)
Stems (ha)	2008 (±664)	1748 (±300)	1765 (±454)	1825 (±264)	1610 (±484)	1616 (±344)	1482 (±305)
BA (ha)	18.85 (±6.32)	23.17 (±3.19)	16.46 (±6.05)	25.38 (±7.61)	19.11 (±5.32)	22.99 (±8.45)	27.33 (±8.47)
Stems Recruitment (ha/year)	44 (±34)	45 (±30)	42 (±25)	33 (±16)	66 (±39)	35 (±18)	30 (±27)
BA Recruitment (ha/year)	0.11 (±0.09)	0.12 (±0.08)	0.13 (±0.09)	0.08 (±0.05)	0.20 (±0.12)	0.10 (±0.06)	0.08 (±0.07)
Stems Mortality (ha/year)	42 (±48)	42 (±34)	44 (±32)	37 (±24)	33 (±20)	29 (±21)	37 (±25)
BA Mortality (ha/year)	0.35 (±0.50)	0.42 (±0.38)	0.41 (±0.30)	0.40 (±0.34)	0.38 (±0.47)	0.32 (±0.30)	0.45 (±0.47)
Land use history (% of plots)							
Agricultural production	40	0	0	0	15	0	0
Deforestation	0	0	81.25	0	0	0	0
Eucalyptus	0	0	0	0	35	0	0
Selective logging	60	100	18.75	100	50	100	100
Landscape							
Forest Size	106.00	21.80	264.00	37.30	44.11	38.40	17.00
Elevation (m)	959.80 (±88.09)	719.40 (±5.55)	267.75 (±19.29)	267.83 (±25.02)	696.85 (±21.26)	846.67 (±56.93)	726.30 (±13.23)
Declivity (%)	30.52 (±9.22)	27.29 (±6.40)	26.25 (±13.27)	29.07 (±3.94)	33.86 (±10.24)	36.35 (±15.01)	25.46 (±8.84)
Slope Angle	161.27 (±88.05)	298.69 (±33.77)	265.53 (±36.29)	109.46 (±28.02)	156.10 (±39.99)	179.92 (±26.83)	162.74 (±19.17)
Forest cover (ha)	264.97	148.97	289.61	302.07	66.71	277.78	109.91

	(±11.25)	(±3.74)	(±9.73)	(±2.11)	(±3.41)	(±6.49)	(±10.72)
Edge distance (m)	101.90 (±69.26)	64.70 (±7.04)	88.58 (±76.65)	74.16 (±35.07)	89.74 (±33.78)	120.71 (±68.75)	103.80 (±39.83)
Soil							
pH (H_2O)	4.52 (±0.25)	4.00 (±0.06)	3.94 (±0.15)	3.76 (±0.08)	4.19 (±0.28)	4.23 (±0.55)	4.11 (±0.28)
K (cmolc/dm³)	0.16 (±0.11)	0.07 (±0.01)	0.08 (±0.01)	0.07 (±0.01)	0.07 (±0.02)	0.09 (±0.03)	0.07 (±0.02)
Ca^{2+} (cmolc/dm³)	0.49 (±0.45)	0.29 (±0.10)	0.31 (±0.08)	0.16 (±0.07)	0.50 (±0.45)	0.49 (±0.56)	0.46 (±0.59)
Mg^{2+} (cmolc/dm³)	0.24 (±0.27)	0.10 (±0.01)	0.17 (±0.06)	0.11 (±0.01)	0.25 (±0.31)	0.16 (±0.10)	0.18 (±0.16)
Al^{3+} (cmolc/dm³)	1.06 (±0.43)	1.36 (±0.14)	1.40 (±0.38)	1.82 (±0.17)	1.62 (±0.59)	0.91 (±0.34)	1.71 (±0.73)
P-res (mg/L)	10.69 (±2.07)	17.45 (±1.34)	21.54 (±2.32)	21.85 (±1.65)	15.36 (±2.96)	21.23 (±5.17)	18.03 (±4.26)
SOM (dag/kg)	5.64 (±1.75)	3.85 (±0.35)	3.89 (±0.57)	3.00 (±0.28)	4.26 (±0.66)	5.14) (±1.56)	4.20 (±0.86)
Climate							
Precipitation (mm)	1112.74 (±214.50)	1254.77 (±260.52)	1194.38 (±298.96)	1194.38 (±298.96)	1254.77 (±260.52)	1263.92 (±226.07)	1254.77 (±260.52)
CWD (mm)	-1021.36 (±432.50)	-881.24 (±299.13)	-2020.77 (±563.16)	-1499.09 (±502.88)	-864.35 (±354.39)	-940.19 (±395.45)	-941.14 (±248.97)
Precipitation driest months (mm)	30.84 (±33.57)	39.44 (±34.82)	33.99 (±30.13)	33.99 (±30.13)	39.44 (±34.82)	39.84 (±30.99)	39.44 (±34.82)
Precipitation Less 100mm	7.82 (±1.17)	7.47 (±1.24)	7.53 (±1.28)	7.53 (±1.28)	7.47 (±1.24)	7.38 (±1.18)	7.47 (±1.24)
Mean Temperature (ºC)	20.85 (±1.59)	20.11 (±0.40)	24.15 (±1.73)	24.15 (±1.73)	20.11 (±0.40)	19.51 (±1.97)	20.11 (±0.40)

