ALINY APARECIDA DOS REIS

PREDICTING *EUCALYPTUS* STAND ATTRIBUTES IN MINAS GERAIS STATE, BRAZIL: AN APPROACH USING MACHINE LEARNING ALGORITHMS WITH MULTI-SOURCE DATASETS

LAVRAS – MG
2018

ALINY APARECIDA DOS REIS

PREDICTING *EUCALYPTUS* STAND ATTRIBUTES IN MINAS GERAIS STATE, BRAZIL: AN APPROACH USING MACHINE LEARNING ALGORITHMS WITH MULTI-SOURCE DATASETS

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia Florestal, área de concentração em Ciências Florestais, para a obtenção do título de Doutor

Prof. Dr. José Márcio de Mello
Orientador

Prof. Dr. Fausto Weimar Acerbi Júnior
Prof. Dr. Antonio Carlos Ferraz Filho
Coorientadores

LAVRAS – MG
2018

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

Reis, Aliny Aparecida dos.
Predicting *Eucalyptus* stand attributes in Minas Gerais state, Brazil: an approach using machine learning algorithms with multi-source datasets / Aliny Aparecida dos Reis. – 2018.
188 p. : il.

Orientador: José Márcio de Mello.
Coorientadores: Fausto Weimar Acerbi Júnior, Antonio Carlos Ferraz Filho.
Tese (Doutorado) - Universidade Federal de Lavras, 2018.
Bibliografia.

1. Forest management. 2. Remote sensing. 3. Random Forest. I. Mello, José Márcio de. II. Acerbi Júnior, Fausto Weimar. III. Ferraz Filho, Antonio Carlos. IV. Título.

ALINY APARECIDA DOS REIS

PREDICTING *EUCALYPTUS* STAND ATTRIBUTES IN MINAS GERAIS STATE, BRAZIL: AN APPROACH USING MACHINE LEARNING ALGORITHMS WITH MULTI-SOURCE DATASETS

PREDIÇÃO DE ATRIBUTOS DENDROMÉTRICOS EM TALHÕES DE *EUCALYPTUS* NO ESTADO DE MINAS GERAIS, BRASIL: UMA ABORDAGEM UTILIZANDO ALGORITMOS DE APRENDIZAGEM DE MÁQUINA COM DADOS DE MÚLTIPLAS FONTES

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia Florestal, área de concentração em Ciências Florestais, para a obtenção do título de Doutor.

APROVADA em 26 de Novembro de 2018.

Prof. Dr. Fausto Weimar Acerbi Júnior	UFLA
Prof. Dr. Lucas Rezende Gomide	UFLA
Prof. Dr. Sérgio Henrique Godinho Silva	UFLA
Prof. Dr. Antonio Carlos Ferraz Filho	UFPI

Prof. Dr. José Márcio de Mello
Orientador

LAVRAS – MG
2018

Aos meus pais, Antônio e Maria Santa, às irmãs, Wal e Joyce e à sobrinha, Lavínia.

DEDICO.

AGRADECIMENTOS

Primeiramente, a Deus, por me iluminar e proteger sempre.

À minha família, pelo amor e apoio incondicionais e pela compreensão quando a minha ausência se fez necessária em muitos momentos importantes.

Ao meu orientador, José Marcio de Mello, pelo apoio, amizade, confiança e disponibilidade em todos os momentos.

Ao professor Fausto Weimar Acerbi Júnior, pela amizade, confiança, companheirismo, ensinamentos, incentivo e pelo seu exemplo, durante os 7 (sete) anos no LEMAF/UFLA

À Universidade Federal de Lavras e ao Programa de Pós-Graduação em Engenharia Florestal, pela oportunidade concedida para a realização deste trabalho.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq pela concessão das bolsas de estudos.

Aos engenheiros Leonardo M. Pires, Vinícius Rabelo e Thais Rosa, pela disponibilização da base de dados.

Aos membros da banca examinadora, que gentilmente aceitaram o meu convite e pelas valiosas contribuições.

Aos meus amigos do LEMAF, em especial, Duda, Marcela, Mônica, Thiza e Inácio por toda parceria e cumplicidade, principalmente, nos últimos anos de doutorado; e aos que já passaram pelo LEMAF, em especial, Juju, Pedro e Rafaella, pela amizade apesar da distância.

Ao meu namorado Weslley, pelo companheirismo, amor, paciência, cumplicidade e motivação.

Aos amigos de minha turma de Engenharia Florestal 2008/01, em especial, Lara, Thaís, Sérgio, Rafael, Thiago e Lucas pela amizade sincera; e às

amigas-irmãs da República Doce Deleite pela amizade verdadeira e pelos preciosos momentos que compartilhamos durante esses mais de 10 anos.

A todos que me ajudaram de alguma forma nesta jornada. Meus sinceros agradecimentos!

RESUMO GERAL

Informações espaciais e quantitativas sobre atributos da floresta são fundamentais ao manejo florestal, sendo importantes indicadores de processos biofísicos, da dinâmica florestal e da provisão de serviços e bens. Neste cenário, esta tese tem como objetivo principal testar a eficácia da integração de dados de campo, imagens óticas multiespectrais obtidas dos satélites Landsat 5 TM, Landsat 8 OLI e Sentinel-2A, dados de radar de abertura sintética (SAR), adquiridos pelo satélite Sentinel-1B, atributos de terreno e dados climáticos, para estimar e mapear características dendrométricas de plantios de *Eucalyptus*, ao Norte do estado de Minas Gerais, Brasil, utilizando métodos paramétricos e não paramétricos de predição espacial. Para tanto, esta tese foi organizada em quatro artigos. No primeiro artigo (1), os métodos Regressão Linear Múltipla (RLM), *Random Forest* (RF), *Support Vector Machine* (SVM) e Redes Neurais Artificiais (RNA) foram avaliados, para estimar e mapear área basal e volume, usando dados espectrais das imagens Landsat 5 TM. Entre os métodos de predição avaliados, o RF apresentou o melhor desempenho, para predizer e mapear área basal e volume, e, por esse motivo, foi utilizado nos artigos seguintes para a predição e mapeamento de características dendrométricas. No segundo artigo (2), diferentes combinações de dados (idade do povoamento, dados óticos multiespectrais Landsat 8 OLI, dados de radar Sentinel-1B e atributos de terreno) foram testadas para a estimativa volumétrica em um plantio de *Eucalyptus*. Os resultados mostraram que uma maior acurácia é obtida a partir da combinação de todos os conjuntos de dados. O terceiro artigo (3) investigou se a melhor resolução espacial das imagens multiespectrais do Sentinel-2A melhorariam as estimativas de características dendrométricas, comparada com as imagens multiespectrais Landsat-8 OLI, quando ambas as imagens foram combinadas com dados de radar Sentinel-1B e atributos de terreno. Como esperado, a melhor resolução espacial das imagens Sentinel-2A resultou em maior acurácia na previsão de características dendrométricas de plantios de *Eucalyptus*. No quarto artigo (4), o RF foi utilizado, para a classificação da produtividade florestal e para predizer o incremento médio anual máximo (IMA_{max}) de plantios de *Eucalyptus*, com base em atributos de terreno e dados climáticos em escala regional. Os atributos de terreno e as variáveis bioclimáticas mostraram bom potencial para classificar a produtividade florestal e para predizer o IMA_{max}. A grande e crescente área de plantações de eucalipto no Brasil e no mundo sugere que as novas abordagens aqui propostas, para a estimativa de atributos do povoamento e da produtividade florestal, apresentam grande potencial de suporte para o monitoramento e manejo de florestas plantadas.

Palavras-chave: Manejo florestal. Sensoriamento remoto. *Random Forest*. Atributos de terreno.

GENERAL ABSTRACT

Quantitative spatial information on forest attributes is critical in forest management as an important indicator of biophysical processes, forest dynamics and the provision of services and goods. In this thesis, the effectiveness of integrating field data, multispectral optical imagery obtained from Landsat 5 Thematic Mapper (TM), Landsat 8 Operational Land Imager (OLI), and Sentinel-2A satellites, synthetic aperture radar (SAR) data acquired by the Sentinel-1B satellite, digital terrain attributes derived from a digital elevation model (DEM) and climate data was tested using parametric and nonparametric methods of spatial prediction for estimating and mapping forest stand attributes in *Eucalyptus* plantations in northern Minas Gerais state, Brazil. For this purpose, this thesis was organized in four articles. In the first one (1), Multiple Linear Regression (MLR), Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Networks (ANN) methods were assessed to estimate and map basal area and volume using Landsat 5 TM data. RF showed the best performance for spatial prediction and mapping of stand attributes in *Eucalyptus* stands, and for this reason, it was used in the next three articles. In the second article (2), different combinations of stand age with variables extracted from three different digital datasets (*i.e.*, Landsat 8 OLI multispectral optical data, Sentinel-1B SAR data, and DEM-derived terrain attributes) were tested to estimate volume. The results showed that the best data combination corresponds to the integration of all datasets (*i.e.*, stand age and the selected variables of Landsat 8 OLI and Sentinel-1B SAR imagery, and DEM-derived terrain attributes). The third article (3) investigated the potential of Sentinel-2A multispectral information for improving forest attribute estimates compared with Landsat-8 OLI imagery when both multispectral optical imagery (*i.e.*, Sentinel-2A and Landsat 8 OLI) were combined with Sentinel-1B SAR data and DEM-derived terrain attributes. As expected, the Sentinel-2A optical data appeared to have a greater explanatory power in predicting forest attributes of *Eucalyptus* plantations compared to Landsat 8 OLI imagery. In the fourth article (4), a nonparametric modeling approach was used to examine relationships between terrain attributes and climate data on forest site productivity and maximum mean annual increment (MAI_{max}) of *Eucalyptus* plantations at a regional scale. Terrain attributes and bioclimatic variables showed good potential to classify site productivity and to predict MAI_{max} in our study area. The large and increasing area of *Eucalyptus* forest plantations in Brazil and elsewhere suggest that the new approaches developed here to estimate forest stand attributes and productivity have great potential to support *Eucalyptus* plantation monitoring and forest management practices.

Keywords: Forest Management. Remote Sensing. Random Forest. Terrain Attributes.

LISTA DE FIGURAS
LIST OF FIGURES

SEGUNDA PARTE – ARTIGOS

ARTIGO 1

ARTIGO 2

ARTIGO 3

ARTIGO 4

LISTA DE TABELAS
LIST OF TABLES

SEGUNDA PARTE – ARTIGOS

ARTIGO 1

ARTIGO 2

ARTIGO 3

ARTIGO 4

SUMÁRIO

FIRST PART

1 INTRODUCTION

The Brazilian forestry sector has intensely increased the productivity of *Eucalyptus* plantations through intensive silviculture, genetic selection, and breeding techniques (CAMPOE et al., 2013; GONÇALVES et al., 2013). The first commercial *Eucalyptus* plantations in Brazil were established in the early 1900s and, along with recent large-scale afforestation and reforestation efforts, such plantations have now expanded rapidly to cover more than 5.7 million hectares. These plantations produce around 17% of the harvested wood in the world and are estimated to have the capacity to absorb approximately 1.2 billion tons of carbon dioxide annually (BRAZILIAN TREE INDUSTRY - IBÁ, 2017).

Traditionally, monitoring of *Eucalyptus* forest growth in Brazilian forest plantations is conducted through annual field-based forest inventories (RAIMUNDO et al., 2017). However, in fast-growing forests like *Eucalyptus* plantations, field-based inventory surveys might not be sufficient to capture productivity differences across the entire area, such as those arising from losses due to pest and disease attacks (COOPS et al., 2006), or from climatic anomalies (GONZÁLEZ-GARCÍA et al., 2015; SCOLFORO et al., 2016). Furthermore, field-based forest inventory measurements are expensive, time consuming, and labour intensive.

In the past decade, advances in remote sensing technologies have allowed increasing detailed data collection, from which spatially-explicit information can be extracted to supplement the field-based forest inventory data collection, and to predict forest stand attributes (BAGHDADI et al., 2014, 2015; BERRA et al., 2012; CANAVESI; PONZONI; VALERIANO, 2010; DUBE; MUTANGA, 2015, 2016; GAMA; SANTOS; MURA, 2010, 2016;

GEBRESLASIE; AHMED; VAN AARDT, 2010, 2011; MURA et al., 2018; PULITI et al., 2018). In parallel with the advances in remote sensing technologies, computational techniques, such as machine-learning algorithms (MLA), have been increasingly used to model spectral and biological data, and have proven to yield high prediction accuracy in analysis of complex variable datasets in forestry and multi-source remote sensing (DUBE et al., 2015; FRANKLIN; AHMED, 2017; GAO et al., 2016; NOVELLI et al., 2016; ZHANG et al., 2018). These techniques overcome the difficulties of classical statistical methods such as spatial correlation, non-linearity of data, and overfitting (WERE et al., 2015), and have been used in several studies due to their stability and efficiency in terms of handling a large number of predictor variables (CASTILLO et al., 2017; HAWRYŁO; WĘŻYK, 2018; LÓPEZ-SERRANO et al., 2016; SHAO; ZHANG; WANG, 2017).

In this context, the main objective of this thesis was to investigate the effectiveness of integrating field data, multispectral optical imagery obtained from Landsat 5 Thematic Mapper (TM), Landsat 8 Operational Land Imager (OLI), and Sentinel-2A satellites, synthetic aperture radar (SAR) data acquired by the Sentinel-1B satellite, digital terrain attributes derived from a digital elevation model (DEM) and climate data using parametric and nonparametric methods of spatial prediction for estimating and mapping forest stand attributes in *Eucalyptus* plantations in northern Minas Gerais state, Brazil. For this purpose, this thesis was organized in four articles. In the first one (1), the performance of three machine-learning algorithms (*i.e.*, Random Forest, Support Vector Machine, and Artificial Neural Networks) were compared with the performance of multilinear regression model to estimate and map basal area and volume of a *Eucalyptus* plantation using Landsat 5 TM data. Furthermore, the first article investigated whether the addition of residual kriging in the spatial prediction methods leads to accuracy improvement in basal area and volume

estimates. In the second article (2), variables extracted from three different digital datasets (*i.e.*, Landsat 8 OLI multispectral optical data, Sentinel-1B SAR data, and terrain attributes) were used separately to estimate *Eucalyptus* plantation volume. Nextly, stand age was combined with the best multispectral optical, SAR, and terrain attribute variables to predict volume using the Random Forest (RF) machine learning algorithm, and the final estimated forest volumes were compared to the field-observations of volume in the independent forest inventory sample. The various models were interpreted and the best overall model was compared with the results obtained when using each dataset alone. The third article (3) investigated whether the object-based approach improves the estimation accuracy of *Eucalyptus* stand attributes compared with the pixel-based approach, and whether the better spatial resolution of Sentinel-2A optical images results in better predictions than when using Landsat 8 OLI data. In the fourth article (4), a nonparametric modeling approach was used to examine relationships between terrain attributes and climate data on forest site productivity and maximum mean annual increment (MAI_{max}) of *Eucalyptus* plantations at regional scale. The main objective of the fourth article was to determine whether geomorphometrics derived from a DEM and/or bioclimatic variables could be used as predictors of productive potential of *Eucalyptus* plantations across a large area of heterogeneous terrain and climate conditions in Minas Gerais state, Brazil.

2 LITERATURE REVIEW

This section reviews *Eucalyptus* plantation stand attribute estimation using multi-source remotely sensed data and ancillary data, such as DEM-derived terrain attributes and climate data, in comparison to field observations. Typically, the general approach is to extract relevant multispectral and SAR satellite data variables, such as vegetation indices and texture derivatives, and DEM-derived terrain attributes, such as slope and aspect, and develop a regression or statistical model.

In southern Brazil, Berra et al. (2012) estimated the volume of a *Eucalyptus* plantation using vegetation indices derived from Landsat 5 TM images with a coefficient of determination (R^2) equal to 0.7. Employing eight Hyperion EO-1 (Earth Observing-1) (total bands = 220) derived vegetation indices and multiple-linear regression models to estimate *Eucalyptus* stand volume in Brazil, Canavesi, Ponzoni and Valeriano (2010) obtained a Root Mean Square Error (RMSE) equal to 43.7 $m^3\ ha^{-1}$. Gama, dos Santos, and Mura (2010) used airborne interferometric and polarimetric SAR data in X and P bands to estimate the volume of *Eucalyptus* plantations and obtained RMSE = 33.6 $m^3\ ha^{-1}$.

Ismail et al. (2015) used ALOS PALSAR backscatter and SPOT 4 multispectral optical data to predict the volume in *Eucalyptus* plantations in Zululand, South Africa. Stand age – measured as an annual increment since planting date – was used as an independent variable in a regression model of volume, with one optical band (short-wave infrared or SWIR) and one SAR band (HV cross-polarization backscatter). The best model produced a RMSE of 31.7 $m^3\ ha^{-1}$, which was considered appropriate for operational forest management in this region. The combination of stand age and multi-source – optical and SAR – satellite remotely sensed data outperformed the use of the

different satellite datasets alone, especially in older stands with closed or high density canopies. Using ASTER satellite data, age, and site index as independent variables, Gebreslasie, Ahmed and Van Aardt (2010) estimated *Eucalyptus* volume with R^2 = 0.9. These results confirmed earlier work that suggested that, since *Eucalyptus* plantations are typically a single species monoculture and the undergrowth (shrubs and small trees) is sparse, a high proportion of tree stem-scattering was produced and represented in the SAR image dataset (GAMA; SANTOS; MURA, 2016).

The use of texture measures derived from multispectral optical images to improve volume estimates in *Eucalyptus* plantations has also been documented (DUBE; MUTANGA, 2015; GEBRESLASIE; AHMED; VAN AARDT, 2011). The general principle is that image texture derivatives can simplify and define complex forest canopy structures even in closed canopies, thus reducing saturation effects (SARKER; NICHOL, 2011). Recently, Dube and Mutanga (2015) demonstrated that certain small-window (or high frequency) texture derivatives from Landsat 8 OLI data helped improve aboveground biomass estimation in both *Eucalyptus* and *Pinus* plantations when compared to the use of multispectral reflectance data alone. They also successfully incorporated simple band ratios and spectral vegetation indices in their models. The use of SAR image texture in *Eucalyptus* plantation volume estimates has not yet been reported, though an earlier forest biomass estimation study with integrated multispectral and SAR data showed that textures were effective (CUTLER et al., 2012).

Spatial and temporal variations of *Eucalyptus* plantation growth are strongly affected by topography and climate conditions. In general, topography controls the hydrologic dynamics of catchments, nutrient, and soil conditions, while climate influences growth through variability of incoming solar radiation, soil moisture, and temperature (ADAMS; BARNARD; LOOMIS, 2014; DUBE

et al., 2017). Scolforo et al. (2017) illustrated that rainfall regime was the principal driver for *Eucalyptus* growth as implied by their fitted regression equations in their study in Bahia, northeast of Brazil. In a recent study, Dube and Mutanga (2016) improved aboveground biomass prediction accuracy (R^2 = 0.8 and RMSE= 19.6 t ha^{-1}) by integrating environmental variables (rainfall and temperature) and five DEM-based geomorphometric variables (slope, aspect, topographic wetness index, elevation, and insolation) with a WorldView-2 multispectral image of *Eucalyptus* and *Pinus* forest plantations in Umgeni Catchment, South Africa. In another South Africa study, Dube et al. (2017) demonstrated that integrating multispectral SPOT 5 image, stand age, and rainfall metrics significantly improved volume estimation in *Eucalyptus* plantations (R^2 = 0.8 and RMSE = 36.0 m^3 ha^{-1}).

Many of these studies used regression techniques or machine learning algorithms in forest volume or biomass model development. Among them, the Random Forest (RF) algorithm is a nonlinear and non-parametric ensemble decision-tree method (BREIMAN, 2001) that provides flexible, robust, and accurate predictive capabilities for high-dimensional datasets (BELGIU; DRĂGU, 2016). RF models for forest attribute estimation (GAO et al., 2016; LATIFI; NOTHDURFT; KOCH, 2010; LÓPEZ-SERRANO et al., 2016; WU et al., 2016) often outperform alternative methods, such as k-Nearest Neighbor (kNN), Support Vector Machine (SVM), Back Propagation Neural Networks (BPNN), and Stepwise Linear Regression (LMSTEP).

3 FINAL CONSIDERATIONS

Remotely sensed datasets have already been used to predict stand attributes of *Eucalyptus* plantations; however, the increasing availability of ancillary data offers new opportunities for improving forest attribute estimation accuracy. In addition, more research is required to investigate the capabilities of new high spectral, spatial and temporal resolution satellites, such as Sentinel-1B and Sentinel-2A combined with machine learning algorithms in predicting stand attributes of *Eucalyptus* plantations.

REFERENCES

ADAMS, H. R.; BARNARD, H. R.; LOOMIS, A. K. Topography alters tree growth-climate relationships in a semi-arid forested catchment. **Ecosphere**, New York, v. 5, n. 11, p. 1-16, Nov. 2014.

BAGHDADI, N. et al. Evaluation of ALOS/PALSAR L-band data for the estimation of *Eucalyptus* plantations aboveground biomass in Brazil. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, New York, v. 8, n. 8, p. 3802-3811, Aug. 2015.

BAGHDADI, N. et al. Testing different methods of forest height and aboveground biomass estimations from ICESat/GLAS data in *Eucalyptus* plantations in Brazil. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, New York, v. 7, n. 1, p. 290-299, Jan. 2014.

BELGIU, M.; DRĂGU, L. Random forest in remote sensing: a review of applications and future directions. **ISPRS Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 144, p. 24-31, 2016.

BERRA, E. F. et al. Estimativa do volume total de madeira em espécies de eucalipto a partir de imagens de satélite Landsat. **Ciência Florestal**, Santa Maria, v. 22, n. 4, p. 853-864, 2012.

BRAZILIAN TREE INDUSTRY. **Brazilian tree industry:** annual report. São Paulo, 2017.

BREIMAN, L. Random forests. **Machine Learning**, Boston, v. 45, n. 1, p. 5-32, 2001.

CAMPOE, O. C. et al. Stem production, light absorption and light use efficiency between dominant and non-dominant trees of *Eucalyptus grandis* across a productivity gradient in Brazil. **Forest Ecology and Management**, Amsterdam, v. 288, p. 14-20, 2013.

CANAVESI, V.; PONZONI, F. J.; VALERIANO, M. M. Estimativa de volume de madeira em plantios de *Eucalyptus* spp. utilizando dados hiperespectrais e dados topográficos. **Revista Árvore**, Viçosa, MG, v. 34, n. 3, p. 539-549, jun. 2010.

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

COOPS, N. C. et al. Assessment of QuickBird high spatial resolution imagery to detect red attack damage due to mountain pine beetle infestation. **Remote Sensing of Environment**, New York, v. 103, n. 1, p. 67-80, 2006.

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

DUBE, T. et al. Predicting *Eucalyptus* spp. stand volume in Zululand, South Africa: an analysis using a stochastic gradient boosting regression ensemble with multi-source data sets. **International Journal of Remote Sensing**, Basingstoke, v. 36, n. 14, p. 3751-3772, July 2015.

DUBE, T. et al. Stand-volume estimation from multi-source data for coppiced and high forest *Eucalyptus* spp. silvicultural systems in KwaZulu-Natal, South Africa. **ISPRS Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 132, p. 162-169, Oct. 2017.

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

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

FRANKLIN, S.; AHMED, O. Object-based Wetland characterization using Radarsat-2 Quad-Polarimetric SAR Data, Landsat-8 OLI Imagery, and airborne lidar-derived geomorphometric variables. **Photogrammetric Engineering & Remote Sensing**, Falls Church, v. 83, n. 1, p. 27-36, Jan. 2017.

GAMA, F. F.; SANTOS, J. R. dos; MURA, J. C. Continuous monitoring of biophysical Eucalyptus sp. parameters using interferometric synthetic aperture

radar data in P and X bands. **Journal of Applied Remote Sensing**, Orlando, v. 10, n. 2, p. 26002-1-26002-15, Apr./June 2016.

GAMA, F. F.; SANTOS, J. R. dos; MURA, J. C. *Eucalyptus* biomass and volume estimation using interferometric and polarimetric SAR Data. **Remote Sensing**, Basel, v. 2, p. 939-956, 2010.

GAO, T. et al. Timber production assessment of a plantation forest: an integrated framework with field-based inventory, multi-source remote sensing data and forest management history. **International Journal of Applied Earth Observation and Geoinformation**, Enschede, v. 52, p. 155-165, 2016.

GEBRESLASIE, M. T.; AHMED, F. B.; VAN AARDT, J. A. N. Extracting structural attributes from IKONOS imagery for eucalyptus plantation forests in Kwazulu-Natal, South Africa, using image texture analysis and artificial neural networks. **International Journal of Remote Sensing**, Basingstoke, v. 32, n. 22, p. 677-7701, 2011.

GEBRESLASIE, M. T.; AHMED, F. B.; VAN AARDT, J. A. N. Predicting forest structural attributes using ancillary data and ASTER satellite data. **International Journal of Applied Earth Observation and Geoinformation**, Enschede, v. 12S, p. S23-S26, 2010.

GONÇALVES, J. L. de M. et al. Integrating genetic and silvicultural strategies to minimize abiotic and biotic constraints in Brazilian eucalypt plantations. **Forest Ecology and Management**, Amsterdam, v. 301, p. 6-27, 2013.

GONZÁLEZ-GARCÍA, M. et al. Dynamic growth and yield model including environmental factors for *Eucalyptus nitens* (Deane & Maiden) Maiden short rotation woody crops in Northwest Spain. **New Forests**, Dordrecht, v. 46, n. 3, p. 387-407, 2015.

HAWRYŁO, P.; WĘŻYK, P. Predicting growing stock volume of Scots Pine stands using Sentinel-2 satellite imagery and airborne image-derived point clouds. **Forests**, Basel, v. 9, n. 5, p. 274-289, May 2018.

ISMAIL, R. et al. Assessing the utility of ALOS PALSAR and SPOT 4 to predict timber volumes in even-aged Eucalyptus plantations located in Zululand, South Africa. **Southern Forests:** A Journal of Forest Science, London, v. 77, n. 3, p. 203-211, 2015.

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

LÓPEZ-SERRANO, P. M. et al. A comparison of machine learning techniques applied to Landsat-5 TM spectral data for biomass estimation. **Canadian Journal of Remote Sensing**, Ottawa, v. 42, n. 6, p. 690-705, 2016.

MURA, M. et al. Exploiting the capabilities of the Sentinel-2 multi spectral instrument for predicting growing stock volume in forest ecosystems. **International Journal of Applied Earth Observation and Geoinformation**, Enschede, v. 66, p. 126-134, 2018.

NOVELLI, A. et al. Performance evaluation of object based greenhouse detection from Sentinel-2 MSI and Landsat 8 OLI data: a case study from Almería (Spain). **International Journal of Applied Earth Observation and Geoinformation**, Enschede, v. 52, p. 403-411, 2016.

PULITI, S. et al. Combining UAV and Sentinel-2 auxiliary data for forest growing stock volume estimation through hierarchical model-based inference. **Remote Sensing of Environment**, New York, v. 204, p. 485-497, 2018.

RAIMUNDO, M. R. et al. Geostatistics applied to growth estimates in continuous forest inventories. **Forest Science**, Bethesda, v. 63, n. 1, p. 29-38, Feb. 2017.

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

SCOLFORO, H. F. et al. Incorporating rainfall data to better plan *Eucalyptus* clones deployment in Eastern Brazil. **Forest Ecology and Management**, Amsterdam, v. 391, p. 145-153, 2017.

SCOLFORO, H. F. et al. Modeling dominant height growth of eucalyptus plantations with parameters conditioned to climatic variations. **Forest Ecology and Management**, Amsterdam, v. 380, p. 182-195, 2016.

SHAO, Z.; ZHANG, L.; WANG, L. Stacked sparse autoencoder modeling using the synergy of airborne LiDAR and satellite optical and SAR data to map forest above-ground biomass. **IEEE Journal of Selected Topics in Applied Earth**

Observations and Remote Sensing, New York, v. 10, n. 12, p. 1-14, Dec. 2017.

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

WU, C. et al. Comparison of machine-learning methods for above-ground biomass estimation based on Landsat imagery. **Journal of Applied Remote Sensing**, Orlando, v. 10, n. 3, p. 35010-1-35010-17, July/Sept. 2016.

ZHANG, J. et al. Estimating aboveground biomass of *Pinus densata***-**dominated forests using Landsat time series and permanent sample plot data. **Journal of Forestry Research**, London, p. 1-18, 2018.

SECOND PART – ARTICLES

ARTICLE 1 - SPATIAL PREDICTION OF BASAL AREA AND VOLUME IN *EUCALYPTUS* STANDS USING LANDSAT TM DATA: AN ASSESSMENT OF PREDICTION METHODS

Aliny Aparecida dos Reis[1*], Mônica Canaan Carvalho[1], José Marcio de Mello[1], Lucas Rezende Gomide[1], Antônio Carlos Ferraz Filho[1] and Fausto Weimar Acerbi Junior[1]

[1]Department of Forest Science, Federal University of Lavras – UFLA, PO Box 3037, Lavras, Minas Gerais, Brazil, Zip Code 37200-000.
E-mails: alinyreis@hotmail.com, monicacanaan@gmail.com, josemarcio@dcf.ufla.br, lucasgomide@dcf.ufla.br, acferrazfilho@gmail.com, fausto@dcf.ufla.br

Publication status: Published in the **New Zealand Journal of Forestry Science**
DOI: 10.1186/s40490-017-0108-0

Background: In fast-growing forests such as *Eucalyptus* plantations, the correct determination of stand productivity is essential to aid decision making processes and ensure the efficiency of the wood supply chain. In the past decade, advances in remote sensing and computational methods have yielded new tools, techniques, and technologies that have led to improvements in forest management and forest productivity assessments. Our aim was to estimate and map the basal area and volume of *Eucalyptus* stands through the integration of forest inventory, remote sensing, parametric, and nonparametric methods of spatial prediction.

Methods: This study was conducted in twenty 5-year-old clonal stands (362 ha) of *Eucalyptus urophylla* S.T.Blake x *Eucalyptus camaldulensis* Dehnh. The stands are located in the northwest region of Minas Gerais state, Brazil. Basal area and volume data were obtained from forest inventory operations carried out in the field. Spectral data were collected from a Landsat 5 TM satellite image, composed of spectral bands and vegetation indices. Multiple linear regression (MLR), Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Networks (ANN) methods were used for basal area and volume estimation. Using ordinary kriging, we spatialised the residuals generated by the spatial prediction methods for the correction of trends in the estimates and more detailing of the spatial behavior of basal area and volume.

Results: The ND54 index was the spectral variable that had the best correlation values with basal area (r = -0.91) and volume (r = -0.52), and was also the variable that most contributed to basal area and volume estimates by the MLR and RF methods. The RF algorithm presented smaller basal area and volume errors when compared to other machine learning algorithms and MLR. The addition of residual kriging in spatial prediction methods did not necessarily result in relative improvements in the estimations of these methods.

Conclusions: Random Forest was the best method of spatial prediction and mapping of basal area and volume in the study area. The combination of spatial prediction methods with residual kriging did not result in relative improvement of spatial prediction accuracy of basal area and volume in all methods assessed in this study, and there is not always a spatial dependency structure in the residuals of a spatial prediction method. The approaches used in this study provide a frame work for integrating field and multispectral data, highlighting methods that greatly improve spatial prediction of basal area and volume estimation in *Eucalyptus* stands. This frame work has potential to support fast growth plantation monitoring, offering options for a robust analysis of high-dimensional data.

Keywords: forest inventory, machine-learning algorithms, multiple linear regression, random forest, support vector machine, artificial neural networks.

Background

The Brazilian forestry sector represents an important share of the products, taxes, jobs, and income generation of the country, and accounts for 3.5% of the national GDP (IBÁ 2015). This is in large part due to the successful establishment of fast-grown plantations of *Eucalyptus* species, which currently occupy around 5.6 million hectares (71.9% of the total planted forest area in Brazil) and represent 17% of the harvested wood in the world (IBÁ 2014, IBÁ 2015).

The *Eucalyptus* genus has more than 500 species, and a subset of which are used in fast-growing plantations (Barrios et al., 2015), commonly located in tropical and sub-tropical regions, and more recently in temperate regions. Spain (González-García et al. 2015), Portugal (Lopes et al. 2009), Uruguay (Barrios et al. 2015), Chile (Watt et al. 2014), South Africa (Dye et al. 2004), Australia (Verma et al., 2014), and the United States (Wear et al., 2015) are some

examples of productive *Eucalyptus* plantations in temperate regions that have cutting cycles ranging from 8 to 12 years. In tropical regions such as Brazil, the cutting cycles of *Eucalyptus* plantations range from 5 to 7 years (Guedes et al. 2015, Scolforo et al. 2016).

Timber production is the main ecosystem service of planted forests and the main management objective for these plantations (Gao et al., 2016). In the case of fast-growing plantations, the correct determination of stand productivity is essential to support forest management planning strategies (González-García et al. 2015, Retslaff et al. 2015). Traditionally, productivity assessments of a plantation are carried out based on field measurements of the diameter at breast height (DBH) and tree height via forest inventory. However, in fast-growing plantations, field-based inventory programmes may not be sufficient to capture productivity differences across the entire area, such as those arising from losses due to pest and disease attacks (Coops et al. 2006), or from climatic anomalies (González-García et al. 2015, Scolforo et al. 2016).

In the past decade, advances in Geographical Information Systems (GIS), Global Positioning Systems (GPS), and remote sensing have provided new tools, techniques, and technologies to support forest management. Thus, low-cost and accurate forest productivity assessment can be made, as well as allowing the collection of information in areas not sampled by forest inventory (Morgenroth and Visser 2013). The analysis of remote sensing information combined with field data has been used by several authors to fill the information gap left by data collected only in the field (Watt et al. 2016, Boisvenue et al. 2016, Moreno et al. 2016, Fayad et al. 2016, Vicharnakorn et al. 2014). Ponzoni et al. (2015) used data collected from Landsat 5 TM images for spectral-temporal characterisation of *Eucalyptus* canopies. Berra et al. (2012) estimated the volume of a *Eucalyptus* plantation in the southern region of Brazil from Landsat 5 TM images. Canavesi et al. (2010) used hyperspectral data from the

Hyperion EO-1 sensor for the volume estimation of *Eucalyptus* plantations under different relief conditions. The results found by these authors corroborate the potential use of data collected by remote sensing to estimate the productivity of *Eucalyptus* plantations.

In parallel to the advances in remote sensing, computational techniques, such as machine-learning algorithms (MLA), have been increasingly used to model spectral and biological data. These techniques overcome the difficulties of classical statistical methods such as spatial correlation, non-linearity of data, and overfitting (Were et al. 2015). In addition, these algorithms allow the use of categorical data, with statistical noise and incomplete data, and therefore are able to address needs under different dataset scenarios (Breiman 2001).

Several studies have shown the superiority of machine-learning algorithms in relation to classical statistics in several areas, such as in forest management. For instance, Ahmed et al. (2015) modelled a Landsat time-series data structure in conjunction with LiDAR data and found that the Random Forest algorithm achieved better results than multiple regression for all forest classes. In another study, García-Gutiérrez et al. (2015) found that machine-learning algorithms (mainly Support Vector Machine) were superior for modelling a range of forest variables (viz., aboveground biomass, basal area, dominant height, mean height, and volume) compared with multiple linear regression. Machine-learning algorithms have also been shown to provide an economical and accurate way to estimate aboveground biomass in forests from Landsat satellite images (Wu et al. 2016). These studies highlight the benefits of applying more robust techniques in solving problems previously resolved by traditional statistical modeling.

In this context, the aims of this study were: (i) to estimate and map basal area and volume of a *Eucalyptus* plantation through the integration of forest inventory, remote sensing, and parametric and nonparametric methods of spatial

prediction; (ii) to compare the performance of machine-learning algorithms (Random Forest, Support Vector Machine, and Artificial Neural Networks) with the linear regression model, and (iii) to assess the improvement in basal area and volume estimation with the addition of residual kriging in spatial prediction methods.

Methods

Study area

The study area is located in Minas Gerais state, the fourth-largest state in Brazil, with an area of 586,521 km². Minas Gerais state has the largest area occupied by plantations of the *Eucalyptus* genus in the country (1,400,232 ha), corresponding to 25.2% of Brazilian *Eucalyptus* plantations. The wood from these plantations is mainly used for the production of charcoal, as well as pulp, lumber, and panels (IBÁ 2015).

The *Eucalyptus* clonal stands under study are located in Lagoa Grande municipality, in the northwest of Minas Gerais state (lat. 17º43'00"S - 17º44'00"S, long. 46° 32'00"W - 46°33'00"W, elevation 560 m a.s.l.) (Figure 1). According to the Köppen climatic classification system, the climate in this region is Aw, classified as a tropical savanna climate, with drier months during the winter, high annual precipitation in the summer and average temperature of all months greater than 18 °C (Alvares et al 2013). The average annual rainfall and the average monthly rainfall of the dry and wet seasons are 1,430 mm, 8 mm, and 257 mm, respectively.

