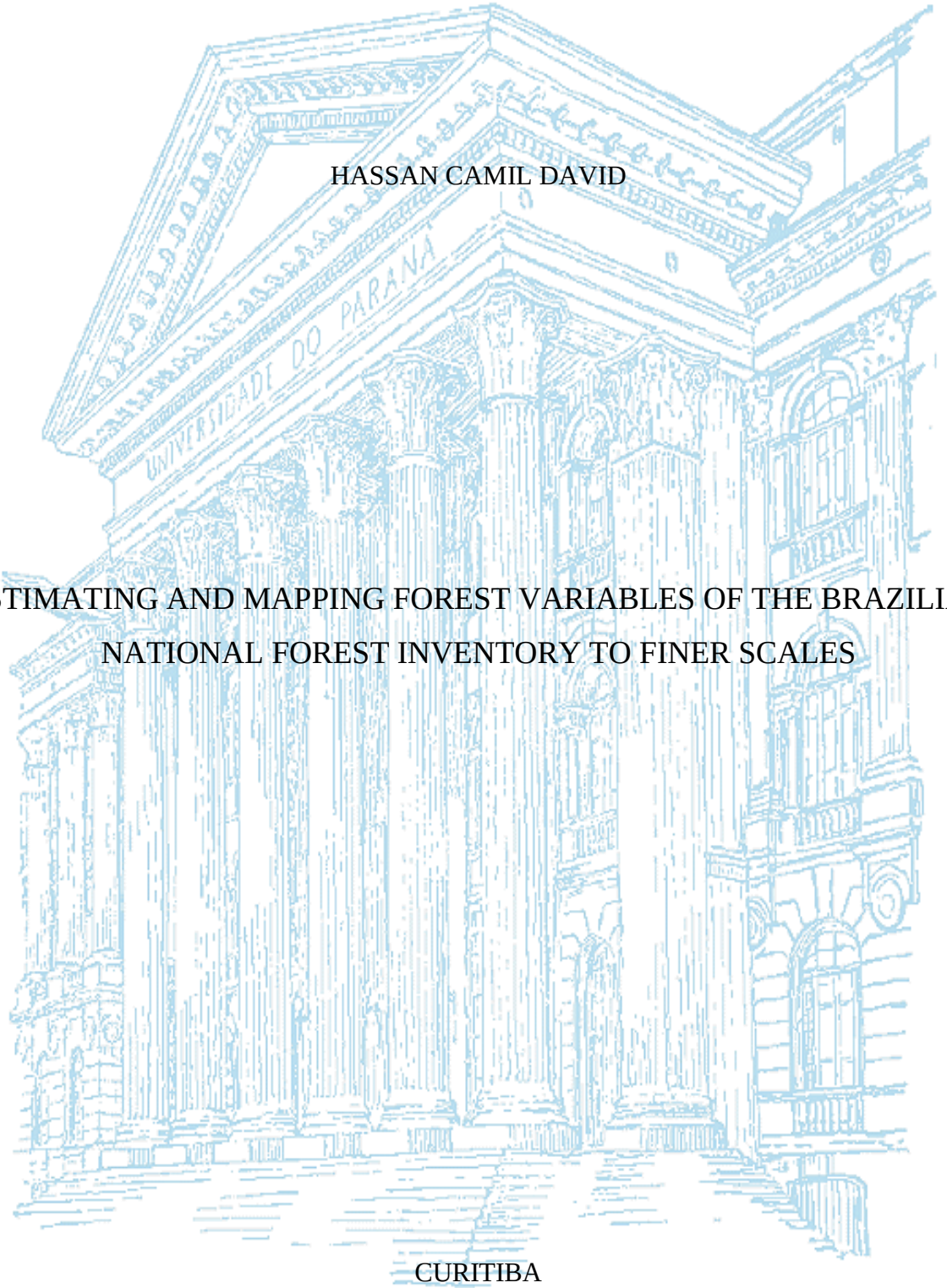
UNIVERSIDADE FEDERAL DO PARANÁ



HASSAN CAMIL DAVID

ESTIMATING AND MAPPING FOREST VARIABLES OF THE BRAZILIAN NATIONAL FOREST INVENTORY TO FINER SCALES

CURITIBA

2018

HASSAN CAMIL DAVID

ESTIMATING AND MAPPING FOREST VARIABLES OF THE BRAZILIAN NATIONAL FOREST INVENTORY TO FINER SCALES

Tese apresentada ao Programa de Pós-graduação em Engenharia Florestal, Área de Concentração em Manejo Florestal, Setor de Ciências Agrárias, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Doutor em Engenharia Florestal.

Orientador:
Professor Dr. Sylvio Péllico Netto

Coorientador:
Professor(a) Drª. Ana Paula Dalla Corte
Dr. Carlos Roberto Sanquetta

CURITIBA

2018

Ficha catalográfica elaborada pela
Biblioteca de Ciências Florestais e da Madeira - UFPR

David, Hassan Camil

Estimating and mapping forest variables of the Brazilian National Forest Inventory to finer scales / Hassan Camil David. – Curitiba, 2018.
116 f. : il.

Orientador: Prof. Dr. Sylvio Péllico Netto
Coorientadores: Profa. Dra. Ana Paula Dalla Corte
Prof. Dr. Carlos Roberto Sanquetta
Tese (Doutorado) - Universidade Federal do Paraná, Setor de Ciências Agrárias, Programa de Pós-Graduação em Engenharia Florestal. Defesa: Curitiba, 23/02/2018.
Área de concentração: Manejo Florestal.

1. Inventário florestal - Brasil. 2. Dendrometria. 3. Biomassa. 4. Sensoriamento remoto. 5. Carbono. 6. Teses. I. Netto, Sylvio Péllico. II. Dalla Corte, Ana Paula. III. Sanquetta, Carlos Roberto. IV. Universidade Federal do Paraná, Setor de Ciências Agrárias. V. Título.

CDD – 634.9
CDU – 634.0.524.6(816)

Bibliotecária: Berenice Rodrigues Ferreira – CRB 9/1160

MINISTÉRIO DA EDUCAÇÃO
SETOR SETOR DE CIENCIAS AGRARIAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO ENGENHARIA FLORESTAL

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ENGENHARIA FLORESTAL da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **HASSAN CAMIL DAVID** intitulada: **ESTIMATING AND MAPPING FOREST VARIABLES OF THE BRAZILIAN NATIONAL FOREST INVENTORY TO FINER SCALES**, após terem inquirido o aluno e realizado a avaliação do trabalho, são de parecer pela sua aprovação no rito de defesa.

A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

Curitiba, 23 de Fevereiro de 2018.

SYLVIO PÉLLICO NETTO
Presidente da Banca Examinadora

ALEXANDER CHRISTIAN VIBRANS
Avaliador Externo

YEDA MARIA MALHEIROS DE OLIVEIRA
Avaliador Externo

SEBASTIÃO DO AMARAL MACHADO
Avaliador Interno

DANIEL PIOTTO
Avaliador Externo

MINISTÉRIO DA EDUCAÇÃO U.F.P.R. PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA FLORESTAL MESTRADO DOUTORADO SETOR DE CIÊNCIAS AGRÁRIAS

To my parents and
my brother

AUTHOR'S BIOGRAPHY

Hassan Camil David was born in Nanuque, MG, Brazil (May 16, 1989). He is son of Camil Cheab David and Vera L. R. dos Santos, respectively born in Carlos Chagas and Mayrink, both cities in the State of MG, Brazil. Hassan lived in his hometown until February 2007, after this he moved to Alegre, ES, Brazil, to course Forest Engineering at the Federal University of the State of Espirito Santo (UFES). He lived temporally (during six months) in Posto da Mata, BA, Brazil, being intern in the forestry company Fibria Celulose. He got the degree of Forest Engineer in February 2012. In this same year, he moved to Curitiba, PR, Brazil to course MBA in Business Management, at the UNINTER, and Master in Forest Engineer, at the Federal University of the State of Parana (UFPR). He got the MBA's degree in 2013 and the Master's degree in 2014. During the 2012-2014 period, he constantly worked with projects related to forest inventory, including the National Forest Inventory (NFI) in the State of Parana. In 2014, started his PhD in Forest Engineer, also at the UFPR. Hassan C. David became Assistant Professor at the Federal Rural University of Amazonia (UFRA), having to move from Curitiba to Capitão Poço, PA, Brazil, in May 2016. Hassan got the PhD's degree in February 2018.

ACKNOWLEDGMENT

I have a great affection for the Federal University of Parana (UFPR), especially for the Graduate Program in Forestry, where I got my Master and PhD degrees. I have to thank a lot the Federal Rural University of Amazonia (UFRA), for providing me a period of license to conclude my PhD thesis, allowing me to enhance the quality of my work.

I owe a lot to my advisor, Dr. Sylvio Péllico Netto, a friend that inspires me with all his dedication to science, besides being a Professor that taught me not only about forestry (and languages, and mathematics, and statistics…), but also about life; I will bear in mind your lessons forever, thank you! I am also very grateful to my co-advisor, Dr. Ana Paula Dalla Corte, who saved no effort to help me out with great discussions and exchange of ideas. Thank you for the great support given for developing my PhD thesis. To my second co-advisor, Dr. Carlos R. Sanquetta, thank you for enhancing this work with your great knowledge about the subject. In addition, I thank the members of my evaluation board: Dr. Alexander C. Vibrans (FURB), Dr. Daniel Piotto (UFSB), Dr. Yeda M. Malheiros de Oliveira (Embrapa Florestas), and Dr. Sebastião do Amaral Machado, for reading and giving important contributions to enhance this work.

To Dr. Joberto Veloso de Freitas (Brazilian Forest Service), for making available the database necessary to develop this work. My sincere thanks! To Mr. Tadeu Weidlich Motta, and Mr. Osmar Ribas, for making available herbarium data from the first phase of the Parana's NFI. To Dr. Canty Morty and Dr. Flávio J. Ponzoni, for giving special attention to explain important issues regarding to radiometric normalization, which ended up not being necessary.

To my girlfriend Alessandra Calegari, who has stayed with me during great experiences and shared knowledge essential to life. Thank you for making me a more mature person.

Last, but not least, I would like to thank all my friends and colleagues that, somehow, had participation in my personal and professional growth while I was PhD student, including: Alemão, Alexandre, Allan, Ana, Ângelo, Antônio, Ataídes, Badeco, Bila, Bonazza, Brown, Calango, Carioca, Chapolim, Cris, David MacFarlane, Ed, Fafá, Fran, Iza, Luan, Lucas, Luciano, MacFadden, Mailson, Myrcia, Pedroca, Rodrigo, Roraí, Roveda, and Sintia.

My best regards,

Hassan C. David

There’s a big... a big hard sun
beating on the big people
in the big hard world...

Eddie Vedder

RESUMO

Introdução: Brasil deu início recentemente ao seu Inventário Florestal Nacional (IFN) seguindo uma amostragem em larga escala de 1:100.000. Com isso, há uma crescente demanda em desagregar variáveis florestais a uma escala mais precisa, refinada. *Objetivo*: Com auxílio de dados de sensoriamento remoto, o objetivo deste trabalho foi testar dois métodos para estimar e mapear, a uma escala fina, componentes florestais como volume de madeira, biomassa e carbono. *Métodos*: O procedimento analítico é dividido em sete passos. Primeiro: coletar variáveis dendrométricas; dados de parcelas do IFN de três mesorregiões do estado do Paraná foram usados. Segundo: modelagem florestal; necessária para estimar variáveis ao nível da árvore, como volume de fuste, biomassa e carbono de árvores. Terceiro: cálculo do erro de amostragem e intervalo de confiança (IC) do IFN. Quarto: pré-processamento de imagens de satélite; foram usadas sete bandas do Landsat-8 OLI, portanto Número Digital (DN) foi transformado em reflectância da superfície (variável física). Quinto: classificação de imagens: classificações orientadas ao objeto foram efetuadas para delimitar classes de (i) florestas em estágio sucessional inicial e (ii) médio-avançado. Sexto: modelagem a nível de pixel; dois métodos em escala fina foram testados para estimar variáveis florestais ao nível de pixel. O primeiro deles usa um Modelo de Reflectância da Superfície (MRS), que é um modelo linear que ajusta volume de madeira em função da reflectância de superfície de dados do Landsat-8 OLI, incluindo bandas do Vermelho, IFP, IFM1, IFM2, Brilho, e NDVI. O método de *stepwise* foi usado para selecionar as variáveis de reflectância. Componentes principais das variáveis selecionadas corresponderam às variáveis independentes. O segundo método é baseado na técnica geoestatística krigagem com regressão (KR), em que latitude e longitude foram as únicas variáveis independentes. O primeiro método (com MRS) foi aplicado ao volume florestal, biomassa e carbono. O segundo (com KR) foi aplicado ao volume florestal apenas. Sétimo: Volume florestal foi extrapolado a escalas ampla (sem técnicas de sensoriamento remoto) e fina (com tais técnicas), e então comparadas entre si. *Resultados*: O método com MRS teve um desempenho melhor que o método com KR, embora ambos forneceram, no geral, resultados próximos. O MRS estimou um volume médio de 123,8 $m^3\ ha^{-1}$ e um total para a área de estudo de 210.961.589 m^3. A KR estimou média de 115,7 $m^3\ ha^{-1}$ e total de 203.326.674 m^3, ou seja, 3,7% a menos que as estimativas do MRS. Em relação à biomassa florestal, o MRS estimou média e total de 120,1 Mg ha^{-1} e 217.736.862 Mg, respectivamente. Para o carbono, a média foi de 49,2 MgC ha^{-1} e total de 89.272.113 MgC, correspondendo a 327.331.082 Mg de CO_2 equivalente estocado na área de estudo. O volume total estimado por ambos métodos de escala fina enquadram-se no IC obtido a uma escala ampla. *Conclusão*: Apesar do MRS e KR terem sido capazes de produzir, no geral, estimativas precisas do volume de madeira, o MRS representou melhor espacialmente o volume de madeira. O MRS também estimou acuradamente biomassa e carbono florestal.