FR1: Cachoeira das Pombas; FR2: Mata da Garagem; FR3: Ipaba Mata1; FR4: Ipaba Mata2; FR5: Centev; FR6: São José and FR7: Mata da Silvicultura.

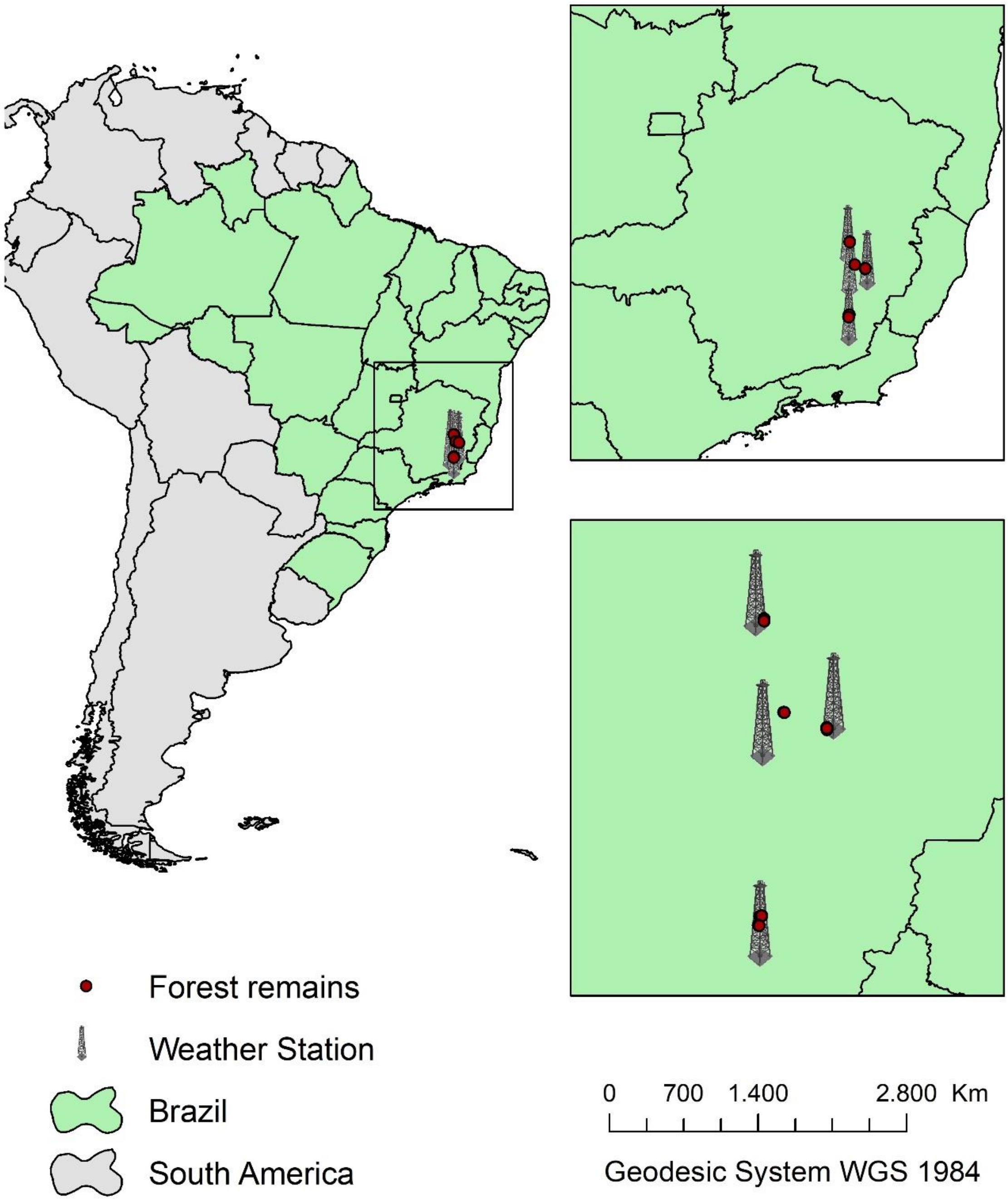

Figure S1. Locations of climatological station and Atlantic Forest fragments located in Minas Gerais, Brazil.