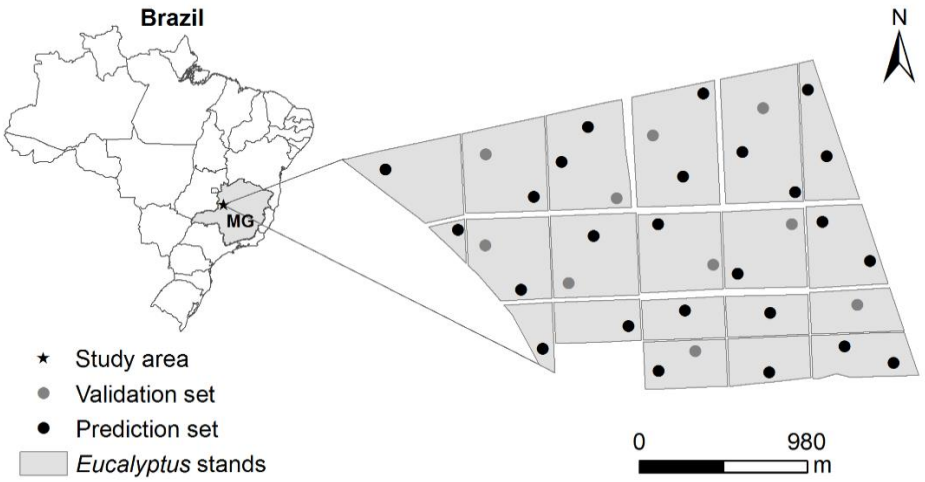


Figure 1. Geographic location of the *Eucalyptus* stands and sampling grid.

Field data description and sampling

This study was undertaken in a set of 20 clonal stands of *Eucalyptus urophylla* S.T.Blake x *Eucalyptus camaldulensis* Dehnh, totaling an area of 362.2 ha. These stands were planted in April and May 2004, with initial spacing of either 3 x 2 m or 3 x 3 m. The forest inventory was carried out in June and July 2009 on a set of 35 georeferenced square plots of 400 m^2. The plots were georeferenced in the field with GPS (Garmin 60CSx, Garmin Ltd., Olathe, Kansas, USA). The sampling procedure adopted was systematic, allocating approximately one plot per 10 ha of forest. In each plot, the diameter at breast height (DBH) of all stems was measured, as well as the total height of the first fifteen trees with normal stems (without bifurcation or any other defect) and height of dominant trees (the 100 largest diameter trees per hectare). Descriptive statistics of the variables collected in the field are shown in Table 1. Estimates of basal area (m^2 ha^{-1}), and total stem volume (m^3 ha^{-1}) were obtained from the information collected in the plots.

Table 1. Descriptive statistics of the variables collected in the field.

Statistic	DBH	H	Hd
Minimum	11.98	16.98	19.40
Maximum	15.45	24.63	26.38
Mean	14.02	21.18	22.98
Standard deviation	0.85	2.33	1.92

Where: DBH = diameter at breast height (cm); H = total height (m); Hd = dominant height (m).

Remote sensing data and processing

Spectral data were obtained from a Landsat 5 TM satellite image, with spatial resolution of 30 m, on the date of 06/25/2009, corresponding with field data collection, in orbit 220, point 072, in bands TM1 (0.45 – 0.52 μm), TM2 (0.52-0.60 μm), TM3 (0.63-0.69 μm), TM4 (0.76-0.90 μm), TM5 (1.55 – 1.75 μm), and TM7 (2.18 – 2.35 μm). The Landsat 5 TM Surface Reflectance Climate Data Record (CDR) was used, which is a Landsat Level-2A product generated by the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) (Masek et al. 2006) obtained from the USGS database (United States Geological Survey) (USGS, 2017). These images already contain radiometric calibration, and geometric and atmospheric corrections.

In addition, vegetation indices using the red, near infrared and short wave infrared spectral bands of Landsat 5 TM (Table 2) were calculated, as described by Lu et al. (2004) and Ponzoni et al. (2012). The Normalized Difference Vegetation Index (NDVI) is the most widely used vegetation index for retrieval of forest biophysical parameters (Rouse et al. 1973, Lu et al. 2004). The Soil-Adjusted Vegetation Index (SAVI) and Modified Soil-Adjusted Vegetation Index (MSAVI) are soil adjusted vegetation indices used to reduce the effect of soil background reflectance (Qi et al. 1994). The Enhanced Vegetation Index (EVI) index was developed to optimise the vegetation signal, correcting reflected light distortions caused by particulate matter suspended in

the air, as well as by influence of background data under the vegetation canopy (Justice et al. 1998). The Global Environment Monitoring Index (GEMI) minimises atmospheric effects, similar to the EVI and minimises observational angular effects in the observed vegetation index signal (Pinty and Verstraete 1992).

Table 2. Vegetation indices used in the spectral characterisation of the *Eucalyptus* stands.

Vegetation indices	Formulation	Reference
NDVI	(TM4 - TM3)/(TM4 + TM3)	Rouse et al. (1973)
ND53	(TM5 - TM3)/(TM5 + TM3)	Huete et al. (2002)
ND54	(TM5 - TM4)/(TM5 + TM4)	Huete et al. (2002)
ND57	(TM5 - TM7)/(TM5 + TM7)	Huete et al. (2002)
SAVI	[(TM4 - TM3)/(TM4 + TM3 + 0.5)].(1.5)	Huete (1988)
MSAVI	$\left[(2TM4+1)-\sqrt{(2TM4+1)^2-8(TM4-2TM3)}\right]/2$	Qi et al. (1994)
EVI	2.5* [(TM4-TM3)/(TM4+6TM3- 7.5TM1+1)]	Justice et al. (1998)
GEMI	$n(1-0.25n).[(TM3-0.125)/(1-TM3)]$ $n=\frac{2(TM4^2-TM3^2)+1.5TM4+0.5TM3}{TM4+TM3+0.5}$	Pinty and Verstraete (1992)

Where: TM= Thematic Mapper, ND= Normalized Difference, NDVI= Normalized Difference Vegetation Index; SAVI= Soil-Adjusted Vegetation Index; MSAVI= Modified Soil-Adjusted Vegetation Index; EVI= Enhanced vegetation Index; GEMI= Global Environment Monitoring Index.

Dataset integration

The choice of an appropriate pixel size is one of the issues to be considered when using remote sensing data to estimate dendrometric characteristics. Due to easy accessibility and affordability, a number of studies have employed Landsat images and found statistically significant correlations between remotely-sensed data and dendrometric characteristics using field plots ranging from 315 m² to 2,500 m² (Dube and Mutanga 2015, López-Sánchez et al. 2014, Zhang et al. 2014, López-Serrano et al. 2016).

Although the size of a single plot (20 x 20 m) in this study does not cover a Landsat pixel, we considered that a plot represents an area larger than its size. As the sampling design was 1 plot per hectare, we ensured that each plot matched with the reference pixel in order to extract reliable data.

Spatial modelling and prediction methods

Exploratory data analysis

Spectral response was extracted from the Landsat TM bands and vegetation indices from the geographical coordinates of the forest inventory plots. Thus, the plot database was composed of basal area ($m^2\ ha^{-1}$), volume ($m^3\ ha^{-1}$), spectral bands values, and vegetation indices values. The total database (35 plots) was systematically divided into two datasets: prediction or fitting set (70% of the database) and validation set (30% of the database). Therefore, 25 plots were used for basal area and volume predictions, and 10 plots were used for validation of the different approaches to estimate basal area and volume in the *Eucalyptus* stands under study.

Pearson correlation analysis was carried out among basal area, volume, values of spectral bands and vegetation indices. From these correlations, the relationship between the dendrometric characteristics of *Eucalyptus* stands and its spectral response in Landsat images was explored.

Multiple linear regression analysis (MLR)

Basal area and volume estimation were accomplished through multiple linear regression analysis (MLR). A stepwise variable elimination method was used in conjunction with the Akaike Information Criterion (AIC) to select only those spectral variables that “best” explained basal area and volume variation. The residuals from regression models were analysed to assess the existence of trends in the errors. The Variance Inflation Factor (VIF) was used to detect

possible correlations between explanatory variables (multicollinearity). The adopted VIF cutoff value was 10.

Random Forest (RF)

The Random Forest (RF) algorithm, initially proposed by Breiman (2001), is an ensemble method that generates a set of individually trained decision trees and combines their results. The greatest advantage of these decision trees as regression methods is that they are able to accurately describe complex relationships among multiple variables, and by aggregating these decision trees, more accurate solutions are generated (Gleason and Im 2012). In addition to these characteristics, RF is an easy parameterisation method (Immitzer et al. 2012). This method has shown great potential in regression studies with integration of spectral data, in some cases generating better results than conventional techniques (Stojanova et al. 2010, Dube et al. 2014, García-Gutiérrez et al. 2015, Görgens et al. 2015, Wu et al. 2016). The RF algorithm fitted in this study is implemented in the open-source software WEKA 3.8 (Frank et al. 2016). Tests were carried out with the exchange of tree numbers and attribute numbers to be drawn. Then, 20 trees with 10 attributes to be drawn by node for basal area and 80 trees and 11 attributes for volume were fixed.

Support Vector Machine (SVM)

Support Vector Machines (SVMs) operate by assuming that each set of inputs will have a unique relation to the response variable, and that the grouping and the relation of these predictors to one another is sufficient to identify rules that can be used to predict the response variable from new input sets. For this, SVMs project the input space data into a feature space with a much larger dimension, enabling linearly non-separable data to become separable in the feature space. This method has been successfully used in forestry classification

problems (Huang et al. 2008, Shao and Lunetta 2012) and more recently in regression problems with the use of spectral data (García-Gutiérrez et al. 2015, Wu et al. 2016). The Kernel function used in the present study was the Gaussian or radial basis function (RBF). The algorithm used is implemented in WEKA 3.8 software under the Sequential Minimal Optimization (SMO) function. Values of parameters C and σ (bandwidth or influence range of each training point in the RBF) were tested within the interval $(10^i)_{i=-3,-2,-1,0,1,2,3}$, where the least squared mean error configuration was chosen for application. For basal area and volume, selected C and σ values were 10 and 0.1, and 100 and 0.01, respectively.

Artificial Neural Networks (ANN)

Artificial Neural Networks (ANNs) are a parallel-distributed information processing system that simulates the working of neurons in the human brain, being able to learn from examples. Artificial Neural Networks are widely used to model complex and non-linear relations between inputs and outputs or to determine patterns in data (Diamantopoulou 2012). The use of this technique in conjunction with remote sensing data is consolidated in several studies (Cluter et al. 2012, García-Gutiérrez et al. 2015, Rodriguez-Galiano et al. 2015, Were et al. 2015). We used the ANN obtained by running the Multilayer Perceptron function (of the multilayer perceptron type) provided by WEKA 3.8 software. The training of neural networks occurred through the back-propagation algorithm, which fit the weights of all the layers of the network from the backpropagation of the error, obtained in the output layer. The weights updating was carried out according to the error, learning rate, and momentum terms (Delta rule). The sigmoidal activation function was employed in all neurons. Determined by previous tests, ANNs were structured with 14 neurons in the input layer (number of variables), 1 neuron in the hidden layer, and 1 neuron in

the output layer, corresponding to estimated basal area or volume. The learning rate, the momentum term, and iteration numbers were fixed at 0.3, 0.5, and 500 for basal area, and 0.2, 0.7, and 500 for volume, respectively.

Relative importance evaluation

The variable importance was assessed for each model with a removal-based approach in order to avoid the limited interpretability of the MLA and to verify how each independent variable contribute to the performance of machine-learning algorithms (RF, SVM, and ANN). All algorithms were adjusted *n* times, with *n* being the number of available variables. At each time, one variable was removed from the training set and then the root mean square error (RMSE) of the algorithm was quantified. At the end, the obtained errors were normalised by the ratio of the largest RMSE so that they were between 0 and 1 and multiplied by 100 (Were et al. 2015). The variable that results in the highest RMSE when removed from the database is the variable with the highest relative importance within the model. This methodology was chosen because it can be consistently applied to all algorithms, allowing comparisons of variable contribution between the methods.

Geostatistical modelling of prediction methods errors

Spatial prediction methods capture the average behavior of the main variable, allowing the identification of its general spatial behavior, without detailing more specific areas or regions. For details of specific regions, estimates obtained exclusively from the auxiliary variables need to be corrected. Thus, residuals generated by spatial prediction methods (MLR, RF, SVM, and ANN) were used for the correction of trends in the estimates and for detailing the spatial behaviour of the main variables (basal area and volume) using ordinary kriging. The interpolated values of the residuals were then added to the estimates

of the spatial prediction methods (MLR, RF, SVM, and ANN). Thus, we obtained the basal area and volume estimates corrected by the ordinary kriging of the residuals for each spatial prediction method.

For the application of ordinary kriging to the spatial prediction method residuals, we considered the stationarity presupposition of the intrinsic hypothesis (Journel and Huijbregts 1978), through fitting of theoretical functions to experimental semivariogram models. Spherical, exponential, and Gaussian models were fitted to the semivariogram of the residuals from each spatial prediction method using Weighted Least Squares. The semivariogram parameters (nugget (τ^2), sill (σ^2), and range (ϕ)) were calculated from the best fitted models, which provided information about the spatial structure as well as input parameters for the kriging interpolation. The nugget represents the minimum semivariance among different sampling intervals. Nugget values greater than zero represent a combination of experimental error and of unresolved spatial variability occurring at scales smaller than inter-sampling lag distance. Sill is the plateau reached by the values of semivariance, and indicates the amount of variation than can be explained by the spatial structure of the data. Range is the distance at which the semivariogram reaches the plateau, indicating the distance which values are spatially correlated. The evaluation of the performance of each semivariogram model and the selection of the best models were based on cross-validation, which estimates the reduced average error (RAE) and the standard deviation of the reduced average error (SRE) (Yamamoto and Landim 2013).

Validation and assessment of the prediction methods

The different approaches to basal area and volume estimation of *Eucalyptus* stands were evaluated by comparing the basic statistics of the predicted maps (mean and standard deviation) with the estimates obtained from

the forest inventory, and through the discrepancies between observed and predicted values in the fitting and validation datasets. These discrepancies were evaluated using the mean error (ME), the mean absolute error (MAE), and the root mean square error (RMSE), as described in Equations 1 – 3.

$$ME=\frac{1}{N}\sum_{i=1}^{N}(X_i - \widehat{X}_i) \quad (1)$$

$$MAE=\frac{1}{N}\sum_{i=1}^{N}\left|X_i - \widehat{X}_i\right| \quad (2)$$

$$RMSE=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \widehat{X}_i)^2} \quad (3)$$

where N is the number of values in the dataset; $\widehat{X}_i$ is the estimated value of the main variable; X_i is the observed value in the prediction and validation sets.

The Relative Improvement (RI) achieved by residual kriging for a particular spatial prediction method was calculated by comparing the change in RMSE when the residual kriging was applied using Equation 4.

$$RI=\frac{RMSE_{spm}- RMSE_{spm\text{-}RK}}{RMSE_{spm}}*100\% \quad (4)$$

where $RMSE_{spm}$ is the root mean square error of a spatial prediction method, $RMSE_{spm\text{-}RK}$ is the root mean square error of the spatial prediction method when residual kriging is added to this method.

Data analysis for this study was performed using the following software: R (R Core Team 2016) with the geoR package (Ribeiro Júnior and Diggle 2001), WEKA 3.8 (Frank et al. 2016), and ArcGis version 10.1 (Esri 2010) with Geostatistical Analyst extension (Esri 2010).

Results

Descriptive statistic of the measured basal area and volume

Basal area ranged from 10.07 to 21.63 $m^2 ha^{-1}$, with average of 16.86 $m^2 ha^{-1}$ and standard deviation of 2.4 $m^2 ha^{-1}$ (Table 3). The average volume was 169.34 $m^3 ha^{-1}$ with a standard deviation of 29.66 $m^3 ha^{-1}$ and range from 95.80

up to 213.85 m^3 ha^{-1}. Basal area had a lower coefficient of variation (CV = 14.26%) compared to volume (CV = 17.51%), demonstrating a considerable homogeneity of this dendrometric characteristic in the evaluated *Eucalyptus* stands.

Table 3. Descriptive statistics obtained from forest inventory processing using the estimators of Simple Random Sampling (SRS).

Estimators	Basal Area		Volume	
	(m^2)*	**(m^2 ha^{-1})**	**(m^3)***	**(m^3 ha^{-1})**
Minimum	0.91	10.07	8.62	95.80
Maximum	1.95	21.63	19.25	213.85
Mean	1.52	16.86	15.24	169.34
Standard deviation	0.22	2.4	2.67	29.66
Coefficient of variation (%)	14.26		17.51	
Sampling error (%)	4.89		6.00	
Total confidence interval	5807.9 – 6405.0		57652.7 – 65018.7	

*Estimates obtained for an area of 900 m² (corresponding to the area of each pixel of the Landsat images).

Correlation among basal area, volume, spectral bands, and vegetation indices

The correlation between plot basal area and the different spectral bands and their ratios (Table 4) ranged from -0.91 (ND54) to 0.15 (TM2). The SAVI, MSAVI, GEMI, and EVI indices were also highly correlated with basal area ($r > 0.85$). The correlation between plot volume and the spectral bands and ratios ranged from -0.52 (ND54) to -0.02 (TM2). The NDVI ($r = 0.49$) and SAVI ($r = 0.47$) indices also had high correlations with volume, but these were lower in magnitude when compared with those for basal area. Many of the spectral bands and ratios were also highly correlated with each other ($r > 0.90$), which can be

considered a drawback due to possible multicollinearity problems in linear regression models.

Table 4. Pearson's correlation coefficient (r) among basal area, volume, and spectral data for the *Eucalyptus* stands.

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. G	1.00															
2. V	0.70*	1.00														
3. TM1	-0.24^{ns}	0.10^{ns}	1.00													
4. TM2	0.15^{ns}	-0.02^{ns}	0.59*	1.00												
5. TM3	-0.20^{ns}	-0.10^{ns}	0.80*	0.72*	1.00											
6. TM4	0.82*	0.41*	-0.05^{ns}	0.43*	0.12^{ns}	1.00										
7. TM5	-0.66*	-0.36^{ns}	0.53*	0.31^{ns}	0.68*	-0.40*	1.00									
8. TM7	-0.68*	-0.40*	0.56*	0.32^{ns}	0.66*	-0.42*	0.90*	1.00								
9. NDVI	0.83*	0.49*	-0.53*	-0.13^{ns}	-0.55*	0.75*	-0.78*	-0.82*	1.00							
10. ND53	-0.60*	-0.32^{ns}	-0.29^{ns}	-0.50*	-0.37^{ns}	-0.66*	0.43*	0.31^{ns}	-0.31^{ns}	1.00						
11. ND54	-0.91*	-0.52*	0.31^{ns}	-0.09^{ns}	0.30^{ns}	-0.86*	0.80*	0.78*	-0.93*	0.65*	1.00					
12. ND57	0.45*	0.27^{ns}	-0.49*	-0.28^{ns}	-0.49*	0.27^{ns}	-0.50*	-0.82*	0.60*	0.00^{ns}	-0.48*	1.00				
13. SAVI	0.88*	0.47*	-0.23^{ns}	0.25^{ns}	-0.12^{ns}	0.97*	-0.57*	-0.60*	0.89*	-0.57*	-0.94*	0.41*	1.00			
14. MSAVI	0.88*	0.45*	-0.36^{ns}	0.13^{ns}	-0.27^{ns}	0.92	-0.65*	-0.67*	0.94*	-0.50*	-0.95*	0.46*	0.99*	1.00		
15. GEMI	0.86*	0.45*	-0.14^{ns}	0.34^{ns}	0.00^{ns}	0.99*	-0.49*	-0.52*	0.83*	-0.62*	-0.91*	0.35^{ns}	0.99*	0.96*	1.00	
16. EVI	0.87*	0.42*	-0.41*	0.12^{ns}	-0.28^{ns}	0.92*	-0.64*	-0.67*	0.94*	-0.48*	-0.94*	0.47*	0.98*	1.00*	0.96*	1.00

Where: V= volume (m^3 ha^{-1}); G= basal area (m^2 ha^{-1}); TM= Thematic Mapper; ND= Normalized Difference; NDVI= Normalized Difference Vegetation Index; SAVI= Soil-Adjusted Vegetation Index; MSAVI= Modified Soil-Adjusted Vegetation Index; GEMI= Global Environment Monitoring Index; EVI= Enhanced vegetation Index; ^{ns}= not significant at 5% and *significant at 5%.

Spatial prediction of basal area and volume by MLR, RF, SVM, and ANN

The spectral data examined had several significant correlations with the basal area and volume data (Table 4). However, they contributed in a reduced form to the regression models due to multicollinearity problems, which resulted in final regression models with few significant explanatory variables (Table 5). The basal area model only included the ND54 vegetation index (Table 5), while the volume model included the TM1 band and NDVI. The coefficient of determination was high for the basal area model (R^2 = 0.81), but was much lower for the volume model (R^2 = 0.37).

Table 5. Regression model fitted for basal area and volume estimation for the *Eucalyptus* stands.

Model	β_0	β_1	β_2	R^2_{aj}	S_{xy}	S_{xy} (%)
$G = \beta_0 + \beta_1 ND54$	0.78***	-1.80***	-	0.81	0.09	5.76
$V = \beta_0 + \beta_1 NDVI + \beta_2 TM1$	-24.11*	42.69***	241.61*	0.37	2.01	13.08

Where: G= basal area (m^2); V= volume (m^3); β_0, β_1, and β_2= coefficients; R^2_{aj}= adjusted coefficient of determination; S_{xy}= residual standard error; TM= Thematic Mapper; ND= Normalized Difference, NDVI= Normalized Difference Vegetation Index; ***significant at 1%, and *significant at 10%.

In the case of basal area and volume predictions using machine-learning algorithms, the increases in RMSEs when the predictors were excluded one by one from the SVM, ANN, and RF models are shown in Figure 2. The variable ranking by relative importance differed for each algorithm. The ND54 index, chosen for basal area model by the MLR, also had the greatest effect on the accuracy of the RF model, both for basal area and volume. The TM2 band had the highest relative importance for the ANN and SVM models of both basal area and volume. The TM1 band, selected by the MLR for volume estimation, also had high importance in the ANN and SVM models of volume.

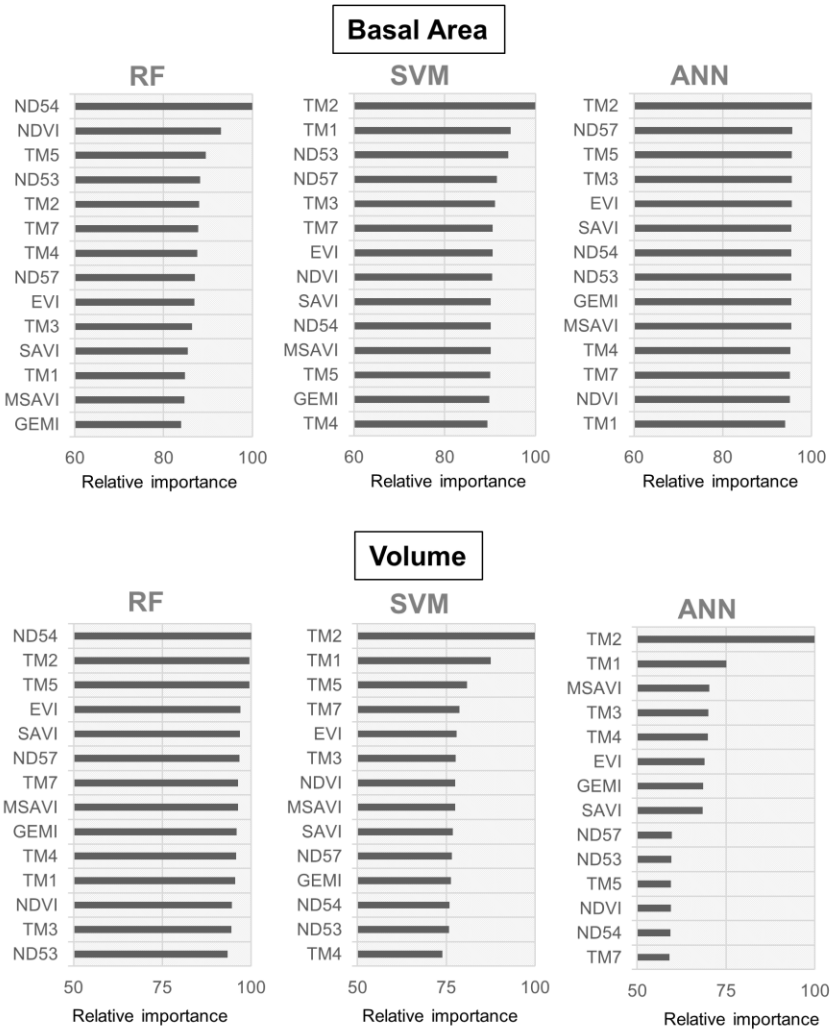


Figure 2. Relative importance of the variables within each machine-learning algorithm: RF, SVM and ANN for basal area and volume.

Comparisons of measured values and estimated values of basal area (Figure 3) showed that basal area was underestimated by the ANN model (Figure 3d). The model fitted using the RF algorithm produced values of basal area that were in closer agreement with measured values (Figure 3b). Similar

results were seen for the volume models, but with a slight overestimation for the plots with small volumes and an underestimation of the plots with high volumes. The model fitted using ANN algorithm did not produce estimates of volume that were consistent with measured values (Figure 3h). The models fitted using the MLR and SVM (Figures 3e and 3g) algorithms produced predicted values that were more closely related to the measured values than those from the ANN algorithm.

Prediction and validation sets of basal area and volume were compared by means of a Student's t test, in order to check if they provided unbiased subsets of the original data (Viana et al., 2012). Average basal area (17.03 $m^2 ha^{-1}$) and volume (171.10 $m^3 ha^{-1}$) obtained from the prediction set did not statistically differ from average basal area (16.45 $m^2 ha^{-1}$) and volume (164.92 $m^3 ha^{-1}$) obtained from the validation set, considering the two-tailed Student's t test (Basal area: $t= 0.629^{ns}$, df= 33, p-value= 0.533; Volume: $t=0.550^{ns}$, df= 33, p-value= 0.585).

The evaluation of spatial prediction methods, based on prediction and validation sets, was done by comparing the statistics presented in Equations 1 through 4 (Table 6). The mean error (ME) should ideally be close to zero if the prediction method is unbiased, and the values of this parameter suggested that all predictions generated impartial estimates when evaluated from both prediction and validation sets. Both the MAE and RMSE showed that basal area estimates were more accurate than volume estimates for all spatial prediction methods. The MAE and RMSE results obtained from the validation set demonstrated that there were no significant differences among the MLR, RF, SVM, and ANN for basal area estimates. For the volume estimates, the models fitted by SVM had the best performance and MLR the poorest performance.

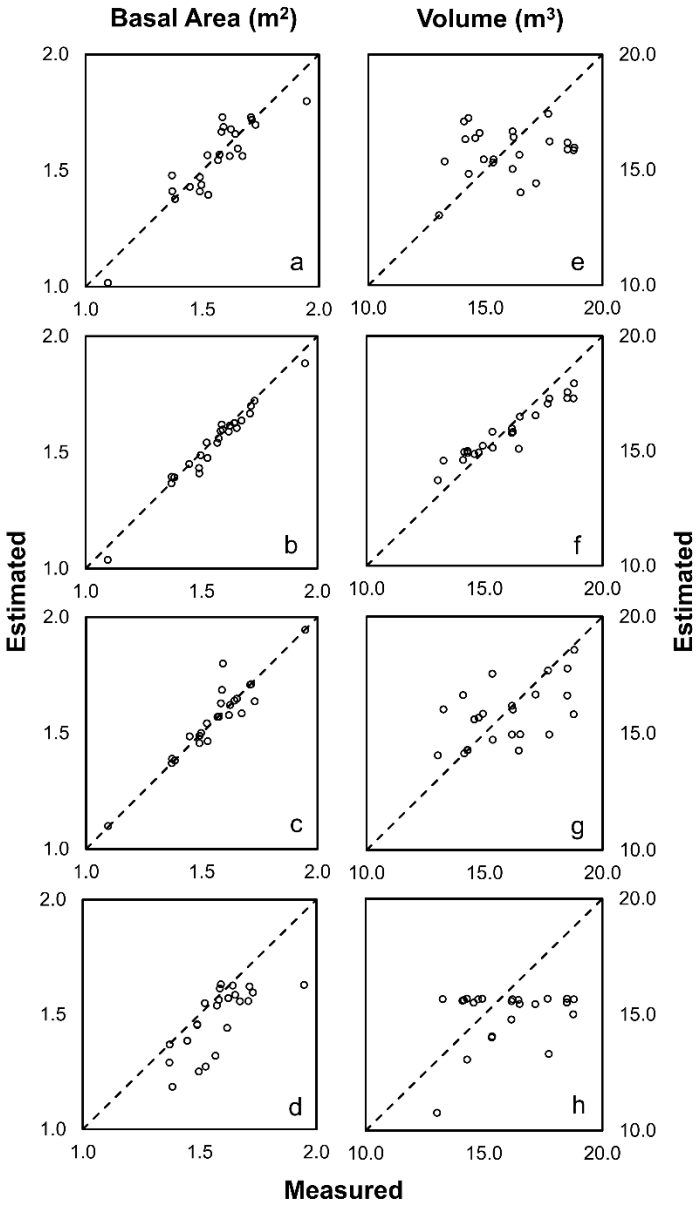


Figure 3. Scatter plots of measured values versus estimated values by: MLR (a) and (e); RF (b) and (f); SVM (c) and (g); and ANN (d) and (h) for basal area and volume, respectively. A 1:1 line (black, dashed) is provided for reference.

Table 6. Prediction methods evaluation using the prediction and validation sets for the *Eucalyptus* stands.

Method	Statistic	Basal area error (m^2)		Volume error (m^3)	
		Prediction set	**Validation set**	**Prediction set**	**Validation set**
MLR	ME	0.00	-0.05	0.00	-0.74
	MAE	0.07	0.09	1.56	2.08
	RMSE	0.08	0.14	1.89	2.48
	RMSE (%)	5.50	9.42	12.27	16.72
RF	ME	0.01	-0.03	0.08	-0.90
	MAE	0.03	0.09	0.62	1.63
	RMSE	0.04	0.14	0.73	2.21
	RMSE (%)	2.48	9.54	4.77	14.91
SVM	ME	-0.01	-0.05	0.00	-0.66
	MAE	0.04	0.09	1.19	1.59
	RMSE	0.06	0.14	1.60	2.02
	RMSE (%)	4.14	9.39	10.41	13.58
ANN	ME	0.09	0.03	0.94	0.45
	MAE	0.10	0.09	1.70	1.68
	RMSE	0.14	0.13	1.98	2.05
	RMSE (%)	8.87	8.52	12.88	13.82

Where: MLR = multiple linear regression; RF = random forest; SVM = support vector machine; ANN = artificial neural networks; ME = mean error; MAE = mean absolute error; RMSE = root mean square error.

Geostatistical modelling of prediction methods errors

The semivariogram models were selected based on RAE and SRE values close to zero and one, respectively (Yamamoto and Landim 2013). The experimental semivariograms constructed from the residuals of the basal area and volume prediction methods had a spatial dependence structure defined in six of the eight analysed situations (Figure 4 and Table 7). The volume residuals from MLR and ANN methods had a pure nugget effect, *i.e.*, no spatial

dependence structure. This result indicated a random spatial distribution of the residuals in these two situations.

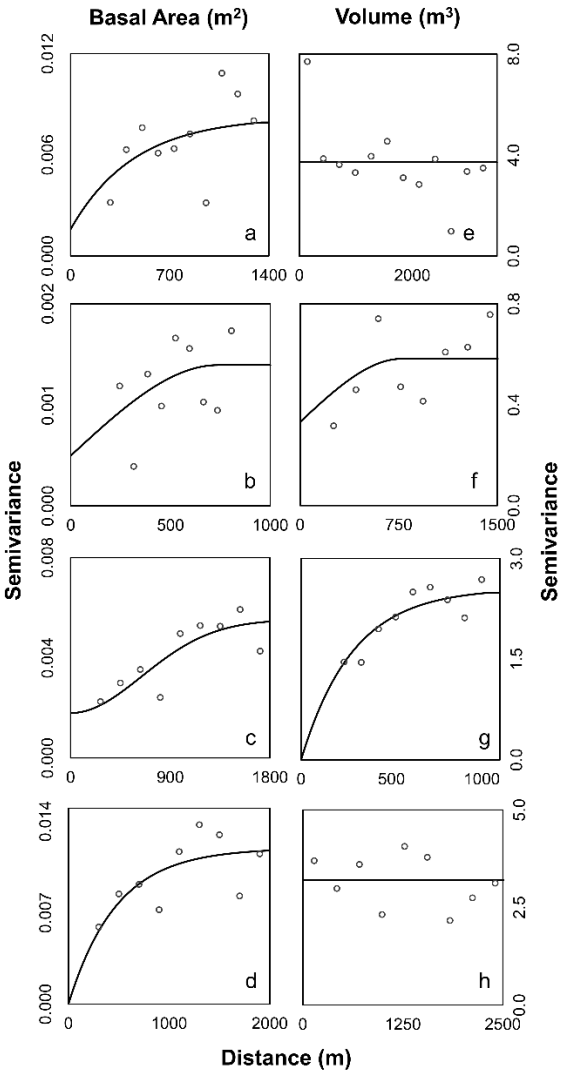


Figure 4. Experimental semivariograms of residuals from: MLR (a) and (e); RF (b) and (f); SVM (c) and (g); and ANN (d) and (h) for basal area and volume, respectively.

Table 7. Nugget (τ^2), sill (σ^2), and range (ϕ) parameters for the selected semivariance function models for each of the variables in study.

Variables	Residual	Selected Model	τ^2	σ^2	ϕ (m)	RAE	SRE
Basal area	MLR	Exponential	0.0016	0.0067	1350	-0.0092	1.0818
	RF	Spherical	0.0004	0.0009	737	-0.0079	1.0586
	SVM	Gaussian	0.0017	0.0037	1577	0.0089	0.9610
	ANN	Exponential	0.0000	0.0119	1430	-0.0303	1.1393
Volume	MLR	Exponential	PNE	PNE	PNE	PNE	PNE
	RF	Spherical	0.3316	0.2505	773	-0.0051	1.0258
	SVM	Exponential	0.0000	2.5582	858	-0.0039	0.9958
	ANN	Exponential	PNE	PNE	PNE	PNE	PNE

Where: MLR = multiple linear regression; RF = random forest; SVM = support vector machine; ANN = artificial neural networks; RAE = reduced average error; SRE = standard deviation of the reduced average error; PNE = pure nugget effect.

The residuals of the spatial prediction methods that had defined spatial dependence structures (Figure 4) were interpolated using ordinary kriging, and their estimates were added to basal area and volume estimates of the respective spatial prediction methods. The Relative Improvement (RI) of the addition of basal area residual kriging by the ANN method was 25%, *i.e.*, there was a reduction from 8.52% to 6.37% in the RMSE (Table 8). For the RF method, the RMSE increased from 9.54% to 10.08%, which corresponds to a 5.7% increase in the error of the basal area estimates by kriging of the residuals. For the volume, the addition of residual kriging improved the precision of SVM estimates, and reduced the precision of the RF estimates.

Table 8. Prediction methods with addition of the residual estimation by ordinary kriging using the prediction and validation sets for the *Eucalyptus* stands.

Method	Statistic	Basal area error (m^2)		Volume error (m^3)	
		Prediction set	**Validation set**	**Prediction set**	**Validation set**
MLR-RK	ME	0.00	-0.03	-	-
	MAE	0.03	0.09	-	-
	RMSE	0.04	0.14	-	-
	RMSE (%)	2.80	9.30	-	-
	RI	49.09	1.27	-	-
RF-RK	ME	0.01	-0.05	0.00	-1.03
	MAE	0.04	0.10	0.63	1.70
	RMSE	0.05	0.15	0.77	2.26
	RMSE (%)	3.08	10.08	5.02	15.25
	RI	-24.19	-5.66	-5.24	-2.28
SVM-RK	ME	0.01	-0.03	-0.32	-0.57
	MAE	0.05	0.10	0.80	1.22
	RMSE	0.06	0.15	1.11	1.74
	RMSE (%)	4.09	9.83	7.19	11.72
	RI	1.21	-4.69	30.93	13.70
ANN-RK	ME	0.02	-0.06	-	-
	MAE	0.04	0.06	-	-
	RMSE	0.09	0.09	-	-
	RMSE (%)	5.79	6.37	-	-
	RI	34.72	25.23	-	-

Where: MLR = multiple linear regression; RF = random forest; SVM = support vector machine; ANN = artificial neural networks; RK = residual estimation by ordinary kriging; ME = mean error; MAE = mean absolute error; RMSE = root mean square error; RI = Relative Improvement.

Mapping of basal area and volume for *Eucalyptus* stands

Basal area and volume estimates obtained by different spatial prediction methods (Table 9) had average values very close to each other, and were in agreement with the forest inventory estimates (Table 3). Only the ANN method

generated underestimated values for both basal area and volume, the total values of basal area and volume were not within the confidence interval generated by the forest inventory.

Table 9. Statistics of basal area and volume maps estimated by spatial predictions methods MLR, RF, SVM, and ANN.