Palavras-chave: Volume de madeira. Biomassa e carbono. Sensoriamento remoto. Modelos derivados de dados de satélite.

ABSTRACT

Introduction: Brazil recently started its National Forest Inventory (NFI) following a broad-scale sampling of 1:100,000. Therewith, there is an increasing need for disaggregating forest variables at a more precise, finer scale. *Objective*: With assistance of remotely sensed data, the aim of this study was to test two methods for estimating and mapping, at a fine scale, forest components as wood volume, biomass and carbon. *Methods*: The analytical procedure has seven steps. First: collecting tree variables; NFI plot data from three meso-regions of the State of Parana were used. Second: forest modelling; necessary for estimating tree-level variables, as stem volume, tree biomass and carbon. Third: calculation of NFI's sampling error and confidence interval (CI). Fourth: satellite imagery pre-processing; seven Landsat-8 OLI bands were used, therefore Digital Number (DN) was transformed into surface reflectance (physical variable). Fifth: imagery classification; object-oriented classifications were performed for delimiting classes of (i) early successional, and (ii) mid- to-late successional forest. Sixth: pixel-level modeling; two fine-scale methods for estimating forest variables at pixel-level were tested. The first of them uses a Surface Reflectance Model (SRM), which is a linear model that fits wood volume as function of surface reflectance from Landsat-8 OLI data, including Red, NIR, SWIR1, SWIR2, Brightness, and NDVI. The stepwise method was used to select the reflectance variables. Principal components of the selected variables were used as input variables. The second method is based on the regression-kriging (RK) geostatistical technique, in which latitude and longitude were the only two input variables. The first method (with SRM) was applied to wood volume, biomass and carbon. The second one (with RK) was applied to wood volume only. Seventh: Wood volume was extrapolated at coarse (without remote sensing techniques) and fine (with such techniques) scales, and then compared with each other. *Results*: The SRM method had a performance better than the RK method, though both of them provided, in general, similar results. The SRM estimated a mean of 123.8 m^3 ha^{-1} and a total volume for the study area of 210,961,589 m^3. The RK estimated a mean of 115.7 m^3 ha^{-1} and total of 203,326,674 m^3, i.e., value 3.7 % lesser than the SRM estimates. In relation to forest biomass, the SRM estimated a mean and total of, respectively, 120.1 Mg ha^{-1} and 217,736,862 Mg. For carbon, the mean was of 49.2 MgC ha^{-1} and total of 89,272,113 MgC, corresponding to 327,331,082 Mg of CO_2-equivalent stocked in the study area. The total wood volumes estimated by both fine-scale methods fit into the CI obtained at a coarse scale. *Conclusion*: Despite SRM and RK were capable of producing overall accurate estimates of wood volume, SRM better represented spatially the wood volume. SRM also estimated accurately forest biomass and carbon.

Keywords: Wood volume. Biomass and carbon. Remote sensing. Satellite data-derived models.

RESUMO EXPANDIDO

1. Introdução

O Inventário Florestal Nacional (IFN) brasileiro está em fases iniciais. Mapeamentos do estoque de madeira, biomassa e carbono, ao nível nacional e regional, aparecerão como uma das primeiras demandas emergentes. Nessas fases iniciais, o papel dos métodos para quantificar recursos florestais a escalas mais finas é de crucial importância no processamento dos dados do IFN. Isso requer uma desagregação de variáveis florestais de escalas amplas a escalas mais refinadas, principalmente para auxiliar o manejo florestal e o planejamento aos níveis regionais (Wilson et al., 2013).

O conceito de desagregação aplicado aqui diz respeito ao refinamento de estimativas de uma escala ampla derivada de dados do IFN, o que pode ser alcançado por meio de técnicas baseadas em sensoriamento remoto. Com o refinamento de estimativas, variáveis florestais são fornecidas a uma escala tão fina quanto for a resolução espacial do sensor remoto empregado. Em geral, quantificar volume de madeira em escalas refinadas é importante para propósitos de planejamento e manejo, enquanto que quantificar biomassa e carbono acima do solo (BAS e CAS) desempenha um papel importante para auxiliar relatórios de emissão de gases de efeito estufa, bem como informar emissões e sequestro de carbono (David et al., 2017; Kim et al., 2013; Wilson et al., 2013).

A maioria das técnicas refinadas para quantificar recursos florestais são baseadas em modelagem e dados de sensoriamento remoto, sendo aplicados tanto aos níveis nacionais, quanto regionais e locais (Brooks et al., 2016; Maack et al., 2016; Dube and Mutanga, 2015; Kim et al., 2013; Lu et al., 2012; Gleason and Im, 2011; Lu, 2006; Cohen and Goward, 2004; Fazakas et al., 1999). Modelos derivados de satélite têm sido amplamente usados na engenharia florestal e em aplicações ecológicas, principalmente a partir do início do acesso livre às imagens Landsat (Cohen and Goward, 2004).

Dada a importância de informar a distribuição espacial de recursos florestais em escalas mais finas que aquela fornecida pelos dados de parcelas do IFN, neste estudo são aplicados dois métodos derivados de dados de satélite para quantificar e mapear volume de madeira, biomassa e

carbono. O primeiro deles é um modelo linear múltiplo que usa dados de reflectância da superfície como variáveis de entrada. O segundo método é baseado em Krigagem com Regressão (KR), em que sua aplicação foi combinada com um modelo linear que usa latitude e longitude dos *pixels* como variáveis de entrada. Esses métodos de escala fina foram comparados entre si, bem como comparados em relação à estimativa a uma escala ampla, fornecida pelos dados de parcelas do IFN.

Os objetivos gerais foram (i) mapear e (ii) desagregar variáveis florestais (volume de madeira, biomassa e carbono) de uma escala ampla para mais fina, usando dados de parcelas do IFN. Como objetivos específicos, tem-se: (i) estimar volume de madeira (por unidade de área) a uma escala ampla, fornecida pelos dados de parcelas de campo do IFN; (ii) estimar volume de madeira (por unidade de área) a uma escala ampla, com assistência de classificação de imagem; (iii) testar e comparar dois métodos de escala fina para estimar e mapear volume de madeira, empregado com assistência de dados de sensoriamento remoto e interpretação de imagem; (iv) comparar volume de madeira estimado tanto a escalas fina e ampla; e (v) testar um método de escala fina para estimar e mapear biomassa e carbono, usando dados de sensoriamento remoto e interpretação de imagem.

2. Material e Métodos

A área de estudo cobre três mesorregiões políticas do estado do Paraná: Centro-Ocidental, Centro-Sul e Sudeste (Fig. 5). Foram usados dados do IFN executado nessas três mesorregiões. A área de estudo cobre ~55.331 km² ou ~28% do estado do Paraná. Essa foi a primeira região no estado a ser inventariada seguindo o desenho amostral padronizado do IFN. Dados de campo foram coletados durante o primeiro semestre de 2013, em que 152 conglomerados foram alocados sistematicamente sobre essas três mesorregiões. A Fig. 5 mostra a área de estudo e a localização de cada conglomerado do IFN.

A primeira variável de interesse desta pesquisa é volume de madeira ($m^3\ ha^{-1}$), resultante da soma de volume de fuste de todas espécies arbóreas observadas nos conglomerados parcialmente e inteiramente florestados. O volume de madeira foi modelado ao nível de árvore, estimado por conglomerado e, então, extrapolado às áreas não-amostradas a escalas ampla e fina. Estimativas em escalas finas, isto é, ao nível de pixel, dependem da aplicação de (i) modelagem por regressão linear em dados de sensoriamento remoto, e (ii) krigagem com regressão. Esses

métodos de escala fina foram comparados com duas alternativas de extrapolação a uma escala ampla; com e sem assistência de classificação de imagens de satélite, para delimitar área florestada.

A segunda variável de interesse é biomassa acima do solo (BAS, Mg ha^{-1}), resultante da soma da biomassa aérea dos compartimentos (fuste, galhos e folhas) de todas as espécies lenhosas observadas nos conglomerados florestados. A terceira variável de interesse é o estoque de carbono (Mg ha^{-1}), diretamente estimado por um fator de conversão de biomassa.

O modelo de regressão linear foi chamado de Modelo de Reflectância da Superfície (MRS), visto que a variável florestal foi estimada em função da reflectância de imagens Landsat cobrindo a área de estudo. O método de KR é auxiliado por um modelo chamado de Modelo de Krigagem com Regressão (MKR), o qual estima a variável florestal em função de latitude e longitude.

As imagens Landsat foram classificadas com base em um algoritmo de classificação orientada ao objeto, em que o *software* eCognition foi usado nessa etapa tanto para segmentar, quanto para classificar as imagens. Um total de sete imagens foram suficiente para cobrir toda a área de estudo.

3. Resultados e Discussão

O método com MRS teve um desempenho melhor que o método com KR, embora ambos forneceram, no geral, resultados próximos. O MRS estimou um volume médio de 120,1 m^3 ha^{-1} e um total para a área de estudo de 210.961.589 m^3. A KR estimou média de 115,7 m^3 ha^{-1} e total de 203.326.674 m^3, ou seja, 3,7% a menos que as estimativas do MRS. Em relação à biomassa florestal, o MRS estimou média e total de 120,1 Mg ha^{-1} e 217.736.862 Mg, respectivamente. Para o carbono, a média foi de 49,2 MgC ha^{-1} e total de 89.272.113 MgC, correspondendo a 327.331.082 Mg de CO_2 equivalente estocado na área de estudo. O volume total estimado por ambos métodos de escala fina enquadra-se no intervalo de confiança obtido a uma escala ampla.

Os resultados revelaram que o método KR foi impossibilitado de estimar volume de madeira dos *pixels* com menores e maiores estoques, precisamente aqueles *pixels* com estoque menor que ~50 m^3 ha^{-1} e maior que 210 m^3 ha^{-1}. Essa limitação tinha sido notada para o modelo de KR, em que o volume de madeira variou de ~50 a 150 m^3 ha^{-1}, enquanto que os volumes observados variaram de ~23 a 223 m^3 ha^{-1}.

Esse fato indica que o procedimento analítico da krigagem com regressão (que adiciona rasters dos resíduos do volume ao volume estimado) foi limitado em corrigir as estimativas do volume de madeira em *pixels* com maiores e menores estoques. Apesar de sua pior performance, o RMSE fornecido pela KR enquadra-se nos resultados de pesquisas combinando dados de sensoriamento remoto e métodos geoestatísticos (Yadav and Nandy, 2015; Scolforo et al., 2015; Scolforo et al., 2016).

Uma das principais razões pela qual a KR tem sido menos confiável que o MRS diz respeito às variáveis de entrada. No MKR testado, o volume de madeira foi limitado à variação de latitude e longitude. Quando uma dessas variáveis não é significante, o volume torna-se então dependente apenas de uma variável geográfica, como ocorreu no caso desta pesquisa (apenas latitude foi significante a 95% de probabilidade). Portanto, o MKR ajustado neste estudo pode atingir volumes extremos apenas sob latitudes extremas, mas esse último é limitado ao tamanho da área. Um outro empecilho subjacente ao MKR é que recursos florestais podem não ser estocados em um gradiente contínuo em função de lat./long., o que prejudicaria a performance dos modelos baseados em localizações geográficas. Os MRSs, por sua vez, foram mais eficientes possivelmente porque eles estimam volume em função da resposta espectral da vegetação, independentemente da localização geográfica dos *pixels* (como nos MKR), sendo capazes de estimar valores inferiores e superiores.

Em relação aos MRS, a grande variação do volume observado ao nível de campo pode ser citada como uma das principais causas da perda de acuracidade. Na base de dados deste estudo, foram encontrados muitos conglomerados inteiramente florestados (isto é, com 40 parcelas classificados em uma classe de floresta) com baixo estoque de madeira, ou mesmo sem estoque algum. Isso ocorre porque o critério de inclusão do IFN (de Dap $\geq$ 10 cm) pode excluir, por exemplo, a maioria das árvores pertencentes a uma floresta em estágio de sucessão inicial. Se uma vegetação jovem como essa é comparada a uma floresta madura, elas são prováveis de ter menos discrepância espectral que de estoque de volume, dada suas similaridades de verde e umidade.

Em segundo lugar, as subunidades dos conglomerados são pequenas demais para serem consideradas como parcelas únicas, devido à resolução espacial de 30 m. Tal condição nos obriga a expandir a coleta de *pixels* pela área total coberta pelo conglomerado, isto é, 4 ha. Nesse caso, a área total é a área formada pelas distâncias entre as subunidades 1 a 3, e 2 a 4 (que é de 200 m, cada). Apesar dessa área total ser um tamanho de amostra razoavelmente suficiente, ela contempla

pixels tanto de manchas florestais amostradas quanto não-amostradas, esperando que elas tenham estoques semelhantes.

Seguindo essa necessária estratégia de expansão, uma possível fonte de erro diz respeito à incompatibilidade entre *pixels* de manchas amostradas e não-amostradas.

4. Conclusões e Recomendações

Os dois métodos de escala fina testados forneceram estimativas próximas para o volume total (diferença de ~8%), porém, o MRS teve a melhor capacidade em mapear a distribuição espacial do volume de madeira. O MRS também forneceu mapeamentos dos estoques de biomassa e carbono acurados. A baixa performance do método KR em mapear espacialmente as variáveis estudadas é devido ao fato de as variáveis geográficas serem limitadas à área de estudo.

Estimativas do volume total derivado do MRS e KR enquadram-se no intervalo de confiança obtido a uma escala ampla, indicando que a intensidade amostral do IFN produz confiável estimativa do volume de madeira. Mesmo com uma ampla intensidade amostral de 1:100.000, os resultados revelam que ambos os métodos para estimar volume de madeira podem ser aplicados aos dados do IFN brasileiro. A extrapolação em escala ampla com auxílio de classificação de imagem subestima o volume total em relação ao método sem tal auxílio. Isso ocorre devido à menor área florestada encontrada no método com classificação de imagem.