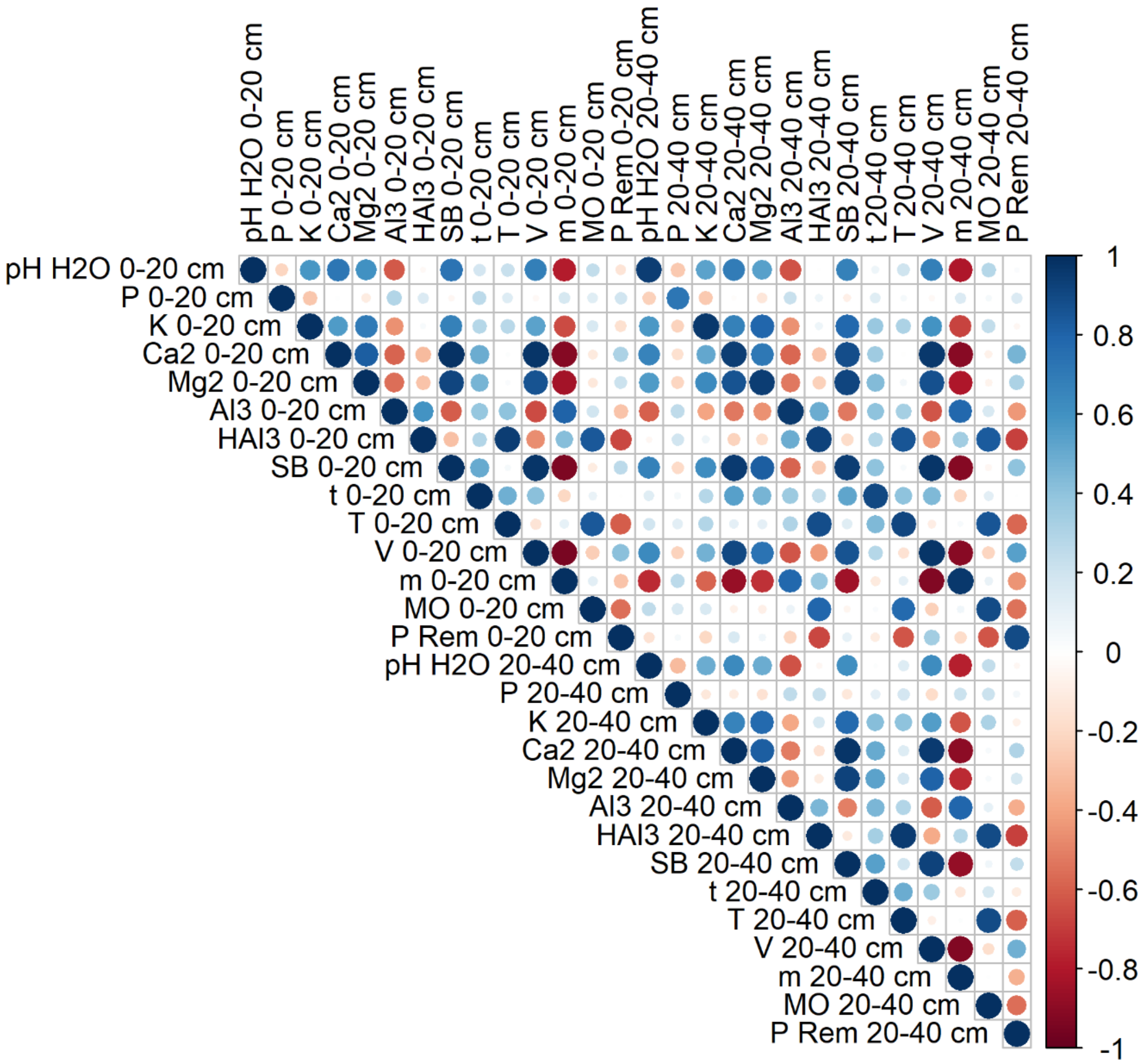

Figure S2. Pearson correlations between soil variables and depth (0–20 cm and 20–40 cm). Dark blue circles indicate positive correlations and dark red circles indicate negative correlations. The size of the circle indicates the strength of the correlation. For abbreviation of soil see Material and Methods.