Method	Basal area (m^2)					Volume (m^3)				
	Min	**Max**	**Mean**	**Standard deviation**	**Total estimate**	**Min**	**Max**	**Mean**	**Standard deviation**	**Total estimate**
MLR	0.62	1.83	1.52	0.20	6151.9	4.51	19.99	15.30	2.30	61739.5
MLR-RK	0.65	1.93	1.52	0.21	6141.0	-	-	-	-	-
RF	0.96	1,89	1.51	0.17	6101.5	9.26	18.08	15.27	1.81	61600.1
RF-RK	0.93	1.93	1.53	0.17	6166.6	9.02	18.37	15.36	1.91	61965.7
SVM	0.88	2.12	1.57	0.18	6326.2	1.36	19.64	15.31	2.57	61760.7
SVM-RK	0.76	2.10	1.56	0.19	6284.2	1.10	21.78	15.29	2.92	61683.8
ANN	0.97	1.65	1.42	0.22	5715.3	8.32	15.68	13.93	2.70	56223.8
ANN-RK	0.90	1.94	1.50	0.23	6070.3	-	-	-	-	-

Where: Min= minimum value; Max= maximum value; MLR= multiple linear regression; RF= random forest; SVM= support vector machine; ANN= artificial neural networks; RK= residual estimation by ordinary kriging.

Maps showing the spatial distribution of basal area and volume identified the same areas with high and low productivity, regardless of the spatial prediction method (Figures 5 and 6). The maps obtained by ANN had a smaller difference between maximum and minimum estimated values for basal area and volume, while the mapping obtained from the SVM models had a greater difference between these values. MLR and RF methods provided similar estimates in the basal area and volume mapping.

The addition of residual kriging in the basal area and volume mapping (Figure 7) resulted in a greater difference between maximum and minimum estimated values in all spatial prediction methods. For ANN, residual kriging

resulted in estimates that were more in agreement with the field observations, correcting the basal area underestimation behavior for the *Eucalyptus* stands under study. However, the addition of residual kriging to the models fitted by RF and SVM methods did not result in significant differences in basal area and volume mapping, and also led to increases in estimation errors in non-sampled areas in the field (Table 8).

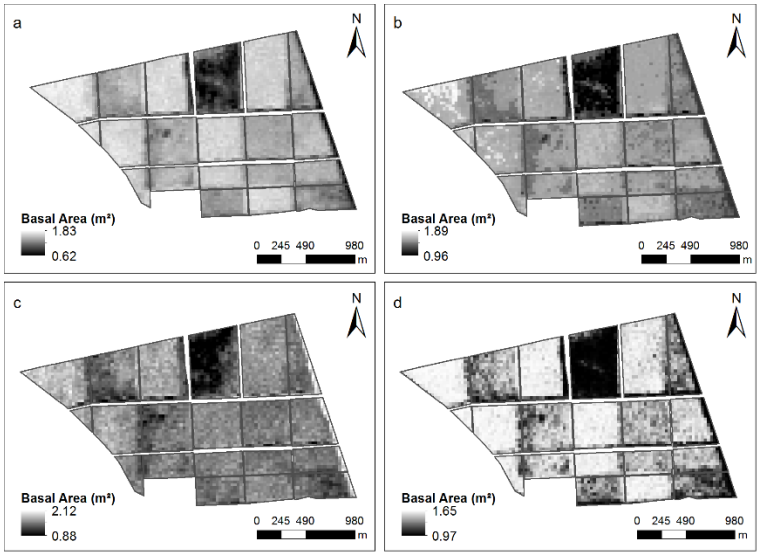


Figure 5. Spatial distribution of the basal area in *Eucalyptus* stands, estimated by: MLR (a), RF (b), SVM (c) and ANN (d).

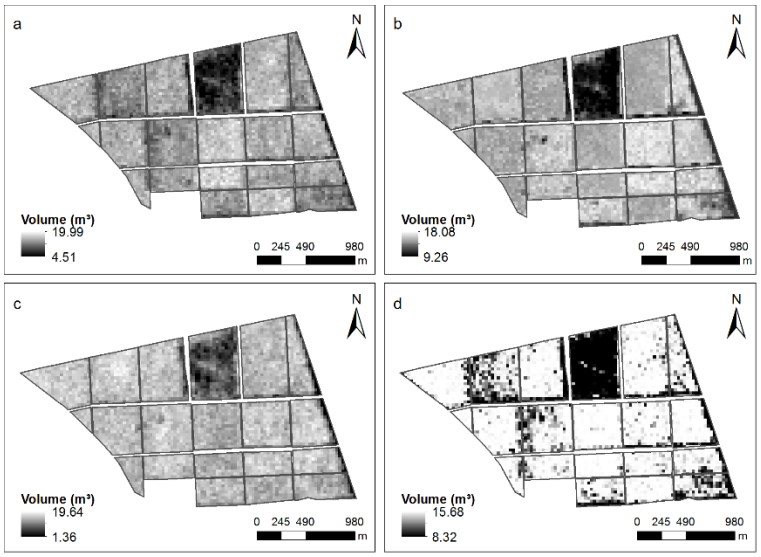


Figure 6. Spatial distribution of the volume in *Eucalyptus* stands, estimated by: MLR (a); RF (b); SVM (c); and ANN (d).

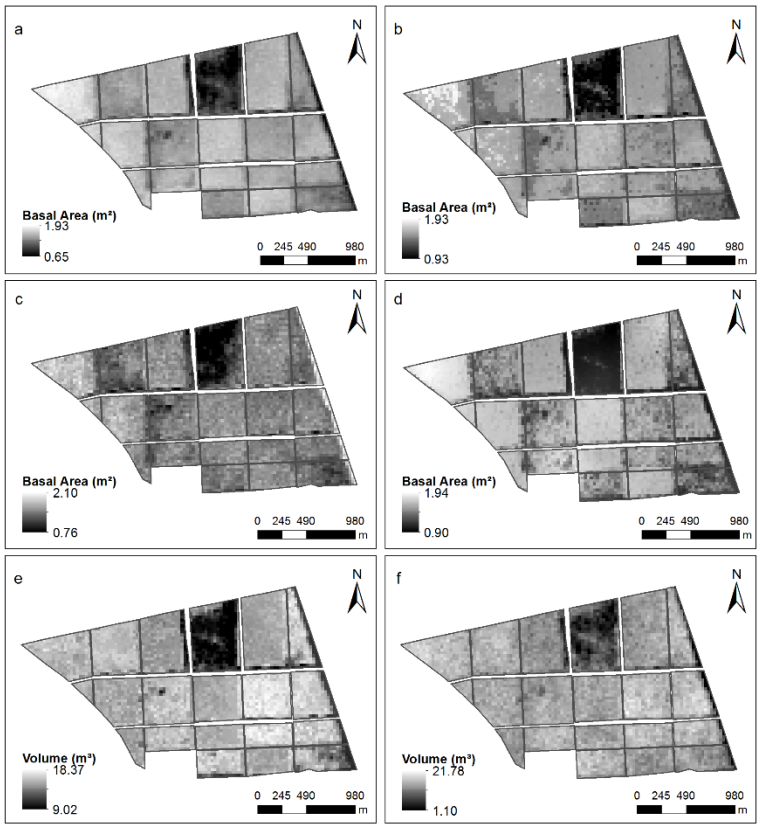


Figure 7. Spatial distribution of the basal area in *Eucalyptus* stands estimated by: MLR (a): RF (b); SVM (c); and ANN (d) with addition of the residual estimation by ordinary kriging; and for volume estimated by RF (e) and SVM (f) with addition of the residual estimation by ordinary kriging.

Discussion

Remote detection of forest canopies is complex due to the size, shape, and dielectric properties of its scatter elements (leaves, branches, and stems) (Galeana-Pizaña et al. 2014). The spatial diversity of forest canopies makes the relationship between forest parameters and remote sensing data a major challenge, although several studies have already demonstrated correlation

between spectral data and forest characteristics of interest (Stojanova et al. 2010, Viana et al. 2012, Castillo-Santiago et al. 2013, Fayad et al. 2016, Gao et al. 2016). For instance, plantations comprised of different *Eucalyptus* species may have very similar values of basal area and volume, but have different spectral characteristics due to differences in spectral behaviour of the species that form the canopies. Also, according to Ponzoni et al. (2015), the canopy reflectance of older *Eucalyptus* plantations (between 4 and 6 years) tend to contain a greater contribution from green leaves and a lower contribution from shadows, the background, and from dry branches inside the canopies than the canopy reflectance of young *Eucalyptus* plantations (<4 years). Thus, the canopy reflectance of older *Eucalyptus* plantations generated highest correlations with bands of the infrared region of the electromagnetic spectrum, and, therefore, with vegetation indices that include these bands in their compositions (Ponzoni et al. 2015). These results are consistent with the best correlations found in this study among the infrared bands, vegetation indices derived from these bands, basal area, and volume. This same behaviour was observed in the studies of Gebreslasie et al. (2008), Canavesi et al. (2010), Berra et al. (2012), and Pacheco et al. (2012).

Basal area was more strongly correlated with the spectral data because this variable is derived from only the diameter of the trees, which is directly related to size of the tree canopies, and determines the canopy reflectance (Ponzoni et al. 2012). On the other hand, volume is derived from the diameter, form factor, and height of the trees. Height estimates are obtained from empirical equations that add errors during the volume estimation process. This acts to reduce the strength of relationships between volume and variables obtained from remotely sensed images. The ND54 index was the spectral variable that had the strongest correlation with basal area ($r = -0.91$) and volume ($r = -0.52$). However, it was also significantly correlated with the other spectral

variables. During multiple linear regression analysis, the fact that two or more explanatory variables are highly correlated may generate multicollinearity problems in the fitted models, since one of the regression assumptions is that no linear relationship may exist between any independent variables or linear combinations of them (Montgomery et al., 2006).

For the MLR method, the best volume estimation model was obtained from the TM1 band and the NDVI index (Table 5), yet was only able to explain approximately 37% of the variation in this stand attribute. Conversely, the best model for basal area estimation used the ND54 index as the predictor variable and was able to explain more than 80% of the variation in this attribute, confirming the explanatory power of spectral data for basal area estimation in *Eucalyptus* stands. Gebreslasie et al. (2010) assessed the suitability of both visible and shortwave infrared ASTER data and vegetation indices for estimating forest structural attributes of *Eucalyptus* species in southern KwaZulu Natal, South Africa. These authors applied a MLR using MSAVI and band 3 as predictor variables and were able to explain slightly more of the variation in basal area ($R^2 = 0.67$) than volume ($R^2 = 0.65$). Although the MLR model for volume does not have a high R^2, the spectral data can efficiently explain the volumetric variations in non-sampled areas in the field. In a similar study for *Eucalyptus* stands located in the southern region of Brazil, Berra et al. (2012) concluded that spectral data obtained from Landsat images were efficient in mapping the volume in the study area, even when the regression models did not present high coefficients of determination ($R^2 <0.70$).

Divergence among variables that were deemed important between the different methods was observed with the machine-learning algorithms. For basal area modeling, the ND54 and NDVI indices had a higher importance value for RF. Statistically, these indices had high correlation values with the variable of interest (r=-0.91 and 0.83, respectively) and high multicollinearity (r=-0.93).

The ND54 index also was the variable that most contributed to the volume estimate by the RF method. The fact that the explanatory variables are correlated does not affect the performance of these algorithms. These methods do not rely on underlying assumptions about the data, which allows them to work with all available explanatory variables, without loss of information in the process of variables selection and reduction (Görgens et al. 2015). For the models fitted using ANN and SVM algorithms, the TM2 band was the most important predictor variable for basal area and volume. The linear correlation between this variable and basal area and volume is low to non-existent (r=0.15 and -0.02, respectively). However, this band is usually applied in vegetation vigor assessment (Meng et al., 2009), a characteristic that is indirectly related to volume and basal area, and which may explain the greater contribution of the TM2 band in the ANN and SVM algorithms, since trees that are more vigorous tend to have higher values of basal area and volume.

The models of basal area and volume developed by the RF algorithm had smaller errors compared with those developed by other machine-learning algorithms and MLR. The performance of this algorithm has been successful in many modeling and remote sensing studies (Lafiti et al. 2010, Rodriguez-Galiano et al. 2015, Wu et al. 2016). In the study by Shataee et al. (2012), volume prediction models developed by RF performed better than those developed using k-Nearest Neighbor (k-NN) and SVM. Employing ASTER satellite data, the relative RMSE obtained for all three volume models was higher than for the models developed in our study: 28.54% for k-NN, 25.86% for SVM, and 26.86% for RF, and only the RF algorithm produced unbiased volume estimations. For basal area, RF produced models with lower RMSE (18.39%) when compared with SVM (RMSE = 19.35%) and k-NN (RMSE = 20.20%); however, only k-NN was able to generate unbiased estimation compared with the other two algorithms used.

One of the positive features of RF is that it achieves satisfactory performance even with a limited number of samples and with many independent variables (attributes), as in the case of this current study. It is an ensemble method, which combines several regression trees to generate an average estimate, in which different attributes are used in each tree, making the results take into account the information of all available attributes. Stojanova et al. (2010) also concluded that ensemble methods (RF) were significantly better in height and canopy cover modeling using remote sensing data than single- and multi-target regression trees. The ANN and SVM algorithms also have proven good performance and robustness in several studies (e.g. Shao and Lunetta 2012, Were et al. 2015). However, the parameterisation of these methods is laborious, and they are very sensitive to the variation of input parameters, with ANN being more sensitive than other methods (Rodriguez-Galiano et al. 2015). This same behaviour was observed in this study, where the use of a restricted dataset by ANN resulted in estimates that were not compatible with the forest inventory estimates (Tables 3 and 9).

The addition of residual kriging in spatial prediction methods did not necessarily result in relative improvements in the estimation of these methods. In the case of MLR and ANN methods, residual kriging contributed to better accuracy of the basal area estimates. These results are consistent with the results of Dai et al. (2014), who reported that the combination of the residual kriging with artificial neural networks provides an improvement in the estimate accuracy of the variables of interest. The combination of MLR with residual kriging also provided improvements in estimates in the studies of Viana et al. (2012), Castillo-Santiago et al. (2013), and Galeana-Pizaña et al. (2014). For basal area and volume estimation, the addition of residual kriging in the RF and SVM methods resulted in a lower precision of the estimates. Hybrid methods are advantagous in the ability to use spatial information (ordinary kriging of

residuals) and non-spatial information (multiple linear regression analysis and machine-learning algorithms). However, in some situations, hybrid methods provide less accurate estimates in regions where the data collected in the field are sparse (Palmer et al., 2009).

The high growth rate of *Eucalyptus* stands in Brazil reinforces the importance of robust methods that consider auxiliary information in the process of estimating variables of interest, such as basal area and volume. The methodologies presented here are powerful tools for estimating basal area and volume from spectral data obtained from Landsat 5 TM or from other multispectral optical sensors. According to Görgens et al. (2015), machine-learning algorithms can continuously learn from new data, and keep all the accumulated knowledge of previous datasets. This fact allows the implementation of these algorithms in other situations where only limited amounts of data are available. The use of all auxiliary variables in the estimation process is another advantage over traditional regression methods, since machine-learning algorithms are not restricted by correlation between input variables, thus avoiding the loss of important information in the estimation process of the variable of interest. Nevertheless, these methods have as disadvantage the lack of transparency of the resulting models, so an alternative to overcome this obstacle is the evaluation of the relative importance of the explanatory variables. Furthermore, the causal relation between inputs and outputs of the estimation process is not clear, which implies a limited biological interpretation (Aertsen et al., 2010, Özçelik et al., 2013).

The results from the current study do need to be interpreted cautiously, as they are limited to a homogenous and relatively small study area. While this work uses a small number of plots, it represents the sampling intensity adopted by most Brazilian forestry companies, *i.e.*, one plot (usually 200–500 m² in size) for each 10 hectares of *Eucalyptus* plantation (Raimundo et al., 2017, Scolforo et

al., 2016) and the results from this research showcase the importance of using remotely sensed data and robust prediction methods for basal area and volume estimation. The data used here were also from a relatively old sensor, Landsat 5 TM. A study by Fassnacht et al. (2014) concluded that predictor data (sensor) type is the most important factor for the accuracy of biomass estimates, and that the prediction method had a substantial effect on accuracy and was generally more important than the sample size. Fassnacht et al. (2014) also suggested that choosing the appropriate statistical method may be more effective than obtaining additional field data for obtaining good biomass estimates.

Considering the cost of improving accuracy of timber production estimates by field measurements in *Eucalyptus* stands, it seems sensible to invest in further studies that focus on more test sites and a wider range of sensor systems (particularly RADAR and LIDAR). This would further increase our understanding of the role of the statistical model set-up in remote sensing-based estimates of forest variables in *Eucalyptus* stands. Further studies could also investigate whether other prediction methods, such as nonlinear regression or Partial least squares regression (PLSR) approaches alter our findings. The integration of additional predictors (e.g., topographic information or climate variables) would be a further possible extension of our work.

Conclusions

Machine-learning algorithms, particularly the random forest (RF) and Support Vector Machine (SVM) algorithms, were able to develop models that estimate basal area and volume in *Eucalyptus* stands using spectral data collected from Landsat 5 TM images. The Artificial Neural Networks (ANN) method did not perform well in this context, in part due to the limited data availability.

Random forest was the best method of spatial prediction and mapping of basal area and volume in *Eucalyptus* stands in Minas Gerais state. However, due to the close performance to the SVM and multiple linear regression methods, we propose that both methods should be tested and then, the best result applied for spatial prediction of basal area and volume in other regions with *Eucalyptus* stands. The approaches used in this study provide a frame work for integrating field and multispectral data, highlighting methods that greatly improve spatial prediction of basal area and volume estimation in *Eucalyptus* stands. Although the sensor TM of Landsat satellites is no longer operational, the concepts presented in this study are expected to be consistent regardless of the sensor. Thus, the approach used in this study can be more broadly applied to basal area and volume estimation in *Eucalyptus* stands using the new optical sensors such as Landsat 8 OLI and Sentinel-2.

The combination of spatial prediction methods with residual kriging should be used with caution, since the relative improvement of spatial prediction accuracy of basal area and volume did not occur in all methods, and there is not always a spatial dependency structure in the residuals of a spatial prediction method.

List of abbreviations

G= basal area; V= volume; MLR= multiple linear regression; RF= random forest; SVM= support vector machine; ANN= artificial neural networks; GDP= gross domestic product; GIS= geographical information systems; GPS= global positioning systems; MLA= machine learning algorithms; USGS= United States Geological Survey; TM= thematic mapper, ND= normalized difference, NDVI= normalized difference vegetation index; SAVI= soil-adjusted vegetation index; MSAVI= modified soil-adjusted vegetation index; EVI= enhanced vegetation Index; GEMI= global environment monitoring index; AIC= akaike information

criterion; VIF= variance inflation factor; RBF= radial basis function; SMO= sequential minimal optimization; RAE= reduced average error; SRE= standard deviation of the reduced average error; S_{xy}= residual standard error; R^2_{aj}= adjusted coefficient of determination; ME= mean error; MAE= mean absolute error; RMSE= root mean square error; RI= Relative Improvement; PNE= pure nugget effect; RK= residual estimation by ordinary kriging.

References

Aertsen, W., Kint, V., Van Orshoven, J., Özkan, K.A., & Muys, B. (2010). Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. *Ecological Modelling*, *221*, 1119–1130. doi:10.1016/j.ecolmodel.2010.01.007.

Ahmed, O. S., Franklin, S. E., Wulder, M. A., & White, J. C. (2015). Characterizing stand-level forest canopy cover and height using Landsat time series, samples of airborne LiDAR, and the Random Forest Algorithm. *ISPRS Journal of Photogrammetry and Remote Sensing*, *101*, 89-101. doi: 10.1016/j.isprsjprs.2014.11.007.

Alvares, C. A., Stape, J. L., Sentelhas, P. C., Gonçalves, J. L. M., & Sparovek, G. (2013). Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift*, *6*, 711 – 728. doi: 10.1127/0941-2948/2013/0507.

Barrios, P. G., Bidegain, M. P., & Gutiérrez, L. (2015). Effects of tillage intensities on spatial soil variability and site-specific management in early growth of *Eucalyptus grandis*. *Forest Ecology and Management*, *346*, 41-50. doi: 10.1016/j.foreco.2015.02.031.

Berra, E. F., Brandelero, C., Pereira, R. S., Sebem, E., Goergen, L. C. G., Benedetti, A. C. P., & Lippert, D. B. (2012). Estimativa do volume total de madeira em espécies de eucalipto a partir de imagens de satélite Landsat. *Ciência Florestal*, *22*:4, 853-864. doi: 10.5902/198050987566.

Boisvenue, C., Smiley, B. P., White, J. C., Kurz, W. A., & Wulder, M. A. (2016). Integration of Landsat time series and field plots for forest productivity estimates in decision support models. *Forest Ecology and Management*, *376*, 284–297. doi: 10.1016/j.foreco.2016.06.022.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. doi:10.1023/A:1010933404324.

Canavesi, V., Ponzoni, F. J., & Valeriano, M. M. (2010). Estimativa de volume de madeira em plantios de *Eucalyptus* spp. utilizando dados hiperespectrais e dados topográficos. *Revista Árvore*, *4:*3, 539-549. doi: 10.1590/S0100-67622010000300018.

Castillo-Santiago, M. A., Ghilardi, A., Oyama, K., Hernández-Stefanoni, J. L., Torres, I., Flamenco-Sandoval, A., Fernández, A., & Mas, J. F. (2013). Estimating the spatial distribution of woody biomass suitable for charcoal making from remote sensing and geostatistics in central Mexico. *Energy for Sustainable Development*, *17*, 177–188. doi: 10.1016/j.esd.2012.10.007.

Cluter, M. E. J., Boyd, D. S., Foody, G. M., & Vetrivel, A. (2012). Estimating tropical forest biomass with a combination of SAR image texture and Landsat TM data: An assessment of predictions between regions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *70*, 66-77. doi: 10.1016/j.isprsjprs.2012.03.011.

Coops, N. C., Johnson, M., Wulder, M. A., & White, J. C. (2006). Assessment of QuickBird high spatial resolution imagery to detect red attack damage due to mountain pine beetle infestation. *Remote Sensing of Environment*, *103*:1, 67–80. doi: 10.1016/j.rse.2006.03.012.

Dai, F., Zhou, Q., Lv, Z., Wang, X., & Liu, G. (2014). Spatial prediction of soil organic matter content integrating artificial neural network and ordinary kriging in Tibetan Plateau. *Ecological Indicators*, *45*, 184–194. doi: 10.1016/j.ecolind.2014.04.003.

Diamantopoulou, M. J. (2012). Assessing a reliable modeling approach of features of trees through neural network models for sustainable forests. *Sustainable Computing: Informatics and Systems*, *2*, 190–197. doi: 10.1016/j.suscom.2012.10.002.

Dube, T., Mutanga, O., Adam, E. & Ismail, R. (2014). Intra-and-inter species biomass prediction in a plantation forest: testing the utility of high spatial resolution Spaceborne Multispectral RapidEye Sensor and Advanced Machine Learning Algorithms. *Sensors*, *14*, 15348-15370. doi:10.3390/s140815348.

Dube, T., & Mutanga, O. (2015). Investigating the robustness of the new Landsat-8 Operational Land Imager derived texture metrics in estimating plantation forest aboveground biomass in resource constrained areas. *ISPRS Journal of Photogrammetry and Remote Sensing, 108,* 12–32. doi: 10.1016/j.isprsjprs.2015.06.002.

Dye, P.J., Jacobs, S., & Drew, D. (2004). Verification of 3-PG growth and water-use predictions in twelve *Eucalyptus* plantation stands in Zululand, South Africa. *Forest Ecology and Management*, *193*, 197-218. doi: 10.1016/j.foreco.2004.01.030.

Environmental Systems Research Institute. (2010). *ArcGIS Desktop: Release 10.1*. Redlands: ESRI.

Fassnacht, F. E., Hartig, F., Latifi, H., Berger, C., Hernández, J., Corvalán, P., & Koch, B. (2014). Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass. *Remote Sensing of Environment, 154*, 102–114. doi: 10.1016/j.rse.2014.07.028.

Fayad, I., Baghdadi, N., Guitet, S., Bailly, J. S., Hérault, B., Gond, V., Hajj, M. E., & Minh, D. H. T. (2016). Aboveground biomass mapping in French Guiana by combining remote sensing, forest inventories and environmental data. *International Journal of Applied Earth Observation and Geoinformation*, *52*, 502–514. doi: 10.1016/j.jag.2016.07.015.

Frank, E., Hall, M. A., & Witten, I. (2016). The WEKA Workbench [Online Appendix] In: I. Witten, E. Frank, M. Hall, & C. Pal *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed., Burlington, MA, USA: Morgan Kaufmann.

Galeana-Pizaña, J. M., López-Caloca, A., López-Quiroza, P., Silván-Cárdenasa, J. L., & Couturier, S. (2014). Modeling the spatial distribution of above-ground carbon in Mexican coniferous forests using remote sensing and a geostatistical approach. *International Journal of Applied Earth Observation and Geoinformation, 30*, 179-189. doi: 10.1016/j.jag.2014.02.005.

Gao, T., Zhu, J., Deng, S., Zheng, X., Zhang, J., Shang, G., & Huang, L. (2016). Timber production assessment of a plantation forest: An integrated framework with field-based inventory, multi-source remote sensing data and forest management history. *International Journal of Applied Earth Observation and Geoinformation*, *52*, 155–165. doi: 10.1016/j.jag.2016.06.004.

García-Gutiérrez, J., Martínez-Álvarez, F., Troncoso, A., & Riquelme, J.C. (2015). A comparison of machine learning regression techniques for LiDAR-derived estimation of forest variables. *Neurocomputing*, *167*, 24–31. doi: 10.1016/j.neucom.2014.09.091.

Gebreslasie, M. T., Ahmed, F. B., & Aardt, J. A. N. (2008). Estimating plot-level forest structural attributes using high spectral resolution ASTER satellite data in even-aged *Eucalyptus* plantations in southern KwaZulu-Natal, South Africa. *Southern Forests*, *70:*3, 227–236. doi: 10.2989/SF.2008.70.3.6.667.

Gebreslasie, M. T., Ahmed, F. B., & Aardt, J. A. N. (2010). Predicting forest structural attributes using ancillary data and ASTER satellite data. *International Journal of Applied Earth Observation and Geoinformation*, 12S, S23–S26. doi:10.1016/j.jag.2009.11.006.

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

González-García, M., Hevia, A., Majada, J., Anta, R. C., & Barrio-Anta, M. (2015). Dynamic growth and yield model including environmental factors for *Eucalyptus nitens* (Deane & Maiden) Maiden short rotation woody crops in Northwest Spain. *New Forests, 46,* 387-407. doi: 10.1007/s11056-015-9467-7.

Görgens, E. B., Montaghi, A., & Rodriguez, L. C. E. (2015). A performance comparison of machine learning methods to estimate the fast-growing forest plantation yield based on laser scanning metrics. *Computers and Electronics in Agriculture*, *116*, 221–227. doi: 10.1016/j.compag.2015.07.004.

Guedes, I. C. L., Mello, J. M., Silveira, E. M. O., Mello, C. R., Reis, A. A., & Gomide, L. R. (2015). Spatial continuity of dendrometric characteristics in clonal cultivated *Eucalyptus* sp. throughout the time. *Cerne*, *21*:4, 527-534. doi: 10.1590/01047760201521041824.

Huang, C., Song, K., Kim, S., Townshend, J. R. G., Davis, P., Masek, J. G., & Goward, S. N. (2008). Use of a dark object concept and support vector machines to automate forest cover change analysis. *Remote Sensing of Environment, 112,* 970–985. doi: 10.1016/j.rse.2007.07.023.

Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment, 25*, 295–309. doi: 10.1016/0034-4257(88)90106-X.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment, 83,* 195-213. doi: 10.1016/S0034-4257(02)00096-2.

Immitzer, M., Atzberger, C. & Koukal, T. (2012). Tree species classification with random forest using very high spatial resolution 8-Band WorldView-2 satellite data. *Remote Sensing*, *4*, 2661-2693. doi: 10.3390/rs4092661.

Indústria Brasileira de Árvores (2014). *Anuário estatístico da indústria brasileira de árvores: ano base 2014*. IBA, Brasília.

Indústria Brasileira de Árvores (2015). *Anuário estatístico da indústria brasileira de árvores: ano base 2015*. IBA, Brasília.

Journel, A.G., & Huijbregts, C.J. (1978). *Mining geostatistics*. London: Academic.

Justice, C.O., Vermote, E., Townshend, J. R. G., Defries, R., Roy, D. O., Hall, D. K., Salomonson, V. V., Privette, J. L., Riggs, G., Strahler, A., Lucht, W., Myneni, R. B., Knyazikhin, Y., Running, S. W., Nemani, R. R., Wan, Z., Huete, A. R., Leeuwen, W. V., Wolfe, R. E., Giglio, L., Muller, J., Lewis, P., & Barnsley, M. J. (1998). The moderate resolution imaging spectroradiometer (MODIS): land remote sensing for global change research. *IEEE Transactions on Geoscience and Remote Sensing*, *36:*4, 1228-1249. doi: 10.1109/36.701075.

Lafiti, H., Nothdurft, A., & Koch, B. (2010). Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: application of multiple optical/LiDAR-derived predictors. *Forestry*, *83*:4, 395-407. doi: 10.1093/forestry/cpq022.

Lopes, D. M., Aranha, J. T., Walford, N., O'Brien, J., & Lucas, N. (2009). Accuracy of remote sensing data versus other sources of information for estimating net primary production in *Eucalyptus globulus* Labill. and *Pinus pinaster* Ait. ecosystems in Portugal. *Canadian Journal of* Remote Sensing, *35*:1, 37-53. doi: 10.5589/m08-078.

López-Sánchez, C. A., García-Ramírez, P., Resl, R., José C Hernández-Díaz, J.C., López-Serrano, P.M., & Wehenkel, C. (2014). Modelling dasometric attributes of mixed and uneven-aged forests using Landsat-8 OLI spectral data in the Sierra Madre Occidental, Mexico. *iForest*, *10*, 288-295. doi: 10.3832/ifor1891-009.

López-Serrano, P.M., Corral-Rivas, J.J., & Díaz-Varela, R.A. (2016). Evaluation of radiometric and atmospheric correction algorithms for aboveground forest biomass estimation using Landsat 5 TM Data. *Remote Sensing, 8:5,* 369. doi: 10.3390/rs8050369.

Lu, D., Mausel, P., Brondízio, E., & Moran, E. (2004). Relationships between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon Basin. *Forest Ecology and Management*, *198*, 149–167. doi: 10.1016/j.foreco.2004.03.048.

Masek, J. G., Vermote, E. F., Saleous, N. E., Wolfe, R., Hall, F. G., Huemmrich, K. F., Gao, F., Kutler, J., & Lim, T.K. (2006). A Landsat surface reflectance dataset for North America, 1990 – 2000. *IEEE Geoscience and Remote Sensing Letters, 3:1*, 68 – 72. doi: 10.1109/LGRS.2005.857030.

Meng, Q., Cieszewski, C., & Madden, M. (2009). Large area forest inventory using Landsat ETM+: A geostatistical approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, *64*, 27–36. doi: 10.1016/j.isprsjprs.2008.06.006.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2006). *Introduction to linear regression analysis*. New York: Wiley.

Moreno, A., Neumann, M., & Hasenauer, H. (2016). Optimal resolution for linking remotely sensed and forest inventory data in Europe. *Remote Sensing of Environment*, *183*, 109–119. doi: 10.1016/j.rse.2016.05.021.

Morgenroth, J., & Visser, R. (2013). Uptake and barriers to the use of geospatial technologies in forest management. *New Zealand Journal of Forestry Science, 43*:16, 1-9. doi: 10.1186/1179-5395-43-16.

Özçelik, R., Diamantopoulou, M. J., Crecente-Campo, F., & Eler, U. (2013). Estimating Crimean juniper tree height using nonlinear regression and artificial neural network models. *Forest Ecology and Management*, *306*, 52–60. doi: 10.1016/j.foreco.2013.06.009.

Pacheco, L. R. F., Ponzoni, F. J., Santos, S. B., Andrades Filho, C. O., Mello, M. P., & Campos, R. C. (2012). Structural characterization of canopies of *Eucalyptus* spp. using radiometric data from TM/Landsat 5. *Cerne*, *18:*1, 105-116. doi: 10.1590/S0104-77602012000100013.

Palmer, D.J., Höck, B.K., Kimberley, M.O., Watt, M.S., Lowe, D.J., & Payn, T.W. 2009. Comparison of spatial prediction techniques for developing *Pinus radiata* productivity surfaces across New Zealand. *Forest Ecology and Management, 258*:9, 2046–2055. doi: 10.1016/j.foreco.2009.07.057.

Pinty, B., & Verstraete, M. M. (1992). GEMI: a non-linear index to monitor global vegetation from satellites. *Vegetatio, 101*:1, 15-20. doi: 10.1007/BF00031911.

Ponzoni, F. J., Shimabukuro, Y. E., & Kuplich, T. M. (2012). *Sensoriamento Remoto da Vegetação*. 2. ed. São Paulo: Oficina de Textos.

Ponzoni, F. J., Pacheco, L. R. F., Santos, S. B., & Andrades Filho, C. O. (2015) Caracterização espectro-temporal de dosséis de *Eucalyptus* spp. mediante dados radiométricos TM/Landsat 5. *Cerne*, *21*:2, 267-275. doi: 10.1590/01047760201521021457.

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment*, *48*, 119-126. doi: 10.1016/0034-4257(94)90134-1.

R Core Team. (2016). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Raimundo, M. R., Scolforo, H. F., Mello, J. M., Scolforo, J. R. S., McTague, J. P., & Reis, A. A. Geostatistics applied to growth estimates in continuous forest inventories. *Forest Science 63:1,* 29–38. doi: 10.5849/FS.2016-056.

Retslaff, F. A. S., Figueiredo Filho, A., Dias, A. N., Bernett, L. G. & Figura, M. A. (2015). Curvas de sítio e relações hipsométricas para *Eucalyptus grandis* na

Região dos Campos Gerais, Paraná. *Cerne*, *2:*2, 199-207. doi: 10.1590/01047760201521021349.

Ribeiro Júnior, P.J., & Diggle, P. J. (2001). GeoR: a package for geostatistical analysis. *R-NEWS 1*(2): 15-18.

Rodriguez-Galiano, V., Castillo, M. S., 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*, 804-818. doi: 10.1016/j.oregeorev.2015.01.001

Rouse, J., Haas, R., Schell, J., Deering. D., & Harlan, J. (1973). *Monitoring the vernal advancements and retrogradation (greenwave effect) of nature vegetation.* NASA/GSFC Final Report., Greenbelt, MD, USA: NASA.

Shao, Y., & Lunetta, R. S. (2012). Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. *ISPRS Journal of Photogrammetry and Remote Sensing*, *70*, 78-87. doi: 10.1016/j.isprsjprs.2012.04.001.

Shataee, S., Kalbi, S., Fallah, A., & Pelz, D. (2012). Forest attribute imputation using machine-learning methods and ASTER data: comparison of k-NN, SVR and random forest regression algorithms. *International Journal of Remote Sensing*, *33*, 6254-6280. doi: 10.1080/01431161.2012.682661.

Scolforo, H. F., Castro Neto, F., Scolforo, J. R. S., Burkhart, H., McTague, J. P., Raimundo, M. R., Loos, R. A., Fonseca, S., & Sartório, R. C. (2016). Modeling dominant height growth of *Eucalyptus* plantations with parameters conditioned to climatic variations. *Forest Ecology and Management*, *380*, 182–195. doi: 10.1016/j.foreco.2016.09.001.

Stojanova, D., Panov, P., Gjorgjioski, V., Kobler, A., & Džeroski, S. (2010). Estimating vegetation height and canopy cover from remotely sensed data with

machine learning. *Ecological Informatics*, *5*, 256-266. doi: 10.1016/j.ecoinf.2010.03.004.

United States Geological Survey (2017). *Landsat imagery*. Available online at: https://earthexplorer.usgs.gov.

Verma, N. K., Lamb, D.W., Reid, N., & Wilson, B. (2014). An allometric model for estimating DBH of isolated and clustered *Eucalyptus* trees from measurements of crown projection area. *Forest Ecology and Management*, *326*,125-132. doi: 10.1016/j.foreco.2014.04.003.

Viana, H., Aranha, J., Lopes, D., & Cohen, W.B. (2012). Estimation of crown biomass of *Pinus pinaster* stands and shrubland above-ground biomass using forest inventory data, remotely sensed imagery and spatial prediction models. *Ecological Modelling*, *226*, 22-35. doi: 10.1016/j.ecolmodel.2011.11.027.

Vicharnakorn, P., Shrestha, R. P., Nagai, M., Salam, A. P., & Kiratiprayoon, S. (2014). Carbon stock assessment using remote sensing and forest inventory data in Savannakhet, Lao PDR. *Remote Sensing*, *6*, 5452-5479. doi:10.3390/rs6065452.

Watt, M. S., Rubilar, R., Kimberley, M. O., Kriticos, D. J., Emhart, V., Mardones, O., Acevedo, M., Pincheira, M., Stape, J., & Fox, T. (2014). Using seasonal measurements to inform ecophysiology: extracting cardinal growth temperatures for process-based growth models of five *Eucalyptus* species/crosses from simple field trials. *New Zealand Journal of Forestry Science*, *44:*9. doi: 10.1186/s40490-014-0009-4.

Watt, M. S., Dash, J. P., Watt, P., & Bhandari, S. (2016). Multi-sensor modelling of a forest productivity index for radiata pine plantations. *New Zealand Journal of Forestry Science*, *46:*9. doi: 10.1186/s40490-016-0065-z.