Como recomendações, deve-se considerar que as subunidades dos conglomerados são pequenas demais (para imagens de 30m de resolução espacial) para compor os pares de dados com os valores de reflectância dos *pixels*, visto que os *pixels* das imagens são de 900 m². Como alternativa, recomenda-se fusionar as imagens para reescalar os *pixels* para 15 m. Essa alternativa pode melhorar a precisão dos MRS.

O método KR tem um grande potencial em ser usado como ferramenta para estimativas a nível de *pixel* no IFN. Outros estudos devem ser feitos testando outras variáveis além de latitude e longitude, como variáveis edáficas e climáticas.

Recomenda-se o uso do *software* eCognition e a classificação orientada ao objeto para classificar imagens Landsat. Para atingir uma boa segmentação e classificação, o usuário deve prestar bastante atenção na escolha de parâmetros de compactação e forma dos objetos, os quais devem formar objetos com forma e tamanho adequados a cada estudo.

LIST OF ABBREVIATIONS

AGB: Aboveground Biomass

AGC: Aboveground Carbon

ALS: Airborne Laser Scanner

AME: Absolute Mean Error

CU: Conservation Unit

FLAASH: Fast Line-of-sight Atmospheric Analysis of Hypercubes

GIS: Geographic Information System

ME: Mean Error

NDVI: Normalized Difference Vegetation Index

NFI: National Forest Inventory

NIR: Near Infra-Red

OLI: Operational Land Imager

PC: Principal Component

PCA: Principal Component Analysis

PDF: Probability Density Function

PPS: Permanent-plot System

RK: Regression-kriging

RKM: Regression-kriging model

RMSE: Root Mean Square Error

SEAB: Secretary of Agriculture and Supplying

SRM: Surface Reflectance Model

SS: Systematic sampling

SV: Semivariogram

SWIR: Short-wave Infra-Red

USGS: United States Geological Survey

LIST OF FIGURES

LIST OF TABLES

SUMMARY

1. INTRODUCTION

Recently, Brazil started its National Forest Inventory (NFI) following a broad-scale sampling suited to meet national demands. Brazil is a continental country that shelters the world's richest tropical forests, but contradictorily, few of its forest resources is known, mainly due to the lack of a national-level forest plot database. The role of the Brazilian NFI is of special importance for supplying a permanent-plot system (PPS), with remeasurements of sampled plots every five years. The PPS is then the master key by which Brazil can produce information about its forest resource stocks at a given strategic level.

Chronologically, the Federal District and the states of the South were the first ones to be inventoried following the NFI's standardized methodology. Among the States of the South, Parana carried out its NFI in three sequential phases: the first of them was conducted during 2013 and composes the data source of this study. The second and third phases occurred over the 2014-2016 period. States from other Brazilian regions have also started and completed their first measurement, but only the State of Santa Catarina had remeasurements concluded.

As Brazil's NFI is on a primary stage, national- and regional-level mappings of wood volume, biomass and carbon stocks will appear as some of the first emerging demands. In a longer perspective, future challenges will arise for disentangling forest changes detected with remeasurements (Goeking, 2015), fixing problems caused by lost samples, among others. In the initial phases, the role of the methods for quantifying forest resources at more precise scales is crucial for NFI data processing. This requires a disaggregation of forest variables from coarse to finer scales, mainly to support forest management and planning at regional level (Wilson et al., 2013). The concept of disaggregation here applied concerns the refinement of broad-scale estimates derived from NFI data, which can be achieved through remote sensing-based techniques. With the refinement of estimates, forest variables are provided at a scale as fine as the spatial resolution from the remote sensor employed. In general, quantifying wood volume at refined scales is important for planning and management purposes, while quantifying aboveground biomass (AGB) and carbon plays an important role to support greenhouse-gas emission reports, as well as to inform carbon sequestration and emissions (David et al., 2017; Kim et al., 2013; Wilson et al., 2013).

Most of the refined techniques for quantifying forest resources are based on modeling and remotely sensed data, being applied both at national, regional, as local levels (e.g., Brooks et al.,

2016; Maack et al., 2016; Dube and Mutanga, 2015; Kim et al., 2013; Lu et al., 2012; Gleason and Im, 2011; Lu, 2006; Cohen and Goward, 2004; Fazakas et al., 1999). Satellite-derived models have been widely used in forestry and ecological applications, mainly from the onset of the free access to Landsat imagery (Cohen and Goward, 2004). Many authors have found strong relations between forest biomass and carbon with visible bands, as well as wood volume with middle infrared bands (Lu et al., 2004). Other variables (e.g., forest crown closure, stand height) have also been accurately modeled with assistance of remote sensing data (Lu et al., 2004). These satisfactory results stem from the good relation between forest variables and spectral information contained in satellite imageries, though spectral responses may considerably change in function of the studied variable.

Hayashi et al. (2015) point out that parametric methods can extrapolate beyond the range of the fitting dataset, however, regression models have limitations on relating field data with remotely sensed data, due to its high sensitivity to influential observations, mathematical relationships existing among sensor data, as well as difficulty in handling several highly-correlated variables. In this scope, Kim et al. (2013) highlight that suitable satellite imagery data used in the regression analysis must be carefully selected, given the variations in forest stand conditions and the complex relationships between spectral reflectance values. Such conditions require cautions on handling the regressed variables, examination of normality of residuals, as well as co-linearity among input variables (Hayashi et al., 2015).

In the forestry field, authors have employed from less to more complex analytical techniques combined with remote sensing (Kim et al., 2013). Among the less complex ones, linear regression models perhaps are the most used (Fassnacht et al., 2014) because they provide accurate results, besides being rapidly processed (Lu et al., 2004). On the other hand, applying techniques as artificial neural network (ANN) in spatializations and mappings, although they often provide a little more of accuracy, they depend on complex programming practices. In general, authors that use such techniques neither expose their underlying algorithms, nor apply them to map the studied variable. In addition to ANN, geostatistical techniques have been increasingly more used, such as ordinary kriging, cokriging, and regression kriging (RK); this last being a hybrid method involving linear model with kriging (Meng et al., 2009) and which often has been considered the most accurate alternative (Kim et al., 2013).

Given the importance of informing the spatial distribution of forest resources at a scale finer than the one provided by NFI plot data, the question raised in this study was whether there is

difference in the forest resource quantification through satellite data-derived methods from different natures. The first of them uses satellite image interpretation and raw NFI data, i.e., without modeling the studied forest variables. Unlike, the second one employs a linear multiple model, the 'SRM', which uses surface reflectance data as input variables. The second method is based on RK, in which its application was combined with a linear model that uses latitude and longitude of pixels as input variables. These methods were compared with each other, as well as compared in relation to the coarse-scale estimation, provided by the NFI plot data. The hypotheses of this study are: i) the fine-scale methods, even being from different natures, produce close estimations of the total stock, because their spatial resolution is the same; ii) fine- and coarse-scale methods provide statistically unequal results of the total stock, because their spatial resolution is not the same.

1.1 OBJECTIVES

1.1.1 General objective

The general objectives are to (i) map and (ii) disaggregate forest variables (wood volume, biomass and carbon) from a coarse to a finer scale, using NFI plot data.

1.1.2 Specific objectives

- Estimating wood volume at a coarse scale, provided by data from NFI field plots.
- Estimating wood volume at a coarse scale supported by image classification.
- Testing and comparing two fine-scale methods for estimating and mapping wood volume, employed with assistance of remotely sensed data and image interpretation.
- Comparing wood volume estimated both at fine and coarse scales.
- Testing a fine-scale method of estimating and mapping forest biomass and carbon, by using remotely sensed data and image interpretation.

2. REVIEW OF LITERATURE

2.1 NATIONAL FOREST INVENTORY (NFI)

2.1.1 Concept and role of NFIs for forest resource assessment

National Forest Inventory (NFI) is a sample-based system that provide a statistically robust information for reporting, capturing of spatial change across landscapes, being of high value for forest management, planning, and scientific purposes (Frazier et al., 2014). NFIs are carried out in several countries around the world aiming to obtain information about forest resource at national scales. The source of data of the NFIs come from ground plots allocated in forested or non-forested lands. NFI plots are generally sampled following a national grid designed for meeting specific demands of the country. Although several countries have already established their NFI program, Tomppo et al. (2010) remark that there is a lack of information of forests at a global scale. From NFI plot data, it is possible to map and quantify forest variables, as wood volume, biomass and carbon.

Given their coverage and complexity on the field collection, NFIs play an important role for accurately informing forest resource stocks at a broad scale, which is useful in strategic planning of governmental activities related to natural forests. This is of special importance for countries that aim to control the use, management, and conservation of its forest resources. NFIs are also important to those countries that have committed to climate accords, mainly the countries that are Parties of the United Nations Framework Climate Change Conference (UNFCCC). Climate agreements often require that nations inform their greenhouse gas (GHG) emissions and estimates of their national forest resources (Deo et al., 2017; McRoberts et al., 2009; Cienciala et al., 2008), which can be informed by means of processing of NFI data.

In the past, before the onset of the UNFCCC in 1994, other purposes rather than climate change and GHG reports have driven countries to conduct forest inventories. Tomppo et al. (2010) cite that the early NFIs in the Nordic countries usually included information about areas, wood volume and increment of the growing stock, as well as structure of forests, silvicultural status, and accomplished and needed cutting and silvicultural regimes. Recently, Vidal et al. (2016) explain

that the latest NFIs have been focused on sustainable ecosystem management, a different vision existent in earlier NFIs, which were more focused on wood production.

Each country has different demands for conducting its NFI, despite certain types of information are common among some countries (Tomppo et al., 2010). NFIs have historical, commercial, and different environmental justification among the countries, meaning that the prospects for standardizing inventories are minimal (McRoberts et al., 2009). Because of this, there is an emerging trend towards standardization of NFI methodologies at least from a same continent, in order to facilitate providing international information (McRoberts et al., 2009).

To meet these needs, Vidal et al. (2016) cite that the data provision mechanisms must be based on a robust and long-term statistical information system. In this sense, following a harmonized and standardized methodology, NFIs could play an important role for international reporting processes, as well as for enhancement of the comparability of national data to discuss forest policies (Vidal et al., 2016).

2.1.2 Approximate standard error of a systematic sampling with clusters

The main books on sampling techniques mostly are written by statisticians and because of this, examples in forestry are unlikely to be addressed on them. Often, NFIs follow sampling structures with features not found in the literature, what ends up implying that methodologies for obtaining statistical estimators (variance, standard error etc.) be adapted to specific cases of the NFIs. Some of the main books on sampling techniques are: Chacko (1965), Cochran (1977), and Sukhatme (1954). On forest inventory, it is highlighted Loetsch et al. (1973), Queiroz (2010), and Péllico Netto & Brena (1997), these two last mostly adopted in Brazil.

Systematic samplings (SS) are commonly preferred in NFIs, mostly adopting clusters as sample units. SS with clusters is a typical two-stage sampling, in which, in the Brazilian NFI, the first and the second stages are systematically designed. A chapter about (equal-size) cluster sampling is addressed in Péllico Netto & Brena (1997), however, the estimators there presented is suited to cases which the first stage is random, unlike the Brazilian NFI.

In the next topics, methods for approximating to variance and standard error were adapted to the Brazilian NFI. The standard error was obtained in order to calculate confidence interval of wood volume ($m^3\ ha^{-1}$). The methods were statistically adapted with assistance from Professor Dr.

Sylvio Péllico Netto (UFPR, Brazil). See highlighted publications about forest inventory and sampling in Péllico Netto et al. (2017), Péllico Netto et al. (2014), Péllico Netto and Brena (1997). The alternative methods of approximation to the standard error for SS with clusters are presented below:

a) Ratio Estimator for SS with clusters (adapted from Cochran, 1977)

This proposed method is an adaptation from Ratio Estimator presented by Cochran (1977, p. 150-157) for simple random sampling. Applying it to SSs requires the assumption that systematic designs are particular cases of the random one. Such method is applied in the State of Santa Catarina (Vibrans et al., 2013) and in the Canadian NFI (Canada's NFI, 2004). Once such method is usually applied, App. 1 shows the results of standard error, mean and total volume from database used in this study, though it was not the chosen method. To begin explaining this method, consider that the Brazilian NFI counts on both forested and non-forested clusters. Firstly let $X_i = 0$ for non-forested clusters, and $X_i = 1$ for forested clusters, assuming that forested clusters may have *p* 100-m² forested plots, with *p* ranging from 1 to 39, for those partially forested ones, or 40 plots, for those entirely forested ones (Fig. 6). The probability (p_i) of a cluster having been allocated within forest is given by [1].

$$p_i = \left(\sum_{j=1}^{p} a_j\right) A_c^{-1} \qquad [1]$$

Where,

a_j: area of the j-th forested plot, i.e., 100 m²; A_c: area of the cluster, i.e., 4,000 m². p: number of forested plots.

Secondly, consider that among *L* sampled clusters, *k* are forested and *L* – *k* are non-forested clusters. Letting A_g to be the area represented by each *k* cluster, i.e., ~40,000 ha, the mathematical expectation of forested area along the sampled area is given by [2].

$$E\left(\sum_{i=1}^{L} X_i A_g\right) = \sum_{i=1}^{L-k}\left[(X_i = 0) p_i A_g\right] + \sum_{i=1}^{k}\left[(X_i = 1) p_i A_g\right] = \sum_{i=1}^{k} p_i A_g =$$

$$\sum_{i=1}^{k}\left(\sum_{i=1}^{p} a_j\right) A_c^{-1} A_g$$

$$E\left(\sum_{i=1}^{L} X_i A_g\right) = \sum_{i=1}^{k}\left(\sum_{i=1}^{p} a_j\right) A_c^{-1} A_g \quad [2]$$

There are two ways to apply the method of Ratio Estimator: (i) considering forest volumes being proportional to the area of the clusters' sub-units, that is 1,000 m²; and (ii) considering forest volumes being proportional to the square area of the clusters' sub-units (Cochran, 1966). Loetsch et al. (1973) clarify theoretically such problem showing that $\sigma_y^2 = 2\sigma_x^2(1+p)$, where σ_y^2 is the variance of a sub-unit with double size in relation to another used for obtaining σ_x^2, and p is the correlation between contiguous sub-units of equal size. Considering the case (i) for the Brazilian NFI case, i.e., adopting volumes proportional to the area of the sub-units, σ_y^2 can be transformed into coefficient of variation (CV) as described in [3]:

$$CV_2^2 = 2\sigma_x^2(1+\rho)(4\bar{X}^2)^{-1} \text{ or simply}$$
$$CV_2 = CV_1[(1+\rho)/2]^{\frac{1}{2}} \quad [3]$$

Where,

CV_i: coefficient of variation of the volume taken in the ith occasion; ρ: correlation between contiguous sub-units of equal size. $\bar{X}$: mean of the volume.