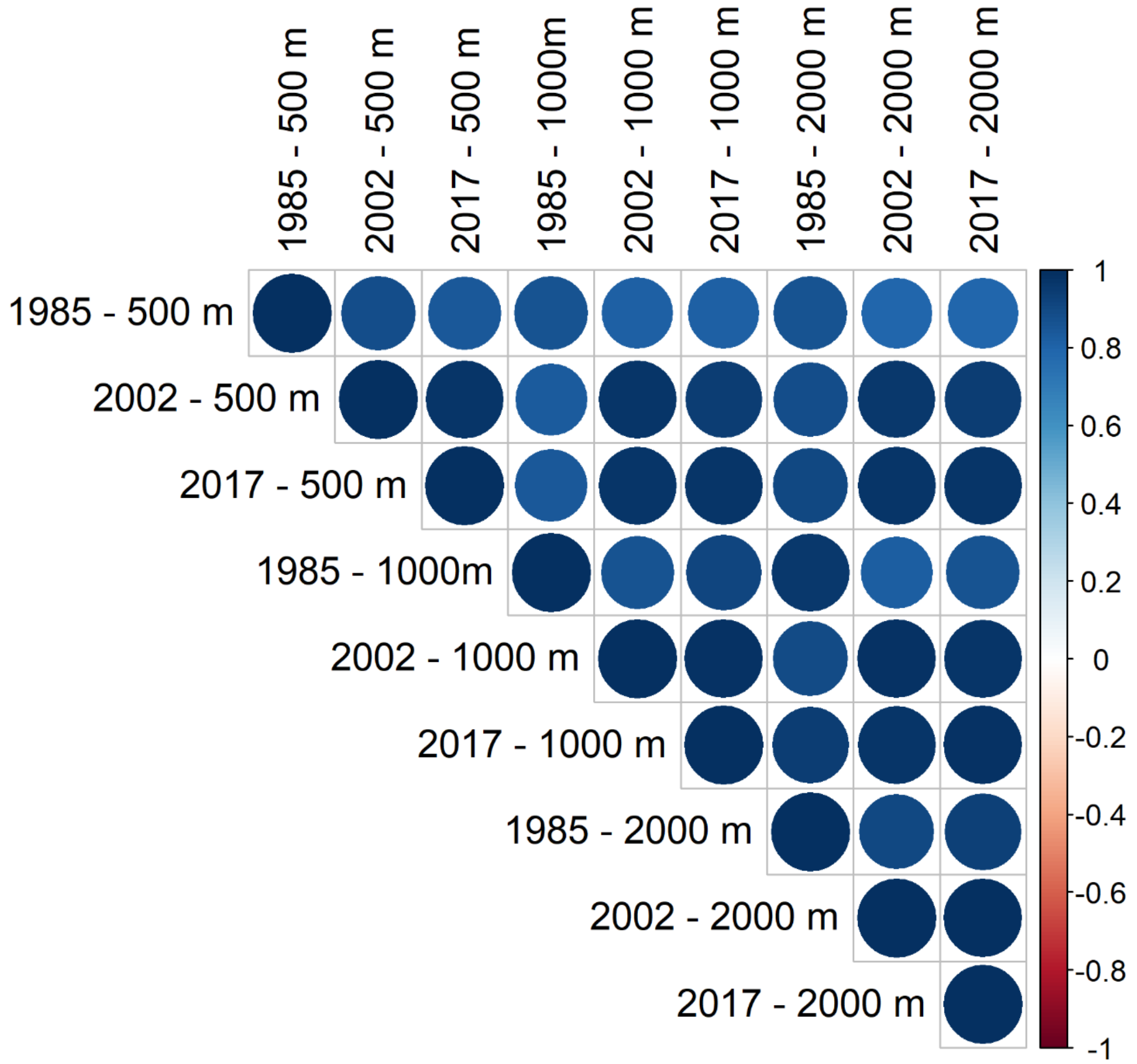

Figure S3. Pearson correlations between forest cover area using radii of 500, 1000, and 2000 m, for the year 1985, 2002 and 2017. Dark blue circles indicate positive correlations and dark red circles indicate negative correlations. The size of the circle indicates the strength of the correlation.