Wear, D. N., Dixon IV, E., Abt, R. C., & Singh, N. (2015). Projecting potential adoption of genetically engineered freeze-tolerant *Eucalyptus* in the United States. *Forest Science*, *61*:3, 466–480. doi: 10.5849/forsci.14-089.

Were, K., Bui, D. T., Dick, O. B., & Singh, B. R. (2015). A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. *Ecological Indicators*, *52*, 394–403. doi: 10.1016/j.ecolind.2014.12.028.

Wu, C., Shen, H., Shen, A., Deng, J., Gan, M., Zhu, J., Xu, H., & Wang, K. (2016). Comparison of machine-learning methods for above-ground biomass estimation based on Landsat imagery. *Journal of Applied Remote Sensing*, *10*,3. doi: 10.1117/1.JRS.10.035010.

Yamamoto, J. K., & Landim, P. M. B. (2013). *Geoestatística: conceitos e aplicações*. São Paulo: Oficina de Textos.

Zhang, J., Huang, S., Hogg, E. H., Lieffers, V., Qin, Y., & He, F. (2014). Estimating spatial variation in Alberta forest biomass from a combination of forest inventory and remote sensing data. *Biogeosciences*, *11*, 2793-2808. doi:10.5194/bg-11-2793-2014.

ARTICLE 2 - VOLUME ESTIMATION IN A *EUCALYPTUS* PLANTATION USING MULTI-SOURCE REMOTE SENSING AND DIGITAL TERRAIN DATA: A CASE STUDY IN MINAS GERAIS STATE, BRAZIL

Aliny Aparecida Dos Reis[1,2], Steven E. Franklin[2], José Marcio de Mello[1], Fausto Weimar Acerbi Junior[1]

1. Department of Forest Science, University of Lavras – UFLA, PO Box 3037, Lavras, Minas Gerais, Brazil, Zip Code 37200-000

2. School of Environment, Trent University, Peterborough, Ontario, Canada K9I 7B8

*Corresponding author email: alinyreis@hotmail.com

Publication status: Published in the **International Journal of Remote Sensing**

DOI: 10.1080/01431161.2018.1530808

Abstract: In this study, we tested the effectiveness of stand age, multispectral optical imagery obtained from the Landsat 8 Operational Land Imager (OLI), synthetic aperture radar (SAR) data acquired by the Sentinel-1B satellite, and digital terrain attributes derived from a digital elevation model (DEM), in estimating forest volume in 351 plots in a 1,498 ha *Eucalyptus* plantation in northern Minas Gerais state, Brazil. A Random Forest (RF) machine learning algorithm was used following Principal Component Analysis (PCA) of various data combinations, including multispectral and SAR texture variables and DEM-based geomorphometric derivatives. Using multispectral, SAR or DEM variables alone (*i.e.*, Experiments (ii)-(iv)) did not provide accurate estimates of volume (RMSE (Root Mean Square Error) > 32 m^3 ha^{-1}) compared to predictions based on age since planting of *Eucalyptus* stands (Experiment (i)). However, when these dataset were individually combined with stand age (i.e., Experiments (v)-(vii)), the RF models resulted in better volume estimates than those obtained when using the individual multispectral, SAR and DEM datasets (RMSE < 28 m^3 ha^{-1}). Furthermore, a model that integrated the selected variables of these data with stand age (Experiment (viii)) improved volume estimation significantly (RMSE = 22.33 m^3 ha^{-1}). The large and increasing area of *Eucalyptus* forest plantations in Brazil and elsewhere suggests that this new approach to volume estimation has potential to support *Eucalyptus* plantation monitoring and forest management practices.

Keywords: Landsat 8 OLI; Sentinel-1B; terrain attributes; Random Forest; forest management.

1. Introduction

Forest plantations cover approximately 277 million hectares globally and have significant capacity to store carbon during growth and in the form of durable forest products after harvest (Payn et al. 2015). In Brazil, *Eucalyptus*

trees are widely planted in part due to fast growth and successful adaptation of *Eucalyptus* cultures (Raimundo et al. 2017; Scolforo et al. 2017), and with recent large-scale afforestation and reforestation efforts, have now expanded rapidly to cover more than 5.6 million hectares (Ibá, 2016). These plantation forests support multiple uses and are estimated to have the capacity to absorb approximately 1.2 billion tons of carbon dioxide annually.

Monitoring of *Eucalyptus* forest growth in Brazilian forest plantations is conducted through annual field-based forest inventories (Raimundo et al. 2017). Forest volume is an important measurement in such inventories to support analysis of wood products and biophysical processes related to forest dynamics (Gao et al. 2016; Dube et al. 2017). However, field-based forest inventory measurements are expensive, time consuming, and labour intensive. Age since planting date is a simple mechanistic way to obtain a general indication of plantation growth and volume in *Eucalyptus* plantations. However, recent studies have suggested that remotely-sensed data can supplement this type of plot-based age estimate of forest volume with spatially-explicit information (Gama, dos Santos, and Mura 2010, 2016; Gebreslasie, Ahmed, and van Aardt 2010, 2011; Baghdadi et al. 2014, 2015; Dube and Mutanga 2015, 2016). For example, Gebreslasie, Ahmed, and van Aardt (2011) and Ismail et al. (2015) in South Africa and Baghdadi et al. (2015) in Brazil reported an accuracy increase in age-based forest structural attribute estimations in *Eucalyptus* plantations incorporating ASTER, SPOT4 and ALOS/PALSAR data, respectively. High spatial resolution sensors, light detection and ranging (LIDAR) and unmanned aerial vehicle (UAV) data have shown great potential for improving forest volume estimates (e.g., Goodbody et al. 2017, Ota et al. 2017, Tompalski et al. 2018). The recent paper by Shinzato et al. (2017) provides the spatial distribution of volume in *Eucalyptus* plantations using high resolution multispectral images and LIDAR data with a RMSE of 17.43 $m^3\,ha^{-1}$. Using a

nonlinear mixed-effect model and LIDAR data, Packalén et al. (2011) obtained a volume estimate accuracy with RMSE between 26 and 45 $m^3 ha^{-1}$ for *Eucalyptus* stands in Bahia, Brazil. On the other hand, the limited availability and high acquisition cost of such data continue to restrict widespread adoption of these technologies in large area applications (McRoberts, Gobakken, and Næsset 2012; Racine et al. 2014; Ota et al. 2017).

A second option is to use freely available public archives of multispectral satellite imagery and synthetic aperture radar (SAR) data to supplement the simple volume-age relationships and in prediction of other forest parameters of interest (e.g., leaf area index (LAI), canopy cover) (Ismail et al. 2015; Gao et al. 2016; Gama, dos Santos, and Mura 2016). Satellite multispectral optical sensor data have been used to distinguish the basic distribution of green vegetation related to photosynthesis of plants (Roy et al. 2014). However, accurate forest volume estimation using such imagery alone has proven difficult to achieve due to spectral saturation problems in dense forests (Lu et al. 2014). Related research has shown that satellite SAR sensor data, which are less influenced by clouds and atmospheric aerosols, are sensitive to plant structure and capable of penetrating vegetation (Torres et al. 2012). Thus, an optimal strategy is to use available multispectral and SAR data together to address *Eucalyptus* plantation forest stand volume estimates and long-term monitoring needs (Ismail et al. 2015; Aslan et al. 2016; Shao and Zhang 2016; Zhao et al. 2016).

The integration of multispectral optical and SAR sensor data has also been reported with digital elevation model (DEM) data to represent terrain conditions or climate variables directly related to forest growth (Dube et al. 2017; López-Sánchez et al. 2017; Scolforo et al. 2017). DEMs are widely available at high spatial resolutions and vertical accuracies, and can be processed for the generation of a wide range of geomorphometric variables, such as slope,

aspect, incidence (or insolation), and hydrological and topographical indices (e.g., slope position, wetness index) (Hengl and Reuter 2009). However, there remains a paucity of good methodological examples and demonstrations of the accuracy that can be obtained by integrating such DEM-based geomorphometric data with optical and SAR remote sensing methods in fast-growing sub-tropical *Eucalyptus* plantations.

In this study, we tested the effectiveness of integrating stand age, multispectral optical data (Landsat 8 OLI), SAR (Sentinel-1B) remote sensing data, and digital terrain attributes or geomorphometrics derived from a DEM to improve the estimation accuracy and mapping of volume in *Eucalyptus* plantations in Brazil. To the best of our knowledge, no study has been carried out to estimate volume of *Eucalyptus* plantations based on integration of the recently launched Sentinel-1B SAR, Landsat-8 OLI, and digital terrain variables. Initially, variables obtained from the three different digital datasets (*i.e.*, multispectral optical data, SAR data, and terrain attributes) were used separately to estimate forest volume measured in a standard *Eucalyptus* plantation plot-based forest inventory protocol for this region. We used Principal Component Analysis (PCA) to reduce the large number of digital variables to fewer, uncorrelated components that explained most of the original dataset variance. Then, stand age was combined with the best multispectral optical, SAR and digital terrain variables in 246 training areas to predict volume in 105 validation plots over three years. The input variables were ranked using the recursive Variable Importance (VI) method within the Random Forest (RF) machine learning algorithm, and the final estimated forest volumes were compared to the field-observations of volume in the independent forest inventory sample. In this paper, we interpret the various models and compare the best overall model with the results obtained when using each dataset alone. A spatial interpretation of the

patterns of *Eucalyptus* plantation volume produced using the best RF results is provided for the final map analysis.

2. Background to *Eucalyptus* plantation volume estimation using remote sensing data

Estimates of *Eucalyptus* plantation volume using multi-source remote sensing and DEM data in comparison to field observations have been reported using a variety of input variables and methods. For example:

i. *Multispectral and SAR images:* Ismail et al. (2015) used ALOS PALSAR backscatter and SPOT 4 multispectral optical data to predict volume in *Eucalyptus* plantations in Zululand, South Africa. Stand age – measured as an annual increment since planting date – was used as an independent variable in a regression model to predict volume, with one optical band (short-wave infrared or SWIR) and one SAR band (HV cross-polarization backscatter). The best model produced a RMSE of 31.71 m^3 ha^{-1}, which was considered appropriate for operational forest management in this region. These results confirmed earlier work that suggested that, since *Eucalyptus* plantations are typically a single species monoculture, and the undergrowth (shrubs and small trees) is sparse, a high proportion of tree stem-scattering was produced and was represented in the SAR image dataset (Gama, dos Santos, and Mura 2016).

ii. *Image texture:* The use of texture measures derived from multispectral optical images to improve volume estimates in *Eucalyptus* plantations has been documented (Gebreslasie, Ahmed, and van Aardt 2011; Dube and Mutanga 2015). The general principle is that image texture derivatives can simplify and define complex forest canopy structures even in closed canopies, thus reducing saturation effects (Sarker and Nichol 2011).

Recently, Dube and Mutanga (2015) demonstrated that certain small-window (or high frequency) texture derivatives from Landsat 8 OLI data helped improve aboveground biomass estimation in both *Eucalyptus* and *Pinus* plantations when compared to the use of multispectral reflectance data alone or vegetation indices. The use of SAR image texture in *Eucalyptus* plantation volume estimates has not yet been reported, although an earlier forest biomass estimation study with integrated multispectral and SAR data showed that textures were effective (Cutler et al. 2012).

iii. *Digital terrain attributes:* Spatial and temporal variations of *Eucalyptus* plantation growth are strongly affected by topography and climate conditions. In general, topography controls the hydrologic dynamics of catchments, nutrient and soil conditions, and climate influences growth through variability in incoming solar radiation, soil moisture, and temperature (Adams, Barnard, and Loomis 2014; Dube et al. 2017; H. F. Scolforo et al. 2017). In one recent study, Dube and Mutanga (2016) improved aboveground biomass prediction accuracy (R^2= 80% and RMSE=19.65 t ha^{-1}) by integrating environmental variables (rainfall and temperature) and five DEM-based geomorphometric variables (slope, aspect, topographic wetness index, elevation, and insolation) with a WorldView-2 multispectral image of *Eucalyptus* and *Pinus* forest plantations in Umgeni Catchment, South Africa.

Many of these studies used regression techniques or machine learning algorithms in forest volume or biomass model development. The RF algorithm is a nonlinear and non-parametric ensemble decision-tree method (Breiman 2001). Such models of forest attribute estimation (Belgiu and Drăgu 2016; Gao et al. 2016; López-Serrano et al. 2016; Wu et al. 2016) often outperform alternative methods, such as k-Nearest Neighbor (kNN), Support Vector

Machine (SVM), Back Propagation Neural Networks (BPNN), and Stepwise Multiple Linear Regression (LMSTEP).

Another key feature of multi-source remote sensing forest volume estimation work is the widespread use of data dimensionality reduction methods; typically, satellite sensor multispectral and SAR imagery, texture derivatives, and DEM-related datasets will enable the production of a very large number of predictive variables. Many of these variables will be correlated or redundant, and multiple variables can overwhelm even the most robust statistical or non-parametric algorithm when the sample sizes are relatively small (not rare in studies relying on plot-based field forest inventory measurements). One common strategy is to employ a data reduction tool, such as Principal Components Analysis (PCA), prior to model development (Fayad et al. 2014; Lee et al. 2015; Lu et al. 2014; Ng 2017).

3. Study area

This study was conducted in a 1,498 ha privately-managed *Eucalyptus* plantation in Diamantina municipality, northern Minas Gerais state, Brazil (Figure 1). The region experiences humid sub-tropical climatic conditions, corresponding to Köppen's climatic type Cwb, with the rainy season occurring between the months of October and March and mean annual rainfall of 1,468 mm. Mean daily temperatures during the hottest month typically do not exceed 22°C (Alvares et al. 2013). Elevations range between 700 m and 1100 m above-mean-sea-level. Soils are relatively uniform and predominately Cambisols and Latosols. The *Eucalyptus* plantation was established between November 2010 and October 2012, using clone species (*E. urophylla x E. grandis*), which typically leads to uniform crown and tree density conditions. *Eucalyptus* trees are harvested every 5–7 years. The stand age is a relatively simple metric to determine with planting and field data collection information. The plantation

was established at approximately 1111 trees per hectare with initial spacing of 3.6 x 2.5 m. Intensive soil preparation and weed control measures are practised. Crown closure occurs quickly within the first 18 months following plantation establishment.

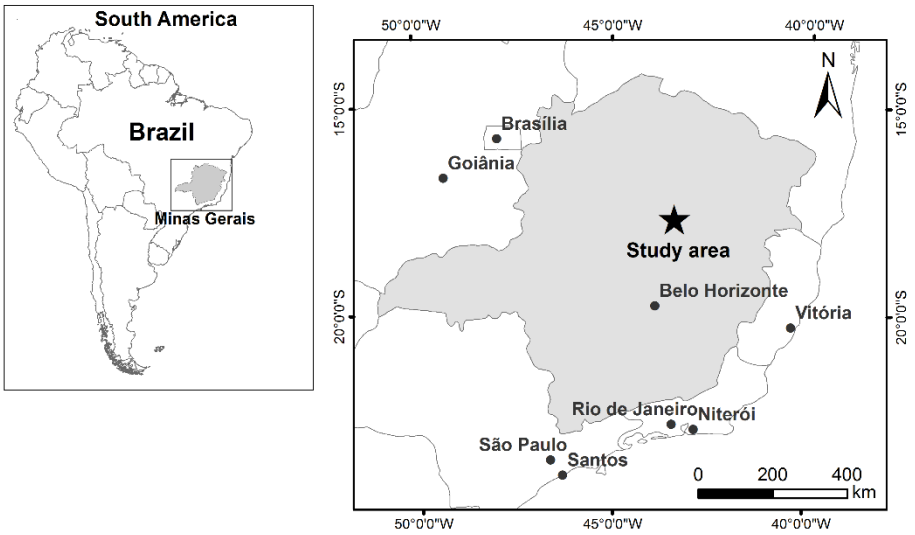


Figure 1. Location of the study area in northern Minas Gerais state, Brazil.

4. Methods

Field measurements

Field data collection in 351 plots across the *Eucalyptus* plantation using a standard continuous forest inventory (CFI) method occurred on August 2015, June 2016, and June 2017 (Figure 2). In the first year of *Eucalyptus* plantation development, field measurements determined survival and mortality within the stands, but detailed forest inventory observations were not made. Subsequent annual field inventory was based on the 351 sample plots distributed systematically and located in the field using survey-grade real time kinetic (RTK) GPS. A plot size of 25 m × 20 m (500 m²) was used. Within each plot,

tree diameter at breast height (DBH) of all stems and the total height of the first 15 trees with normal stems (without bifurcation or any other defect) were measured. From the information collected in the plots, estimates of stand volume (in m^3 ha^{-1}) were estimated using standard forest allometric equations developed for this region. The individual volume equations were based on DBH and total height, fitted for each measurement year and DBH class. All fitted equations showed high coefficient of determination (R^2 > 97%) and low residual standard error (S_{xy} < 0.05 m^3). The plot-based field inventory *Eucalyptus* forest plantation volume for 2015-2017 is summarized in Table 1.

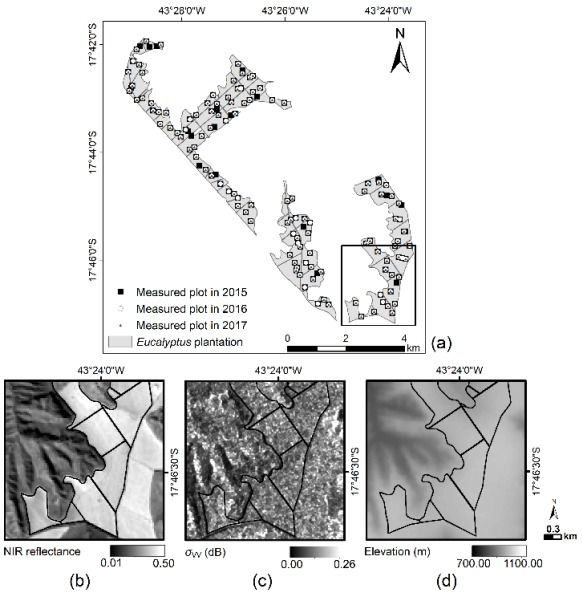


Figure 2. *Eucalyptu*s plantation: (a) Distribution of 351 plots used in field inventory and a small 6.25 km^2 subarea identified in the southeast to show the multi-source remote sensing and digital terrain data used; (b) Landsat 8 OLI NIR band enhancement in greyscale of the subarea, (c) Sentinel-1B C-band SAR VH polarization enhancement in greyscale, and (d) ALOS PALSAR DEM with relative elevation values on a 12.5 m grid.

Table 1. *Eucalyptus* forest plantation description and volume estimates obtained from forest inventory field data collection in 351 plots over three years.

Field data collection		Number of plots	Age (Years)		Volume (m^3 ha^{-1})			
Month	Year		Min	Max	Min	Max	Mean	StdDev
August	2015	137	2.90	4.80	85.55	253.49	146.26	33.86
June	2016	117	3.70	5.60	111.81	277.22	181.13	35.33
June	2017	97	4.70	6.60	131.28	338.57	211.49	38.99

Where: Min = minimum; Max = maximum; StdDev = standard deviation.

Remote sensing data collection and preprocessing

Landsat 8 OLI multispectral imagery, Sentinel-1B SAR data, and ALOS PALSAR DEM data were acquired for this research. The selected Landsat 8 OLI satellite scenes (Path 218 and Row 72) covering the study area were acquired on dates that most closely coincided with the field measurement dates (i.e., June 12, 2015, July 16, 2016, and June 01, 2017, respectively), and were relatively cloud-free and without obvious radiometric issues. These scenes were obtained from the USGS database (United States Geological Survey) as Level 1 Terrain Corrected (L1T) product with a 30 m geometric resolution, less than one-half pixel RMSE locational error, and the following bands: coastal aerosol (C, 0.43–0.45 μm), blue (B, 0.45–0.51 μm), green (G, 0.53–0.59 μm), red (R, 0.64–0.67 μm), near infrared (NIR, 0.85–0.88 μm), shortwave infrared-1 (SWIR1, 1.57–1.65 μm), shortwave infrared-2 (SWIR2, 2.11–2.29 μm). In addition to the spectral bands, six multispectral indices were employed (le Maire et al. 2011; Shao and Zhang 2016; López-Sánchez et al. 2017; Riihimäki, Heiskanen, and Luoto 2017; Vaglio et al. 2017), including: Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Simple Ratio (SR), Soil-Adjusted Vegetation Index (SAVI), Modified Soil-Adjusted Vegetation Index (MSAVI) and Normalized Difference Moisture Index (NDMI).

Sentinel-1B SAR scenes were selected for this study with acquisition dates close to the Landsat scene acquisition dates (i.e., June 13, 2015, July 27, 2016, and June 8, 2017, respectively). We used Ground Range Detected (GRD) Level-1 data acquired in interferometric wide swath (IW) mode with dual vertical-vertical (VV) and vertical horizontal (VH) polarization. These SAR data were calibrated, radiometrically- and geometrically-corrected and filtered using the Sentinel-1 Toolbox implemented in SNAP (Sentinel Application Platform) and have a spatial resolution of 10 m. Additionally, we calculated three SAR indices using the available VV and VH polarizations: i) simple polarization ratios VH/VV and VV/VH, and ii) polarization averaging (VH-VV)/2 (Omar, Misman, and Kassim 2017).

Terrain attributes derived from the DEM (ALOS PALSAR) with a spatial resolution of 12.5 m were used to quantify the topographic conditions in the plantation study area. Basic terrain attributes included general geomorphometric variables of elevation, slope, aspect (sin and cosine), cross-sectional and longitudinal curvatures (Hengl and Reuter 2009). In addition, specific geomorphometrics were computed to characterize local topographic and hydrological variability (Adams, Barnard, and Loomis 2014). First, convergence index, flow accumulations, channel network base level, vertical distance to channel network, valley depth, and relative slope position provided insight into site-specific water-, gravity- and wind-field conditions. Second, direct insolation, diffuse insolation, analytical hillshade (also known as shaded relief or incidence value), and the Topographic Wetness Index (TWI) metric were computed to represent the variability associated with mean solar-field conditions (Mohamedou, Tokola, and Eerikäinen 2017). All terrain attributes were calculated using the SAGA GIS software package (v. 5.0.0). Prior to calculation of the selected terrain attributes, the DEM was preprocessed with a standard gap/sink-filling error detection and noise removal routine.

Extraction of textural images from Landsat 8 OLI and Sentinel-1B SAR data

The Grey Level Co-occurrence Matrix (GLCM) statistical texture approach (Haralick, Shanmugam, and Dinstein 1973) was used to generate texture images (Lu et al. 2014) from the green and near infrared Landsat bands, and the VV and VH Sentinel-1B SAR backscatter data. These images appeared visually to contain high image contrast in the *Eucalyptus* plantations of interest. Eight GLCM texture features were then generated with 64-bit quantization, over small windows (3 x 3), and averaged for directional texture: mean (MEA), variance (VAR), homogeneity (HOM), contrast (CON), dissimilarity (DIS), entropy (ENT), second moment (2M), and correlation (COR). A small window size was used in order to preserve high frequency spatial information, which was shown in earlier work to decrease when using larger windows as a result of over-smoothing of fine-scale textural variations (Franklin, Wulder, and Lavigne 1996). Recent research has also documented improved performance of small window sizes when applied to forest attribute estimation in *Eucalyptus* plantations (Dube and Mutanga 2015). Note that the intent was not a test of all possible texture methods and spatial operators, but rather an initial step using statistical textures to augment the spectral domain of the multi-source remotely sensed imagery with high frequency spatial information (López-Serrano et al. 2016). The goal was to characterize the fine-scale spatial heterogeneity of *Eucalyptus* forest plantation structures. Texture analysis was performed using software ENVI Version 4.7 (Exelis Visual Information Solutions, Boulder, Colorado).

Principal Component Analysis (PCA) of predictive variables

Principal Component Analysis (PCA), a commonly-used and effective data dimensionality reduction technique (Lee et al. 2015; Maack et al. 2015; Ng 2017), was applied in this study. All data were normalized to reduce potential

scaling issues due to differing measurement units. Seven multispectral bands, six multispectral indices, sixteen optical texture derivatives, five SAR variables (two backscatter coefficients and three image-derived indices) and sixteen SAR texture derivatives were processed separately to create orthogonal components. We then named those components based on the variables with the highest individual component loadings. We selected PCs that together represented no less than 98% of the data variance in each dataset. In total, 16 components were generated to represent the variance in the Landsat 8 OLI multispectral bands (PC1-4), multispectral vegetation indices (PC1-3) and optical textures (PC1-9); and 13 components represented the SAR backscatter and derived SAR indices (PC1-2) and SAR textures (PC1-11). The final sets of PCs were used as input variables with the digital terrain variables in the volume estimation procedures described in the next section.

Random Forest regression algorithm

The Random Forest (RF) (Breiman 2001) machine learning algorithm is well suited for analysis of complex variable datasets in forestry and multi-source remote sensing (Dube et al. 2015; Gao et al. 2016; Novelli et al. 2016; Franklin and Ahmed 2017; Zhang et al. 2018). While more complex machine learning routines are increasingly available, RF is often recommended as a good first choice for use in proof-of-concept analysis, such as that undertaken here (Daroczi 2015). In this study, we used 300 decision trees in each RF implementation following preliminary tests of model performance. This number of decision trees ensured that every plot observation was predicted at least several times for each model output. The number of variables randomly sampled at each split was equal to the square root of the number of variables. Finally, the total field sample of plots (351) was divided randomly into 70% (246 plots) and 30% (105 plots) for training and validation of the RF models, respectively.

We applied a backward stepwise predictor selection approach for each of eight experiments described in the next section to ensure that only predictor variables were selected that decreased the overall model uncertainties. Accordingly, in each run, 20% of the lower importance predictors were removed and the resulting Root Mean Square Error (RMSE) based on comparisons to the field inventory calculations in the 105 validation plots of volume was assessed. Initial tests revealed that this method produced the best overall predictive accuracy and allowed us to simplify the modelling process by identifying the minimum number of predictors to offer the best predictive accuracy (see also Ismail and Mutanga 2010). The RF Variable Importance metric was computed, based on recursive substitution, enabling the most important variables in each model run to be interpreted based on the percentage increase of the mean square error (IncMSE) resulting when that variable was removed from the prediction. These analyses were performed using the R software package (R Core Team, 2017).

RF Experiments

Figure 3 shows that volume of *Eucalyptus* stands increases with the stand age and is highly variable. This relationship and variability is of interest in this study.

Eight Random Forest experiments were conducted to predict plantation volume using the predictive variables extracted from the three different digital datasets: the developed PCs of Landsat 8 OLI and Sentinel-1B and the digital terrain attributes (Table 2). The first Experiment (i) used stand age alone to predict volume, and improvement of this simple decision-tree relationship was expected based on the addition of the new remote sensing and terrain explanatory variables.

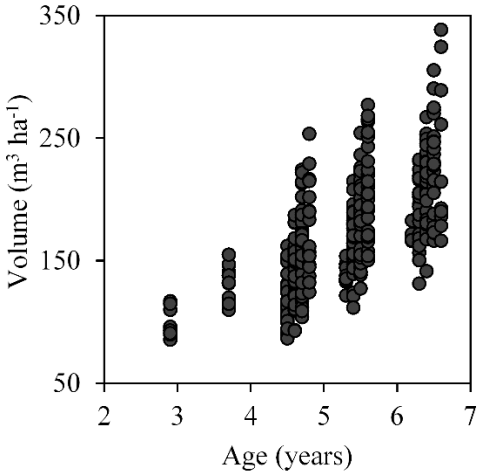


Figure 3. *Eucalyptus* plantation volume as a function of stand age measured in the field.

Accuracy assessment

The accuracies of predicted *Eucalyptus* plantation forest volume using the eight developed RF models were evaluated using adjusted coefficient of determination (R^2) and Root Mean Square Error (RMSE) in m^3 ha^{-1} calculated based on field-based inventory volume estimates with the validation data (105 plots). In addition, we calculated the relative RMSE (RMSEr) expressed as percentage, as described in Equations (1) and (2):

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

$$RMSE_r=\frac{RMSE}{\bar{X}}\times 100 \quad (2)$$

where n is the number of plots in the dataset; $\hat{X}_i$ is the estimated value of volume; X_i is the observed value of volume in the validation dataset; and $\bar{X}$ is the mean of the validation dataset. Finally, we mapped the estimated *Eucalyptus* plantation forest volume for the area of the plantation.

Table 2. Experimental design for Random Forest models using various combinations of stand age, Landsat 8 OLI, Sentinel-1B, and ALOS PALSAR DEM-derived predictor variables for *Eucalyptus* forest plantation volume estimation.

Data type	Data source	Details	Experiment
Stand age	Forest Inventory	Independently determined based on plantation records indicating the dates of tree planting and field measurement dates	(i)
Landsat 8 OLI multispectral variables	Landsat 8 OLI	Multispectral bands (PC1-4), multispectral indices (PC1-3) and textures (PC1-9)	(ii)
Sentinel-1B SAR variables	Sentinel-1B SAR	Backscatter coefficients and image-derived indices (PC1-2) and textures (PC1-11)	(iii)
Digital terrain variables	ALOS PALSAR DEM	Elevation, slope, aspect (sin and cosine), cross-sectional curvature, longitudinal curvature, convergence index, flow accumulation, channel network base level, vertical distance to channel network, valley depth, relative slope position, direct insolation, diffuse insolation, analytical hillshade, and TWI	(iv)
Age + Landsat 8 OLI multispectral selected variables	Forest Inventory + Landsat 8 OLI	Stand age plus the best multispectral image predictors (PCs) as indicated by the VI measure in Experiment (ii)	(v)
Age + SAR selected variables	Forest Inventory + Sentinel-1B SAR	Stand age plus the best SAR image predictors (PCs) as indicated by the VI measure in Experiment (iii)	(vi)
Age + digital terrain selected attributes	Forest Inventory + ALOS PALSAR DEM	Stand age plus the best digital terrain attributes as indicated by the VI measure in Experiment (iv)	(vii)
Age + Landsat 8 OLI multispectral + Sentinel-1B SAR + digital terrain selected variables	Forest Inventory + Landsat 8 OLI + Sentinel-1B SAR + ALOS PALSAR DEM	Stand age plus the best multispectral and SAR image predictors (PCs) and digital terrain attributes as determined in the preceding steps	(viii)

Where: TWI= Topographic Wetness Index, PC= Principal Component, VI= Variable Importance; DEM= Digital Elevation Model, OLI= Operational Land Imager, SAR= Synthetic Aperture Radar.

5. Results and Discussion

Table 3 shows the adjusted coefficient of determination (R^2), RMSE and RMSEr obtained using the eight RF models developed using the 246 training plots and then applied to the validation plot data (105).

Table 3. Performance of the RF volume models in *Eucalyptus* plantation based on different predictor group combinations using 351 plots separated into 246 training (RF model development) and 105 validation samples (RF model assessment).

Experiment	Predictor Data Type	R^2 (%)	RMSE (m^3 ha^{-1})	RMSEr (%)
(i)	Stand age	49.08	30.98	17.87
(ii)	Landsat 8 OLI Multispectral variables	41.05	32.06	18.49
(iii)	Sentinel-1B SAR variables	33.88	34.54	19.92
(iv)	Digital terrain variables	9.70	39.77	22.94
(v)	Age + Landsat 8 OLI multispectral variables	64.04	25.10	14.48
(vi)	Age + Sentinel-1B SAR variables	54.65	27.84	16.05
(vii)	Age + digital terrain variables	69.19	22.40	12.92
(viii)	Age + Landsat 8 OLI Multispectral + Sentinel-1B SAR + digital terrain variables	71.43	22.33	12.88

Where: R^2 = adjusted coefficient of determination; RMSE = root mean square error and RMSEr = relative root mean square error

As expected, stand age predicted *Eucalyptus* plantation forest volume in these plots with approximately 49% accuracy and a relatively low RMSE (30.98 m^3 ha^{-1}) (Experiment (i)). Using RF Variable Importance in this study, stand age was always selected as the most important variable. Le Marie et al. (2011) and Baghdadi et al. (2015) also found that stand age was the most important variable used in their volume estimation models. In those studies, age alone was able to explain more than 80% of volume variability in highly productive *Eucalyptus* plantations in São Paulo State, Brazil. The stand age influences the rate of photosynthetic activity and the growth rate of *Eucalyptus* trees, mainly due to the accelerated rate of growth of these trees. These results confirm earlier findings that have shown stand age to be a good initial estimator of volume in *Eucalyptus* plantations (Ismail et al. 2015; Gebreslasie, Ahmed, and van Aardt 2010; Dube et al. 2017). However, the large amount of variation not explained by age, and the resulting RMSE (30.98 m^3 ha^{-1}) for such plantation volume estimates, confirmed the need to consider a more sophisticated approach employing more accurate and detailed field or remotely-sensed data. Although stand age was able to predict the average volume per development phase of trees in the *Eucalyptus* plantation (i.e., age since tree planting), stand age was not sensitive to differences in volume expressed spatially and potentially caused by mortality, growth on less or high productive sites, potential nutrient/moisture limitations or pest-affected areas.

In subsequent models, based on the use of Landsat 8 OLI multispectral data, Sentinel-1B SAR image data, and terrain attributes alone, the estimates of volume were generally poor. In each case, the RMSE > 32 m^3 ha^{-1} and the R^2 < 41%. Employing eight Hyperion EO-1 (total bands = 220)-derived vegetation indices and multiple-linear regression models to estimate *Eucalyptus* stand volume in Brazil, Canavesi, Ponzoni, and Valeriano (2010) obtained a RMSE equal to 43.73 m^3 ha^{-1}, a higher value than the RMSE obtained in our study

using Landsat multispectral optical data alone. Gama, dos Santos, and Mura (2010) used airborne interferometric and polarimetric SAR data in X and P bands to estimate volume of *Eucalyptus* plantations, and obtained RMSE=33.56 m^3 ha^{-1}. The RMSE obtained by these authors is similar to our results using SAR data to estimate volume. In our study, the least accurate model performance was reported when using the terrain attributes alone as predictor variables (RMSE=39.77 m^3 ha^{-1}, and R^2 = 9.70%). The use of Sentinel-1B SAR variables alone led to higher RMSE and lower variance explained than the use of Landsat 8 OLI optical data alone; but both of these remote sensing datasets outperformed the use of terrain attributes alone. Note that the differences in multispectral and SAR data performance were not large; RMSE and R^2 differences with the best Landsat 8 OLI optical result and the best Sentinel-1B SAR result were within 2.48 m^3 ha^{-1} and 7%, respectively. Saturation problems are considered to be one of the main causes of the relatively poor performance when using multispectral and SAR image data alone in our study. Using ALOS/PALSAR L-band data, Bagidadi et al. (2015) reported a backscatter saturation aboveground biomass level of 50 t ha^{-1} reached at age 3 years in *Eucalyptus* plantations in Brazil. Dos Reis et al. (2018) assessed the influence of age in the relationship between volume of *Eucalyptus* stands and the spectral response in Landsat spectral bands and multispectral indices, and found that the relationship reached the maximum value at age 5 years. After this age, the correlation between volume and Landsat spectral data was affected by saturation issues.

The use of the combined stand age with multispectral image, SAR image and terrain attributes (*i.e.*, Experiments (v) - (vii)) in RF models resulted in better volume estimates than those obtained when using the individual multispectral, SAR and digital terrain datasets (RMSE < 28 m^3 ha^{-1} and R^2 > 55%). While the remote sensing images and digital terrain datasets did not perform as well as the stand age estimates when used alone, the best model

performance (Experiment (viii)) was obtained when using a combination of the multispectral, SAR, terrain attributes in Experiments (v), (vi) and (vii) with stand age. The improvement in the strength of the models and the higher accuracy of volume estimates, which integrated stand age, multispectral and SAR remotely sensed data (Landsat 8 OLI and Sentinel-1B images) with terrain attributes in *Eucalyptus* plantations, were significant and resulted in the best performance RF model (R^2 = 71.43% and RMSE = 22.33 m^3 ha^{-1}) (Experiment (viii)). In a South Africa study, Dube et al. (2017) demonstrated that integrating multispectral SPOT 5 image and ancillary data (age and rainfall metrics) significantly improved volume estimation in *Eucalyptus* plantations (R^2 = 77% and RMSE = 36.02 m^3 ha^{-1}). Others have used ASTER satellite data, age and site index as independent variables; for example, Gebreslasie, Ahmed, and van Aardt (2010) estimated *Eucalyptus* volume with R^2= 88% (no area estimates were provided). These results confirmed the increased performance of a multi-source dataset in explaining *Eucalyptus* volume, which was also noted in the present study.

Figure 4 illustrates the one-to-one relationship between observed and predicted volume of *Eucalyptus* plantation as obtained using the two best-performing models (i.e., Experiments (vii) and (viii)). Both experiments resulted in a predictable dispersion of the observed values in relation to the field-measured values close to the axis of 45°, indicating increased predicted value precision in comparison to observed values. Figure 5 contains maps showing the spatial distribution of stand volume within the study area obtained from these two models corresponding with the date of the most recent (June 2017) field inventory measurement of volume. These two maps captured similar patterns in areas of higher and lower stand volumes, however, Experiment (vii) using stand age and terrain attributes (Figure 5 (a)) resulted in a smoother volume map than Experiment (viii) using the combined multispectral, SAR and digital terrain data

with stand age (Figure 5 (b)). One interpretation of the patterns in these two maps is that the multispectral and SAR imagery depict a higher level of volume spatial variability within the plantation. The use of the spatially-coherent remote sensing variables served to adjust the estimates of volume across the plantation when compared to plot-based field inventory estimates. Note that volume estimates were reasonably well distributed compared to the original range of field-measured volumes (Table 1, measurement of 2017).