As its name suggests, the method of Ratio Estimator takes into account ratios between volume and area, which in the NFI case, such volume and area are from sub-units from the partially and entirely forested clusters. Assuming that Brazil's NFI counts on *k* forested clusters of *M* sub-units (where *M* = 4) each one, the total number of sub-units is *N* = *kM*. Also, assuming that X_{ij} is the wood volume and Y_{ij} is the forested area from the N^{th} sub-unit, the average ratio is then defined by [4]:

$$\bar{r} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{M} X_{ji}}{\sum_{j=1}^{k}\sum_{i=1}^{M} Y_{ji}} = \frac{\bar{x}}{\bar{y}} \quad [4]$$

The variance of the average ratio ($\bar{r}$) is:

$$s_{\bar{r}}^2 = \frac{\bar{r}^2}{k-1}\left[\frac{\sum_{j=1}^{k}\left(\sum_{i=1}^{M} X_{ji}\right)^2}{\bar{x}^2} + \frac{\sum_{j=1}^{k}\left(\sum_{i=1}^{M} Y_{ji}\right)^2}{\bar{y}^2} - 2\frac{\sum_{j=1}^{k}\left(\sum_{i=1}^{M} X_{ji}\right)\left(\sum_{i=1}^{M} Y_{ji}\right)}{\bar{x}\,\bar{y}}\right] \quad [5]$$

The coefficient of variation of the average ratio ($\bar{r}$) is:

$$CV = \frac{s_{\bar{r}}}{\bar{r}} \quad [6]$$

The variance of the mean of the average ratio ($\bar{r}$) is:

$$s_{\bar{r}}^2 = \frac{s_{\bar{r}}^2}{n} \quad \text{for an infinite population} \quad [7]$$

Finally, the standard error of the average ratio ($\bar{r}$) is given by:

$$s_{\bar{r}} = \sqrt{\frac{s_{\bar{r}}^2}{n}} \quad \text{for an infinite population} \quad [8]$$

b) Method of the First Differences for SS with clusters (adapted from Chacko, 1965)

The Method of the First Differences (MFD) was originally proposed for single units (Chacko, 1965); here, an adaptation of the MFD to clusters is presented. To begin explaining how to obtain the sampling error, keep assuming that Brazil's NFI is structured with *k* clusters of *M* sub-units (where *M* = 4), so that the total number of sub-units is *N* = *kM*. The following mathematical formulations consider X_{ij} associated to each sub-unit, such that *X* = variable of interest [in the present case, wood volume ($m^3\ ha^{-1}$)], *i* = 1, 2, … *M* and *j* = 1, 2, ... *k*. The mean for the entire population is then given by:

$$\bar{X} = N^{-1}\sum_j^k \sum_i^M X_{ij}$$

The sampling error in the MFD is calculated with the sums of the differences between pairs of successive units (i.e., cluster's sub-units). Considering that we have *M* = 4 units per cluster, each cluster will count on 4 pairs: (i) first sub-unit is paired with the second one, (ii) second with the third, (iii) third with fourth, and (iv) the fourth sub-unit paired with the first one. This last pair is

included because the first and last sub-units are arranged side-by-side as the other pairs. Note that such arrangement is only possible between neighbor sub-units, and not between side-by-side plots.

Fig. 1 gives an example of the arrangement of paired sub-units in clusters, which is necessary for computing sampling error in the Brazilian NFI, following the Method of the First Difference.

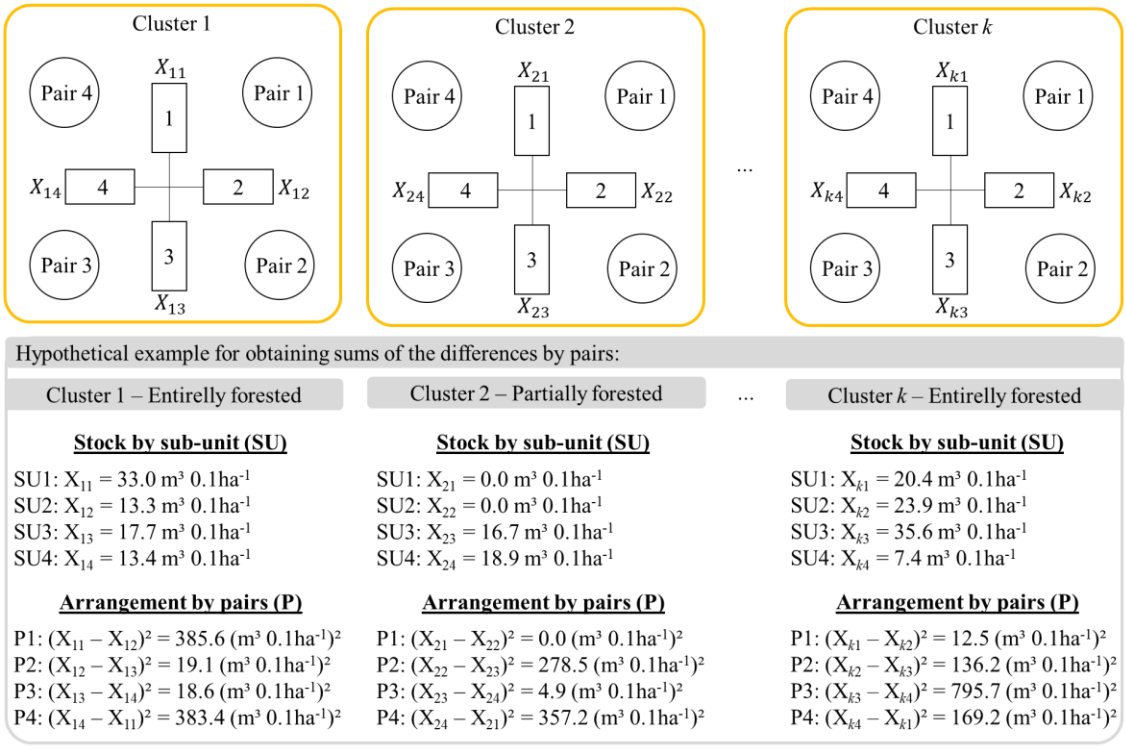


FIG. 1. STEPS FOR OBTAINING SUMS OF THE DIFFERENCES BY PAIRS, NECESSARY FOR COMPUTING SAMPLING ERROR IN THE BRAZILIAN NFI, FOLLOWING THE METHOD OF THE FIRST DIFFERENCE.

To calculate the variance, first it is necessary to obtain the difference between the wood volume of a sub-unit and of its paired sub-unit (Fig. 1). Let d_{ji} to be the difference of the volumes observed in paired sub-units, i.e., $X_{ji} - X_{j(i+1)}$. Hence, the variance is given by the sum of squares of the differences divided by twice the number of sub-units, since each sub-unit repeats twice. The variance of X is given by:

$$S_X^2 \cong \frac{\sum_{j=1}^{k}\sum_{i=1}^{M}\left[X_{ji} - X_{j(i+1)}\right]^2}{2kM} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{M}\left[X_{ji} - X_{j(i+1)}\right]^2}{2N} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{M}\left[d_{ji}\right]^2}{2N}$$

The variance of the mean ($\bar{X}$) is then obtained dividing S_X^2 by N, as follows:

$$S_{\bar{X}}^2 \cong \frac{\sum_{j=1}^{k}\sum_{i=1}^{M}\left[d_{ji}\right]^2}{N(2N)} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{M}\left[d_{ji}\right]^2}{2N^2}$$

The standard error is given by:

$$S_{\bar{X}} \cong \sqrt{\frac{\sum_{j=1}^{k}\sum_{i=1}^{M}\left[d_{ji}\right]^2}{2N^2}}$$

c) Method of the Second Differences for SS with clusters (adapted from Chacko, 1965)

As in the MFD, the Method of the Second Differences (MSD) was originally proposed for single units (Chacko, 1965) and an adaption to SS with clusters was necessary. As in the MFD, consider d_{ji} being the difference of the volumes observed in paired sub-units ($X_{ji} - X_{j(i+1)}$). The difference between MFD and MSD is that, in the MSD, it is considered the second difference $\Delta_{ji} = d_{ji} - d_{j(i+1)}$, instead of simply d_{ji}, which is the first difference of the volumes observed in the paired sub-units.

Therefore, the idea of paired sub-units still remains, but now involving two paired sub-units at once, totaling three differences (Δ_{ji}) by cluster: (i) difference between $d_{j1} - d_{j2}$; (ii) difference between $d_{j2} - d_{j3}$; and (iii) difference between $d_{j3} - d_{j4}$, where: $d_{j1} = X_{j1} - X_{j2}$; $d_{j2} = X_{j2} - X_{j3}$; $d_{j3} = X_{j3} - X_{j4}$; and $d_{j4} = X_{j4} - X_{j1}$, for the j^{th} cluster.

The variance of X is given by:

$$S_X^2 \cong \frac{\sum_{j=1}^{k}\sum_{i=1}^{M}\left[d_{ji} - d_{j(i+1)}\right]^2}{2k(M-1)} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{M}\left[d_{ji} - d_{j(i+1)}\right]^2}{2(N-k)} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{M}\left[\Delta_{ji}\right]^2}{2(N-k)}$$

The variance of the mean ($\bar{X}$) is obtained by:

$$S_{\bar{X}}^2 \cong \frac{\sum_{j=1}^{k}\sum_{i=1}^{M}\left[\Delta_{ji}\right]^2}{2N(N-k)} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{M}\left[\Delta_{ji}\right]^2}{2(N^2-Nk)}$$

The standard error is given by:

$$S_{\bar{X}} \cong \sqrt{S_{\bar{X}}^2}$$

2.2 SATELLITE IMAGE CLASSIFICATION

Image classification is a technique that produces thematic classes from multi-band images. According to Tempfli et al. (2009), the principle of image classification is that an image pixel (or object) is assigned to a class based on its feature vector, by comparing it to predefined clusters or classes in the feature space. A proper definition of classes is one of the main factors responsible for obtaining a well accurate image classification. The definition of classes is an interactive process and is carried out during the 'training process', whereas the comparison of the individual pixels (or objects) with the classes is performed through 'classification algorithms'. Tempfli et al. explain that the image classification process typically involves five steps:

1) *Selection and preparation of the images*. Satellite sensors, spectral bands, and date of acquisition should be carefully selected, according to the land cover types to be classified.
2) *Definition of the clusters (classes) in the feature space*. Two approaches are possible: supervised classification (SC) and unsupervised classification (USC). In the SC, the user defines the clusters during the training process, i.e., process of sample selection. Selected samples are also known as training data. In the USC, in turn, an algorithm automatically defines the clusters and the user must choose the number of clusters. Therefore, selecting data for training is not required in the USC.
3) *Selection of classification algorithm*. The choice of the algorithm depends on the purpose of the classification, on the characteristics of the image, as well as on the training data. It is cited as the main SC algorithms: *Minimum Distance*; *Mahalanobis Distance; Maximum Likelihood*; *Neural Net-based* and *Fuzzy logic-based* algorithms. It is cited as the main USC algorithms: IsoData; and *k-means*.
4) *Execution of the classification*. To execute the image classification, besides the three steps above cited, the user must choose classification criteria that are used by the algorithms for finding similarities among image pixels/objects and the predefined clusters.

5) *Validation of the result*. Once the classification is performed, its accuracy should be assessed by comparing it to reference data, i.e., ground truth.

In the step of data training, either samples of pixel or samples of objects can be used, from which two methods are derived: (i) pixel-oriented classification (ii) and object-oriented classification. Thus, the definition of clusters (step 3 described above) may be composed by samples of image pixels or objects, depending on the chosen classification method.

2.2.1 Pixel-oriented classification

Pixel-based classification may be executed both through supervised as unsupervised classifiers. According to Meneses and Sano (2012), most of the classifiers have implemented pixel-oriented classification algorithms, from which only the spectral information of pixels is used to find homogeneous regions to assign pixels into clusters (classes). Regionalization of homogenous regions is performed through distance measures or odds of pixels belonging to a certain class. In this method, the image pixels are assigned to a class that better satisfies the determined criteria.

2.2.2 Object-oriented classification

Object-oriented classification was the method used in this study, with assistance from the software eCognition Developer 8.7. To execute the object-oriented classification, the process of image segmentation is necessary to group pixels into objects as homogenous as possible. According to Brites et al. (2012), the segmentation can be performed both manually as automatically. The manual procedure depends on delimitation of polygons that represent homogeneous regions along the image, therefore such alternative is often unfeasible for large areas. The automatic procedure, in turn, is widely used and depends on a (i) segmentation algorithm, and (ii) quantitative data, such as spectral bands, parameters of scale and shape of objects, among others.

In this study, the algorithm ‘classification’ available in eCognition Developer 8.7 was used. This algorithm evaluates the membership value of an image object against a list of selected classes. The classification result of the image object is updated according to the class evaluation result. The three best classes are stored in the image object classification result. Classes without a class

description are assumed to have a membership value of 1. The software also provides several segmentation algorithms and executes fuzzy logic-based supervised classification. Among the algorithms available in the software, the ‘multiresolution segmentation’ algorithm was used, which is described in further detail as follows.