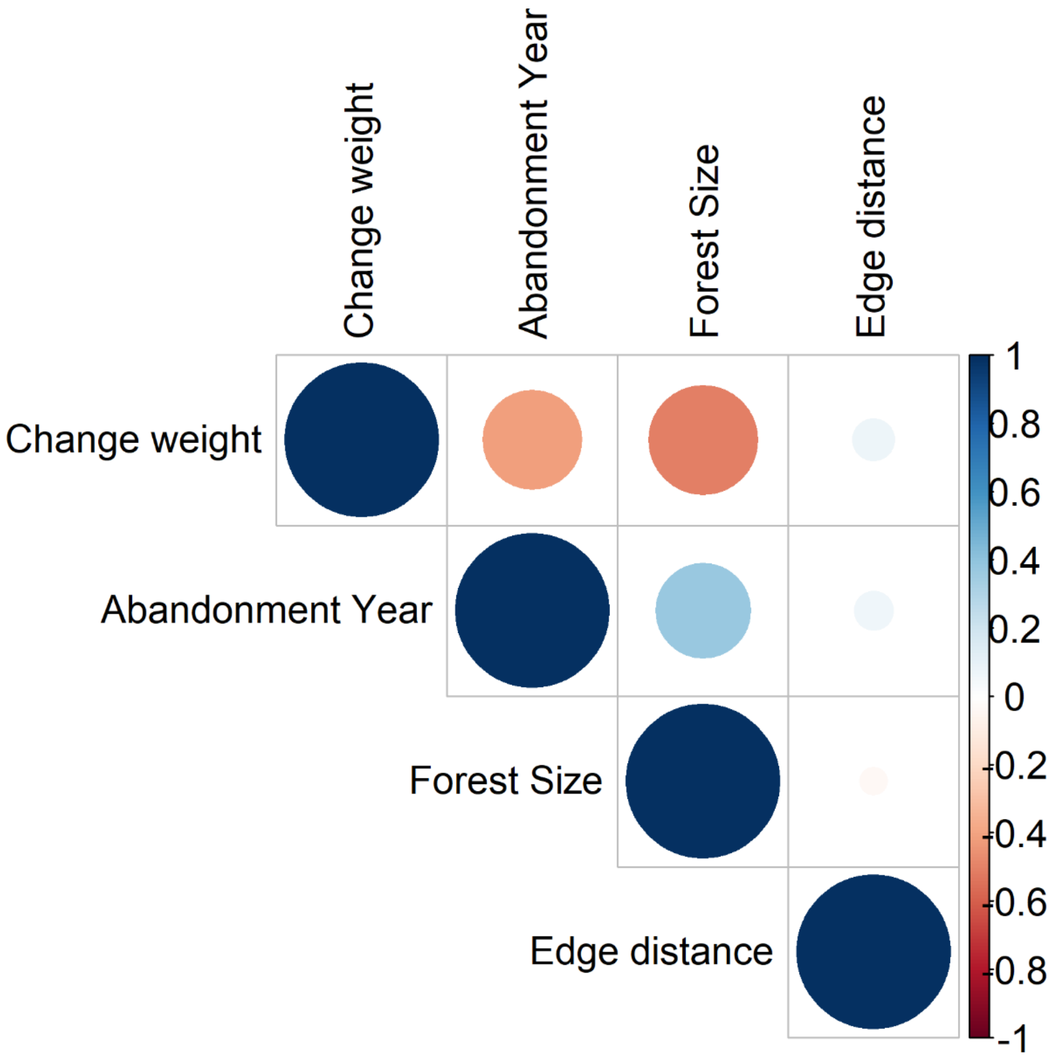

Figure S4. Pearson correlations between forest cover area, Edge distance and Forest Size, Abandonment Year. Dark blue circles indicate positive correlations and dark red circles indicate negative correlations. The size of the circle indicates the strength of the correlation.