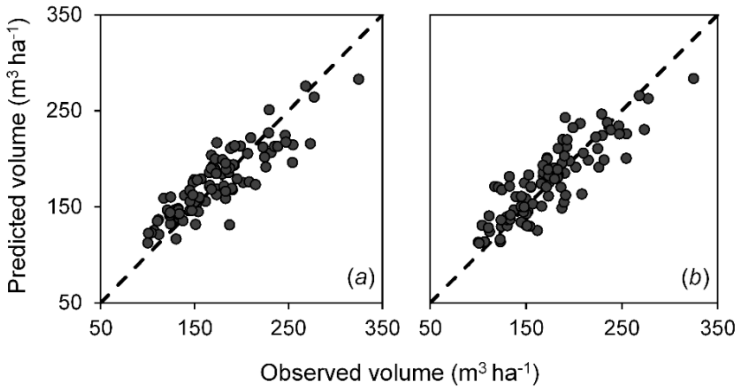


Figure 4. Scatterplots of the predicted vs. observed volume: (a) Experiment (vii) based on stand age with terrain attributes in validation samples (105 plots), and (b) Experiment (viii), based on the best combination of stand age with Landsat 8 OLI multispectral data, Sentinel-1B SAR image data, terrain attibutes in validation samples (105 plots). A 1:1 line (black, dashed) is provided for reference.

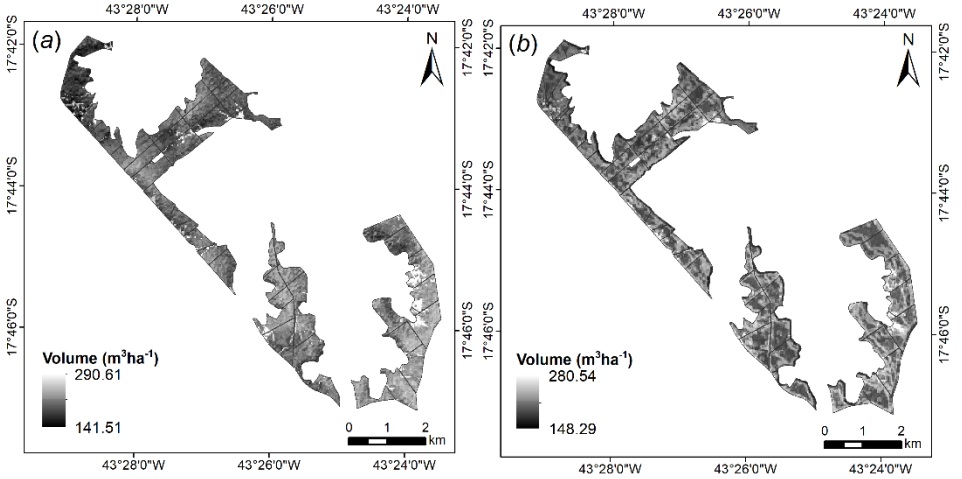


Figure 5. Spatial distribution of volume estimates in the 1,498 ha *Eucalyptus* plantation obtained with the best-performing models: (a) Experiment (vii) based on stand age with terrain attibutes; and (b) Experiment (viii) based on stand age with the best combination of Landsat 8 OLI multispectral data, Sentinel-1B SAR image data, terrain attibutes together.

The estimated volumes using the RF model based on stand age plus DEM geomorphometrics in Experiment (vii) varied between 141.51-290.61 m^3 ha^{-1}, with a mean value of 208.97 m^3 ha^{-1}. In Experiment (viii), the RF model based on stand age plus multispectral, SAR and DEM data, the variability was 148.29-280.54 m^3 ha^{-1}, with a mean value of 197.38 m^3 ha^{-1}. In most of the study area, the estimated volumes predicted in Experiments (vii) and (viii) ranged between 180-230 m^3 ha^{-1}. The total volume estimation across the 1,498 ha of *Eucalyptus* plantation when applying stand age plus the best available predictive variables (i.e., selected from the Landsat 8 OLI, Sentinel-1B and DEM-based datasets) was 294,751.45 m^3. The mean annual increment in volume ranged from 20.84 to 46.49 m^3 ha^{-1} $year^{-1}$ depending on plot location. These values are consistent with mean annual increments in *Eucalyptus* plantations in Brazil

compiled from historical plantation records across the entire country (Gonçalves et al. 2013).

The interpretation of RF Variable Importance confirmed that the use of stand age with multi-source remotely sensed data increased *Eucalyptus* volume estimation accuracy in a predictable pattern that was consistent with the idea of multi-source remote sensing data capturing plantation spatial variability in volume (Table 4). For example, while stand age was always the most important variable (IncMSE = 45%), subsequent variables were typically selected in a predictable pattern; first, variables from the Landsat 8 OLI multispectral image (IncMSE = 25%), second, variables from the Landsat 8 OLI and Sentinel-1B SAR textures (IncMSE = 12% and 11%, respectively), third, the terrain attributes (IncMSE = 9%), and fourth, indices from the Landsat 8 OLI multispectral image (IncMSE = 7%). This pattern is represented in Table 4 for the best-performing RF model. The PC descriptions in Table 4 were based on the importance of original variables in the loadings of PCA, *i.e.*, the variables with higher contribution in each component. Landsat 8 OLI G, NIR, SWIR and texture were the most important selections (after stand age), followed by Sentinel-1B SAR textures, and then terrain attributes (in particular, slope, relative slope position and incidence value). Each of the selected variables contributed at least with 7% to the explanation of variance in the best performing RF model of volume.

Table 4. Generalized ranking and description of most important groups and individual multi-source variables as measured by RF Variable Importance metric in the best RF model predicting stand volume with 71% accuracy (RMSE = 22.33 m^3 ha^{-1}) for 351 plots (246 training dataset, 105 validation dataset). (Note that a threshold of 7% of the variance explained was used to rank variables in this table).

Generalized ranking of most important group of remotely sensed variables[1]	**RF Selected variables**	**Description**
Landsat 8 OLI multispectral bands	PC1	Brightness across all spectral bands (except NIR)
	PC2	Contrast between visible and NIR bands
	PC3	Contrast between NIR and green bands
	PC4	Contrast between green and SWIR bands
Landsat 8 OLI multispectral image texture	PC5	Contrast between mean green and SWIR high and low spatial frequency texture
	PC7	Contrast between mean green and SWIR high spatial frequency texture
Sentinel-1B SAR texture	PC5	Mean texture of VH backscatter (high spatial frequency)
	PC7	Mean texture of VH backscatter (low spatial frequency)
Terrain attributes	Analytical hillshade	Incidence value
	Relative slope position	Slope position (in percent)
	Vertical distance to channel network	Relative slope/drainage position
Landsat 8 OLI multispectral indices	PC1	Equally weighted combination of multispectral indices

Where: PC = principal component, DEM= Digital Elevation Model, SAR= Synthetic Aperture Radar. [1]listed from top to bottom.

The use of principal components instead of raw remotely sensed variables (*i.e.*, spectral bands, multispectral indices, textures measures, and so on) in combination with RF reduced data dimensionality and potential uncertainties in data selection based on RF variable importance rank (Fayad et al. 2014; Maack et al. 2015; Fedrigo et al. 2018; Pourrahmati et al. 2018). Remotely-sensed data are typically highly inter-correlated, and resulting multicollinearity can cause singularities in the RF models. Consequently, variable importance scores may not be reliable (Strobl et al. 2008). In one study by Pourrahmati et al. (2018), principal components were less error-prone than waveform metrics for estimation of canopy height using ICESat/GLAS and optical images data. In another study, Maack et al. (2015) used spectral, textural and photogrammetric information from very-high resolution (VHR) stereo satellite data (Pléiades and WorldView-2) to estimate forest biomass across two test sites located in Chile and Germany. They demonstrated that the implementation of PCA as predictor selection procedure minimizes predictor inter-correlation and aggregates information content of the respective predictor type, as was also observed in our study. Nonetheless, further research may explore the benefits of using PCA compared with other techniques separately (e.g., Variance Inflation Factor) for predictor selection in data with multicollinearity using RF models.

To summarize, predictive RF models of *Eucalyptus* forest plantation volume using stand age alone, or stand age with multispectral, SAR and digital terrain data (in separate individual estimations), were less accurate than models constructed using an integrated multi-source dataset that yielded the best selection or combinations of these variables. The multi-source remote sensing approach can mitigate image acquisition issues, such as cloud cover restrictions on multispectral image acquisitions, and may also be effective in quantifying

expected productivity variability in plantations that result from mortality, growth limitations, pests, disease, and silviculture interventions.

6. Conclusion

This research examined the accuracy of forest volume estimation in a 1,498 ha *Eucalyptus* plantation in Minas Gerais state, Brazil, using combinations of stand age, Landsat 8 OLI multispectral imagery, Sentinel-1B SAR data, and ALOS PALSAR DEM-derived terrain attributes. A Random Forest machine learning algorithm employed annual forest inventory field measurements in 246 training plots for the years 2015-2017. Overall, the use of stand age alone provided 49% prediction accuracy in an independent set of 105 plots (RMSE = 30.98 m^3 ha^{-1}). A combination of Landsat 8 OLI multispectral, Sentinel-1B SAR, and digital terrain data with stand age improved the plantation volume estimation to greater than 71% accuracy (RMSE = 22.33 m^3 ha^{-1}). Increased model performance was interpreted to be a result of spatial variability of growth in the plantation, which was captured in the multi-source remote sensing image data. The approach used in this study provides a framework for integrating field data and the recent free-and-readily available multi-source remote sensing data, highlighting an approach that significantly improved spatial prediction of volume estimation in *Eucalyptus* stands over that which could be obtained using age alone. Although established *Eucalyptus* plantations in Brazil are typically a relatively homogenous, single species monoculture, the approach in this study can be expected to provide similar results in different forest conditions. For example, the methods used in this study can be applied to additional forest attribute estimation (e.g., basal area, height, diameter and biomass) in plantations and natural forests in other regions of the world where remote sensing image data availability may be limited. Further research will determine model performance using object-based image analysis, multi-seasonal imagery,

specific terrain models that characterize soil and hydrological variability, and advanced deep learning algorithms.

References

Adams, H.R., H.R. Barnard, and A.K. Loomis. 2014. "Topography Alters Tree Growth–climate Relationships in a Semi-Arid Forested Catchment." *Ecosphere* 5 (11):art148. doi:10.1890/ES14-00296.1.

Alvares, C.A., J.L. Stape, P.C. Sentelhas, J.L.M. Gonçalves, and G. Sparovek. 2013. "Köppen's Climate Classification Map for Brazil." *Meteorologische Zeitschrift* 22 (6):711–28. doi:10.1127/0941-2948/2013/0507.

Aslan, A., A.F. Rahman, M.W. Warren, and S.M. Robeson. 2016. "Mapping Spatial Distribution and Biomass of Coastal Wetland Vegetation in Indonesian Papua by Combining Active and Passive Remotely Sensed Data." *Remote Sensing of Environment* 183: 65–81. doi:10.1016/j.rse.2016.04.026.

Baghdadi, N., G. le Maire, J. S. Bailly, Kenji Ose, Y. Nouvellon, M. Zribi, C. Lemos, and R. Hakamada. 2015. "Evaluation of ALOS/PALSAR L-band data for the estimation of *Eucalyptus* plantations aboveground biomass in Brazil." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8 (8):3802–11. doi:10.1109/JSTARS.2014.2353661.

Baghdadi, N., G. le Maire, I. Fayad, J. S. Bailly, Y. Nouvellon, C. Lemos, and R. Hakamada. 2014. "Testing different methods of forest height and aboveground biomass estimations from ICESat/GLAS data in *Eucalyptus* plantations in Brazil." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7 (1):290–99. doi:10.1109/JSTARS.2013.2261978.

Belgiu, M., and L. Drăgu. 2016. “Random Forest in Remote Sensing: A Review of Applications and Future Directions.” *ISPRS Journal of Photogrammetry and Remote Sensing* 114: 24–31. doi:10.1016/j.isprsjprs.2016.01.011.

Breiman, L. 2001. “Random Forests.” *Machine Learning* 45 (1):5–32. doi:10.1023/A:1010933404324.

Canavesi, V., F.J. Ponzoni, and M.M. Valeriano. 2010. “Estimativa de volume de madeira em plantios de Eucalyptus spp. utilizando dados hiperespectrais e dados topográficos.” *Revista Árvore* 34 (3):539–49. doi:0.1590/S0100-67622010000300018.

Cutler, M.E.J, D.S. Boyd, G.M. Foody, and A. Vetrivel. 2012. “Estimating Tropical Forest Biomass with a Combination of SAR Image Texture and Landsat TM Data: An Assessment of Predictions between Regions.” *ISPRS Journal of Photogrammetry and Remote Sensing* 70: 66–77. doi:10.1016/j.isprsjprs.2012.03.011.

Daroczi, G., 2015. *Mastering Data Analysis with R*, Packt Publishing, OpenSource.

Dos Reis, A.A., F.W. Acerbi Júnior, J. M. de Mello, L.M.T. de Carvalho, and L.R. Gomide. 2018. “Relationship between spectral data and dendrometric variables in *Eucalyptus* sp. stands.” *Floresta e Ambiente* 25 (2):2–9. doi:10.1590/2179-8087.017015.

Dube, T., and O. Mutanga. 2015. “Investigating the Robustness of the New Landsat-8 Operational Land Imager Derived Texture Metrics in Estimating Plantation Forest Aboveground Biomass in Resource Constrained Areas.” *ISPRS Journal of Photogrammetry and Remote Sensing* 108:12–32. doi:10.1016/j.isprsjprs.2015.06.002.

Dube, T., and O. Mutanga. 2016. “The Impact of Integrating WorldView-2 Sensor and Environmental Variables in Estimating Plantation Forest

Species Aboveground Biomass and Carbon Stocks in uMgeni Catchment, South Africa." *ISPRS Journal of Photogrammetry and Remote Sensing* 119: 415–425. doi:10.1016/j.isprsjprs.2016.06.017.

Dube, T., O. Mutanga, E.M. Abdel-Rahman, R. Ismail, and R. Slotow. 2015. "Predicting *Eucalyptus* spp. Stand Volume in Zululand, South Africa: An Analysis Using a Stochastic Gradient Boosting Regression Ensemble with Multi-Source Data Sets." *International Journal of Remote Sensing.* 36(14): 3751–3772. doi:10.1080/01431161.2015.1070316.

Dube, T., M. Sibanda, C. Shoko, and O. Mutanga. 2017. "Stand-Volume Estimation from Multi-Source Data for Coppiced and High Forest *Eucalyptus* spp. Silvicultural Systems in KwaZulu-Natal, South Africa." *ISPRS Journal of Photogrammetry and Remote Sensing* 132:162–69. doi:10.1016/j.isprsjprs.2017.09.001.

Fayad, I., N. Baghdadi, J.S. Bailly, N. Barbier, V. Gond, M. Hajj, F. Fabre, and B. Bourgine. 2014. "Canopy height estimation in French Guiana with LiDAR ICESat/GLAS data using Principal Component Analysis and Random Forest Regressions." *Remote Sensing* 6 (12):11883–914. doi:10.3390/rs61211883.

Fedrigo, M., G.J. Newnham, N.C. Coops, D.S. Culvenor, D.K. Bolton, and C.R. Nitschke. 2018. "Predicting Temperate Forest Stand Types Using Only Structural Profiles from Discrete Return Airborne Lidar." *ISPRS Journal of Photogrammetry and Remote Sensing* 136:106–19. doi:10.1016/j.isprsjprs.2017.11.018.

Franklin, S.E., M.A. Wulder, and M.B. Lavigne. 1996. "Automated Derivation of Geographic Window Sizes for Use in Remote Sensing Digital Image Texture Analysis." *Computers & Geosciences* 22 (6):665–73. doi:10.1016/0098-3004(96)00009-X.

Franklin, S.E., and O.S. Ahmed. 2017. "Object-Based Wetland Characterization Using Radarsat-2 Quad-Polarimetric SAR Data, Landsat-8 OLI Imagery, and Airborne Lidar-Derived Geomorphometric Variables." *Photogrammetric Engineering & Remote Sensing* 83 (1):27–36. doi:10.14358/PERS.83.1.27.

Gama, F.F., J.R. Santos, and J.C. Mura. 2010. "*Eucalyptus* Biomass and Volume Estimation Using Interferometric and Polarimetric SAR Data." *Remote Sensing* 2: 939-956. doi:10.3390/rs2040939.

Gama, F.F., J.R. Santos, and J.C. Mura. 2016. "Continuous Monitoring of Biophysical *Eucalyptus* sp. Parameters Using Interferometric Synthetic Aperture Radar Data in P and X Bands." *Journal of Applied Remote Sensing* 10 (2):26002. doi:10.1117/1.JRS.10.026002.

Gao, T., J. Zhu, S. Deng, X. Zheng, J. Zhang, G. Shang, and L. Huang. 2016. "Timber Production Assessment of a Plantation Forest: An Integrated Framework with Field-Based Inventory, Multi-Source Remote Sensing Data and Forest Management History." *International Journal of Applied Earth Observation and Geoinformation* 52: 155-165. doi:10.1016/j.jag.2016.06.004.

Gebreslasie, M.T., F.B. Ahmed, and J.A.N van Aardt. 2010. "Predicting Forest Structural Attributes Using Ancillary Data and ASTER Satellite Data." *International Journal of Applied Earth Observation and Geoinformation*. 12S: S23–S26. doi.org/10.1016/j.jag.2009.11.006.

Gebreslasie, M.T., F.B. Ahmed, and J.A.N van Aardt. 2011. "Extracting Structural Attributes from IKONOS Imagery for *Eucalyptus* Plantation Forests in Kwazulu-Natal, South Africa, Using Image Texture Analysis and Artificial Neural Networks." *International Journal of Remote Sensing*. 32(22): 7677-7701. doi.org/10.1080/01431161.2010.527392.

Gonçalves, J.L.M., C.A. Alvares, A.R. Higa, L.D. Silva, A.C. Alfenas, J. Stahl, S.F.B. Ferraz, et al. 2013. "Integrating Genetic and Silvicultural Strategies to Minimize Abiotic and Biotic Constraints in Brazilian Eucalypt Plantations." *Forest Ecology and Management* 301:6–27. doi:10.1016/j.foreco.2012.12.030.

Goodbody, T.R.H., N.C. Coops, P.L. Marshall, P. Tompalski, and P. Crawford. 2017. "Unmanned Aerial Systems for Precision Forest Inventory Purposes: A Review and Case Study." *The Forestry Chronicle*. 93(1): 71-81. doi:10.5558/tfc2017-012.

Haralick, R.M., K. Shanmugam, and I. Dinstein. 1973. "Textural Features for Image Classification." IEEE Transactions on Systems, Man, and Cybernetics SMC-3 (6):610–21. doi:10.1109/TSMC.1973.4309314.

Hengl, T., Reuter, H.I. (eds) 2009. *Geomorphometry: Concepts, Software, Applications.* Developments in Soil Science: Elsevier.

IBÁ - Brazilian Tree Industry, 2016. *Brazilian Tree Industry - Annual Report*. Brazilian Tree Industry (IBA), São Paulo.

Ismail, R., and O. Mutanga. 2010. "A Comparison of Regression Tree Ensembles: Predicting Sirex Noctilio Induced Water Stress in *Pinus Patula* Forests of KwaZulu-Natal, South Africa." *International Journal of Applied Earth Observation and Geoinformation* 12:S45–51. doi:10.1016/j.jag.2009.09.004.

Ismail, R., H. Kassier, M. Chauke, F. Holecz, and N. Hattingh. 2015. "Assessing the Utility of ALOS PALSAR and SPOT 4 to Predict Timber Volumes in Even-Aged *Eucalyptus* Plantations Located in Zululand, South Africa." *Southern Forests: A Journal of Forest Science* 77 (3):203–11. doi:10.2989/20702620.2014.1001681.

Le Maire, G., C. Marsden, Y. Nouvellon, C. Grinand, R. Hakamada, J. L. Stape, and J.P. Laclau. 2011. "MODIS NDVI time-series allow the monitoring

of Eucalyptus plantation biomass." *Remote Sensing of Environment* 115 (10):2613–25. doi:10.1016/j.rse.2011.05.017.

Lee, C., S. Youn, T. Jeong, E. Lee, and J. Serra-Sagristà. 2015. "Hybrid Compression of Hyperspectral Images Based on PCA With Pre-Encoding Discriminant Information." *IEEE Geoscience and Remote Sensing Letters* 12 (7):1491–95. doi:10.1109/LGRS.2015.2409897.

López-Sánchez, C.A., P. García-Ramírez, R. Resl, J.C. Hernández-Díaz, P.M. López-Serrano, and C. Wehenkel. 2017. "Modelling Dasometric Attributes of Mixed and Uneven-Aged Forests Using Landsat-8 OLI Spectral Data in the Sierra Madre Occidental, Mexico." *IForest* 10: 288-295. doi:10.3832/ifor1891-009.

López-Serrano, P.M., C.A. López-Sánchez, J.G. Álvarez-González, and J. García-Gutiérrez. 2016. "A Comparison of Machine Learning Techniques Applied to Landsat-5 TM Spectral Data for Biomass Estimation*." Canadian Journal of Remote Sensing* 42 (6):690–705. doi:10.1080/07038992.2016.1217485.

Lu, D., Q. Chen, G. Wang, L. Liu, G.Li, and E. Moran. 2014. "A Survey of Remote Sensing-Based Aboveground Biomass Estimation Methods in Forest Ecosystems." *International Journal of Digital Earth* 9: 63-105. doi:10.1080/17538947.2014.990526.

Maack, J., T. Kattenborn, F.E. Fassnacht, F. Enßle, J. Hernández, P. Corvalán, and B. Koch. 2015. "Modeling Forest Biomass Using Very-High-Resolution Data - Combining Textural, Spectral and Photogrammetric Predictors Derived from Spaceborne Stereo Images." *European Journal of Remote Sensing* 48(1): 245-261. doi:10.5721/EuJRS20154814.

McRoberts, R.E., T. Gobakken, and E. Næsset. 2012. "Post-Stratified Estimation of Forest Area and Growing Stock Volume Using Lidar-

Based Stratifications." *Remote Sensing of Environment* 125: 157–66. doi:10.1016/j.rse.2012.07.002.

Mohamedou, C., T. Tokola, and K. Eerikäinen. 2017. "LiDAR-Based TWI and Terrain Attributes in Improving Parametric Predictor for Tree Growth in Southeast Finland." *International Journal of Applied Earth Observation and Geoinformation* 62:183–91. doi:10.1016/j.jag.2017.06.004.

Ng, S.C. 2017. "Principal Component Analysis to Reduce Dimension on Digital Image." *Procedia Computer Science* 111:113–19. doi:10.1016/j.procs.2017.06.017.

Novelli, A., M.A. Aguilar, A. Nemmaoui, F.J. Aguilar, and E. Tarantino. 2016. "Performance Evaluation of Object Based Greenhouse Detection from Sentinel-2 MSI and Landsat 8 OLI Data: A Case Study from Almería (Spain)." *International Journal of Applied Earth Observation and Geoinformation* 52: 403–411. doi:10.1016/j.jag.2016.07.011.

Omar, H., M.A. Misman, and A.R. Kassim. 2017. "Synergetic of PALSAR-2 and Sentinel-1A SAR Polarimetry for Retrieving Aboveground Biomass in Dipterocarp Forest of Malaysia." *Applied Sciences* 7 (7):675. doi:10.3390/app7070675.

Ota, T., M. Ogawa, N. Mizoue, K. Fukumoto, and S. Yoshida. 2017. "Forest Structure Estimation from a UAV-Based Photogrammetric Point Cloud in Managed Temperate Coniferous Forests." *Forests* 8 (9):343. doi:10.3390/f8090343.

Packalén, P., L. Mehtätalo, and M. Maltamo. 2011. "ALS-based estimation of plot volume and site index in a *Eucalyptus* plantation with a nonlinear mixed-effect model that accounts for the clone effect." *Annals of Forest Science* 68 (6):1085–92. doi:10.1007/s13595-011-0124-9.

Payn, T., J.M. Carnus, P. Freer-Smith, M. Kimberley, W. Kollert, S. Liu, C. Orazio, L. Rodriguez, L.N. Silva, and M.J. Wingfield. 2015. "Changes

in Planted Forests and Future Global Implications." *Forest Ecology and Management* 352:57–67. doi: 10.1016/j.foreco.2015.06.021.

Pourrahmati, M.R., N. Baghdadi, A.A Darvishsefat, M. Namiranian, V. Gond, J. S. Bailly, and N. Zargham. 2018. "Mapping Lorey's Height over Hyrcanian Forests of Iran Using Synergy of ICESat/GLAS and Optical Images." *European Journal of Remote Sensing* 51 (1):100–115. doi:10.1080/22797254.2017.1405717.

Racine, E.B., N.C. Coops, B. St-Onge, and J. Begin. 2014. "Estimating Forest Stand Age from LiDAR-Derived Predictors and Nearest Neighbor Imputation." *Forest Science* 60(1):128–136. doi:10.5849/forsci.12-088.

R Core Team. (2017). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Raimundo, M.R., H.F. Scolforo, J.M. de Mello, J.R.S. Scolforo, J.P. McTague, and A.A. dos Reis. 2017. "Geostatistics Applied to Growth Estimates in Continuous Forest Inventories." *Forest Science* 63(1):29–38. doi:10.5849/FS-2016-056.

Riihimäki, H., J. Heiskanen, and M. Luoto. 2017. "The Effect of Topography on Arctic-Alpine Aboveground Biomass and NDVI Patterns." *International Journal of Applied Earth Observation and Geoinformation* 56: 44–53. doi:10.1016/j.jag.2016.11.005.

Roy, D.P., M.A. Wulder, T.R. Loveland, C.E. Woodcock, R.G. Allen, M.C. Anderson, D. Helder, et al. 2014. "Landsat-8: Science and Product Vision for Terrestrial Global Change Research." *Remote Sensing of Environmen*t 145:154–72. doi:10.1016/j.rse.2014.02.001.

Sarker, L.R., and J.E. Nichol. 2011. "Improved Forest Biomass Estimates Using ALOS AVNIR-2 Texture Indices." *Remote Sensing of Environment* 115(4):968–77. doi:10.1016/j.rse.2010.11.010.

Scolforo, H.F., J.R.S. Scolforo, J.L. Stape, J.P. McTague, H. Burkhart, J. McCarter, F. Castro Neto, R.A. Loos, and R.C. Sartorio. 2017. "Incorporating Rainfall Data to Better Plan *Eucalyptus* Clones Deployment in Eastern Brazil." *Forest Ecology and Management* 391: 145-153. doi:10.1016/j.foreco.2017.02.025.

Shao, Z., and L. Zhang. 2016. "Estimating Forest Aboveground Biomass by Combining Optical and SAR Data: A Case Study in Genhe, Inner Mongolia, China." *Sensors* 16: 834. doi:10.3390/s16060834.

Shinzato, E.T., Y.E. Shimabukuro, N.C. Coops, P. Tompalski, and E.A. Gasparoto. 2017. "Integrating area-based and individual tree detection approaches for estimating tree volume in plantation inventory using aerial image and airborne laser scanning data." *iForest* 10 (1):296–302. doi:10.3832/ifor1880-009.

Strobl, C., A.-L. Boulesteix, T. Kneib, T. Augustin; and A. Zeileis. 2008. "Conditional variable importance for random forests." *BMC bioinformatics* 9 (1):307. doi: 10.1186/1471-2105-9-307.

Tompalski, P., N. Coops, P. Marshall, J. White, M. Wulder, and T. Bailey. 2018. "Combining multi-date airborne laser scanning and digital aerial photogrammetric data for forest growth and yield modelling." *Remote Sensing* 10 (3):347. doi:10.3390/rs10020347.

Torres, R., P. Snoeij, D. Geudtner, D. Bibby, M. Davidson, E. Attema, P. Potin, et al. 2012. "GMES Sentinel-1 Mission." *Remote Sensing of Environment* 120: 9–24. doi:10.1016/j.rse.2011.05.028.

Vaglio, G.L., F. Pirotti, M. Callegari, Q. Chen, G. Cuozzo, E. Lingua, C. Notarnicola, and D. Papale. 2017. "Potential of ALOS2 and NDVI to Estimate Forest above-Ground Biomass, and Comparison with Lidar-Derived Estimates." *Remote Sensing* 9:18. doi:10.3390/rs9010018.

Wu, C., H. Shen, A. Shen, J. Deng, M. Gan, J. Zhu, H. Xu, and K. Wang. 2016. "Comparison of Machine-Learning Methods for above-Ground Biomass Estimation Based on Landsat Imagery." *Journal of Applied Remote Sensing* 10 (3):35010. doi:10.1117/1.JRS.10.035010.

Zhang, C., S. Denka, H. Cooper, and D.R. Mishra. 2018. "Quantification of Sawgrass Marsh Aboveground Biomass in the Coastal Everglades Using Object-Based Ensemble Analysis and Landsat Data." *Remote Sensing of Environment* 204:366-379. doi:10.1016/j.rse.2017.10.018.

Zhao, P., D. Lu, Guangxing Wang, Lijuan Liu, Dengqiu Li, Jinru Zhu, and Shuquan Yu. 2016. "Forest Aboveground Biomass Estimation in Zhejiang Province Using the Integration of Landsat TM and ALOS PALSAR Data." *International Journal of Applied Earth Observation and Geoinformation* 53:1–15. doi:10.1016/j.jag.2016.08.007.

ARTICLE 3 – REMOTE SENSING-BASED MULTI-DATA APPROACH FOR ESTIMATION OF FOREST ATTRIBUTES: AN APPLICATION ON *EUCALYPTUS* PLANTATION IN MINAS GERAIS STATE, BRAZIL

Aliny Aparecida Dos Reis[1], Fausto Weimar Acerbi Junior[1], Steven E. Franklin[2], José Marcio de Mello[1], Antonio Carlos Ferraz Filho[3]

1. Department of Forest Science, Federal University of Lavras – UFLA, PO Box 3037, Lavras, Minas Gerais, Brazil, Zip Code 37200-000

2. School of Environment, Trent University, Peterborough, Ontario, Canada K9I 7B8

3. CPCE, Federal University of Piauí, Manoel Gracindo Avenue, km 1, 64900-000, Bom Jesus, Piauí, Brazil

*Corresponding author email: alinyreis@hotmail.com

Publication status: In preparation for submission to **International Journal of Applied Earth Observation and Geoinformation**

Abstract: Freely-available multi-source remotely-sensed data, such as Landsat 8 OLI, Sentinel-1B, and Sentinel-2A, offer a new opportunity for forest attribute estimation. The higher spectral and temporal resolution of such datasets are especially important for monitoring fast-growing forests, such as *Eucalyptus* plantations in Brazil. Using the random forest machine learning algorithm, we investigated the potential use of Sentinel-2A multispectral data to predict volume, basal area, and diameter at breast height (DBH) of *Eucalyptus* plantation compared with the previously available Landsat 8 OLI imagery when both multispectral data (*i.e.*, Sentinel-2A and Landsat 8 OLI) were combined with Sentinel-1B SAR data and Digital Elevation Model (DEM)-derived terrain attributes. Additionally, we assessed the benefits of object-based approach for forest attribute modelling and mapping compared with the pixel-based approach. Sentinel-2A multispectral data had similar to superior capabilities for *Eucalyptus* stand attributes estimates compared to Landsat 8 OLI when combined with Sentinel-1B SAR data and DEM-derived terrain attributes. Sentinel-2A performed better in predicting volume (Root Mean Square Error (RMSE) = 7.23%) and DBH (RMSE = 2.65%) in the pixel-based approach, while Landsat 8 OLI performed better in predicting basal area (RMSE = 5.09%) in the object-based approach. Thus, this study highlights the efficiency of freely-accessible multi-source datasets, such as Sentinel-1B, Sentinel-2A, ALOS/PALSAR DEM, and Landsat 8 OLI, for *Eucalyptus* plantation attributes estimation and also provides new insights into the opportunities and limitations related to the use of object-based approach for attributes estimation in planted forests.

Keywords: Sentinel-1B SAR, Sentinel-2A, object-based forest attribute modeling.

1. Introduction

In recent decades, remotely-sensed data have proven their usefulness to accurately provide information on forest ecosystems extent and characteristics (Chrysafis et al. 2017). Estimation of forest stand attributes using remotely sensed data has considerable significance for supporting activities related to forest inventory, planning, management, and monitoring from a local to a global scale (Chrysafis et al. 2017; Puliti et al. 2018; Reuveni et al. 2018). For such purposes, data from active or passive sensors have been used as predictor variables in combination with measures taken in the field from sampling plots. High spatial resolution sensors, light detection and ranging (LIDAR) and unmanned aerial vehicle (UAV) data have shown good potential for improving forest attribute estimates (e.g., Goodbody et al., 2017; Ota et al., 2017; Tompalski et al., 2018). However, their widespread application is hampered by cost limitations (Hawryło and Wężyk 2018).

On the other hand, open access to medium-resolution multispectral satellite images (e.g. Landsat-7, Landsat-8, and Resourcesat-2) have created an opportunity to estimate forest attributes through remote sensing data even more applicable (Puliti et al. 2018). Despite the constraints arising from scene heterogeneity, these images have long been used for forest attribute estimation and mapping in different forest types around the world (Baghdadi et al. 2015; Dube and Mutanga 2015; López-Sánchez et al. 2017; López-Serrano et al. 2016; Zhao et al. 2016; J. Zhang et al. 2018). All of these studies focused on single-sensor imagery; however, recent studies have shown that models combining multispectral optical and synthetic aperture radar (SAR) imagery performed substantially better than models using single-sensor imagery for predicting and mapping forest attributes (Ismail et al. 2015; Vaglio et al. 2017; Shao and Zhang 2016; Vafaei et al. 2018; Pham et al. 2018). In addition, the increasing availability of ancillary data also offer new alternatives for improving forest

attribute estimation accuracy. Recent studies have integrated attributes derived from digital elevation models (DEM) with optical and SAR data to represent terrain conditions directly related to forest growth in the modeling of forest attributes with good results (López-Sánchez et al. 2017; López-Serrano et al. 2016).

The recent deployment of European Space Agency (ESA)'s Sentinel operational satellites has made high spatial resolution C-band SAR and multispectral optical data freely available at global scale, and along with the Landsat 8 OLI and ALOS/PALSAR datasets, these satellites have strongly facilitated forest monitoring (Roy et al. 2014; Torres et al. 2012; Li and Roy 2017). The higher spectral and temporal resolutions of such datasets (*i.e.* Sentinel-1, Sentinel-2, and Landsat 8) are especially important for monitoring fast-growing forest, such as *Eucalyptus* plantations.

Eucalyptus plantations are the most widely distributed planted forest in Brazil. They cover more than 5.6 million hectares in the country, contribute with 17% of the harvested wood in the world (Ibá, 2016), and are managed with short cutting cycles ranging from 5 to 7 years (Scolforo et al. 2017; Raimundo et al. 2017). Most of these plantations are established for wood production as even-aged single-species monocultures, which leads to several distinct spectral, structural, and temporal characteristics that may affect their response in remotely sensed data (Ismail et al. 2015; Dube and Mutanga 2015; Dube et al. 2017).

Previous studies have examined data obtained from sensors on a number of different satellites, including SPOT, IKONOS, ASTER, ALOS/PALSAR, ICESat/GLAS, and Landsat 8 for forest attributes estimation in *Eucalyptus* plantations (Gebreslasie, Ahmed, and van Aardt 2010, 2011; Dube and Mutanga 2015; Ismail et al. 2015; Baghdadi et al. 2015; Dube et al. 2017). Although these remotely-sensed datasets have been proved useful in estimating forest attributes in general, more research is required to investigate the capabilities of new high

spectral, spatial, and temporal satellite resolution, such as Sentinel-1B and Sentinel-2A, in predicting forest attributes in *Eucalyptus* plantations.

Furthermore, the majority of studies estimating forest attributes from remote sensing data estimation have been conducted on a pixel-based approach by matching in-situ measurements with individual pixels (Dube and Mutanga 2015; Ismail et al. 2015; Dube et al. 2017; Gebreslasie, Ahmed, and van Aardt 2011). However, an object-based image approach can combine spatial and spectral information within image analysis using image objects as basic units instead of individual pixels to extract information about forest canopies (Chen et al. 2011, 2018; Kajisa et al. 2009), and have not been thoroughly explored for forest attribute estimation.

To the best of our knowledge, no study has been carried out aiming at estimating forest attributes based on the integration of the recently launched Sentinel-2A, Sentinel-1B SAR, Landsat-8 OLI with DEM-derived terrain attributes using both pixel-based and object-based approaches. For this reason, we investigated the potential use of Sentinel-2A multispectral data for *Eucalyptus* plantation attributes estimation compared with the previously available Landsat 8 OLI imagery when both multispectral data (*i.e.*, Sentinel-2A and Landsat 8 OLI) were combined with Sentinel-1B SAR data and terrain attributes. Additionally, we assessed the benefits of object-based approach for forest attribute modelling and mapping compared with the pixel-based approach by employing a nonparametric modelling approach.