The multiresolution segmentation is an optimization procedure that, for a given number of image objects, minimizes the average heterogeneity, maximizing thus their average homogeneity. In eCognition Developer 8.7, the segmentation procedure works following the six rules summarized below:

1) The segmentation procedure starts with single image objects of one pixel and repeatedly merges them in several loops in pairs to larger units as long as an upper threshold of homogeneity is not exceeded locally. This homogeneity criterion is defined as a combination of spectral and shape homogeneity. Higher values for the scale parameter result in larger image objects, smaller values in smaller image objects.
2) The seed looks for its best-fitting neighbor for a potential merger.
3) If best fitting is not mutual, the best candidate image object becomes the new seed image object and finds its best fitting partner.
4) When best fitting is mutual, image objects are merged.
5) In each loop, every image object in the image object level will be handled once.
6) The loops continue until no further merger is possible.

The software eCognition Developer 8.7 allows users to set algorithm parameters that have to be carefully analyzed. Fig. 2 presents window displaying parameters available for the ‘multiresolution segmentation’ algorithm. The algorithm parameters available in eCognition Developer 8.7 (Fig. 2) help to improve the segmentation depending on user’s specific aims, satellite bands, and targets to be classified. Each parameter highlighted with balloons in Fig. 2 is briefly described below. Torres-Sánchez et al. (2015) assess how the values in the parameters of the software eCognition affect segmentation results.

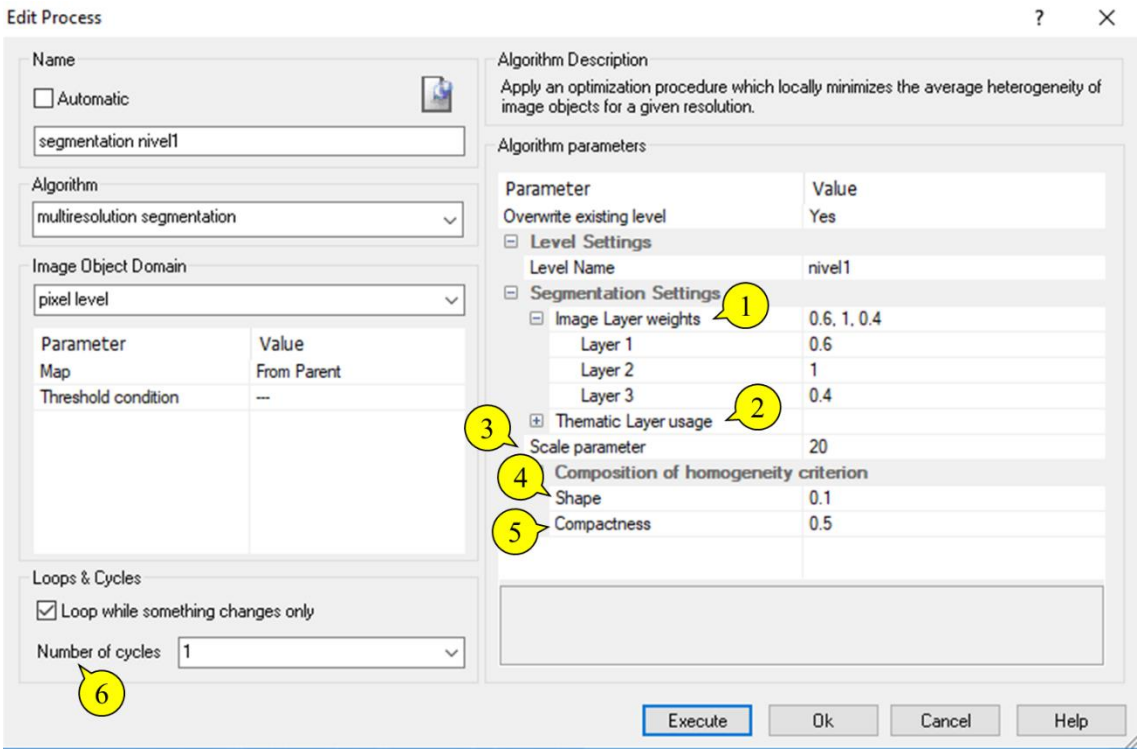


FIG. 2. ECOGNITION DEVELOPER 8.7 WINDOW DISPLAYING SEGMENTATION PARAMETERS, HIGHLIGHTING WEIGHTS FOR IMAGE LAYERS (BALLOON 1), THEMATIC LAYER USAGE (2), PARAMETERS OF SCALE (3), SHAPE (4) AND COMPACTNESS (5), AS WELL AS NUMBER OF CYCLES OF SEGMENTATION (6).

i. Segmentation Settings

- Image Layer Weights (Fig. 2, Balloon 1): Weighting values are assigned obeying the contribution of image layers. The higher the weight assigned to an image layer, the more weight will be given to that layer's pixel information during the segmentation process. Users can use a variable as a layer weight.

- Thematic Layer Usage (Fig. 2, Balloon 2): Users have to specify the thematic layers to be candidates for segmentation. Each thematic layer that is used for segmentation will lead to additional splitting of image objects while enabling consistent access to its thematic information. Users also can segment an image using more than one thematic layer. The results are image objects representing proper intersections between the thematic layers.

- Scale Parameter (Fig. 2, Balloon 3): The parameter is an abstract term that determines the maximum allowed heterogeneity for the resulting image objects. For heterogeneous data, the resulting objects for a given scale parameter will be smaller than in more homogeneous data. By modifying the value in the Scale Parameter value, users can vary the size of image objects.

ii. Composition of Homogeneity Criterion

- Shape (Fig. 2, Balloon 4): The value of the Shape field modifies the relationship between shape and color criteria. By modifying the Shape criterion, users can define the color criteria. In effect, by decreasing the value assigned to the Shape field, users can define to which percentage the spectral values of the image layers will contribute to the entire homogeneity criterion. This is weighted against the percentage of the shape homogeneity, which is defined in the Shape field. Changing the weight for the Shape criterion to 1 will result in objects more optimized for spatial homogeneity. However, the shape criterion cannot have a value larger than 0.9, due to the fact that without the spectral information of the image, the resulting objects would not be related to the spectral information at all. The slider bar adjusts the amount of Color and Shape to be used for the segmentation. In addition to spectral information, the object homogeneity is optimized with regard to the object shape, defined by the Compactness parameter.

- Compactness (Fig. 2, Balloon 5): The compactness criterion is used to optimize image objects with regard to compactness. This criterion should be used when different image objects which are rather compact but are separated from non-compact objects only by a relatively weak spectral contrast. Users can use the slider bar to adjust the degree of compactness to be used for the segmentation.

iii. Loops and cycles (Fig. 2, Balloon 6)

This option allows that users choose the number of cycles in the segmentation process. Leaving the algorithm to loop while no further object is segmented also is possible.

2.3 IMAGE PROCESSING FOR VEGETATION ASSESSMENT

2.3.1 Spectral response of vegetation

Spectral response consists of the energy detected by remote sensors and recorded as digital measurements, therefore it is related to the spectral resolution of a remote sensor. The spectral response can be determined for any material/target, as vegetation, soil, roads, water resources, etc. (Lillesand et al., 2014; Wang and Weng, 2013). The term spectral signatures is preferably used instead of spectral responses, because spectral response measured by remote sensors over various features often allow assessing the type or condition of the features (Lillesand et al., 2014). Franklin (2001) summarizes (in order of importance) the factors affecting remote sensing spectral response as follows:

1. The spectral properties (reflectance, absorption, transmittance) of the target;
2. The illumination geometry, including topographic effects;
3. The atmosphere
4. The radiometric properties of the sensor (e.g., signal-to-noise ratio);
5. The geometrical properties of the target (e.g., leaf inclination).

Atmosphere, temporal and spatial effects influence spectral response because the energy recorded by a sensor is always modified to some extent by the atmosphere between the sensor and the ground (Lillesand et al., 2014). In this scope, Wang and Weng (2013) explain that surveys that examine spectral response of vegetation, such as natural forests, are always challenging because the spectral signature of trees is affected not only by the variation in crown illumination, but also by several other background effects. Therefore, understanding the basic pattern of reflectance and absorption can help with the interpretation of remote sensing imagery in forestry applications

(Franklin, 2001). Fig. 3 gives an illustration of the spectral response curve of chlorophylls in green leaves and of water.

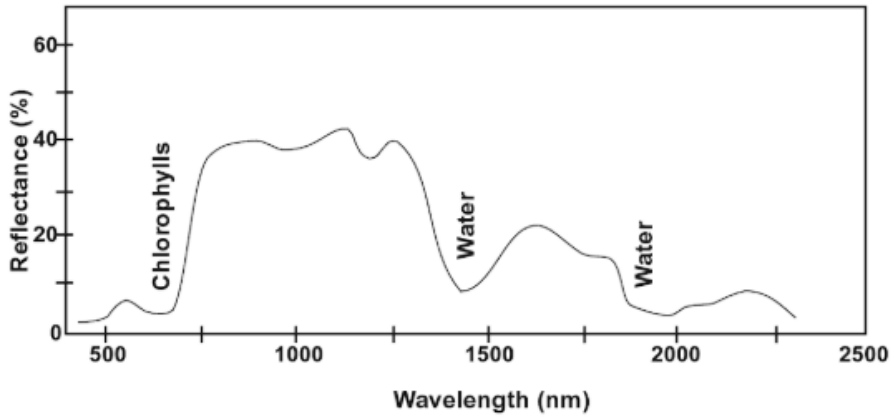


FIG. 3. SPECTRAL RESPONSE CURVES OF VEGETATION ILLUSTRATING THE PORTIONS OF THE SPECTRUM IN WHICH ABSORPTION AND REFLECTANCE DOMINATE. IMAGE SOURCED FROM FRANKLIN (2001).

Franklin (2001) explains that for a green leaf, it is normally observed a small green peak reflectance (at approximately 550 nm), and that the rapid rise in reflectance in the near-infrared (before 1000 nm) occurs in the red-edge (Fig. 3). The notable variations of leaves among tree species, and even variations in a same tree, affect the reflectance and radiance captured by spectral bands from remote sensors.

In the visible range (400 nm – 720 nm), leaf pigments (chlorophyll, carotene, and xanthophyll) dominate the reflectance and therefore they define the reflectance curve in this spectral region. In the near infrared (NIR) range (720 nm – 1,100 nm), there is few absorptions of radiation in leaves, so that the larger the gap of the leaves' internal structure, the larger the reflectance. In the shortwave infrared (SWIR) range (1,100 nm – 3,200 nm), in turn, the reflectance is more affected by the water in the leaves (Ponzoni et al., 2012). Fig. 4 shows the behavior of types of vegetation employing Landsat-8 OLI bands and compositions, including the bands (Red, Infra-red, short-wave infra-red) cited above.

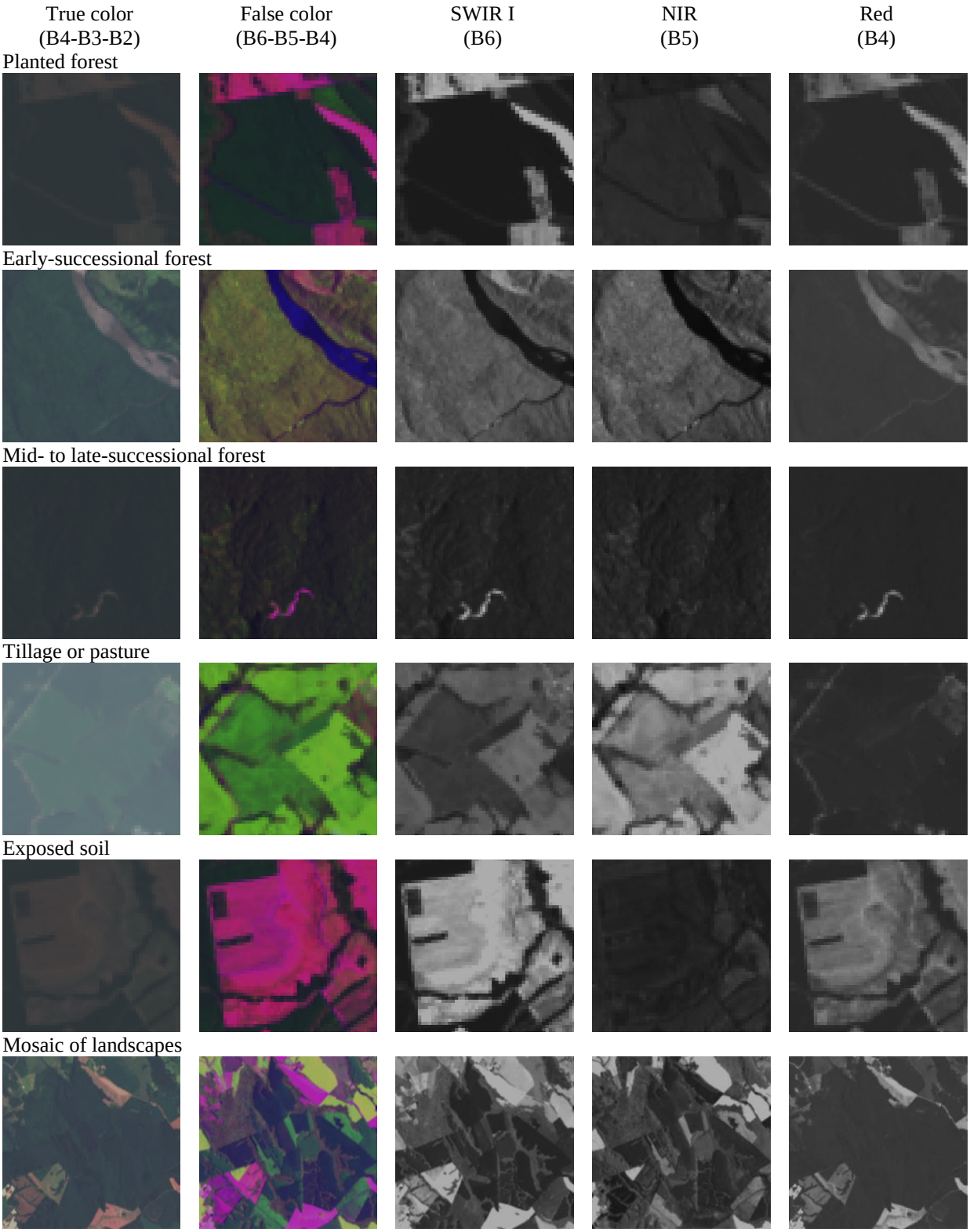


FIG. 4. KEY OF INTERPRETATION OF VEGETATION TARGETS BY MEANS OF COMPOSITIONS AND SINGLE-BAND IMAGES OF LANDSAT-8 OLI. SWIR I: SHORT-WAVE INFRA-RED. NIR: NEAR INFRA-RED. SOURCE: THE AUTHOR.