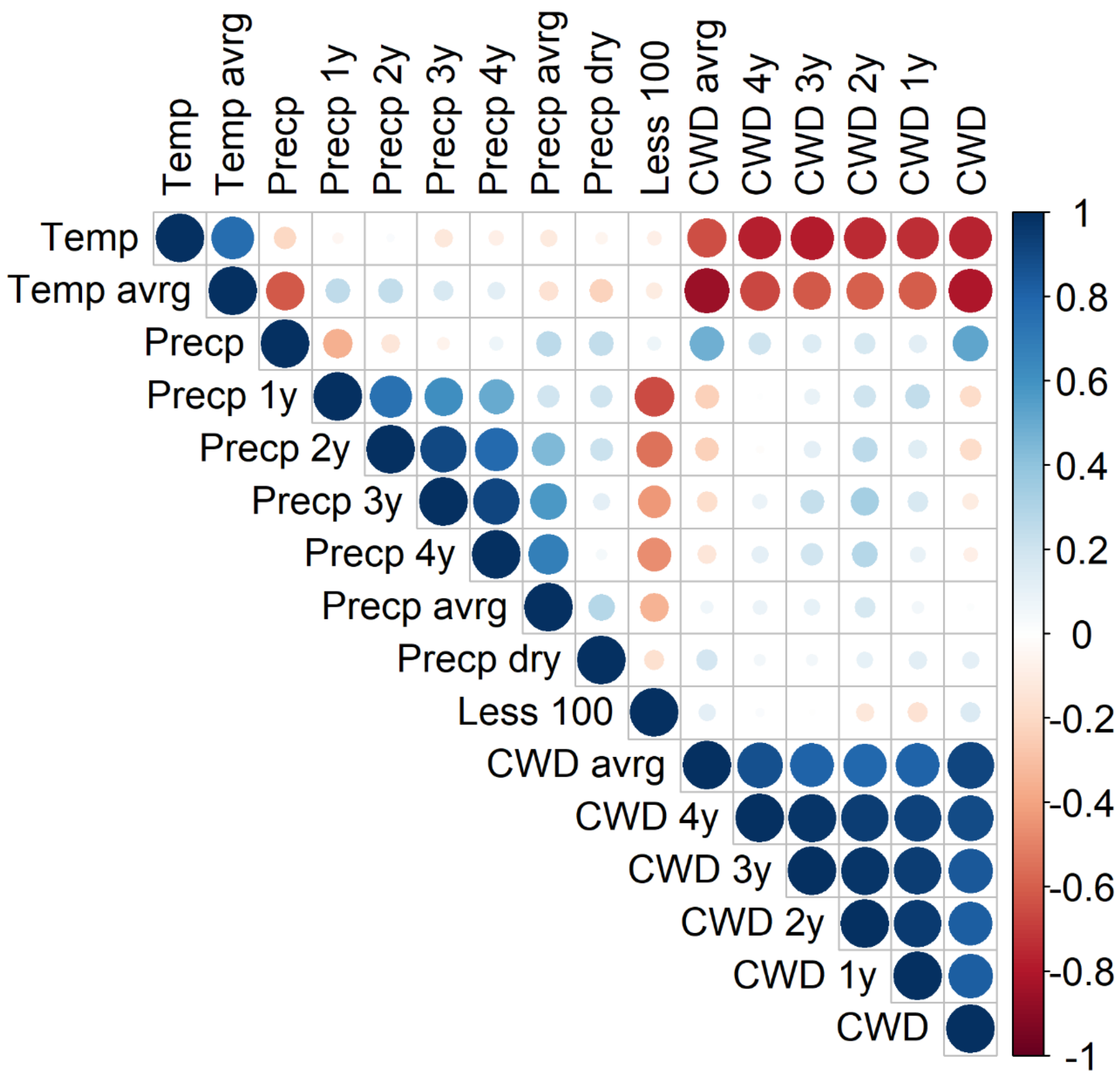

Figure S5. Pearson correlations between for annual precipitation (Precp) and climatic water deficit (CWD) for 1, 2, 3 and 4 years before the measurement year, number of months with less than 100 mm of rainfall (less 100), precipitation in the three driest months (Precp dry), and average annual temperature (Temp). Blue circles indicate positive