2. Study area

This study was conducted in a 1,498 ha privately-managed *Eucalyptus* plantation in Diamantina municipality, northern Minas Gerais state, Brazil (Figure 1). The region experiences humid sub-tropical climatic conditions, corresponding to Köppen's climatic type Cwb, with the rainy season occurring

between the months of October and March, and mean annual rainfall of 1,468 mm. Mean daily temperatures during the hottest month typically do not exceed 22°C (Alvares et al. 2013). Elevations range between 700 m and 1100 m above-mean-sea-level. Soils are predominately Cambisols and Latosols.

The *Eucalyptus* plantation was established between November 2010 and October 2012, using clones of the species *E. urophylla* x *E. grandis*, which typically lead to uniform crown and tree density conditions. The stands were established with approximately 1100 trees per hectare, using initial spacing of 3.6 x 2.5 m. Intensive soil preparation and weed control measures have been practised since plantation establishment, with crown closure often occurring within the first 18 months.

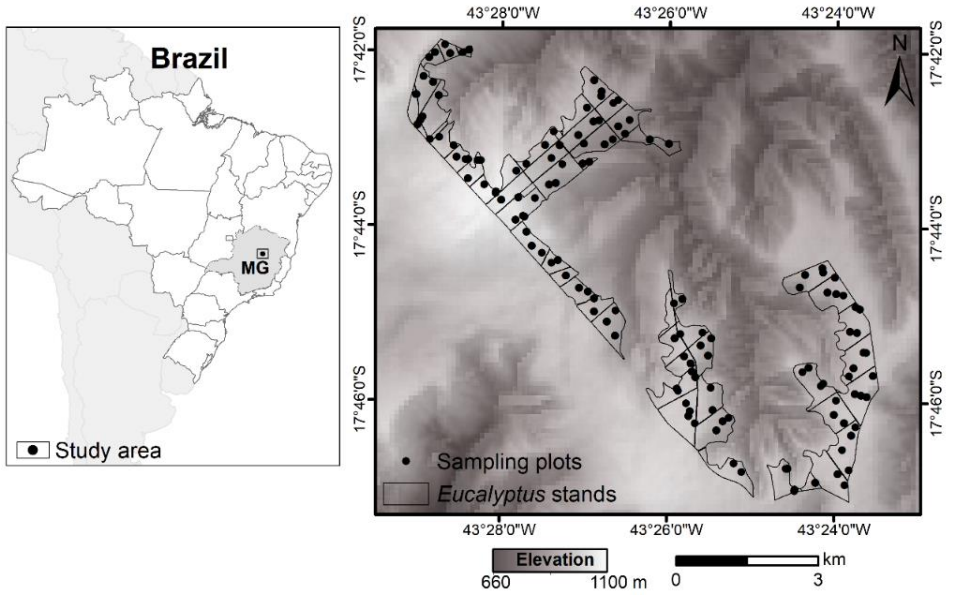


Figure 1. Location of the study area in northern Minas Gerais state, Brazil, and spatial distribution of sampling plots used in the field inventory.

3. Methods

3.1 Field data

Field data collection was performed in June 2017 in 97 plots across the *Eucalyptus* plantation using a standard continuous forest inventory (CFI) system. The sample plots were distributed systematically and located in the field using survey-grade real time kinetic (RTK) GPS. A plot size of 25 m × 20 m (500 m^2) was used. Within each plot, tree diameter at breast height (DBH) of all stems and the total height of the first 15 trees with normal stems (without bifurcation or any other defect) were measured. Estimates of basal area (m^2 ha^{-1}) and total stem volume (m^3 ha^{-1}) were obtained from the information collected in the plots. The individual volume equations were based on DBH and total height. All fitted equations showed high coefficient of determination (R^2 > 97%) and low residual standard error (S_{xy} < 0.05 m^3). The plot-based field inventory values for the *Eucalyptus* plantation attributes are summarized in Table 1.

Table 1. *Eucalyptus* plantation description and attribute estimates obtained from forest inventory field data collection in 97 sampling plots.

Variable	**Mean**	**Sd**	**Range**
Stand age (years)	6.44	0.21	4.70 – 6.60
Total height (m)	23.43	2.11	18.63 – 28.62
DBH (cm)	15.01	0.71	12.71 – 17.48
Basal area (m^2 ha^{-1})	19.81	2.01	13.36 – 24.96
Volume (m^3 ha^{-1})	211.49	38.99	131.28 – 338.57

Where DBH = Diameter at breast height and Sd = Standard Deviation.

3.2 *Remote sensing data collection and preprocessing*

The remotely-sensed data acquired for this study were: Landsat 8 OLI and Sentinel-2A multispectral imagery, Sentinel-1B SAR data, and ALOS/PALSAR (Advanced Land Observing Satellite/Phase Array type L-band Synthetic Aperture Radar) DEM data.

The selected Landsat 8 OLI satellite scene (Path 218 and Row 72) covering the study area was acquired on the closest date to the field measurement dates (i.e., August 20, 2017), and was relatively cloud-free and without obvious radiometric issues. This scene was obtained from the USGS database (United States Geological Survey) as Landsat level 2 surface reflectance collection 1 product with a 30 m geometric resolution, less than one-half pixel RMSE locational error, and the following bands: coastal aerosol (C, 0.43–0.45 μm), blue (B, 0.45–0.51 μm), green (G, 0.53–0.59 μm), red (R, 0.64–0.67 μm), near infrared (NIR, 0.85–0.88 μm), shortwave infrared-1 (SWIR1, 1.57–1.65 μm), shortwave infrared-2 (SWIR2, 2.11–2.29 μm).

Sentinel data were acquired from Copernicus Open Access Hub (https://scihub.copernicus.eu/). Atmospherically-corrected bottom-of-atmosphere (BoA) Sentinel-2 data was produced using the Sen2Cor processor (currently version 2.5.5), available on the Sentinel Application Platform (SNAP) and developed by ESA to perform atmospheric, terrain, and cirrus correction of top-of-atmosphere Level-1C input data. We selected an image with acquisition date close to the Landsat scene acquisition date, with relatively cloud-free coverage, and taken on August 30, 2017. Sentinel-2A satellite multispectral image contains 13 spectral bands in the visible and near-infrared (VNIR) to shortwave infrared (SWIR) spectral range: four bands have a spatial resolution of 10 m (blue (B): 490 nm, green (G): 560 nm, red (R): 665 nm and near-infrared (NIR): 842 nm); six bands have a spatial resolution of 20 m (red edge 1: 705 nm, red edge 2: 740 nm, red edge 3: 783 nm, narrow NIR: 865 nm, SWIR1:

1610 nm and SWIR2: 2190 nm); and three bands have a spatial resolution of 60 m (coastal aerosol: 443 nm, water vapor: 940 nm and SWIR cirrus: 1375 nm). In this study, we used only the Sentinel-2A bands with spatial resolutions of 10 m and 20 m.

The selected Sentinel-1B SAR scene was acquired on the date that most coincided with the field measurement dates (i.e., August 19, 2017, respectively). We used Ground Range Detected (GRD) Level-1 data acquired in interferometric wide swath (IW) mode with dual vertical-vertical (VV) and vertical horizontal (VH) polarization. These SAR data were calibrated, radiometrically- and geometrically-corrected and filtered using the Sentinel-1 Toolbox implemented in the software SNAP and have a spatial resolution of 10 m.

A high resolution (12.5 meters) DEM from ALOS/PALSAR DEM data source was used in this study to quantify the geomorphometric conditions in the *Eucalyptus* plantation area. Prior to calculation of the terrain attributes, the DEM was preprocessed with a standard gap/sink-filling error detection and noise removal routine. Basic terrain attributes included the following terrain variables: slope, aspect (converted to linear aspect with a zero value for north facing slopes), total curvature, relative slope position, and slope length (Hengl and Reuter 2009). In addition, several specific terrain attributes were computed to characterize more complex geomorphometric and hydrological variability (Adams, Barnard, and Loomis 2014). First, derivatives of slope were combined with roughness and distance measures to create 9 specific terrain attributes: Terrain Roughness Index (TRI), Surface Relief Ratio (SRR), Surface Area Ratio (SAR), Compound Topographic Index (CTI), Topographic Wetness Index (TWI), SAGA Topographic Wetness Index (SAGA TWI), Integrated Moisture Index (IMI), vertical distance to channel network (VDCN), and Topographic Position Index (TPI). These variables were selected to provide insight into site-

specific water-, gravity- and wind-field conditions expected to vary across the large geographic region of interest. Second, slope derivatives were combined with estimates of insolation to create 3 attributes to characterize mean solar-field conditions: Topographic Openness, Site Exposure Index (SEI), and Heat Load Index (HLI) (Mohamedou, Tokola, and Eerikäinen 2017). All geomorphometric attributes were calculated using either the SAGA GIS software package (v. 5.0.0) or the Geomorphometry and Gradient Metrics Toolbox v2.0 (Evans et al., 2014).

3.3 Extraction of textural images from Landsat 8 OLI, Sentinel-2A and Sentinel-1B SAR data

The Grey Level Co-occurrence Matrix (GLCM) statistical texture approach (Haralick, Shanmugam, and Dinstein 1973) was used to generate texture images (Lu et al. 2016) from the near-infrared (NIR) bands of Landsat 8 OLI and Sentinel-2A multispectral images, and the VV Sentinel-1B SAR backscatter data. These images appeared visually to contain high image contrast in the *Eucalyptus* plantations of interest. Eight GLCM texture features were then generated with 64-bit quantization, over small windows (3 x 3), and averaged for directional texture: mean (MEA), variance (VAR), homogeneity (HOM), contrast (CON), dissimilarity (DIS), entropy (ENT), second moment (2M), and correlation (COR). A small window size was used in order to preserve high frequency spatial information, which was shown in earlier work to decrease when using larger windows as a result of over-smoothing of fine-scale textural variations (Franklin, Wulder, and Lavigne 1996). Recent research has also documented improved performance of small window sizes when applied to forest attribute estimation in *Eucalyptus* plantations (Dube and Mutanga 2015). Note that our goal was to characterize the fine-scale spatial heterogeneity of *Eucalyptus* plantation structures and not to test all possible texture methods and

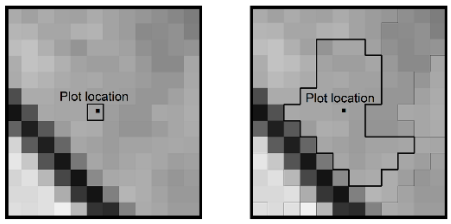
spatial operators. Our intent was to present an initial step using statistical textures to augment the spectral domain of the multi-source remotely sensed imagery with high frequency spatial information (López-Serrano et al. 2016). Texture analysis was performed using software ENVI Version 4.7 (Exelis Visual Information Solutions, Boulder, Colorado).

3.4 Image segmentation

Segmentation is the process of clustering neighbouring pixels with similar spectral and spatial characteristics to minimize the internal spectral and spatial heterogeneity of samples (Blaschke 2010). Image segmentation was performed in eCognition (Definies, 2009) using the multiresolution segmentation algorithm (Baatz and Schäpe, 2000) independently for the Landsat 8 OLI and Sentinel-2A multispectral images. For both images, the atmospherically corrected multispectral bands of the visible and near-infrared (VNIR) to shortwave infrared (SWIR) spectral range were used for segmentation. A scale factor of 15 was used for the Landsat 8 OLI segmentation, which produced a total of 2079 objects. A scale factor of 30 was used for the Sentinel-2A segmentation, creating 2692 objects. Segmentation scale factors were selected based on previous test of segmentation for predicting forest attributes. For both segmentations, to the shape factor was assigned the value of 0.3. The compactness factor was set as 0.5 and 0.6 for the Landsat 8 OLI and Sentinel-2A segmentations, respectively. The shape factor determines the proportionate weighting between colour and shape criteria in the segmentation, while the compactness factor controls the compactness of resulting objects (Baatz and Schäpe, 2000). For this criterion, values close to one allow for more compact objects. The compactness factor used was determined by visual inspection.

3.5 Estimating forest attributes

We used the pixel-based and object-based approaches to model and map volume (V), basal area (BA), and diameter at breast height (DBH) in a *Eucalyptus* plantation. Two data combination were used: (I) Landsat 8 OLI multispectral optical data combined with Sentinel-1B SAR, and ALOS/PALSAR DEM-derived data; and (II) Sentinel-2A multispectral optical data combined with Sentinel-1B SAR, and ALOS/PALSAR DEM-derived data. In the pixel-based and the object-based approaches, data values were collected from a single pixel and a single object (the mean reflectance of pixels within the object) on the field-measured plot locations, respectively, as exemplified in Figure 2.



(a) (b)

Figure 2. Exemplification of data collection in the pixel-based (a) and the object-based (b) approaches.

For both approaches, Random Forest (RF) machine learning algorithm (Breiman 2001) was used to model and predict volume (V), basal area (BA), and diameter at breast height (DBH). Random Forest (RF) regression algorithm is an ensemble method that combines multiple decision trees and obtains results by averaging the predictions from all individual regression trees. RF has proven to yield high prediction accuracy in analysis of complex variable datasets in forestry and multi-source remote sensing (Dube et al. 2015; Gao et al. 2016;

Novelli et al. 2016; Franklin and Ahmed 2017; C. Zhang et al. 2018), and was chosen in this study due to its stability and efficiency in terms of handling a large number of predictor variables. The total of 97 field-sampled plots was randomly divided into 70% (68 plots) and 30% (29 plots) for training and validation of the RF models, respectively. We used 500 decision trees in each RF implementation following preliminary tests of model performance for each forest attribute estimation, whereas the number of variables randomly sampled at each split was equal to the square root of the number of variables.

A backward stepwise predictor selection was applied for each RF model to ensure that only predictor variables that decreased the overall model uncertainties were used. Accordingly, in each run, 20% of the least important predictors were removed and the resulting Root Mean Square Error (RMSE) based on comparisons to the field inventory calculations in the validation plots was assessed. Initial tests revealed that this method produced the best overall predictive accuracy and allowed us to simplify the modelling process by identifying the minimum number of predictors to offer the best predictive accuracy (see also Ismail and Mutanga 2010). Finally, we computed the RF Variable Importance metrics, based on recursive substitution, enabling the most important variables in each model run to be interpreted based on the percentage increase of the mean square error (IncMSE) occurring when each variable was removed from the model. Next, the IncMSE was normalised by the ratio of the largest IncMSE (resulting in values between 0 and 1), and multiplied by 100 (Were et al. 2015); the higher an average accuracy decrease following variable removal, the higher the relative importance of that variable (Breiman 2001). All RF analyses were performed using the R software package *randomForest* (R Core Team, 2017).

3.6 Accuracy assessment

The accuracies of predicted *Eucalyptus* plantation attributes were evaluated using RMSE calculated based on field-based inventory estimates with the prediction values of the validation data (29 plots). In addition, we calculated the RMSE expressed as percentage (RMSE (%)), as described in Equations (1) and (2).

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

$$RMSE\ (\%) = \frac{RMSE}{\bar{X}} \times 100 \quad (2)$$

where n is the number of plots in the dataset; $\hat{X}_i$ is the estimated value of an *Eucalyptus* plantation attribute; X_i is the observed value of an *Eucalyptus* plantation attribute in the validation dataset; and $\bar{X}$ is the mean value of each *Eucalyptus* attribute in the validation dataset.

The estimated *Eucalyptus* plantation attribute maps were computed using a combination of the best modelling approach (pixel-based or object-based) and datasets.

4. Results

Table 2 summarizes the prediction results of the *Eucalyptus* plantation attributes using different sensors and prediction approaches. Sentinel-2A data appeared to have a greater explanatory power in predicting volume (RMSE = 15.23 m^3 ha^{-1}) and DBH (RMSE = 0.39 cm) compared to Landsat 8 OLI in the pixel-based modelling approach, as one would expect due to high spatial resolution of Sentinel-2A imagery. On the other hand, Landsat 8 OLI result in improved basal area estimates in the object-based approach (RMSE = 0.97 m^2 ha^{-1}).

Table 2. Performance of the RF models for predicting volume (m^3 ha^{-1}), basal area (m^2 ha^{-1}), and DBH (cm) in a *Eucalyptus* plantation based on different remotely-sensed datasets.

Modelling approach	Remotely-sensed data combination*	RMSE					
		Volume		Basal Area		DBH	
		m^3 ha^{-1}	%	m^2 ha^{-1}	%	cm	%
Pixel-based	(I)	19.74	9.33	1.16	6.04	0.40	2.67
	(II)	15.23	7.23	1.12	5.91	0.39	2.65
Object-based	(I)	19.55	9.41	0.97	5.09	0.41	2.75
	(II)	16.69	7.97	1.11	5.79	0.51	3.39

*(I)= Landsat 8 OLI multispectral optical data combined with Sentinel-1B SAR, and ALOS/PALSAR DEM data; (II) = Sentinel-2A multispectral optical data combined with Sentinel-1B SAR, and ALOS/PALSAR DEM data. RMSE= Root Mean Square Error; DBH = diameter at breast height.

Figures 3 and 4 show a comparison of measured versus predicted forest attributes, where values closer to the axis of 45° indicate higher accuracy. On average, *Eucalyptus* stand attributes were estimated with a coefficient of determination (R^2) lower than 0.70. The large range of coefficients of determination values (0.18–0.68) reveals the complexity of modelling forest attributes using remotely-sensed data. In our study, volume was the best predicted variable among the forest attributes, whereas basal area appears as the most difficult to predict using remotely-sensed data.

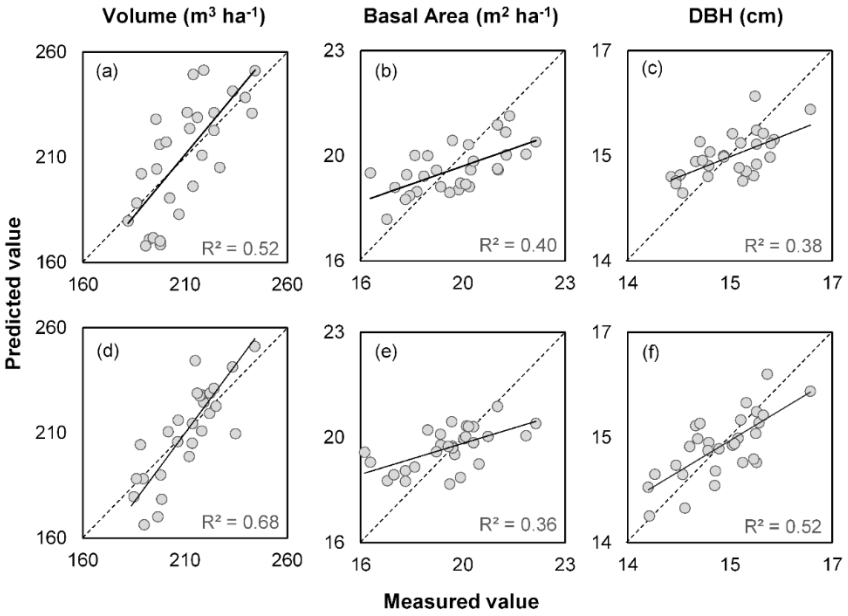


Figure 3. Scatterplots of the predicted versus measured values of forest attributes in the pixel-based approach: volume (a), basal area (b), and diameter at breast height (DBH) (c) based on Landsat 8 OLI multispectral optical data combined with Sentinel-1B SAR, and DEM-derived terrain attributes; and volume (d), basal area (e), and DBH (f) based on Sentinel-2A multispectral optical data combined with Sentinel-1B SAR, and DEM-derived terrain attributes.

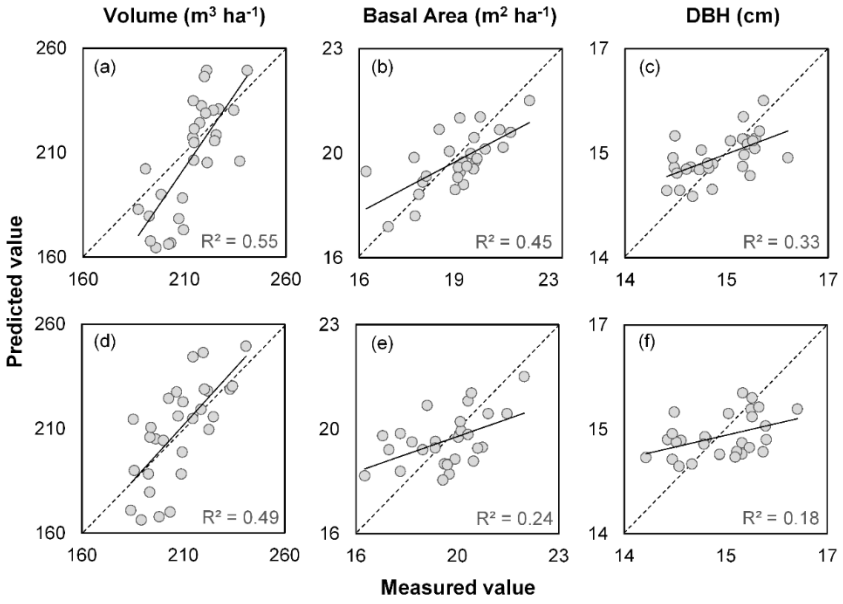


Figure 4. Scatterplots of the predicted versus measured values of forest attributes in the object-based approach: volume (a), basal area (b), and diameter at breast height (DBH) (c) based on Landsat 8 OLI multispectral optical data combined with Sentinel-1B SAR, and DEM-derived terrain attributes; and volume (d), basal area (e), and DBH (f) based on Sentinel-2A multispectral optical data combined with Sentinel-1B SAR, and DEM-derived terrain attributes.

Figures 5, 6, and 7 show the forest attribute estimation maps of the *Eucalyptus* plantation, which were obtained by both object-based (basal area) and pixel-based (volume and DBH) approaches. The mean DBH in the study area was equal to 15.00 cm with values ranging from 13.86 cm to 16.52 cm. For basal area, the estimated values ranged from 15.97 m^2 ha^{-1} to 22.34 m^2 ha^{-1}, with a mean value of 19.15 m^2 ha^{-1}. The estimated volume varied between 160.00 m^3 ha^{-1} and 298.26 m^3 ha^{-1}, with a mean value of 208.97 m^3 ha^{-1}. The total volume estimation across the 1,498 ha of *Eucalyptus* plantation was

314,389.08 m^3, resulting in a range of mean annual increment in volume from 24.84 m^3 ha^{-1} $year^{-1}$ to 46.31 m^3 ha^{-1} $year^{-1}$, depending on the plot location.

Figure 8 shows the relative importance of each selected variable for predicting *Eucalyptus* stand attributes in our study. The spectral reflectance in the shortwave infrared (SWIR1) region, from both Landsat 8 OLI and Sentinel-2A images, was the most influential variable for predicting volume, basal area, and DBH. The spectral reflectance in the green (G) and red (R and Red edge 1) regions were also important variables for predicting the forest attributes under study. Terrain attributes (especially HLI, relative slope position, total curvature, aspect, and TRI) contributed significantly for the prediction of *Eucalyptus* stand attributes: these terrain attributes corresponded to 54.5%, 41.6%, and 53.8% of the selected variables by RF models to predict volume, basal area, and DBH, respectively. Texture measures were more important to predict basal area (25.0% of the selected variables by RF model) than volume (9.1% of the selected variables by RF model) and DBH (7.7% of the selected variables by RF model). SAR data contributed only for the prediction of volume and DBH in our study.

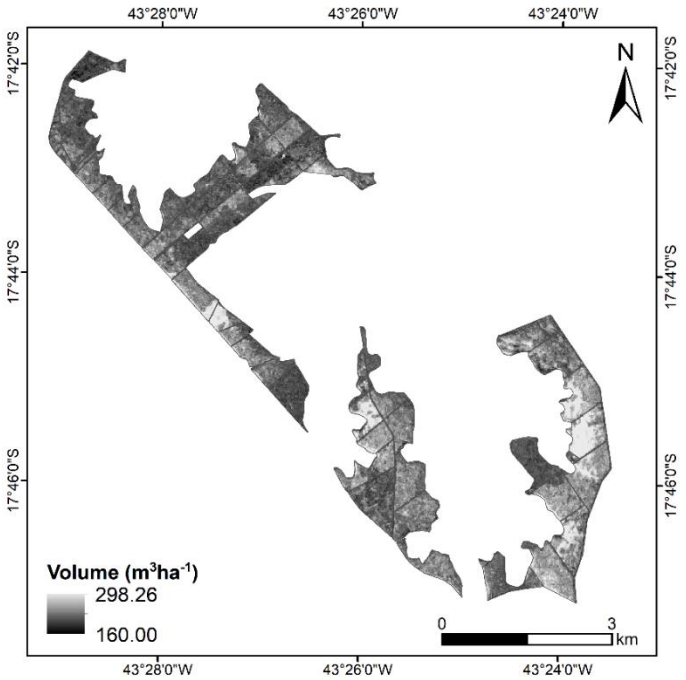


Figure 5. Spatial distribution of volume estimates in the 1,498 ha *Eucalyptus* plantation obtained with the best-performing RF model in the pixel-based approach based on Sentinel-2A multispectral optical data combined with Sentinel-1B SAR and DEM-derived terrain attributes.

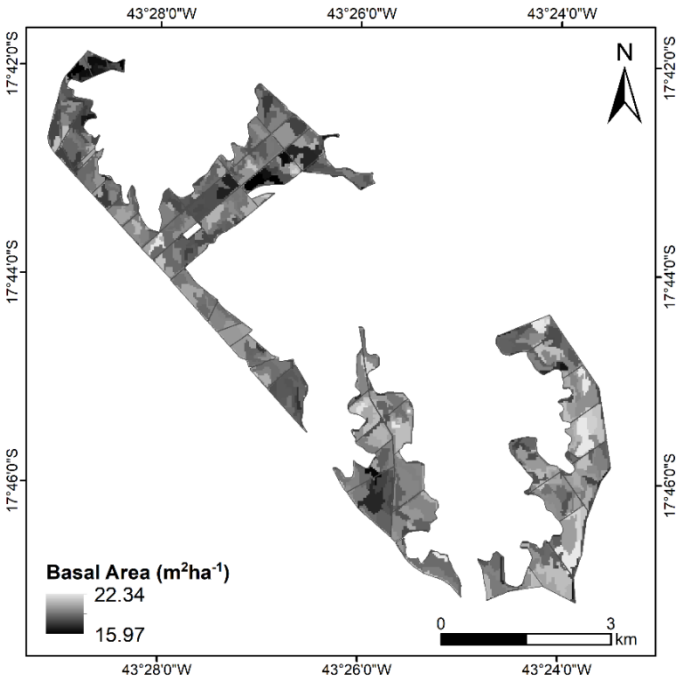


Figure 6. Spatial distribution of basal area estimates in the 1,498 ha *Eucalyptus* plantation obtained with the best-performing RF model in the object-based approach based on Landsat 8 OLI multispectral optical data combined with Sentinel-1B SAR and DEM-derived terrain attributes.

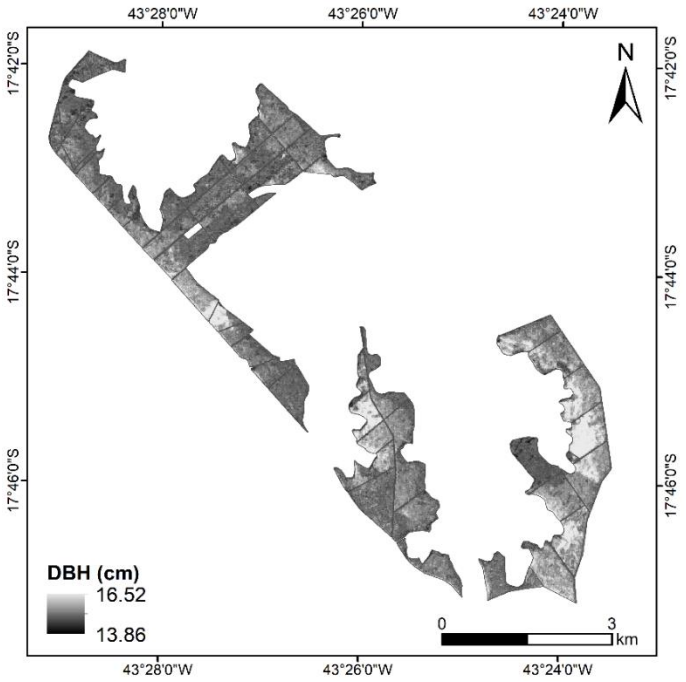


Figure 7. Spatial distribution of diameter at breast height (DBH) estimates in the 1,498 ha *Eucalyptus* plantation obtained with the best-performing RF model in the pixel-based approach based on Sentinel-2A multispectral optical data combined with Sentinel-1B SAR and DEM-derived terrain attributes.

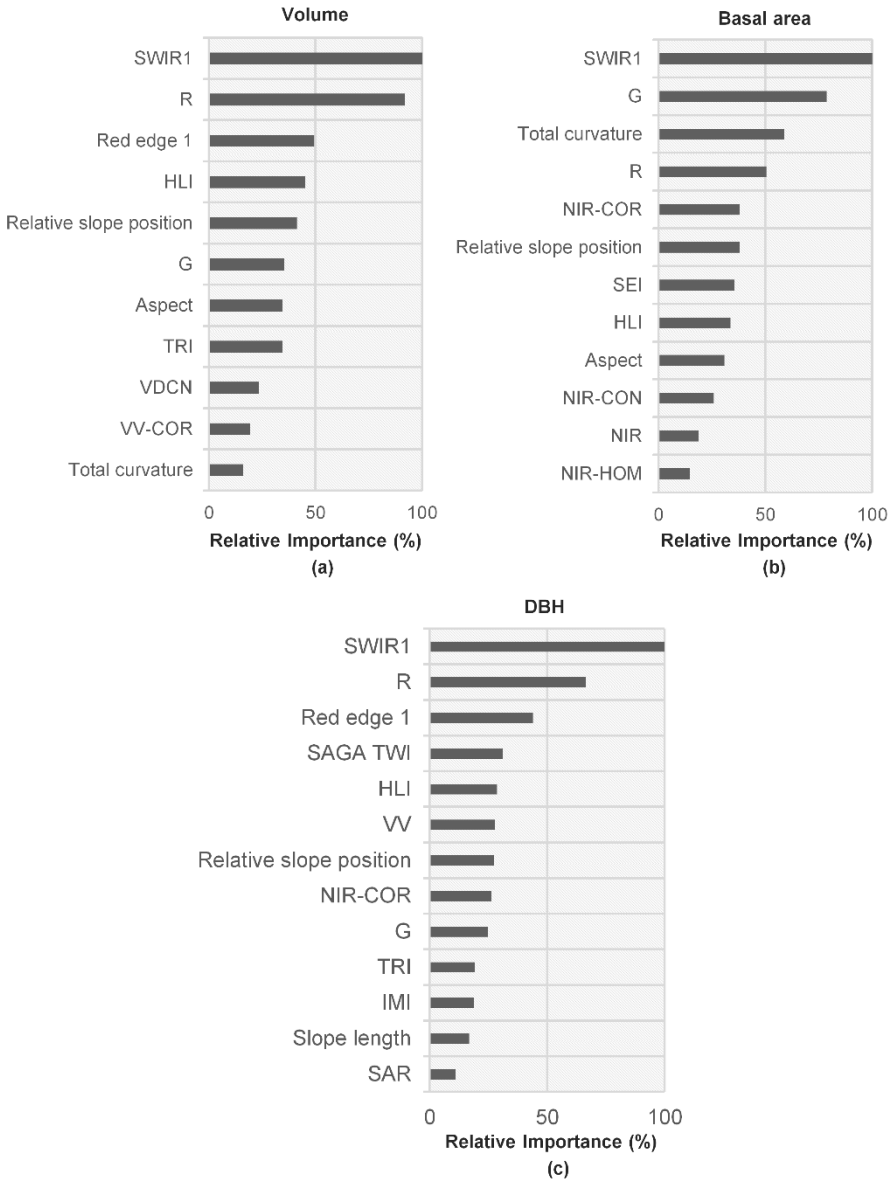


Figure 8. Relative importance of the variables as measured by the Variable Importance metric in the Random Forest algorithm predicting volume (a), basal area (b) and diameter at breast height (DBH) (c).

5. Discussion

The use of remotely-sensed data for forest attributes estimation is a major challenge due to numerous factors influencing the relationship between forest attributes and remote sensing variables, as for instance, topography, soil conditions, and forest structures (Lu et al. 2016; Reuveni et al. 2018; Galidaki et al. 2017). Therefore, investigating new prediction approaches and new remotely-sensed data, including their combinations with ancillary data, for forest attributes estimation is still required. Our primary objective was to test whether the Sentinel-2A sensor with its new red edge bands can help to improve the accuracy of forest attributes estimation compared to the use of Landsat 8 OLI data, when multispectral optical imagery are combined with SAR data and DEM-derived terrain attributes.

We found that using Sentinel-1 SAR and Sentinel-2A multispectral imagery data can give satisfactory results in predicting volume of *Eucalyptus* stands independently of the prediction approach used, especially with the inclusion of DEM-derived terrain attributes as ancillary data. These results are in agreement with the findings of previous studies that reported slightly better performance using the Sentinel-2A optical imagery compared with the Landsat 8 OLI in vegetation analysis (Chrysafis et al. 2017; Korhonen et al. 2017). The better performance of the RF model that included Sentinel-2A data may be related to the Red edge 1 band, which was frequently selected as a predictor variable. Similar performance was also observed, in a lesser degree, in predicting basal area and DBH of *Eucalyptus* stands when the pixel-based approach was used. However, the accuracy estimates of DBH were close to each other in terms of RMSE (Table 2). In addition, for basal area prediction, the best result was obtained when Landsat 8 OLI data was combined with Sentinel-1B SAR and DEM-derived terrain attributes in the object-based approach.

This study also demonstrates that the combination of freely-accessible multi-source datasets (Sentinel-1B, Sentinel-2A, ALOS/PALSAR DEM, and Landsat 8 OLI) could significantly improve the overall accuracy of forest attributes estimation based on remotely-sensed data. We obtained more accurate estimates than previous studies using other remotely-sensed data combination for predicting forest attributes in *Eucalyptus* plantations. Ismail et al. (2015) combined ALOS/PALSAR backscatter with SPOT 4 multispectral optical data and stand age as ancillary data to predict volume in *Eucalyptus* plantations in Zululand, South Africa. The best regression model of volume fitted by the authors produced a RMSE of 31.71 m^3 ha^{-1}, which was considered appropriate for operational forest management in this region.

In another study in South Africa, Dube et al. (2017) demonstrated that integrating multispectral SPOT 5 image, stand age, and rainfall metrics significantly improved volume estimation in *Eucalyptus* plantations (RMSE = 36.02 m^3 ha^{-1}). The results found by these authors correspond almost twice the error estimates found in our study for volume estimation (RMSE = 15.23 m^3 ha^{-1}), thus highlighting the potential of the combination of Sentinel-2A spectral data with Sentinel-1B SAR and DEM-derived terrain attributes for predicting *Eucalyptus* stand attributes. Although most of the studies have focused in predicting aboveground biomass and volume of *Eucalyptus* plantations (Dube et al. 2014, 2015, 2017; Ismail et al. 2015; Dube and Mutanga 2015), few studies investigated the potential of remotely-sensed data to predict DBH and basal area of these plantations.

Using optical IKONOS satellite data (multispectral and panchromatic) as independent variables, Gebreslasie, Ahmed, and van Aardt (2011) estimated DBH and basal area of *Eucalyptus* stands using Artificial Neural Networks (ANN) models with RMSE = 5.07% and RMSE = 8.71%, respectively. In another study, the same authors predicted basal area of *Eucalyptus* plantations

using ASTER satellite data, age and site index as independent variables with a RMSE of 7.39 m^2 (Gebreslasie, Ahmed, and van Aardt 2010).

The object-based approach did not improve the estimation accuracy of all tested forest attributes. However, this approach was especially important for predicting basal area, whereas the pixel-based approach result in more accurate estimates of volume and DBH. Chen et al. (2011) compared the pixel-based and object-based approaches to estimate forest canopy heights using Quickbird imagery data. These authors found that most of the object-based models results in a higher accuracy for the estimated canopy height than the pixel-based model; however, the pixel-based approach still showed comparable or better performance than the object-based approach at some segmentation scales.

C. Zhang et al. (2018) assessed the benefits of object-based approach for sawgrass biomass modelling and mapping in the Everglades, Florida, compared with the pixel-based approach. The results found by these authors indicate that object-based modelling approach is a promising alternative to the traditional pixel-based approach, since the former approach provides the capability to match in-situ data to an image object to reduce the positional discrepancy between image and field data. The positional discrepancy of sample plots is one of the major uncertainties in forest attributes estimation using remotely sensed data (Lu et al. 2016), and there is a higher probability that a field plot be located in an object than in a single pixel.

On the other hand, selecting an appropriate object scale for matching in-situ measures is a difficult task (C. Zhang et al. 2018). In theory, a field-measured plot should match a “pure” image object which can represent well the vegetation structure of this plot. This indicates that selecting an appropriate object scale is crucial for applying an object-based approach, since the selection of arbitrary scales may result in poor accuracy of forest attribute estimates (Chen et al. 2011). In this study, a trial-and-error procedure was applied to select the

scale parameter, *i.e.*, selecting a scale parameter, running the RF models, and then assessing the forest attributes estimates until the best results were achieved.