Reflectance (and radiance) are physical units used to facilitate comparison between the same or different sensors at different times, and the comparison between satellites (Franklin, 2001). This is a required task because remotely sensed data are typically presented to the user as digital number (DN). In fact, DNs are consistent internally within the image and between different bands (or wavelengths), but they have to be transformed into reflectance (or radiance) by standardization and comparison reasons (Franklin, 2001).

The at-sensor radiance equation is given by [9]:

$$L_s = a_0 + a_1 DN \quad [9]$$

Where:

L_s: at-sensor radiance ($W\ m^{-2}\ \mu m^{-1}\ sr^{-1}$); *DN*: digital number; a_i absolute calibration coefficients of a given satellite.

The coefficients a_0 and a_1 would be equivalent to a simple gain and offset, based on a scaled measure of the range of DN in the image, plus a spectral reference. However, the measurement most useful in forestry is reflectance, which is a property of the target alone (Franklin, 2001). At-sensor reflectance is given by [10]:

$$p = \frac{\pi\ d^2\ L_s}{E_0\ cos\theta_z} \quad [10]$$

Where:

p: apparent reflectance; d: normalized Earth/Sun distance; L_s: at-sensor radiance ($W\ m^{-2}\ \mu m^{-1}\ sr^{-1}$); E_0: irradiance ($W\ m^{-2}\ \mu m^{-1}$); θ_z: solar zenith angle.

2.3.2 Combined use of NFI and satellite imagery data for quantifying forest resources

Perhaps, wood volume, biomass and carbon are the most important forest variables for the economic and conservation points of view. Estimating stocks and increments of these variables at a large scale provides a broad vision of resources available in a country, allowing it to better control timber (firewood, saw log, etc.) flow and prices, as well as to estimate carbon sequestration and emission.

At a national level, NFI plots are the most important source of data for estimating forest variables by means of different methods (McRoberts et al., 2009). Remotely sensed data are often combined with NFI data in order to estimate and map forest attributes. Several studies have shown that such combination produces more accurate and precise estimates, being therefore useful for large area forest attribute mapping (Bohlin et al., 2017).

Among the main cited forest variables, studies testing satellite data-derived estimates (e.g., Mäkelä and Pekkarinen, 2004; Mohammadi et al., 2010) have shown that wood volume is more difficult to model than AGB and carbon. Lu et al. (2004) cite examples of authors who found strong relations between the visible bands and biomass, and negative relations between the middle infrared bands and wood volume. Other forest variables, such as crown closure, basal area, mean height of trees etc, also have been accurately estimated by using remotely sensed data (Lu et al., 2004; Thenkabail et al., 2003). Thenkabail et al. (2003) used satellite data even to estimate diversity indices and species associations, which are variables that can be easily estimated in NFI data.

Gobakken et al. (2012) used NFI field plots to estimate AGB at various geographical scales. These authors combined field data with airborne laser scanner (ALS) data, testing two estimation strategies; a model-dependent and a model-assisted. Gobakken et al. found that the two laser-ground estimation strategies differed from the NFI AGB estimates by about 2% and 8%. The results indicated a proper application of the methods for estimating AGB at different scales. Ene et al. (2016), Maack et al. (2016), Margolis et al. (2015), and Maltamo et al. (2009) also obtained accurate results in AGB and/or wood volume estimations, by integrating NFI data with ALS data.

Deo et al. (2017) estimated AGB but combining NFI plot data with medium spatial-resolution satellite data, as well as with a statewide LiDAR dataset. The authors obtained accuracy considered acceptable for mapping AGB. Also, with satellite data, Fazakas et al. (1999) estimated AGB and wood volume through NFI plot data. Fazakas et al obtained RMSE of up to 8.7% for biomass and 4.6% for wood volume when they increased the aggregation of cells (pixels) to 510 ha. Tomppo et al. (2002) also obtained accurate results in estimating these two variables (wood volume and AGB) combining Landsat-TM data and IRS-1C WiFS data, together with NFI plot data.

The present study also applies a combined use of remotely sensed and NFI data, testing modeling of forest variables through two pixel-level methods: (i) linear regression models, with use of surface reflectance of pixels; and (ii) regression-kriging, with use of geographical variables.

Regression models have been widely used to estimate and map forest variables, as function of pixel information, including reflectance, radiance, digital number, vegetation index, combined or not. The following surveys are examples of such application: Blackard et al. (2008), Deo et al. (2017), Dube and Mutanga (2015), Foody et al. (2003), Gara et al. (2017), Hall et al. (2006), Mohammadi et al. (2010), Powell et al. (2010), Schroeder et al. (2008), Tomppo et al. (2002), Zheng et al. (2004), and Zheng et al. (2007). Comparisons among regression models and other estimation techniques derived from satellite imagery data are addressed in Lu (2006).

RK is a geostatistical, hybrid method that combines kriging with a regression-kriging model (RKM) (Odeh et al., 1995). The role of the RKM is to introduce regression residuals (also called uncertainty) into the spatial analysis (Meng et al., 2009). Thus, the RK method has the advantage of generating estimates for the spatial distribution of the variables of interest, and its uncertainty for the study area. The uncertainty maps allow the evaluation of the reliability of estimates by identifying the locals with major uncertainties, which can be useful, for example, to select different estimation methods for those areas. Among the RK applications, it can be cited: Viana et al. (2012) used RK for estimating forest variables. Hengl et al. (2007) combined RK and satellite imagery data. Mello et al. (2015), Scolforo et al. (2015), and Angulo-Martínez et al. (2009) employed geographical variables in RKMs, such as done in the present study.

3. MATERIAL AND METHODS

3.1 STUDY AREA AND DATA SOURCE

The study area covers three political regions of the State of Parana: Mid-West, Mid-South, and Southeast (Fig. 5). These regions are located across the Plateau of Ponta Grossa (second plateau) and the Plateau of Guarapuava (third plateau), with altitude ranging from 320 to 1,300 m. This area fits into the Atlantic Forest biome; forest types as Rainforest, followed by Savanna, and Seasonal Forest, cover the major part of the study area. According to Köppen classification, Cfa and Cfb are the predominant climate in the region, i.e., a sub-tropical climate without dry seasons, with mean temperatures higher than 22 ºC (Cfa) in some regions, and lower than 22 ºC (Cfb) in other ones.

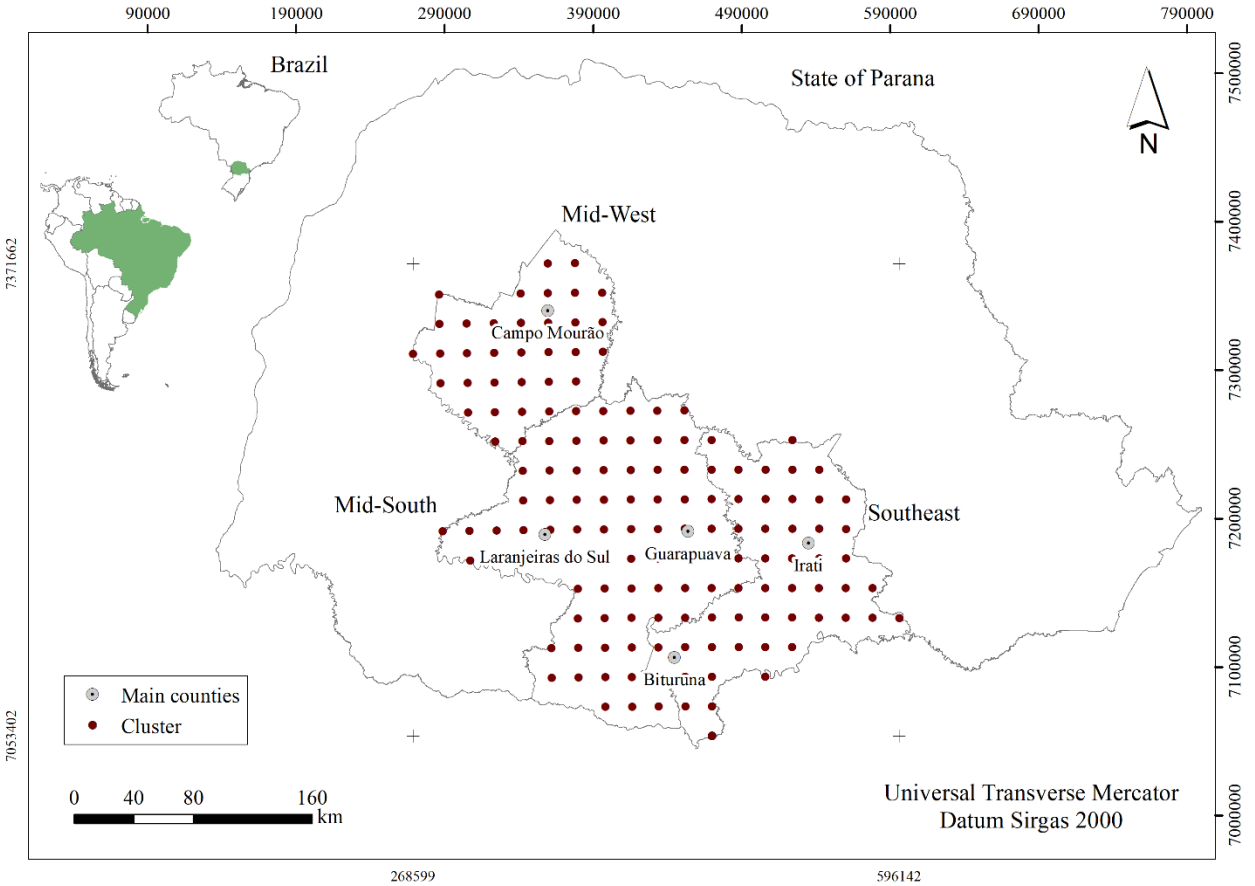


FIG. 5. STUDY AREA AND LOCATION OF CLUSTERS ALLOCATED FOLLOWING STANDARDS OF THE NFI SAMPLING OVER THE STUDY AREA.

Data from the National Forest Inventory (NFI) carried out in these regions of Parana were used. The study area covers ~55,331 km² or ~28% of the State of Parana and it was the first region in the state to be inventoried by following the standardized NFI sampling design. Field data were collected during the first half of 2013, in which 152 clusters were systematically allocated over these regions. Fig. 5 shows the study area and location of each cluster of the NFI.

As in many other countries, the Brazilian NFI follows a systematic sampling design with regular grid of 20 x 20 km, but such distance reduces insofar as it moves away from the Equator line to the polar region, due to the curvature of the Earth. By following this grid, fixed-area clusters are allocated throughout Brazil for collecting forest variables, including tree dimensions and sanity, dead and alive biomass, soil samples, identification of forest species, among others. The clusters consist of structures with four 1,000-m² sub-units; each sub-unit is still split into ten 100-m² plots, in which the tenth plot of each sub-unit is split twice again into smaller sub-plots. Fig. 6 shows cluster dimensions and criteria for inclusion of trees for measurement.

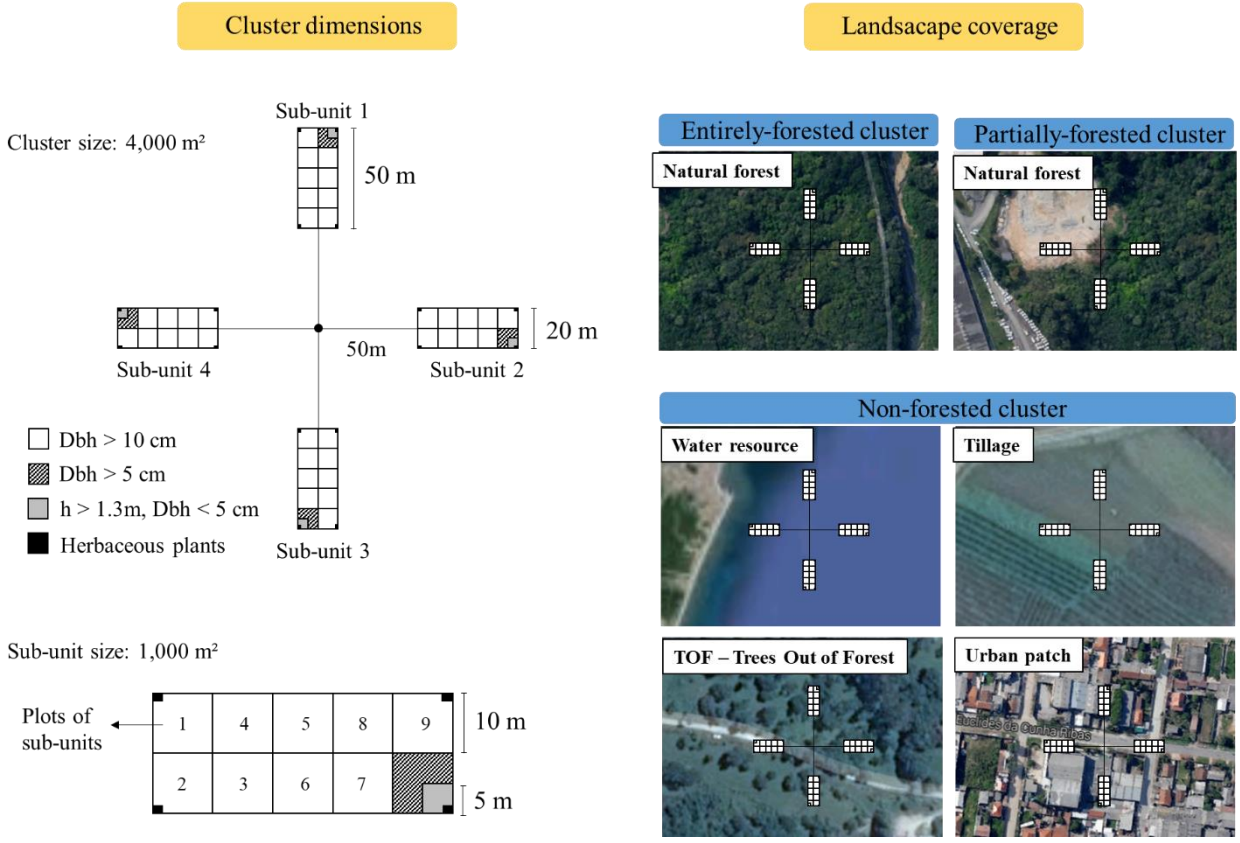


FIG. 6. CLUSTER DIMENSIONS AND LANDSCAPES OBSERVED IN THE NFI SAMPLING.