correlations and red circles indicate negative correlations. The size of the circle indicates the strength of the correlation.

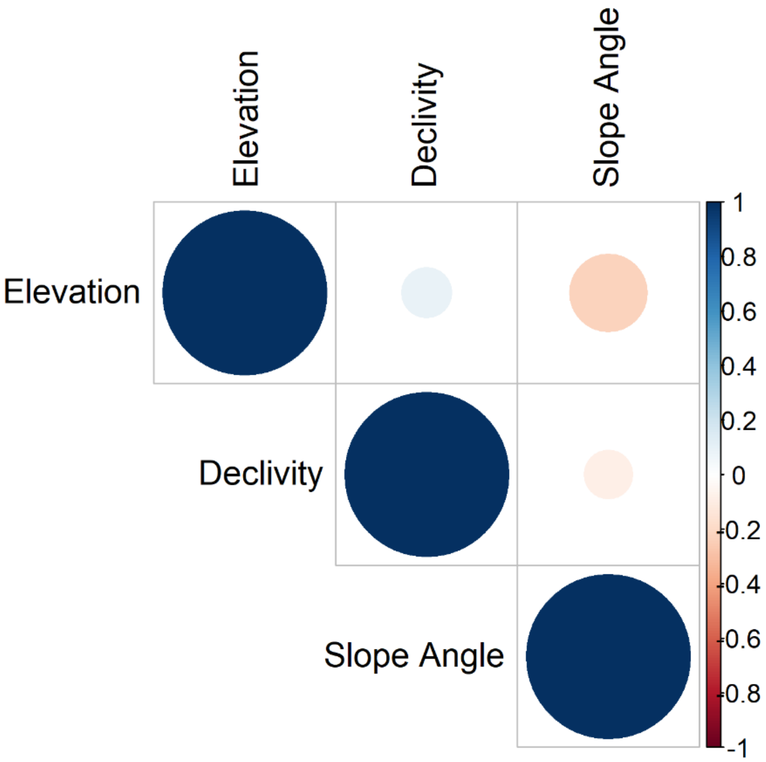

Figure S6. Pearson correlations between for landscape variables. Blue circles indicate positive correlations and red circles indicate negative correlations. The size of the circle indicates the strength of the correlation.