The current study serves as a useful illustration of the potentials of using freely-available multi-source remotely-sensed data (*i.e.*, Sentinel-2A, Sentinel-1B SAR, Landsat 8 OLI, and ALOS/PALSAR DEM) for estimation of *Eucalyptus* stand attributes. It also provided some new and encouraging results of the potential of the proposed remotely-sensed data combination to support monitoring and management of planted forests without the need for commercial satellite imagery. Regarding the employed object-based approach, general conclusions can hardly be drawn on the basis of the results provided by this study; there are many specific limitations that preclude general inference of the benefits of using object-based approach for forest attribute estimation compared with the traditional pixel-based approach.

6. Conclusions

The main merit of this study is the illustration of a new remotely-sensed data combination for forest attribute prediction in even-aged single-species monocultures as *Eucalyptus* plantation using the freely-available and recently launched Sentinel-2A and Sentinel-1B SAR data integrated with DEM-derived terrain attributes. Sentinel-2A multispectral optical imagery have similar to superior capabilities of Landsat 8 OLI in estimating *Eucalyptus* stand attributes when combined with Sentinel-1B SAR data and DEM-derived terrain attributes. Sentinel-2A performed better in predicting volume and DBH in the pixel-based approach, while Landsat 8 OLI outperformed in predicting basal area in the object-based approach. This study provides new insights into the opportunities and limitations related to the use of object-based approach for forest attribute estimation in planted forests; however, it is important to acknowledge the intrinsic limitations that object-based approach have when applied for forest

attribute modelling. Further research will provide additional insight into appropriate object scale to achieve more accurate estimates of forest attributes. Also further research should investigate whether multi-seasonal imagery may have greater benefit in predicting forest attributes of *Eucalyptus* plantations.

References

Adams, H.R., H.R. Barnard, and A.K. Loomis. 2014. "Topography Alters Tree Growth–climate Relationships in a Semi-Arid Forested Catchment." *Ecosphere* 5 (11): art148. https://doi.org/10.1890/ES14-00296.1.

Alvares, C.A., J.L. Stape, P.C. Sentelhas, J.L.M. Gonçalves, and G. Sparovek. 2013. "Köppen's Climate Classification Map for Brazil." Meteorologische Zeitschrift 22 (6):711–28. https://doi.org/10.1127/0941-2948/2013/0507.

Baatz, M., and A. Schäpe. 2000. Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation. In *Angewandte Geographische Informations-Verarbeitung, XII*; Herbert Wichmann Verlag: Berlin, Germany. (58): 12–23.

Baghdadi, N., G. Le Maire, J.S. Bailly, Kenji Ose, Yann Nouvellon, Mehrez Zribi, Cristiane Lemos, and Rodrigo Hakamada. 2015 Baghdadi, N., G. le Maire, J. S. Bailly, Kenji Ose, Y. Nouvellon, M. Zribi, C. Lemos, and R. Hakamada. 2015. "Evaluation of ALOS/PALSAR L-band data for the estimation of *Eucalyptu*s plantations aboveground biomass in Brazil." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8 (8):3802–11. https://doi.org/10.1109/JSTARS.2014.2353661.

Blaschke, T. 2010. "Object Based Image Analysis for Remote Sensing." *ISPRS Journal of Photogrammetry and Remote Sensing* 65 (1): 2–16. https://doi.org/10.1016/j.isprsjprs.2009.06.004.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45 (1):5–32.

https://doi.org/10.1023/A:1010933404324.

Chen, G., G.J. Hay, G. Castilla, B. St-Onge, and R. Powers. 2011. "A Multiscale Geographic Object-Based Image Analysis to Estimate Lidar-Measured Forest Canopy Height Using Quickbird Imagery." *International Journal of Geographical Information Science* 25 (6):877–93. https://doi.org/10.1080/13658816.2010.496729.

Chen, G., J.C. Thill, S. Anantsuksomsri, N. Tontisirin, and R. Tao. 2018. "Stand age estimation of rubber (*Hevea brasiliensis*) plantations using an integrated pixel- and object-based tree growth model and annual landsat time series." *ISPRS Journal of Photogrammetry and Remote Sensing* 144: 94–104. https://doi.org/10.1016/j.isprsjprs.2018.07.003.

Chrysafis, I., G. Mallinis, S. Siachalou, and P. Patias. 2017. "Assessing the relationships between growing stock volume and Sentinel-2 imagery in a Mediterranean forest ecosystem." *Remote Sensing Letters* 8 (6):508–17. https://doi.org/10.1080/2150704X.2017.1295479.

Definies A.G. 2009. *Definiens eCognition Developer 8 User Guide*, Definiens AG: Munich, Germany.

Dube, T., and O. Mutanga. 2015. "Investigating the robustness of the new Landsat-8 Operational Land Imager derived texture metrics in estimating plantation forest aboveground biomass in resource constrained areas." *ISPRS Journal of Photogrammetry and Remote Sensing* 108:12–32. https://doi.org/10.1016/j.isprsjprs.2015.06.002.

Dube, T., O. Mutanga, E.M. Abdel-Rahman, R. Ismail, and R. Slotow. 2015. "Predicting *Eucalyptus* spp. stand volume in Zululand, South Africa: An analysis using a stochastic gradient boosting regression ensemble with multi-source data sets." *International Journal of Remote Sensing*. 36 (14): 3751‑3772. https://doi.org/10.1080/01431161.2015.1070316.

Dube, T., O. Mutanga, A. Elhadi, and R. Ismail. 2014. "Intra-and-Inter Species

Biomass Prediction in a Plantation Forest: Testing the Utility of High Spatial Resolution Spaceborne Multispectral Rapideye Sensor and Advanced Machine Learning Algorithms." *Sensors:* 14 (8):15348–70. https://doi.org/10.3390/s140815348.

Dube, T., M. Sibanda, C. Shoko, and O. Mutanga. 2017. "Stand-Volume Estimation from Multi-Source Data for Coppiced and High Forest Eucalyptus Spp. Silvicultural Systems in KwaZulu-Natal, South Africa." *ISPRS Journal of Photogrammetry and Remote Sensing* 132:162–69. https://doi.org/10.1016/j.isprsjprs.2017.09.001.

Evans, J.S., J. Oakleaf, S.A. Cushman, and D. Theobald, 2014. *An ArcGIS Toolbox for Surface Gradient and Geomorphometric Modeling*, Version 2.0-0. URL: http://evansmurphy.wix.com/evansspatial (last date accessed: 24 July 2018).

Franklin, S.E., M.A. Wulder, and M.B. Lavigne. 1996. "Automated Derivation of Geographic Window Sizes for Use in Remote Sensing Digital Image Texture Analysis." *Computers & Geosciences* 22 (6):665–73. https://doi.org/10.1016/0098-3004(96)00009-X.

Franklin, S.E., and O.S. Ahmed. 2017. "Object-Based Wetland Characterization Using Radarsat-2 Quad-Polarimetric SAR Data, Landsat-8 OLI Imagery, and Airborne Lidar-Derived Geomorphometric Variables." *Photogrammetric Engineering & Remote Sensing* 83 (1):27–36. https://doi.org/10.14358/PERS.83.1.27.

Galidaki, G., D. Zianis, I. Gitas, K. Radoglou, V. Karathanassi, M. Tsakiri–Strati, I. Woodhouse, and G. Mallinis. 2017. "Vegetation Biomass Estimation with Remote Sensing: Focus on Forest and Other Wooded Land over the Mediterranean Ecosystem." *International Journal of Remote Sensing* 38 (7):1940–66. https://doi.org/10.1080/01431161.2016.1266113.

Gao, T., J. Zhu, S. Deng, X. Zheng, J. Zhang, G. Shang, and L. Huang. 2016. "Timber Production Assessment of a Plantation Forest: An Integrated Framework with Field-Based Inventory, Multi-Source Remote Sensing Data and Forest Management History." *International Journal of Applied Earth Observation and Geoinformation* 52: 155–165. https://doi.org/10.1016/j.jag.2016.06.004.

Gebreslasie, M. T., F. B. Ahmed, and J. A N van Aardt. 2010. "Predicting Forest Structural Attributes Using Ancillary Data and ASTER Satellite Data." *International Journal of Applied Earth Observation and Geoinformation* 12S: S23–S26. https://doi.org/10.1016/j.jag.2009.11.006.

Gebreslasie, M. T., F. B. Ahmed, and Jan A N van Aardt. 2011. "Extracting Structural Attributes from IKONOS Imagery for Eucalyptus Plantation Forests in Kwazulu-Natal, South Africa, Using Image Texture Analysis and Artificial Neural Networks." *International Journal of Remote Sensing* 32 (22): 7677-7701. https://doi.org/10.1080/01431161.2010.527392.

Goodbody, T.R.H., N.C. Coops, P. Tompalski, P. Crawford, and K.J.K. Day. 2017. "Updating Residual Stem Volume Estimates Using ALS- and UAV-Acquired Stereo-Photogrammetric Point Clouds*." International Journal of Remote Sensing* 38(8-10): 2938-2953. https://doi.org/10.1080/01431161.2016.1219425.

Haralick, R.M., K. Shanmugam, and I.H. Dinstein. 1973. "Textural Features for Image Classification." *IEEE Transactions on Systems, Man, and Cybernetics* SMC-3 (6):610–21. https://doi.org/10.1109/TSMC.1973.4309314.

Hawryło, P., and P. Wężyk. 2018. "Predicting Growing Stock Volume of Scots Pine Stands Using Sentinel-2 Satellite Imagery and Airborne Image-Derived Point Clouds." *Forests* 9 (5):274. https://doi.org/10.3390/f9050274.

Hengl, T., Reuter, H.I. (eds) 2009. *Geomorphometry: Concepts, Software, Applications.* Developments in Soil Science: Elsevier.

IBÁ - Brazilian Tree Industry, 2016. *Brazilian Tree Industry - Annual Report*. Brazilian Tree Industry (IBA), São Paulo.

Ismail, R., and O. Mutanga. 2010. "A Comparison of Regression Tree Ensembles: Predicting Sirex Noctilio Induced Water Stress in Pinus Patula Forests of KwaZulu-Natal, South Africa." *International Journal of Applied Earth Observation and Geoinformation* 12:S45–51. https://doi.org/10.1016/j.jag.2009.09.004.

Ismail, R., H. Kassier, M. Chauke, F. Holecz, and N. Hattingh. 2015. "Assessing the Utility of ALOS PALSAR and SPOT 4 to Predict Timber Volumes in Even-Aged Eucalyptus Plantations Located in Zululand, South Africa." *Southern Forests: A Journal of Forest Science* 77 (3):203–11. https://doi.org/10.2989/20702620.2014.1001681.

Kajisa, T., T. Murakami, N. Mizoue, N. Top, and S. Yoshida. 2009. "Object-Based Forest Biomass Estimation Using Landsat ETM+ in Kampong Thom Province, Cambodia." *Journal of Forest Research* 14 (4):203–11. https://doi.org/10.1007/s10310-009-0125-9.

Korhonen, L., Hadi, P. Packalen, and M. Rautiainen. 2017. "Comparison of Sentinel-2 and Landsat 8 in the Estimation of Boreal Forest Canopy Cover and Leaf Area Index." *Remote Sensing of Environment* 195: 259–274. https://doi.org/10.1016/j.rse.2017.03.021.

Li, J., and D.P. Roy. 2017. "A Global Analysis of Sentinel-2A, Sentinel-2B and Landsat-8 Data Revisit Intervals and Implications for Terrestrial Monitoring." *Remote Sensing* 9 (9):902. https://doi.org/10.3390/rs9090902.

López-Sánchez, C.A., P. García-Ramírez, R. Resl, J.C. Hernández-Díaz, P.M. López-Serrano, and C. Wehenkel. 2017. "Modelling Dasometric Attributes of Mixed and Uneven-Aged Forests Using Landsat-8 OLI Spectral Data in

the Sierra Madre Occidental, Mexico." *iForest* 10: 288-295. https://doi.org/10.3832/ifor1891-009.

López-Serrano, P.M., C.A. López-Sánchez, J.G. Álvarez-González, and J. García-Gutiérrez. 2016. "A Comparison of Machine Learning Techniques Applied to Landsat-5 TM Spectral Data for Biomass Estimation." *Canadian Journal of Remote Sensing* 42 (6):690–705. https://doi.org/10.1080/07038992.2016.1217485.

Lu, D., Q. Chen, G. Wang, L. Liu, G. Li, and E. Moran. 2016. "A Survey of Remote Sensing-Based Aboveground Biomass Estimation Methods in Forest Ecosystems." *International Journal of Digital Earth*. 9(1): 63-105. https://doi.org/10.1080/17538947.2014.990526.

Mohamedou, C., T. Tokola, and K. Eerikäinen. 2017. "LiDAR-Based TWI and Terrain Attributes in Improving Parametric Predictor for Tree Growth in Southeast Finland." *International Journal of Applied Earth Observation and Geoinformation* 62:183–91. https://doi.org/10.1016/j.jag.2017.06.004.

Novelli, A., M.A. Aguilar, A. Nemmaoui, F.J. Aguilar, and E. Tarantino. 2016. "Performance Evaluation of Object Based Greenhouse Detection from Sentinel-2 MSI and Landsat 8 OLI Data: A Case Study from Almería (Spain)." *International Journal of Applied Earth Observation and Geoinformation* 52: 403–411. https://doi.org/10.1016/j.jag.2016.07.011.

Ota, T., M. Ogawa, N. Mizoue, K. Fukumoto, and S. Yoshida. 2017. "Forest Structure Estimation from a UAV-Based Photogrammetric Point Cloud in Managed Temperate Coniferous Forests." *Forests* 8 (9):343. https://doi.org/10.3390/f8090343.

Pham, T.D., K. Yoshino, N.N. Le, and D.T. Bui. 2018. "Estimating Aboveground Biomass of a Mangrove Plantation on the Northern Coast of Vietnam Using Machine Learning Techniques with an Integration of ALOS-2 PALSAR-2 and Sentinel-2A Data." *International Journal of*

Remote Sensing 1–28. https://doi.org/10.1080/01431161.2018.1471544.

Puliti, S., S. Saarela, T. Gobakken, G. Ståhl, and E. Næsset. 2018. "Combining UAV and Sentinel-2 Auxiliary Data for Forest Growing Stock Volume Estimation through Hierarchical Model-Based Inference." *Remote Sensing of Environment* 204: 485–497. https://doi.org/10.1016/j.rse.2017.10.007.

R Core Team. (2017). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Raimundo, M.R., H.F. Scolforo, J.M. de Mello, J.R.S. Scolforo, J.P. McTague, and A.A. dos Reis. 2017. "Geostatistics Applied to Growth Estimates in Continuous Forest Inventories." *Forest Science* 63 (1):29–38. https://doi.org/10.5849/FS-2016-056.

Reuveni, Y., E. Dahan, Y. Anker, and M. Sprintsin. 2018. "Estimating Forest Parameters Using Landsat ETM+ Spectral Responses and Monocultured Plantation Fieldwork Measurements Data." *International Journal of Remote Sensing* 39 (8): 2620–36. https://doi.org/10.1080/01431161.2018.1430400.

Roy, D.P., M.A. Wulder, T.R. Loveland, C.E. Woodcock, R.G. Allen, M.C. Anderson, D. Helder, et al. 2014. "Landsat-8: Science and Product Vision for Terrestrial Global Change Research." *Remote Sensing of Environment* 145: 154–72. https://doi.org/10.1016/j.rse.2014.02.001.

Scolforo, H.F., J.R.S. Scolforo, J.L. Stape, J.P. McTague, H. Burkhart, J. McCarter, F. Castro Neto, R.A. Loos, and R.C. Sartorio. 2017. "Incorporating Rainfall Data to Better Plan Eucalyptus Clones Deployment in Eastern Brazil." *Forest Ecology and Management* 391: 145-153. https://doi.org/10.1016/j.foreco.2017.02.025.

Shao, Z., and L. Zhang. 2016. "Estimating Forest Aboveground Biomass by Combining Optical and SAR Data: A Case Study in Genhe, Inner Mongolia, China." *Sensors* 16: 834. doi:10.3390/s16060834.

Tompalski, P., N. Coops, P. Marshall, J. White, M. Wulder, and T. Bailey. 2018. "Combining multi-date airborne laser scanning and digital aerial photogrammetric data for forest growth and yield modelling." *Remote Sensing* 10 (3):347. https://doi:10.3390/rs10020347.

Torres, R., P. Snoeij, D. Geudtner, D. Bibby, M. Davidson, E. Attema, P. Potin, et al. 2012. "GMES Sentinel-1 Mission." *Remote Sensing of Environment* 120: 9–24. doi:10.1016/j.rse.2011.05.028.

Vafaei, S., J. Soosani, K. Adeli, H. Fadaei, H. Naghavi, T.D. Pham, and D.T. Bui. 2018. "Improving Accuracy Estimation of Forest Aboveground Biomass Based on Incorporation of ALOS-2 PALSAR-2 and Sentinel-2A Imagery and Machine Learning: A Case Study of the Hyrcanian Forest Area (Iran)." *Remote Sensing* 10 (2). https://doi.org/10.3390/rs10020172.

Vaglio, G.L., F. Pirotti, M. Callegari, Q. Chen, G. Cuozzo, E. Lingua, C. Notarnicola, and D. Papale. 2017. "Potential of ALOS2 and NDVI to Estimate Forest Above-Ground Biomass, and Comparison with Lidar-Derived Estimates." *Remote Sensing* 9 (18): 1 – 16. https://doi.org/10.3390/rs9010018.

Were, K., D.T. Bui, Ø.B. Dick, and B.R. Singh. 2015. "A Comparative Assessment of Support Vector Regression, Artificial Neural Networks, and Random Forests for Predicting and Mapping Soil Organic Carbon Stocks across an Afromontane Landscape." *Ecological Indicators* 52: 394–403. https://doi.org/10.1016/j.ecolind.2014.12.028.

Zhang, C., S. Denka, H. Cooper, and D.R. Mishra. 2018. "Quantification of Sawgrass Marsh Aboveground Biomass in the Coastal Everglades Using Object-Based Ensemble Analysis and Landsat Data." *Remote Sensing of Environment* 204: 366–79. https://doi.org/10.1016/j.rse.2017.10.018.

Zhang, J., C. Lu, H. Xu, and G. Wang. 2018. "Estimating Aboveground Biomass of Pinus Densata-Dominated Forests Using Landsat Time Series

and Permanent Sample Plot Data." *Journal of Forestry Research*. https://doi.org/10.1007/s11676-018-0713-7.

Zhao, P., D. Lu, G. Wang, L. Liu, D. Li, J. Zhu, and S. Yu. 2016. "Forest Aboveground Biomass Estimation in Zhejiang Province Using the Integration of Landsat TM and ALOS PALSAR Data." *International Journal of Applied Earth Observation and Geoinformation* 53: 1–15. https://doi.org/10.1016/j.jag.2016.08.007.

ARTICLE 4 - REGIONAL-SCALE FOREST SITE PRODUCTIVITY PREDICTION FROM CLIMATE AND TERRAIN DATA FOR *EUCALYPTUS* PLANTATIONS IN BRAZIL

Aliny Aparecida Dos Reis[a,*], Steven E. Franklin[b], Fausto Weimar Acerbi Junior[a], Antonio Carlos Ferraz Filho[c], José Marcio de Mello[a]

[a] *Department of Forest Science, Federal University of Lavras – UFLA, PO Box 3037, 37200-000, Lavras, Minas Gerais, Brazil*

[b] *School of Environment, Trent University, K9I 7B8, Peterborough, Ontario, Canada*

[c] *CPCE, Federal University of Piauí, Manoel Gracindo Avenue, km 1, 64900-000, Bom Jesus, Piauí, Brazil*

*Corresponding author email: alinyreis@hotmail.com

Publication status: In preparation for submission to **Forestry**

Abstract: Forest site productivity classification is a fundamental forest management tool for strategic planning of either new or established stands, definition of silvicultural treatments, and quantification of productive potential of forests. Typically, site productivity classification is achieved through field observations collected at sample plots. In some cases, appropriate field measures are possible (for example, in new or young stands, over large geographic regions, and in areas with accessibility constraints or a high degree of spatial and temporal variability). In this research, we describe an indirect and coherent spatial approach to predict forest productivity at regional scale based on terrain attributes derived from an ALOS PALSAR (Advanced Land Observing Satellite - Phase Array type L-band Synthetic Aperture Radar) digital elevation model (DEM) and bioclimatic variables for *Eucalyptus* plantations across an area of heterogeneous terrain and climate conditions, located in Minas Gerais state, Brazil. Terrain attributes derived from a DEM and bioclimatic variables correctly classified site productivity, with an overall accuracy of 91.6%, and predicted maximum mean annual increment (MAI_{max}) with a Root Mean Square Error (RMSE) of 6.1 $m^3\ ha^{-1}\ yr^{-1}$ (16.2 %). Temperature seasonality (BIO04), precipitation of the driest month (BIO14), elevation, and annual mean temperature (BIO01) emerged as major determinants of potential productivity for *Eucalyptus* plantations in our study area, enabling us to discriminate three site productivity classes, and predict the potential MAI_{max} for an area of approximately 100,690 km^2 in the northeast, north and central regions of Minas Gerais state. The approach developed here can be used to estimate site productivity and potential MAI_{max} at any geographical location, given the relevant terrain attributes and bioclimatic variables, thereby supporting forest management decisions. This approach could be also linked to climate models to predict how productivity of *Eucalyptus* plantations would change under climate change.

Keywords: bioclimatic variables; terrain attributes; Random Forest; forest management.

1. Introduction

The Brazilian forestry sector has intensely increased the productivity of *Eucalyptus* plantations through intensive silviculture, genetic selection, and breeding techniques (Gonçalves et al. 2013; Campoe et al. 2013). The first commercial *Eucalyptus* plantations in Brazil were established in the early 1900s, and due to the recent large-scale afforestation and reforestation efforts, such plantations have now expanded rapidly to cover more than 5.7 million hectares in the country (Ibá, 2017). Most of these plantations are located in the southeast and central-west regions of Brazil, and are managed in short rotations (6–8 years) with the mean annual increment varying between 25 to 60 m^3 ha^{-1} $year^{-1}$ of roundwood, depending on the environmental conditions (Gonçalves et al. 2013). The major determinants of productivity in *Eucalyptus* plantations include tree genetic material and local site environment (Resende et al. 2016; Marcatti et al. 2017). Thus, the high yields achieved by genetically identical clonal *Eucalyptus* stands may be limited by the spatial variability of climate and terrain conditions (e.g. temperature, precipitation, solar radiation, and slope). This can lead to heterogeneity in tree growth and reductions in productivity at the stand and plantation levels (Stape et al. 2010; Gonçalves et al. 2013; Campoe et al. 2016).

Forest site productivity is defined as the potential of a site to produce wood biomass (Skovsgaard and Vanclay 2008; Skovsgaard and Vanclay 2013) and is usually determined based on characteristics of forest vegetation (e.g. height of dominant trees at a given reference age) or the site (e.g. topography and soils), or both combined (Véga and St-Onge 2009; Parresol et al. 2017; Huang et al. 2017; Sharma and Parton 2018; Díaz-Varela et al. 2011). Forest site

productivity classification is a widely used forest management tool for strategic planning of new or established stands, definition of appropriate silvicultural treatments (such as when to start thinning and/or when to clearfell the forest), and to quantify the productivity of stands and plantations (Burkhart and Tomé 2012). The most common and applied approach of evaluating site productivity is the site index (SI) (Skovsgaard and Vanclay 2008; Skovsgaard and Vanclay 2013), which is based on the strong correlation between volume and height growth (Sharma, Brunner, and Eid 2012), and is defined as the height of dominant trees of a stand at a predetermined reference age (Assmann 1970; Burkhart and Tomé 2012). This approach relies on the establishment of detailed stand or plot samples, which are not always available for measurement (Bravo-Oviedo et al. 2008). Also, such field observations are relatively costly, time-consuming and often spatially-limited based on sample plot distribution and intensity (Bueis et al. 2016). An alternative approach is to estimate potential forest site productivity from site variables describing location, topography, or climate (Laamrani et al. 2014; Sharma and Parton 2018; Parresol et al. 2017; Huang et al. 2017). Increasingly, these variables can be readily obtained over large geographic regions from existing maps, aerial and satellite remote sensing imagery, and digital geographic information system (GIS) databases (Sharma, Brunner, and Eid 2012).

Previous studies have estimated the productive potential of forests based on topography and climate variables (e.g. Chen et al., 2002; Díaz-Varela et al., 2011; Laamrani et al., 2014). Most of these studies have used basic terrain information, such as elevation, slope, aspect, and curvature (Latta, Temesgen, and Barrett 2009; Díaz-Varela et al. 2011; Sharma, Brunner, and Eid 2012; Laamrani et al. 2014; Huang et al. 2017). Recently, advances in remote sensing technologies have allowed for increasingly detailed land surface representation with dense-grid Digital Elevation Models (DEMs), from which more complex

terrain derivatives can be created to classify forest sites and estimate the productive potential of forests. The basis for using these complex terrain attributes is the expected relationship between site conditions and growth potential (Laamrani et al. 2014; Bueis et al. 2016; Parresol et al. 2017); for example, it has been shown that terrain attributes, in certain forest ecosystems, influence local soil conditions and microclimates through the pattern of insolation, precipitation, temperature, and relative humidity (Bravo et al. 2011; Bueis et al. 2016; Berrill and O'Hara 2016; Adams, Barnard, and Loomis 2014).

In this study, we used a nonparametric modeling approach to examine relationships between terrain attributes and climate data on forest site productivity and maximum mean annual increment (MAI_{max}) at regional scale. Our objective was to determine whether terrain attributes derived from a DEM and/or bioclimatic variables could be used as predictors of productive potential of *Eucalyptus* plantations across an area of heterogeneous terrain and climate conditions in Minas Gerais state, Brazil. Although many studies have reported the efficacy of soil attributes in improving forest productivity prediction based on site factors (Chen, Krestov, and Klinka 2002; Bueis et al. 2016; Bravo et al. 2011; Mohamed et al. 2014; Parresol et al. 2017; Sharma, Brunner, and Eid 2012), soil data were not included in our approach because of the lack of accurate data for the study area. Our findings support management decision-making by identifying the terrain attributes and climate conditions most strongly correlated with *Eucalyptus* plantation productivity. In addition, improved forestry management practices (e.g., sampling, experimental design, growth modelling, plantation silviculture and establishment) can be specified to aid medium and long-term planning.

2. Study area

The research area includes 180 *Eucalyptus* stands in 19 privately-managed *Eucalyptus* plantations located in the northeast, north and central regions of Minas Gerais state, Brazil (Figure 1). These *Eucalyptus* plantations are located in tropical and sub-tropical climatic areas, corresponding to Köppen's climatic types Aw, Cwa, and Cwb, with dry winter and the rainy season occurring between the months of October and March. During the hottest months, mean daily temperatures typically do not exceed 22°C for Aw climatic types, but it exceeds 30°C for Cwa and Cwb climatic types (Alvares et al. 2013). The mean annual rainfall varies between 650 and 1500 mm. Elevations range between 500 and 1100 m above-mean-sea-level in the *Eucalyptus* stands.

The *Eucalyptus* stands used in this study were established between October 2003 and October 2012, using clones of the species *E. urophylla* x *E. grandis*, which typically lead to uniform crown and tree density conditions. The stands were established with approximately 1100 trees per hectare, using initial spacing of either 3.0 x 3.0 or 3.6 x 2.5 m, and cover more than 6000 ha. Intensive soil preparation and weed control have been practised since plantation establishment, with crown closure often occurring within the first 18 months after planting.

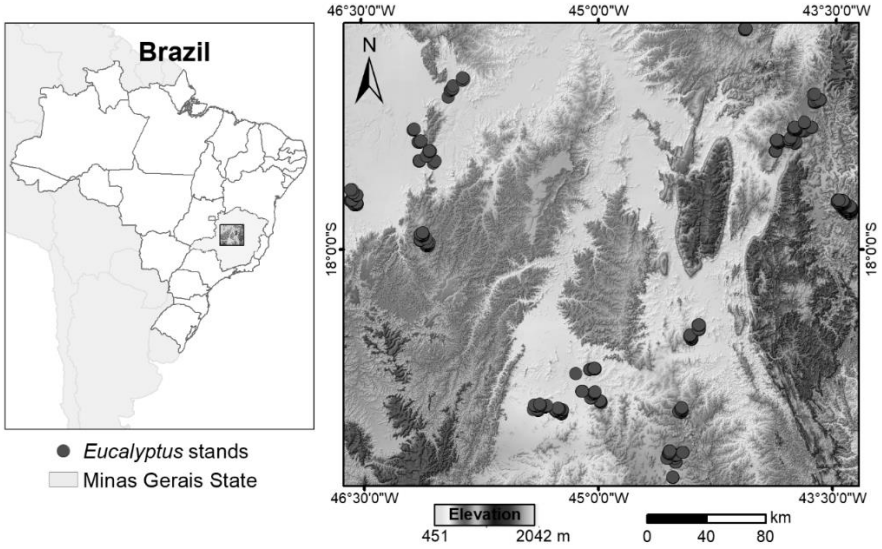


Figure 1. Location of the *Eucalyptus* stands used in this study, Minas Gerais state, Brazil.

3. Materials and Methods

3.1. Field data

Field data collection was performed between November 2005 and June 2017 in 245 plots across the *Eucalyptus* stands using standard continuous forest inventory (CFI) method. The plots were measured annually between 2 and 6 times, resulting in a total of 1119 plot measurements. The sample plots were distributed systematically and located in the field using survey-grade real time kinetic (RTK) GPS. A plot size of 25 × 20 m (500 m²) was standard. Within each plot, tree diameter at breast height (DBH) of all stems and the total height of the first 15 trees with normal stems (without bifurcation or any other defect) were measured, as well as the height of dominant trees (the 100 largest diameter trees per hectare). From the information collected in the plots, estimates of stand volume (in $m^3\ ha^{-1}$) were obtained using standard forest mensuration equations for this region. The individual tree volume equations were based on DBH and

total height, fitted for each measurement year. All fitted equations showed high coefficient of determination ($R^2 > 97\%$) and low residual standard error ($S_{xy} < 0.05$ m³). The mean annual increment (MAI) was calculated by dividing volume at stand level (in $m^3\ ha^{-1}$) by age since tree planting. The plot-based field inventory values for the *Eucalyptus* forest plantation variables are summarized in Table 1.

Table 1. Descriptive statistics of the age and growth variables based on 1119 plot measurements in 19 *Eucalyptus* plantations in Minas Gerais state, Brazil.

Variable	**Mean**	**Sd**	**Range**
Stand age (years)	5.30	1.95	2.00 – 13.25
Dominant height (m)	23.32	6.40	7.43 – 39.50
Mean annual increment ($m^3\ ha^{-1}\ yr^{-1}$)	34.67	12.64	6.32 – 73.77

Where Sd = Standard Deviation.

The growth variables used in this study to characterize site productivity were maximum mean annual increment (MAI_{max}) and stand site index (SI). The MAI_{max} can be defined as the maximum MAI value achieved by the stand. The SI of a forest stand is defined as the expected dominant height of the stand at a given base age, determined as age 7 years for this study (Packalén, Mehtätalo, and Maltamo 2011; Scolforo et al. 2016). First, we tested different age-height models to predict SI; the Schumacher anamorphic model (Equation 1) form fitted using the Algebraic Difference Approach (ADA) yielded biologically accurate predictions across the full range of the available plot data (1119 pairs of age-height):

$$SI = H_{d_i}\ e^{\left[2.2456\left(\frac{1}{A_i} - \frac{1}{A_{ref}}\right)\right]} \qquad \text{Eq. 1}$$

S_{xy} = 1.64 m and 7.13%

where SI is the Site Index (m), H_{d_i} is the dominant height (m) at Age (A^i) in years, and A_{ref} is the selected reference age for this study (7 years). The 245 field-sampled plots were then classified into three forest site productivity classes (high, medium, and low). These classes were defined by convention with high productivity corresponding to SI = 33.5 m (expected dominant height of the plots at 7 years varying between 30 and 37 m); medium productivity corresponding to SI = 26.5 m (expected dominant height of the plots at 7 years varying between 23 and 30 m), and low productivity corresponding to SI = 19.5 m (expected dominant height of the plots at 7 years varying between 16 and 23 m).

3.2. DEM data

ALOS PALSAR (Advanced Land Observing Satellite - Phase Array type L-band Synthetic Aperture Radar) DEM data with a spatial resolution of 12.5 m were used in this study to quantify the terrain conditions in the *Eucalyptus* plantation areas. Prior to calculation of the terrain attributes, the DEM was preprocessed with a standard gap/sink-filling error detection and noise removal routine. Several basic terrain attributes, which are indirectly related to causative factors controlling tree growth (e.g. heat, moisture, light, and aeration) were selected for the site productivity classification. These basic terrain attributes included general variables of elevation, slope, aspect (converted to linear aspect with a zero value for north facing slopes), total curvature, slope position, and slope length (Hengl and Reuter 2009). In addition, several specific terrain attributes were computed to characterize more complex geomorphometry and hydrological variability (Table 2) (Adams, Barnard, and Loomis 2014).

First, derivatives of slope were combined with roughness and distance measures to create 9 specific terrain attributes: Terrain Roughness Index (TRI), Surface Relief Ratio (SRR), Surface Area Ratio (SAR), Compound Topographic

Index (CTI), Topographic Wetness Index (TWI), SAGA Topographic Wetness Index (SAGA TWI), Integrated Moisture Index (IMI), vertical distance to channel network, and Topographic Position Index (TPI). These variables were selected to provide insight into site-specific water-, gravity- and wind-field conditions expected to vary across the large geographic region of interest. Second, slope derivatives were combined with estimates of insolation to create 3 attributes to characterize mean solar-field conditions: Topographic Openness, Site Exposure Index (SEI), and Heat Load Index (HLI) (Mohamedou, Tokola, and Eerikäinen 2017).

Overall, TRI, SRR, and SAR are interpreted as measures of landscape topographic roughness and heterogeneity (Riley, DeGloria, and Elliot 1999; Jenness 2004), whereas CTI, TWI, SAGA TWI, and IMI are used as measures to quantify the balance between water accumulation and drainage conditions at the local scale, and express topographic control on hydrological processes and water availability (Hengl and Reuter 2009). Topographic Openness is a non-local summary of the topographic dominance or enclosure of any location in an irregular surface (Yokoyama, Shlrasawa, and Pike 2002), and can be interpreted for convex and concave forms (Prima and Yoshida 2010). For example, local ridges are represented by positive topographic openness and local valleys by negative topographic openness (Yokoyama, Shlrasawa, and Pike 2002). SEI represents relative solar exposure (Balice et al., 2000), and HLI is a measure of the local topography and an indirect approximation of temperature related to insolation conditions (McCune and Keon 2002). All geomorphometric attributes were calculated using either the SAGA GIS software package (v. 5.0.0) or the Geomorphometry and Gradient Metrics Toolbox v2.0 (Evans et al., 2014).

Table 2. List of specific terrain attributes used to predict forest site productivity and maximum mean annual increment (MAI_{max}) (Adapted from Franklin and Ahmed (2017)).

Variables	**Description (Equation)**
Terrain Roughness Index	$TRI=Y\left[\sum\left(X_{ij}\text{-}X_{00}\right)^2\right]^{\frac{1}{2}}$
Surface Relief Ratio	$SRR=\left(X(mean)\text{-}X(min)\right)/\left(X(max)\text{-}X(min)\right)$
Surface Area Ratio	SAR= Surface area/planimetric area
Topographic Openness	Linear slope, aspect, roughness index
Site Exposure Index	$SEI=slope*\cos\left(\pi\frac{(aspect\text{-}180)}{180}\right)$
Compound Topographic Index	$CTI=\ln\left(\frac{\text{Area Value}}{\tan(\beta)}\right)$
Topographic Wetness Index	$TWI=\ln\left(\frac{\text{Specific Catchment Area}}{\tan(\beta)}\right)$
SAGA Topographic Wetness Index	$\text{SAGA TWI}=\ln\left(\frac{\text{Modified Catchment Area}}{\tan(\beta)}\right)$
Heat Load Index	$HLI=\frac{1-\cos(aspect-45)}{2}$
Integrated Moisture Index	IMI=SRP(w0.4)+FA(w0.3)+SWC(w0.2)+CL(w0.1)
Topographic Position Index	$TPI=X_0\text{-}X(mean)$

Where: X_{ij} = elevation of each neighbor cell to cell (00); X_0 = elevation at cell (0); X(mean), X(max), and X(min) = mean, maximum, and minimum elevations, respectively; β = local slope gradient, in degrees; SRP = solar radiation potential; FA = flow accumulation of water downslope; SWC = total available water capacity of soil; CL = curvature of the landscape; w = weight.

3.3. Climate data

Climate data were obtained from the WorldClim dataset (Fick and Hijmans 2017), which consists of gridded annual mean values covering the period between 1970 and 2000. The dataset provides interpolated climate layers with a spatial resolution of approximately 1 km^2 for 19 bioclimatic variables

based on historical data: Annual mean temperature (BIO1 – °C), Mean diurnal range (BIO2 – °C), Isothermality (BIO3 – °C), Temperature seasonality (BIO4 – °C), Maximum Temperature of Warmest Month (BIO5 – °C), Minimum Temperature of Coldest Month (BIO6 – °C), Temperature Annual Range (BIO7 – °C), Mean Temperature of Wettest Quarter (BIO8 – °C), Mean Temperature of Driest Quarter (BIO9 – °C), Mean Temperature of Warmest Quarter (BIO10 – °C), Mean Temperature of Coldest Quarter (BIO11 – °C), Annual Precipitation (BIO12 – mm), Precipitation of Wettest Month (BIO13 – mm), Precipitation of Driest Month (BIO14 – mm), Precipitation Seasonality (BIO15 – mm), Precipitation of Wettest Quarter (BIO16 – mm), Precipitation of Driest Quarter (BIO17 – mm), Precipitation of Warmest Quarter (BIO18 – mm), and Precipitation of Coldest Quarter (BIO19 – mm). These variables represent patterns found in monthly weather station data and the environmental gradient of the study area (Fick and Hijmans 2017).