As the NFI is established in a systematic sampling design, clusters can be allocated in any landscape, i.e., both in forested as in non-forested areas. Clusters in natural forests can be categorized into entirely-forested or partially-forested clusters, meaning that can be allocated entirely or partially within forests, respectively. Fig. 6 also shows the main landscapes observed in the inventory carried out in this part of the State of Parana.

According to the NFI's methodology, each 100-m² plot is categorized in relation to its predominant land-use, i.e., the one that occupies the larger extension of a plot. This categorization includes nineteen land-uses of which seven of them relate to natural forest: 1) mature forest; 2) mid-successional forest with palm trees; 3) mid-successional forest without palm trees; 4) mid-successional palm tree forest; 5) early-successional forest with palm trees; 6) early-successional forest without palm trees; and 7) early-successional palm tree forest.

Entirely-forested clusters correspond to the case in which all the 100-m² plots are classified into one of these seven classes of natural forest. Likewise, partially-forested cluster is that one in which at least one 100-m² plot is classified into one of these classes.

3.2 ANALYTICAL PROCEDURE

The first variable of interest of this research is wood volume ($m^3\ ha^{-1}$), resulting from the sum of stem volume of all woody trees observed in the entirely- or partially-forested clusters. Wood volume was modeled at tree level, estimated by cluster, and then extrapolated to non-sampled areas at coarse and fine scales. Fine-scale estimations, i.e., at pixel level, rely on application of (i) linear regression modeling on remotely sensed data, and (ii) regression kriging. These fine-scale methods were compared to the extrapolation at a coarse scale.

The second variable of interest is aboveground forest biomass (AGB, $Mg\ ha^{-1}$), resulting from the sum of biomass of aerial compartments (stem, branches, and leaves) of all woody trees observed in forested clusters. The third variable is forest carbon stock ($Mg\ ha^{-1}$), directly estimated by a conversion factor of biomass. Such as wood volume, these two variables also were modeled at tree level, estimated by cluster, and extrapolated to non-sampled areas, however, they were extrapolated at fine scale only, so no comparison between extrapolations at fine and coarse scales was done for these two variables. Models and procedures used to estimate tree-level wood volume,

biomass and carbon are presented in section [3.2.1 (ii)]. Fig. 7 and the following sections describe steps of the analytical procedure.

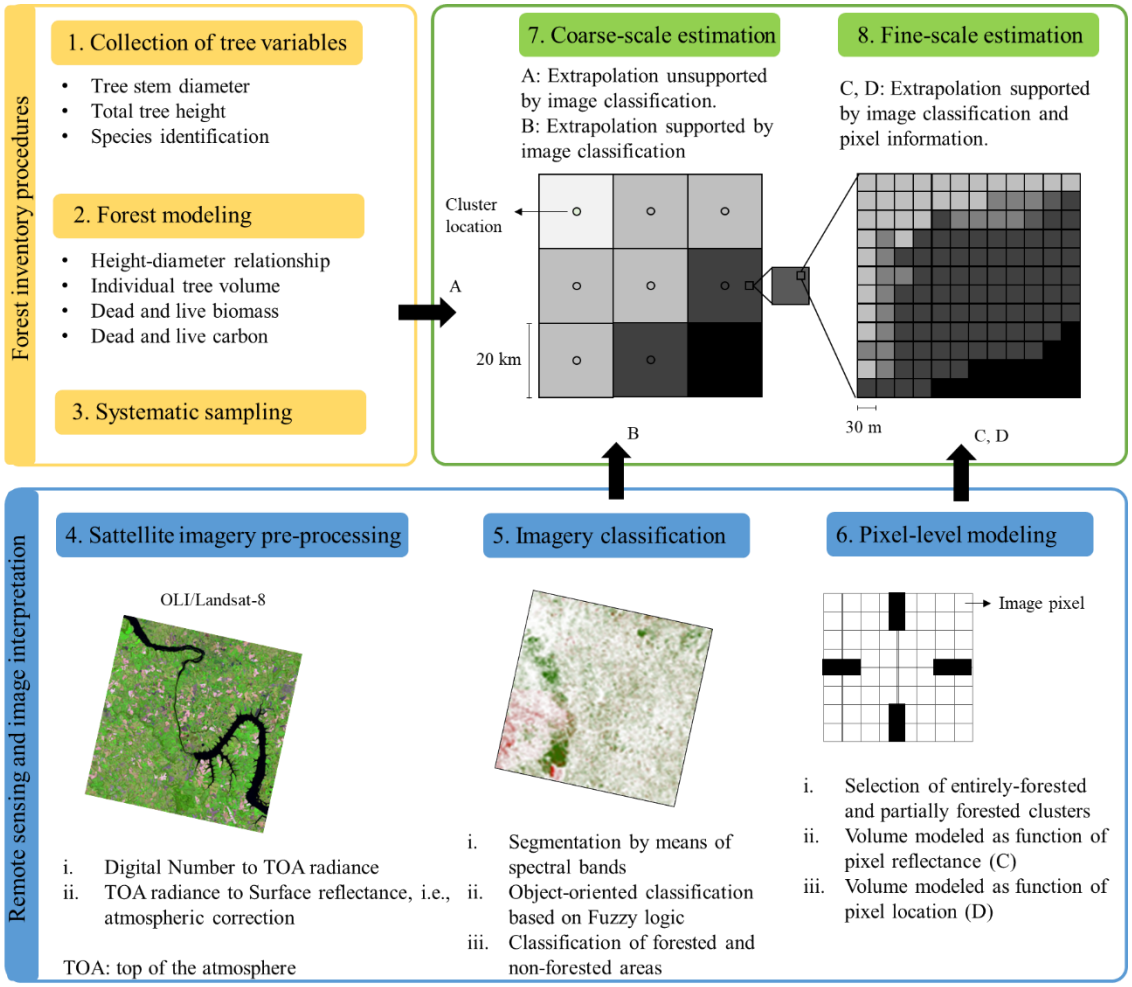


FIG. 7. ANALYTICAL PROCESS STEPS FOR ESTIMATING AND EXTRAPOLATING WOOD VOLUME IN NATURAL FORESTS AT COARSE AND FINE SCALES.

3.2.1 Forest inventory procedures

i. Collection of tree variables

The first step of the analytical procedure consists of collecting the field data necessary to forest modeling (Fig. 7, step 1). As described, tree data were collected in 152 clusters systematically allocated along the study area. Every woody tree, palm tree, and tree fern, observed in the 100-m² plots, dead and alive, with diameter at breast height (Dbh) ≥10 cm (Fig. 6), had its Dbh measured with a diameter tape, besides the total height of some selected trees measured by

using an electronic clinometer; total and merchantable heights of the remaining trees were visually estimated. Despite both palm trees and tree ferns were measured in the NFI, the analyses took into account data of woody alive trees only, first because the volume/biomass of palm trees and tree ferns are ~1.1% over the total, and second because this study focuses on wood forest product.

ii. Tree-level modeling

Tree-level modeling (Fig. 6, step 2) was applied to estimate the following variables: (i) tree height, (ii) stem volume, (iii) tree biomass, and (iv) tree carbon.

- Height-diameter modeling

In the Brazilian NFI, few trees have their total height measured through clinometer; the remaining trees have total height visually estimated. In order to avoid errors caused by naked-eye estimations, the total height of trees measured without clinometer was estimated by the Naslund model [11].

$$h - 1.3 = \frac{Dbh^2}{(a + bDbh)^2} \qquad [11]$$

Given the experiences and great results obtained by Péllico Netto et al. (2015), data were split into four datasets obeying the following strata of height/diameter ratio (h/d): (i) $0.2 \leq h/d < 0.7$, (ii) $0.7 \leq h/d < 1.2$, (iii) $1.2 \leq h/d < 1.7$, and (iv) $h/d \geq 1.7$. The Naslund model was fitted to each dataset.

- Individual tree volume estimation

This step consists of estimating volume of the stem of trees, considering stem as the section from zero to total height of the tree, disregarding volume of crown and roots. Here is proposed an alternative method based on equations from the project “Forest Inventory of Pine in the Southern Brazil”, carried out during the 1970s (IDBF, 1978). Such study applied the model [12] was used for estimating diameters along stems (IDBF, 1978, p. 130).

$$D_{h_i} = a + b\ h_i \qquad [12]$$

In [12], D_{h_i} is the diameter at the *i*-th height, h_i is a given height of the tree, and the coefficients *a* and *b* are derived from two sub-equations: one fitted for broad-leaved species [13], and another for the species popularly known as Pine-of-Parana (*Araucaria angustifolia* (Bert.) O. Kuntze) [14]. Due to its high ecological importance in Parana, a species-specific equation was used to the species Araucaria.

Broad-leaved species

$$a = 1.2124\ Dbh^{0.9975}\ h^{-0.0606} \qquad [13.a]*$$

$$b = 0.8994\ Dbh^{-0.0584}\ h^{0.0317} \qquad [13.b]*$$

Araucaria species

$$a = 1.2063\ Dbh^{0.9855}\ h^{-0.0531} \qquad [14.a]*$$

$$b = 0.8986\ Dbh^{-0.0360}\ h^{0.0283} \qquad [14.b]*$$

Where,

Dbh: diameter at breast height (over-bark), in cm. h: tree total height, in m. * Equations taken from IDBF, 1978, p. 130.

Knowing that any diameter along the stem can be estimated by inputting the sub-equations [13] and [14] into the model [12], the diameter at ground height (Dgh) can be estimated if h_i is set to zero; consequently, the estimation of Dgh disregards the sub-equation b. The equations for estimating the Dgh of broad-leaved [15] and Araucaria [16] species can be described as follows.

$$D_{h_i=0} = Dgh = 1.2124\ Dbh^{0.9975}\ h^{-0.0606} \qquad [15]$$

$$D_{h_i=0} = Dgh = 1.2063\ Dbh^{0.9855}\ h^{-0.0531} \qquad [16]$$

The proposal for estimating individual tree volume is based on a form factor that relates the half of the sectional area at ground height, and the sectional area at breast height. By a matter of semantics, the proposed form factor [17] was named as the 'sectional form factor', symbolized by $f_{0.5g}$. Its formulation is given as follows:

$$f_{0.5g} = \frac{\pi}{40000}\frac{Dgh^2}{2} \Big/ \frac{\pi}{40000} Dbh^2$$

$$f_{0.5g} = \frac{Dgh^2}{2Dbh^2} \qquad [17]$$

Replacing the equations [15] and [16] in [17], we have the sectional form factor equations for broad-leaved species and for the Araucaria species, respectively. Table 1 gives the equations used to estimate individual tree volume of these species.

TABLE 1. EQUATIONS APPLIED TO ESTIMATE STEM VOLUME IN FORESTED CLUSTERS.

Volume equations			
$v = s_{1.3}\, h\, f_{0.5g}$,	where	$f_{0.5g} = \frac{(1.2124\, Dbh^{0.9975}\, h^{-0.0606})^2}{2Dbh^2}$	Broad-leaved species
$v = s_{1.3}\, h\, f_{0.5g}$,	where	$f_{0.5g} = \frac{(1.2063\, Dbh^{0.9855}\, h^{-0.0531})^2}{2Dbh^2}$	Araucaria species

v: stem volume (over-bark), in m³. $s_{1.3}$: sectional area (over-bark) at breast height, in m². $f_{0.5g}$: sectional form factor. Dbh: diameter at breast height (over-bark), in cm. h: tree total height, in m.

That alternative method has two main differences from the traditional form factors: first because here sectional areas are used rather than diameters, which was chosen after trials involving both broad-leaved species and Araucaria from inventory databases. Secondly, and perhaps the main reason for this proposal, is that the Dgh has not to be collected, even because no diameter besides Dbh is collected in the NFI. To who wants to adjust the proposed methodology, the accessibility of the Dgh is cited as another great advantage in relation to the traditional form factors, which generally require measurements of inaccessible diameters.

- Tree biomass and carbon estimation

A biomass equation and a biomass-carbon conversion factor fitted by Ratuchne (2010) were applied to estimate, respectively, tree biomass and carbon (Table 2). Ratuchne considered biomass of stem, dead and alive branches, trunk bark, and foliage of the trees. The biomass equation (Table 2) provided RMSE of 47.24% and R^2 of 95.5%. The average conversion factor, proportional to the average weights of each compartment (Ratuchne, 2010, p. 44), was 41.0%. Ratuchne employed data from forests belonging to the Mixed Ombrophlilous Forest, in Parana, Brazil.

TABLE 2. EQUATIONS APPLIED TO ESTIMATE TREE BIOMASS AND CARBON IN FORESTED CLUSTERS.

Biomass and carbon equations	
$bm = -3.025\ Dbh + 0.425\ Dbh^2 + 0.006\ Dbh^2 h$	Mixed Ombro-phlilous Forest
$c = 0.41\ bm$	

bm: total tree biomass, in kg. c: total tree carbon, in kg. Dbh: diameter at breast height (over-bark), in cm. h: tree total height, in m. Source: Ratuchne (2010).