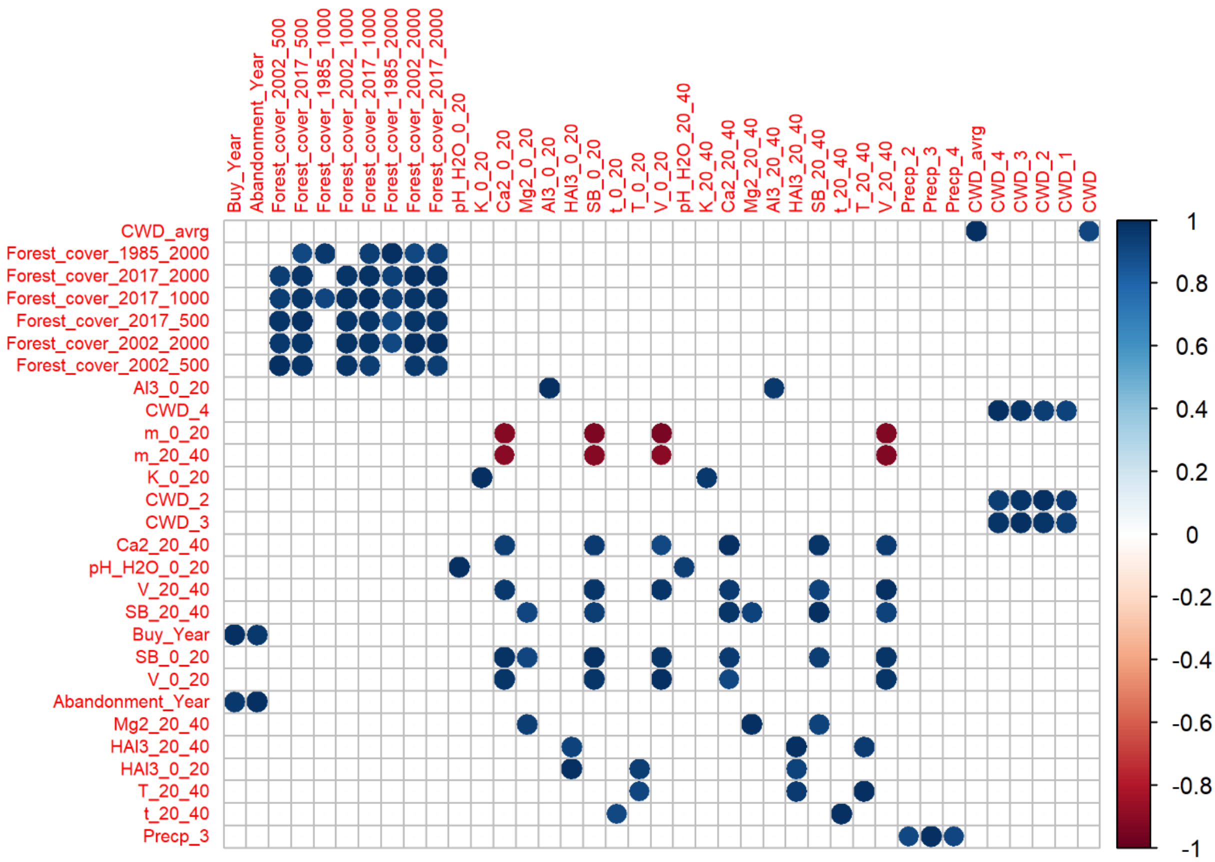

Figure S7. Pearson correlations highly correlated (greater than 90). Blue circles indicate positive correlations and red circles indicate negative correlations. The size of the circle indicates the strength of the correlation.