3.4. Random Forest algorithm

The Random Forest algorithm (RF) (Breiman 2001) was used in both regression and classifier form in this study. This routine provides fast, flexible, robust and accurate predictive capabilities for high-dimensional digital datasets (Belgiu and Drăgu 2016; Ismail and Mutanga 2010; Gao et al. 2016; Zhang et al. 2018).

3.4.1. Site productivity classification

The Random Forest (RF) classification algorithm was implemented to develop a regression tree ensemble model to predict forest site productivity for the study area using the DEM-derived terrain attributes in addition to the WorldClim bioclimatic variables. The total of 245 field-sampled plots was divided randomly into 70% (172 plots) and 30% (73 plots) for training and

validation of the RF classifier, respectively. We used 500 decision trees in each RF implementation following preliminary tests of classifier performance, whereas the number of variables randomly sampled at each split was equal to the square root of the number of variables. Site productivity classification accuracy was evaluated in standard confusion matrices with overall and individual class accuracies and estimates of omission/commission errors for each classifier run (Congalton and Green, 1999). The overall accuracy is computed by dividing the total number of correct results by the total number of samples in the error matrix. The producer's accuracy indicates the probability of a reference area being correctly classified and is a measure of omission error, whereas the user's accuracy indicates the probability of an area classified on the map actually represents that class on the ground, and is a measure of commission error.

3.4.2. Maximum mean annual increment estimation

The Random Forest (RF) regression algorithm was used to model and predict the maximum mean annual increment (MAI_{max}) for the study area using as predictive variables the DEM-derived data (terrain attributes) and WorldClim dataset (bioclimatic variables). The total of 245 field-sampled plots was again randomly divided into 70% (172 plots) and 30% (73 plots) for training and validation of the RF model, respectively. 400 decision trees were used in each RF implementation following preliminary tests of model performance for MAI_{max} estimation, whereas the number of variables randomly sampled at each split was equal to the square root of the number of variables. The accuracy of predicted MAI_{max} was evaluated using the Root Mean Square Error (RMSE), in $m^3\ ha^{-1}\ yr^{-1}$ and in percentage, calculated based on the field-based inventory MAI_{max} estimates with the validation dataset (73 plots).

3.4.3. Random Forest Variable Selection and Variable Importance

A backward stepwise predictor selection was applied for each RF model to ensure only predictor variables that decreased the overall model uncertainties were used in the models. Accordingly, in each run, 20% of the lower importance predictors were removed. Initial tests revealed that this method produced the best overall predictive accuracy and allowed us to simplify the modelling process by identifying the minimum number of predictors to offer the best predictive accuracy (see also Ismail and Mutanga (2010)). Finally, we computed the RF Variable Importance metric, based on recursive substitution, enabling the most important variables in each model run to be interpreted based on either the percentage increase of the mean square error (IncMSE) or the mean decrease of accuracy, occurring when that variable was removed from the prediction or classification, respectively. Next, the IncMSE or the mean decrease of accuracy was normalised by the ratio of the largest IncMSE or mean decrease of accuracy (resulting in values between 0 and 1), and multiplied by 100 (Were et al. 2015); the higher the average accuracy decrease following variable removal, the higher the relative importance of that variable (Breiman 2001). All RF analyses were performed using the R software package *randomForest* (R Core Team, 2017).

4. Results and Discussion

4.1. *Site productivity classification*

The classification of forest site productivity based on terrain extracted from a DEM and bioclimatic variables in *Eucalyptus* plantations in Minas Gerais state, Brazil (Figure 2) yielded good results based on the field validation dataset (Table 3). The best RF classifier achieved an overall accuracy of 91.6% and a margin of error of 3.3%, which, based on recent comparative studies, are excellent results in prediction of site productivity. For example, Bueis et al.

(2016) reported an accuracy of 71% using soil, climatic and physiographic parameters to predict site index for Scots pine plantations in northern Spain.

Table 3. Table summary of forest site productivity classification accuracy.

		Site productivity					
		High		Medium		Low	
Overall Accuracy (%)	**Margin of Error (%)**	**P.A. (%)**	**U.A. (%)**	**P.A. (%)**	**U.A. (%)**	**P.A. (%)**	**U.A. (%)**
91.55	3.30	95.8	88.5	96.7	93.5	76.5	92.9

Where: P.A. = Producer's accuracy, U.A. = User's accuracy.

Medium productivity sites were the major class in the study area (67%) and, as expected, that class was more accurately mapped (producer's accuracy = 96.7% with an inclusion error of 6.5%) than the low or high productivity sites. High productivity sites were the minor class in the study area (10%). This class also presented high producer's accuracy (95.8%), with an omission error of 4.2%. Low productivity sites occurred in approximately 23% of the study area, and were classified with lower accuracy than the medium and high productivity classes, with an omission error of 23.5%. Low omission and commission errors are more important in the high and low productivity classes than in the medium productivity class, since errors in productivity classification for low and high productivity stands generate higher bias in the growth and yield predictions than for medium productivity stands (Bravo et al. 2011). Also, the correct classification of the poorest stands is considered more important, since such areas represent the worst sites to establish *Eucalyptus* plantations. Spatially-explicit site productivity maps based on these classes can help foresters to avoid these areas as plantation sites or take precautions to implement adequate silvicultural strategies for these sites, such as wider spacings, water conservation

techniques and reduction of forest productivity (Theron and Bredenkamp, 2004; Bueis et al., 2016; Gonçalves et al., 2017).

Figure 2 shows the site productivity map created based on terrain attributes and bioclimatic variables obtained with the best-performing RF classifier. The predominance of the medium productivity class and the minority of the high productivity class in the study area, respectively, can be seen. Generally, most of the southern and a small portion of the eastern side of the study area were characterized by high productivity sites, whereas from the central portion to the northern side of the study area, the site productivity was frequently low. Medium productive sites were interspersed throughout the study area, and occurred especially in the northwestern side of the study area.

A maximum mean annual increment (MAI_{max}) map of the study area covering the 19 *Eucalyptus* plantations (Figure 3) was based on the best-performing RF regression model. The comparison between field and predicted MAI_{max} (Figure 4) shows that the data were generally distributed along the 1:1 line, with an RMSE of 6.12 m^3 ha^{-1} yr^{-1} (16.15%); however, the predicted MAI_{max} tended to overestimate small values and underestimate large values. This behavior can also be observed in the MAI_{max} values per productivity class (Table 4) based on field and predicted data. These results indicate that forest productivity is strongly related to site quality, and that terrain attributes and bioclimatic variables are useful for estimating potential production and the limits imposed by topographic and climatic aspects on productivity of *Eucalyptus* plantations. Estimates of actual production must be based on more detailed local data (e.g. vegetation structure and soil attributes) since our approach is based on relatively high-resolution spatially-variable geographic datasets to predict potential forest productivity at regional scale.

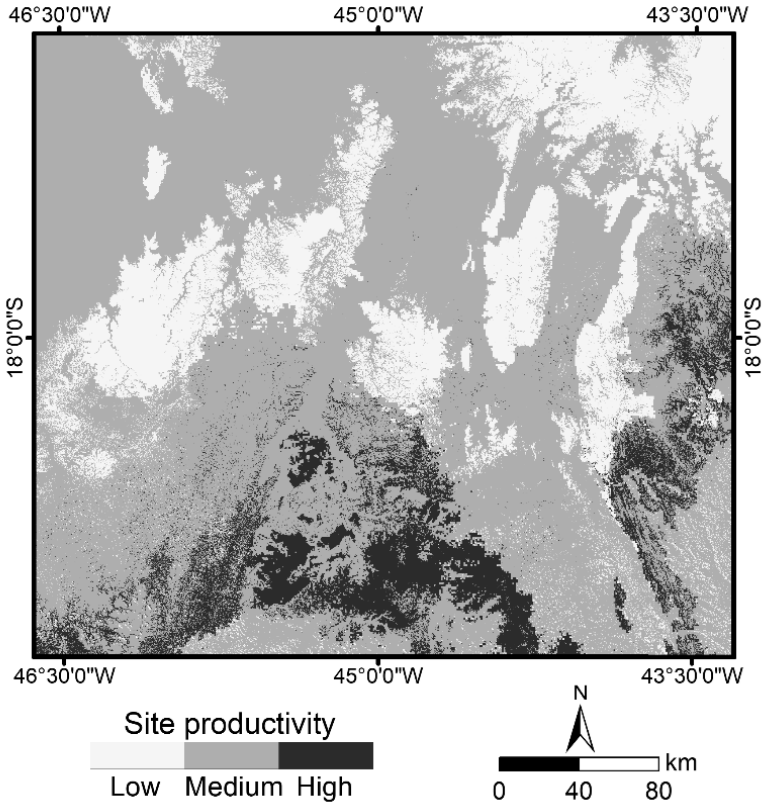


Figure 2. Forest site productivity map predicted from terrain attributes and bioclimatic variables for the study area located in Minas Gerais state, Brazil.

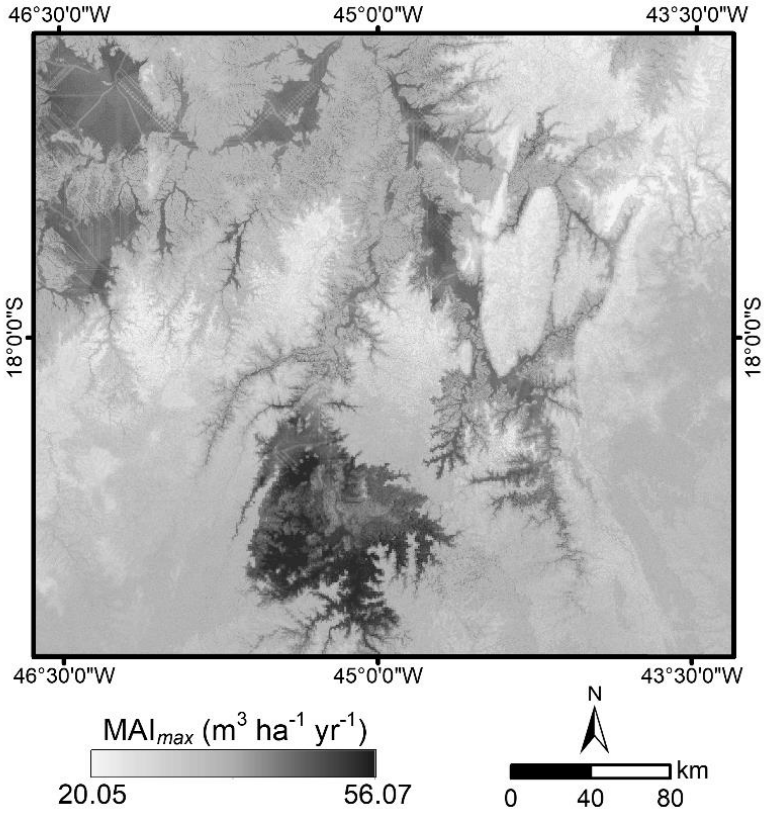


Figure 3. Maximum mean annual increment (MAI_{max}) map predicted from terrain attributes and bioclimatic variables for the study area located in Minas Gerais state, Brazil.

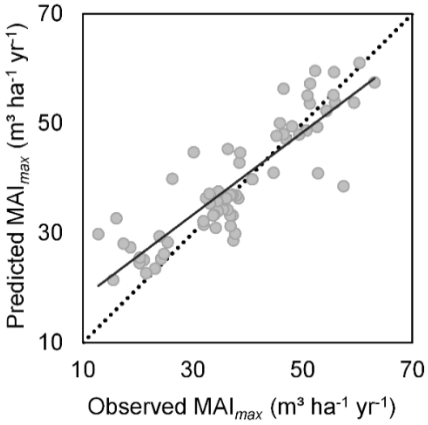


Figure 4. Scatterplot of the predicted versus observed maximum mean annual increment (MAI_{max}) obtained with the best-performing Random Forest model based on terrain attributes and bioclimatic variables in validation samples (73 plots). A 1:1 line (black, dashed) is provided for reference. The trendline (black, solid) is displayed for illustrative purposes only.

Table 4. Descriptive statistics of maximum mean annual increment (MAI_{max}) values per productivity class based on field (1119 plot measurements) and predicted data for the study area.

Site Productivity	**MAI_{max} (m³ ha^{-1} yr^{-1})**			
	Field data		**Predicted data**	
	Mean	**Range**	**Mean**	**Range**
Low	21.41	12.79 - 26.30	31.10	20.05 - 48.79
Medium	36.01	19.18 - 48.43	37.41	25.13 - 52.79
High	54.02	44.36 - 73.77	41.75	30.10 - 56.07

The most important variables for site productivity classification and maximum mean annual increment (MAI_{max}) prediction in *Eucalyptus* plantations, as measured by the Variable Importance (VI) output of the RF

algorithm (Figure 5), were Temperature seasonality (BIO4) and Precipitation of Driest Month (BIO14), respectively. These specific climate variables are directly related to water availability and periods of drought that play significant roles in survival and growth of *Eucalyptus* trees (Gonçalves et al. 2013; Binkley et al. 2017). These results are also in accordance with earlier findings where temperature and precipitation were among one of the most important drivers for the determination of forest site productivity at stand level (Albert and Schmidt 2010; Yue et al. 2016; Scolforo et al. 2017; Sharma and Parton 2018).

Ferraz Filho et al. (2011), Scolforo et al. (2013), and Scolforo et al. (2017) found that including climate variables (solar radiation, mean monthly precipitation, and/or temperature) improved the fit and predictive accuracy of their models for SI prediction in *Eucalyptus* plantations. However, these studies do not take into account the effect of topography on water availability. Although precipitation is a dominant factor, it is not the only variable that may affect water availability for plant growth (for example see Almeida et al. (2007)). In particular, topography may also be taken as decisive factor for influencing water availability by impacting the water drainage, air temperature, evapotranspiration, and retention capacities of local soils (Laamrani et al. 2014; Mohamedou, Tokola, and Eerikäinen 2014), creating variable topoclimatic conditions along topographic gradients across complex terrain (Adams, Barnard, and Loomis 2014; Bennie et al. 2008).

Elevation and Annual mean temperature (BIO01) were the next most important variables for both site productivity classification and MAI_{max} prediction. Subsequently, the most important variables included: Negative Topographic Openness, Annual Precipitation (BIO12), Isothermality (BIO03), Precipitation Seasonality (BIO15), Slope length, HLI, Positive Topographic Openness, and TWI (Figure 6). These variables are related with the main drivers controlling the forest growth and productivity: light, water, and temperature

(Adams, Barnard, and Loomis 2014), and were able to indirectly explain the main characteristics of these drivers to determine site productivity in *Eucalyptus* plantations.

In this interpretation, for our study area, high productivity sites were associated with low elevation areas and high annual precipitation, smooth terrain with lower heat load index and temperature seasonality, and the highest soil moisture content, representing cooler, less exposed, and wetter locations. On the other hand, low productivity sites were associated with higher elevation, and precipitation and temperature seasonal variability, rough or more variable terrain with high heat load index and limited soil moisture content, representing more exposed, potentially warmer, and drier locations. In between these extreme productivity classes, medium productivity sites were associated with intermediate elevation combined with low precipitation seasonality and heat load index, representing areas with less variability in water availability to *Eucalyptus* tree growth during the year.

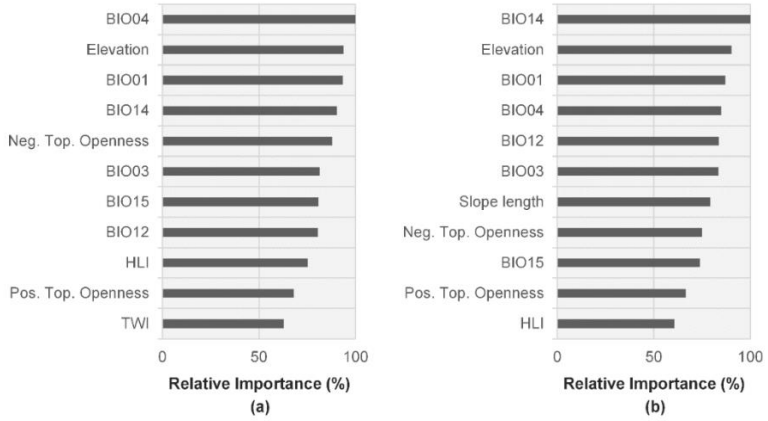


Figure 5. Relative importance of the variables as measured by the Variable Importance metric in the Random Forest algorithm predicting forest site productivity (a) and maximum mean annual increment (MAI_{max}) (b).

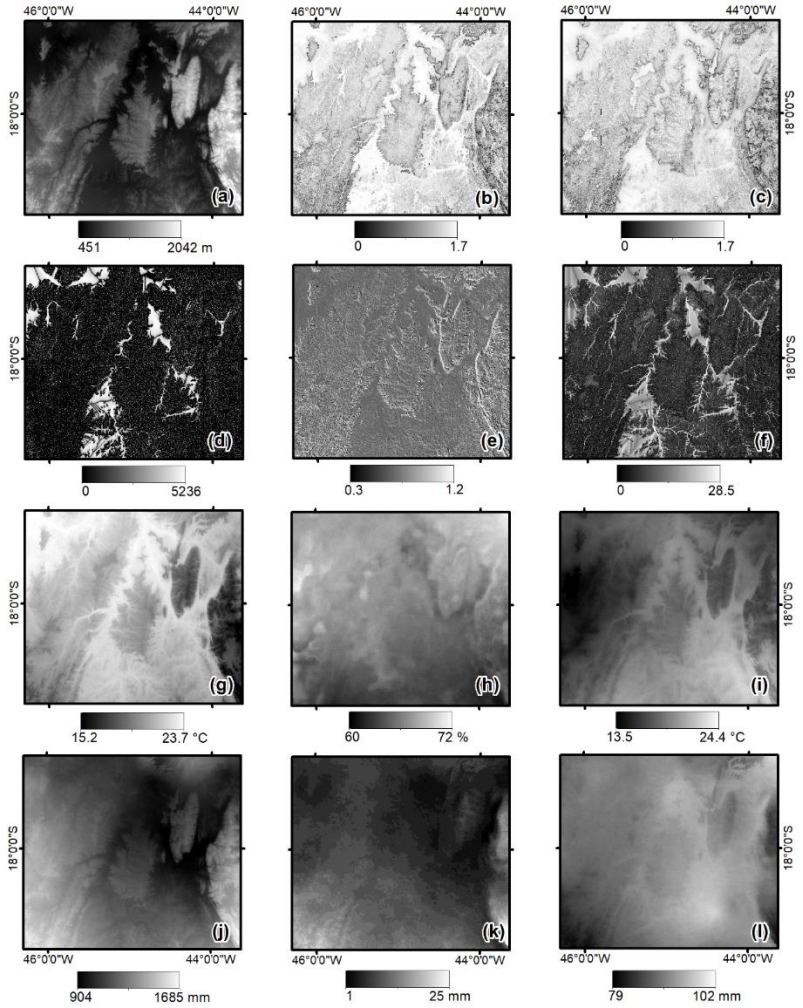


Figure 6. Terrain attributes and bioclimatic variables selected by Random Forest (RF) algorithm for site productivity classification and maximum mean annual increment (MAI_{max}) prediction: elevation (a), positive topographic openness (b), negative topographic openness (c), slope length (d), Heat Load Index - HLI (e), Topographic Wetness Index - TWI (f), annual mean temperature - BIO01 (g), isothermality - BIO03 (h), temperature seasonality - BIO04 (i), annual precipitation - BIO12 (j), precipitation of driest month -BIO14 (k), and precipitation seasonality BIO15 (l).

In this analysis, a nonparametric modeling approach was used to assess large-scale variation in productivity of *Eucalyptus* plantations in Minas Gerais, Brazil. Different statistical methods have been used to predict forest productivity based on site variables: discriminant analysis (Bravo et al. 2011; Bueis et al. 2016), classification and regression trees (Díaz-Varela et al. 2011), multiple regression (Latta, Temesgen, and Barrett 2009; Sharma, Brunner, and Eid 2012; Sharma and Parton 2018). The results of this study demonstrate that it is possible to create a reasonably reliable predictive nonparametric model (RF) of site productivity from many spatially-variable, geographic variables, especially in areas showing high environmental variability, where average productivity variation is consequence of climatic and topographic variations at regional scale. Such approach is particularly useful when dealing with large geographic regions, areas of plantation expansion, newly acquired land, new potential outgrower producers or where there is no information about the culture in the local area of interest (Marcatti et al. 2017), and constituted a significant first step in predicting and mapping forest productivity across Minas Gerais state.

5. Conclusions

Topographic and climatic effects on potential productivity of *Eucalyptus* plantations were analyzed by Random Forest machine learning algorithm. Terrain attributes derived from a DEM and bioclimatic variables correctly classified site productivity, with an overall accuracy of 91.6%, and predicted maximum mean annual increment (MAI_{max}) with a RMSE of 6.1 m^3 ha^{-1} yr^{-1} (16.2%).

Temperature seasonality (BIO04), precipitation of the driest month of the year (BIO14), elevation, and annual mean temperature (BIO01) emerged as major indicators of potential productivity in *Eucalyptus* plantations in our study area, enabling us to discriminate three site productivity classes, and predict the

potential maximum mean annual increment (MAI_{max}) for an area of approximately 100,690 km² in the northeast, north and central regions of Minas Gerais state, Brazil.

The approach developed here can be used to estimate site productivity and potential MAI_{max} at any geographical location, given the availability of terrain attributes and bioclimatic variables worldwide, thereby supporting more informed forest management decisions. This approach could be also linked to climate models to predict how productivity of *Eucalyptus* plantations would change under climate change.

References

Adams, H.R., Barnard, H.R., Loomis, A.K., 2014. Topography alters tree growth–climate relationships in a semi-arid forested catchment. Ecosphere 5, art148. https://doi.org/10.1890/ES14-00296.1

Albert, M., Schmidt, M., 2010. Climate-sensitive modelling of site-productivity relationships for Norway spruce (*Picea abies* (L.) Karst.) and common beech (*Fagus sylvatica* L.). For. Ecol. Manage. 259, 739–749. https://doi.org/10.1016/j.foreco.2009.04.039

Almeida, A.C., Soares, J. V., Landsberg, J.J., Rezende, G.D., 2007. Growth and water balance of *Eucalyptus grandis* hybrid plantations in Brazil during a rotation for pulp production. For. Ecol. Manage. 251, 10–21. https://doi.org/10.1016/j.foreco.2007.06.009

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

Assmann, E., 1970. The Principles of Forest Yield Study. Pergamon Press, Oxford.

Balice, R.G., Miller, J.D., Oswald, B.P., Edminster, C., Yool, S.R., 2000. Forest Surveys and Wildfire Assessment in the Los Alamos Region; 1998–1999. Los Alamos National Lab, NM (US), LA 13714-MS. 12 pp.

Belgiu, M., Drăgu, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS J. Photogramm. Remote Sens. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Bennie, J., Huntley, B., Wiltshire, A., Hill, M.O., Baxter, R., 2008. Slope, aspect and climate: Spatially explicit and implicit models of topographic microclimate in chalk grassland. Ecol. Modell. 216, 47–59. https://doi.org/10.1016/j.ecolmodel.2008.04.010

Berrill, J.P., O'Hara, K.L., 2016. How do biophysical factors contribute to height and basal area development in a mixed multiaged coast redwood stand? Forestry 89, 170–181. https://doi.org/10.1093/forestry/cpv049

Binkley, D., Campoe, O.C., Alvares, C., Carneiro, R.L., Cegatta, Í., Stape, J.L., 2017. The interactions of climate, spacing and genetics on clonal Eucalyptus plantations across Brazil and Uruguay. For. Ecol. Manage. 405, 271–283. https://doi.org/10.1016/j.foreco.2017.09.050

Bravo-Oviedo, A., Tomé, M., Bravo, F., Montero, G., del Río, M., 2008. Dominant height growth equations including site attributes in the generalized algebraic difference approach. Can. J. For. Res. 38, 2348–2358. https://doi.org/10.1139/X08-077

Bravo, F., Lucà, M., Mercurio, R., Sidari, M., Muscolo, A., 2011. Soil and forest productivity: A case study from Stone pine (Pinus pinea L.) stands in

Calabria (southern Italy). IForest 4, 25–30. https://doi.org/10.3832/ifor0559-004

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

Bueis, T., Bravo, F., Pando, V., Turrión, M.B., 2016. Relationship between environmental parameters and *Pinus sylvestris* L. site index in forest plantations in northern Spain acidic plateau. IForest 9, 394–401. https://doi.org/10.3832/ifor1600-008

Burkhart, H.E., Tomé, M., 2012. Modeling Forest Trees and Stands, first ed. Springer, Dordrecht.

Campoe, O.C., Munhoz, J.S.B., Alvares, C.A., Carneiro, R.L., de Mattos, E.M., Ferez, A.P.C., Stape, J.L., 2016. Meteorological seasonality affecting individual tree growth in forest plantations in Brazil. For. Ecol. Manage. 380, 149–160. https://doi.org/10.1016/j.foreco.2016.08.048

Campoe, O.C., Stape, J.L., Nouvellon, Y., Laclau, J.P., Bauerle, W.L., Binkley, D., Le Maire, G., 2013. Stem production, light absorption and light use efficiency between dominant and non-dominant trees of *Eucalyptus grandis* across a productivity gradient in Brazil. For. Ecol. Manage. 288, 14–20. https://doi.org/10.1016/j.foreco.2012.07.035

Chen, H.Y., Krestov, P. V, Klinka, K., 2002. Trembling aspen site index in relation to environmental measures of site quality at two spatial scales. Can. J. For. Res. 32, 112–119. https://doi.org/10.1139/x01-179

Congalton, R.G., and K. Green, 1999. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, Lewis, Boca Raton, Florida, 137 p.

Díaz-Varela, R.-A., Álvarez-Álvarez, P., Diaz-Varela, E., Calvo-Iglesias, S.,

2011. Prediction of stand quality characteristics in sweet chestnut forests in NW Spain by combining terrain attributes, spectral textural features and landscape metrics. For. Ecol. Manage. 261, 1962–1972. https://doi.org/10.1016/j.foreco.2011.02.023

Evans, J.S., J. Oakleaf, S.A. Cushman, and D. Theobald, 2014. An ArcGIS Toolbox for Surface Gradient and Geomorphometric Modeling, Version 2.0-0. URL: http://evansmurphy.wix.com/evansspatial (last date accessed: 24 July 2018).

Ferraz Filho, A.C., Scolforo, J.R.S., Ferreira, M.Z., Maestri, R., Assis, A.L. de, Oliveira, A.D. de, Mello, J.M. de, 2011. Dominant height projection model with the addition of environmental variables. Cerne 17, 427–433. https://doi.org/10.1590/S0104-77602011000300018

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

Franklin, S., Ahmed, O., 2017. Object-based Wetland Characterization Using Radarsat-2 Quad-Polarimetric SAR Data, Landsat-8 OLI Imagery, and Airborne Lidar-Derived Geomorphometric Variables. Photogramm. Eng. Remote Sens. 83, 27–36. https://doi.org/10.14358/PERS.83.1.27

Gao, T., Zhu, J., Deng, S., Zheng, X., Zhang, J., Shang, G., Huang, L., 2016. Timber production assessment of a plantation forest: An integrated framework with field-based inventory, multi-source remote sensing data and forest management history. Int. J. Appl. Earth Obs. Geoinf. https://doi.org/10.1016/j.jag.2016.06.004

Gonçalves, J.L. de M., Alvares, C.A., Higa, A.R., Silva, L.D., Alfenas, A.C., Stahl, J., Ferraz, S.F. de B., Lima, W. de P., Brancalion, P.H.S., Hubner,

A., Bouillet, J.-P.D., Laclau, J.-P., Nouvellon, Y., Epron, D., 2013. Integrating genetic and silvicultural strategies to minimize abiotic and biotic constraints in Brazilian eucalypt plantations. For. Ecol. Manage. 301, 6–27. https://doi.org/10.1016/j.foreco.2012.12.030

Gonçalves, J.L.M., Alvares, C.A., Rocha, J.H.T., Brandani, C.B., Hakamada, R., 2017. Eucalypt plantation management in regions with water stress. South. For. a J. For. Sci. 79, 169–183. https://doi.org/10.2989/20702620.2016.1255415

Hengl, T., Reuter, H.I. (eds) 2009. Geomorphometry: Concepts, Software, Applications. Developments in Soil Science: Elsevier.

Huang, S., Ramirez, C., Conway, S., Kennedy, K., Kohler, T., Liu, J., 2017. Mapping site index and volume increment from forest inventory, Landsat, and ecological variables in Tahoe National Forest, California, USA. Can. J. For. Res. 47, 113–124. https://doi.org/10.1139/cjfr-2016-0209

IBÁ - Brazilian Tree Industry, 2017. Brazilian Tree Industry - Annual Report. Brazilian Tree Industry (IBÁ), São Paulo.

Ismail, R., Mutanga, O., 2010. A comparison of regression tree ensembles: Predicting Sirex noctilio induced water stress in Pinus patula forests of KwaZulu-Natal, South Africa. Int. J. Appl. Earth Obs. Geoinf. 12, S45–S51. https://doi.org/10.1016/j.jag.2009.09.004

Jenness, J.S., 2004. Calculating landscape surface area from digital elevation models. Wildl. Soc. Bull. 32, 829–839. https://doi.org/10.2193/0091-7648(2004)032[0829:CLSAFD]2.0.CO;2

Laamrani, A., Valeria, O., Bergeron, Y., Fenton, N., Cheng, L.Z., Anyomi, K., 2014. Effects of topography and thickness of organic layer on productivity

of black spruce boreal forests of the Canadian Clay Belt region. For. Ecol. Manage. 330, 144–157. https://doi.org/10.1016/j.foreco.2014.07.013

Latta, G., Temesgen, H., Barrett, T.M., 2009. Mapping and imputing potential productivity of Pacific Northwest forests using climate variables. Can. J. For. Res. 39, 1197–1207. https://doi.org/10.1139/X09-046

Marcatti, G.E., Resende, R.T., Resende, M.D. V., Ribeiro, C.A.A.S., dos Santos, A.R., da Cruz, J.P., Leite, H.G., 2017. GIS-based approach applied to optimizing recommendations of Eucalyptus genotypes. For. Ecol. Manage. 392, 144–153. https://doi.org/10.1016/j.foreco.2017.03.006

McCune, B., Keon, D., 2002. Equations for potential annual direct incident radiation and heat load. J. Veg. Sci. 13, 603–606. https://doi.org/10.1111/j.1654-1103.2002.tb02087.x

Mohamed, A., Reich, R.M., Khosla, R., Aguirre-Bravo, C., Briseño, M.M., 2014. Influence of climatic conditions, topography and soil attributes on the spatial distribution of site productivity index of the species rich forests of Jalisco, Mexico. J. For. Res. 25, 87–95. https://doi.org/10.1007/s11676-014-0434-5

Mohamedou, C., Tokola, T., Eerikäinen, K., 2017. LiDAR-based TWI and terrain attributes in improving parametric predictor for tree growth in southeast Finland. Int. J. Appl. Earth Obs. Geoinf. 62, 183–191. https://doi.org/10.1016/j.jag.2017.06.004

Mohamedou, C., Tokola, T., Eerikäinen, K., 2014. Applying airborne γ-ray and DEM-derived attributes to the local improvement of the existing individual-tree growth model for diameter increment. Remote Sens. Environ. 155, 248–256. https://doi.org/10.1016/j.rse.2014.08.033

Packalén, P., Mehtätalo, L., Maltamo, M., 2011. ALS-based estimation of plot volume and site index in a eucalyptus plantation with a nonlinear mixed-effect model that accounts for the clone effect. Ann. For. Sci. 68, 1085–1092. https://doi.org/10.1007/s13595-011-0124-9

Parresol, B.R., Scott, D.A., Zarnoch, S.J., Edwards, L.A., Blake, J.I., 2017. Modeling forest site productivity using mapped geospatial attributes within a South Carolina Landscape, USA. For. Ecol. Manage. 406, 196–207. https://doi.org/10.1016/j.foreco.2017.10.006

Prima, O.D.A., Yoshida, T., 2010. Characterization of volcanic geomorphology and geology by slope and topographic openness. Geomorphology 118, 22–32. https://doi.org/10.1016/j.geomorph.2009.12.005

R Core Team. (2017). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing

Resende, R.T., Marcatti, G.E., Pinto, D.S., Takahashi, E.K., Cruz, C.D., Resende, M.D. V., 2016. Intra-genotypic competition of *Eucalyptus* clones generated by environmental heterogeneity can optimize productivity in forest stands. For. Ecol. Manage. 380, 50–58. https://doi.org/10.1016/j.foreco.2016.08.041

Riley, S.J., DeGloria, S.D., Elliot, R., 1999. A Terrain Ruggedness Index that Qauntifies Topographic Heterogeneity. Intermt. J. Sci. https://doi.org/citeulike-article-id:8858430

Scolforo, H.F., de Castro Neto, F., Scolforo, J.R.S., Burkhart, H., McTague, J.P., Raimundo, M.R., Loos, R.A., da Fonseca, S., Sartório, R.C., 2016. Modeling dominant height growth of eucalyptus plantations with parameters conditioned to climatic variations. For. Ecol. Manage. 380, 182–195. https://doi.org/10.1016/j.foreco.2016.09.001

Scolforo, H.F., Scolforo, J.R.S., Stape, J.L., McTague, J.P., Burkhart, H., McCarter, J., de Castro Neto, F., Loos, R.A., Sartorio, R.C., 2017. Incorporating rainfall data to better plan eucalyptus clones deployment in eastern Brazil. For. Ecol. Manage. https://doi.org/10.1016/j.foreco.2017.02.025

Scolforo, J.R.S., Maestri, R., Ferraz Filho, A.C., de Mello, J.M., de Oliveira, A.D., de Assis, A.L., 2013. Dominant Height Model for Site Classification of Eucalyptus grandis Incorporating Climatic Variables. Int. J. For. Res. 2013, 1–7. https://doi.org/10.1155/2013/139236

Sharma, M., Parton, J., 2018. Climatic Effects on Site Productivity of Red Pine Plantations. For. Sci. 00, 1–11. https://doi.org/10.1093/forsci/fxy013

Sharma, R.P., Brunner, A., Eid, T., 2012. Site index prediction from site and climate variables for Norway spruce and Scots pine in Norway. Scand. J. For. Res. 27, 619–636. https://doi.org/10.1080/02827581.2012.685749

Skovsgaard, J.P., Vanclay, J.K., 2013. Forest site productivity: a review of spatial and temporal variability in natural site conditions. Forestry 86, 305–315. https://doi.org/10.1093/forestry/cpt010

Skovsgaard, J.P., Vanclay, J.K., 2008. Forest site productivity: a review of the evolution of dendrometric concepts for even-aged stands. Forestry 81, 13–31. https://doi.org/10.1093/forestry/cpm041

Stape, J.L., Binkley, D., Ryan, M.G., Fonseca, S., Loos, R.A., Takahashi, E.N., Silva, C.R., Silva, S.R., Hakamada, R.E., Ferreira, J.M. de A., Lima, A.M.N., Gava, J.L., Leite, F.P., Andrade, H.B., Alves, J.M., Silva, G.G.C., Azevedo, M.R., 2010. The Brazil Eucalyptus Potential Productivity Project: Influence of water, nutrients and stand uniformity on wood production. For. Ecol. Manage. 259, 1684–1694.

https://doi.org/10.1016/j.foreco.2010.01.012

Theron, K., Bredenkamp, B.V., 2004. Stand density and stocking in plantations. In: Burley, J., Evans, J., Youngquist, J.A. (Eds.), Encyclopedia of Forest Science, Elsevier, pp. 829 – 836.

Véga, C., St-Onge, B., 2009. Mapping site index and age by linking a time series of canopy height models with growth curves. For. Ecol. Manage. 257, 951–959. https://doi.org/10.1016/j.foreco.2008.10.029

Were, K., Bui, D.T., Dick, Ø.B., Singh, B.R., 2015. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. Ecol. Indic. 52, 394–403. https://doi.org/10.1016/j.ecolind.2014.12.028

Yokoyama, R., Shlrasawa, M., Pike, R.J., 2002. Visualizing Topography by Openness : A New Application of Image Processing to Digital Elevation Models. Photogramm. Eng. Remote Sens. 68, 257–265.

Yue, C., Kahle, H.-P., von Wilpert, K., Kohnle, U., 2016. A dynamic environment-sensitive site index model for the prediction of site productivity potential under climate change. Ecol. Modell. 337, 48–62. https://doi.org/10.1016/j.ecolmodel.2016.06.005

Zhang, C., Denka, S., Cooper, H., Mishra, D.R., 2018. Quantification of sawgrass marsh aboveground biomass in the coastal Everglades using object-based ensemble analysis and Landsat data. Remote Sens. Environ. 204, 366–379. https://doi.org/10.1016/j.rse.2017.10.018