Descriptive statistics of the main tree variables are presented in Table 3, including Dbh, total height, stem volume, tree biomass and carbon estimated by equations (Table 1 and Table 2), and artificial form factor; this last being the ratio between the estimated volume and volume of a cylinder with diameter = Dbh and height = total height of the tree.

TABLE 3. DESCRIPTIVE STATISTICS OF TREE VARIABLES BY SPECIES CLASS, IN NFI DATA.

Tree variable	Minimum	Medium	Maximum	CV (%)	n
Broad-leaved species					
Diameter at breast height (cm)	10.0	18.4	118.3	52.0%	10,822
Total height (m)	1.4	10.4	28.0	34.9%	10,822
Form factor	0.481	0.551	0.696	4.8%	10,822
N. of trees ha^{-1} (Dbh ≥ 10 cm)*	3	237	1,208	91.4%	10,822
Volume ($m^3.tree^{-1}$)*	0.009	0.228	9.512	189.0%	10,822
Total biomass ($kg.tree^{-1}$)*	40.7	213.3	7,372.3	157.0%	10,822
Total carbon ($kg.tree^{-1}$)*	16.7	87.4	3,022.6	157.0%	10,822
Araucaria species					
Diameter at breast height (cm)	10.0	27.1	74.8	53.5%	526
Total height (m)	4.0	13.4	26.0	32.5%	526
Form factor	0.458	0.508	0.587	4.7%	526
N. of trees ha^{-1} (Dbh ≥ 10 cm)*	3	53	1,315	350.0%	526
Volume (m^3)*	0.020	0.580	4.831	128.1%	526
Total biomass ($kg.tree^{-1}$)*	42.5	488.8	3,180.6	112.6%	526
Total carbon ($kg.tree^{-1}$)*	17.4	200.4	1,304.0	112.6%	526

* number of trees by cluster.

iii. Systematic Sampling with Cluster

Brazil's NFI follows a systematic sampling (SS) with cluster in which each cluster is split into four equal-size sub-units (Fig. 6). To accomplish the extrapolation of wood volume at a coarse scale, two important sampling approaches had to be taken into account. Firstly, the variance from a SS is highly dependent on the arrangement from which the units are distributed along the population (Chacko, 1965). Secondly, a SS is likely to be as efficient as homogeneous units are put spatially together, so that samples taken in different occasions are more or less alike (Chacko, 1965).

It is based on the cited theoretical concepts that it was opted to apply the Method of First Difference (MFD) for obtaining sampling error, which is a method suited to SSs. The step-by-step procedure for obtaining sample estimators by MFD is presented in the section 2.1.2. In fact, the MFD approximates the sampling error as the other methods applied in SSs. Opting for the MFD, rather than, for example, a method suited to random sampling, is due to the Maltese cross structure of the NFI's clusters. Such structure favors a spatial homogeneity and dependence of the sub-units, favoring the NFI case to fit into the theoretical concepts of SSs cited in Chacko (1965). Processing data from a systematic sample as if it is from a random sample might be a simpler alternative, but some statistical improprieties would imply in overestimates of the standard error (Chacko, 1965).

With the standard error obtained by the MFD (section 2.1.2), the confidence intervals (α = 0.05) for the mean of the variable of interest, in this case, wood volume ($m^3\ ha^{-1}$), is given by:

$$[\bar{x} - tS_{\bar{X}} \leq \bar{X} \leq \bar{x} + tS_{\bar{X}}] = 95\%$$

For the total, the confidence interval is given by:

$$\left[\hat{X} - NtS_{\bar{X}} \leq X \leq \hat{X} + NtS_{\bar{X}}\right] = 95\%$$

Due to the lack of normality, the variable wood volume had to be transformed by means of the Box-Cox transformation [18]. With an optimized $\lambda = 0.405$, the Kolmogorov-Smirnov test for normality indicated the normalization of data at 95% probability level.

$$Y_i(\lambda) = \frac{X_i^{\lambda} - 1}{\lambda} \qquad [18]$$

As described before, Brazil's NFI possesses both forested as non-forested clusters (Fig. 6), in addition, forested clusters may be either entirely or partially forested. In this research, non-forested clusters were disregarded from the calculations of sampling error. By non-forested cluster, it was considered that cluster without trees with Dbh ≥ 10 cm in the 40 plots, in the four sub-units.

There are important issues to highlight about that inclusion criterion that relies on the Dbh size. First, a cluster may have no tree with Dbh ≥ 10 cm and be classified as entirely-forested cluster. This situation occurs when all its 40 plots fall into (and are classified as) very-young tree vegetation (popularly called *capoeirão*), where there are thin trees only; this is a typical example by which a cluster is disregarded from the statistical analysis, even being entirely within forest. Second, taking that same example (40 plots classified as forest), but considering now that some plots (< 40) have trees with Dbh ≥ 10 cm; this cluster was considered here as partially forested, although it is classified as entirely-forested cluster.

The last issue, but not least, concerns to the bigger variance that partially-forested clusters should produce in relation to the entirely forested ones, resulting in a bigger sampling error. It means that the expected sampling error is bigger than if the statistical analysis had entirely-forested clusters only. To demonstrate such supposition, three datasets were considered for calculating the sampling error: (i) with partially-forested clusters only, (ii) with entirely-forested clusters only, and (iii) with both together, so that single estimators were obtained for all clusters. This last case was employed to compare with the fine-scale methods.

3.2.2 Remote sensing and image interpretation

i. Satellite imagery and pre-processing

Remotely sensed data and image interpretation were necessary to perform the fine-scale estimations (Fig. 7, step 7b). To cover the study area, seven Landsat-8 OLI imagery was downloaded from the U.S. Geological Survey (USGS). The next steps of the analytical process (Fig. 7, steps 4, 5, 6, and 7) then depend on satellite imagery.

The main information about the Landsat-8 OLI imagery is shown in Table 4. The period from 2013-Dec to 2014-Feb was chosen to acquire the satellite imagery because it provides cloud cover lower than 0.25% for all imagery.

TABLE 4. INFORMATION ABOUT LANDSAT-8 OLI IMAGERY ACQUIRED FOR REMOTE SENSING ANALYSIS AND IMAGE INTERPRETATION.

Path/Row	Acquisition date (y-m-d)	Cloud cover
221/077	2013-12-04	0.06%
221/078	2014-01-21	0.12%
222/077	2014-01-28	0.02%
222/078	2013-10-08	0.11%
223/076	2013-12-09	0.01%
223/077	2014-02-04	0.02%
223/078	2014-01-19	0.25%

Imagery pre-processing that is required before performing the subsequent analyses were performed with Envi 5.3, being: (i) conversion of digital number to radiance scale, and (ii) conversion of radiance to surface reflectance (Fig. 7, step 4), this last executed with Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH) algorithm, implemented in Envi 5.3. FLAASH performs atmospheric correction necessary to remove most of the haze and other atmospheric perturbations (Nazeer et al., 2014). After this, the (iii) imagery was reprojected to Sirgas 2000 and (iv) clipped for removing overlapped borders from neighbor scenes.

ii. Imagery classification

The next step refers to imagery classification (Fig. 7, step 5), performed through the software eCognition Developer 8.7. This software executes object-oriented classification, where ‘objects’ means a definite spatially connected region of the satellite image. For this reason, object-oriented classification requires that imagery be segmented into small objects; operation known as image segmentation. eCognition Developer 8.7 allows users to segment imagery by means of seven different algorithms, in which the ‘multiresolution segmentation’ algorithm was chosen for this research. This algorithm applies an optimization procedure for locally minimizing the average heterogeneity of objects. Fig. 8 shows the ‘multiresolution segmentation’ algorithm window displaying the standardized parameters chosen in this study, after an exploratory analysis.

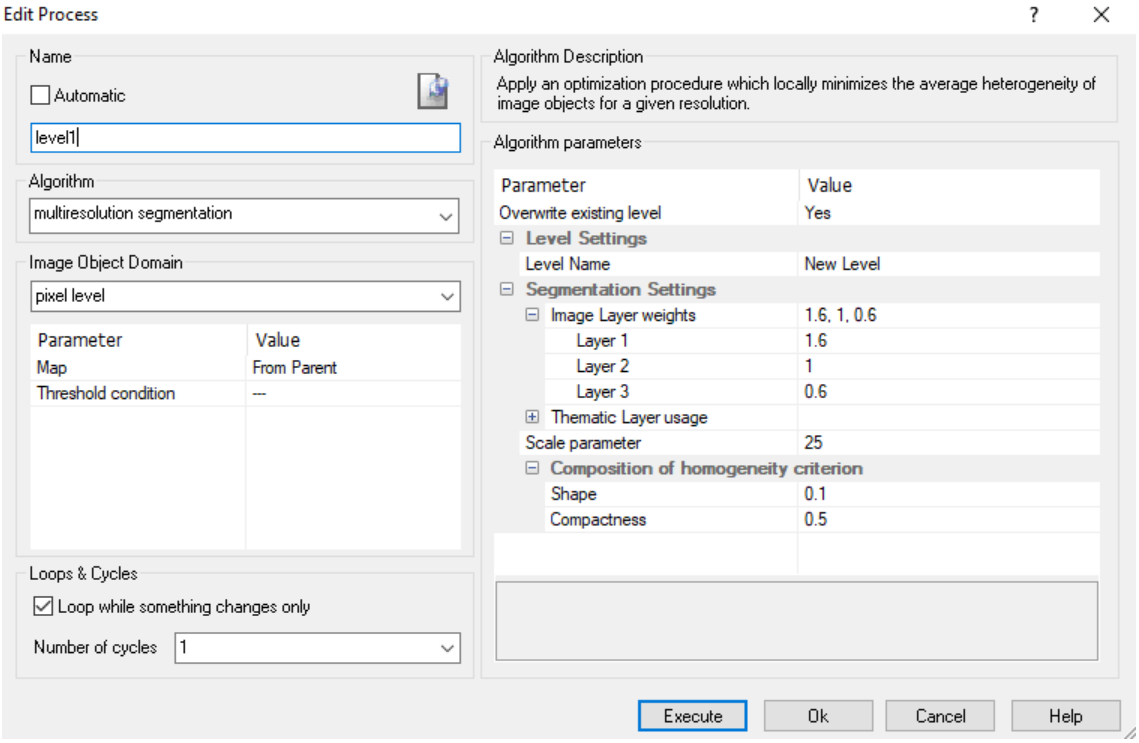


FIG. 8. ECOGNITION DEVELOPER 8.7 WINDOW DISPLAYING SEGMENTATION PARAMETERS AND WEIGHTS FOR IMAGE LAYERS USED FOR CLASSIFYING LANDSAT-8 SCENES.

As seen in Fig. 8, the 'multiresolution segmentation' algorithm uses image layers and weighs for each layer as segmentation criteria, besides parameters referring to size (scale), shape, and compactness of the image objects. For this research, all the seven images were segmented using always the same standardized parameters and image layer weights (w), as well as the same image layers: SWIR1 (Layer1), NIR (L2), and Red (L3).

After performing an exploratory analysis changing weights (w), parameters, and layers, the largest weight (w = 1.6) was assigned to SWIR1 (i.e., Layer 1), followed by NIR (Layer 2, w = 1.0) and Red (Layer 3, w = 0.6) (Fig. 8). The values for the shape and compactness parameters (0.1 and 0.5, respectively, Fig. 8) were defined so that objects be as homogeneous as possible. These segmentation parameters generated so small objects (consequently in so large volume) that it was necessary to split the Landsat imagery into four equal parts, due to the high computational burden.

Once concluded the segmentation process, the imagery was finally classified into land-use classes (Fig. 7, step 5). eCognition Developer 8.7 is a fuzzy rule-based classifier (Gupta and Bhadauria, 2014) that associates image objects with previously determined classes; the process is a supervised classification such that samples (of objects) are collected for representing previously

determined classes. The number of collected samples by class varies, but it should be large enough to compose a frequency distribution representing the class' object population.

The software still allows that classes be hierarchically structured, in which a child class inherits descriptions from its parent class. This arrangement of parents and children is called the 'class hierarchy' and is what bears specific class descriptions. Franklin (2001) addresses the importance of accomplishing forest classification under a hierarchical structure.

After testing various class hierarchies, two parent classes and three child classes were strategically chosen, in order to distinguish as much as possible natural forests from another type of vegetation, such as tillage and pasture. The parent classes were (I) forest and (II) non-forest. The class (II) is the inverted expression given for the class (I), such that objects not classified as forest turns into non-forest automatically (it is a simple option provided by eCgonition Developer 8.7). The child classes associated with (I) were (i) early-successional natural forests, (ii) mid- to late-successional natural forests, and (iii) planted forests. These two classes of natural forests were created basing on the Resolution n. 2 from CONAMA (Brasil, 1994). Eventually it was necessary to create another child class for patches of lowland forests, which are generally areas wetter than the higher ones and consequently have different spectral response from the infra-red bands (NIR, SWIR1, SWIR2).

About 10,000 ~ 20,000 samples (by scene, by class) of image objects (result from image segmentation) were collected to represent parent and child classes. Fig. 4 shows examples of key samples for interpretation of vegetation classes of this study. The software then provides probability distribution functions (pdfs, also called "membership functions") to model frequency distribution of samples belonging to each class. These pdfs are always associated with a certain variable; in this research, five variables from image layers were chosen for the image classification: (i) brightness index, (ii) mean SWIR1, (iii) mean NIR, (iv) NDVI, and (v) maximum difference of the image objects. Here, brightness index plays the role of a proxy variable and is defined as the mean reflectance of SWIR1, NIR, and Red of an image object. This same brightness index was used both in the classifications through eCognition, as in the modeling described later. For the sake of simplicity, brightness index is just called brightness. NDVI is the Normalized Difference Vegetation Index usually applied in vegetation studies with remote sensing data (Powell et al., 2010). Fig. 9 shows examples of pdfs by class used for classifying the Landsat-8 imagery.

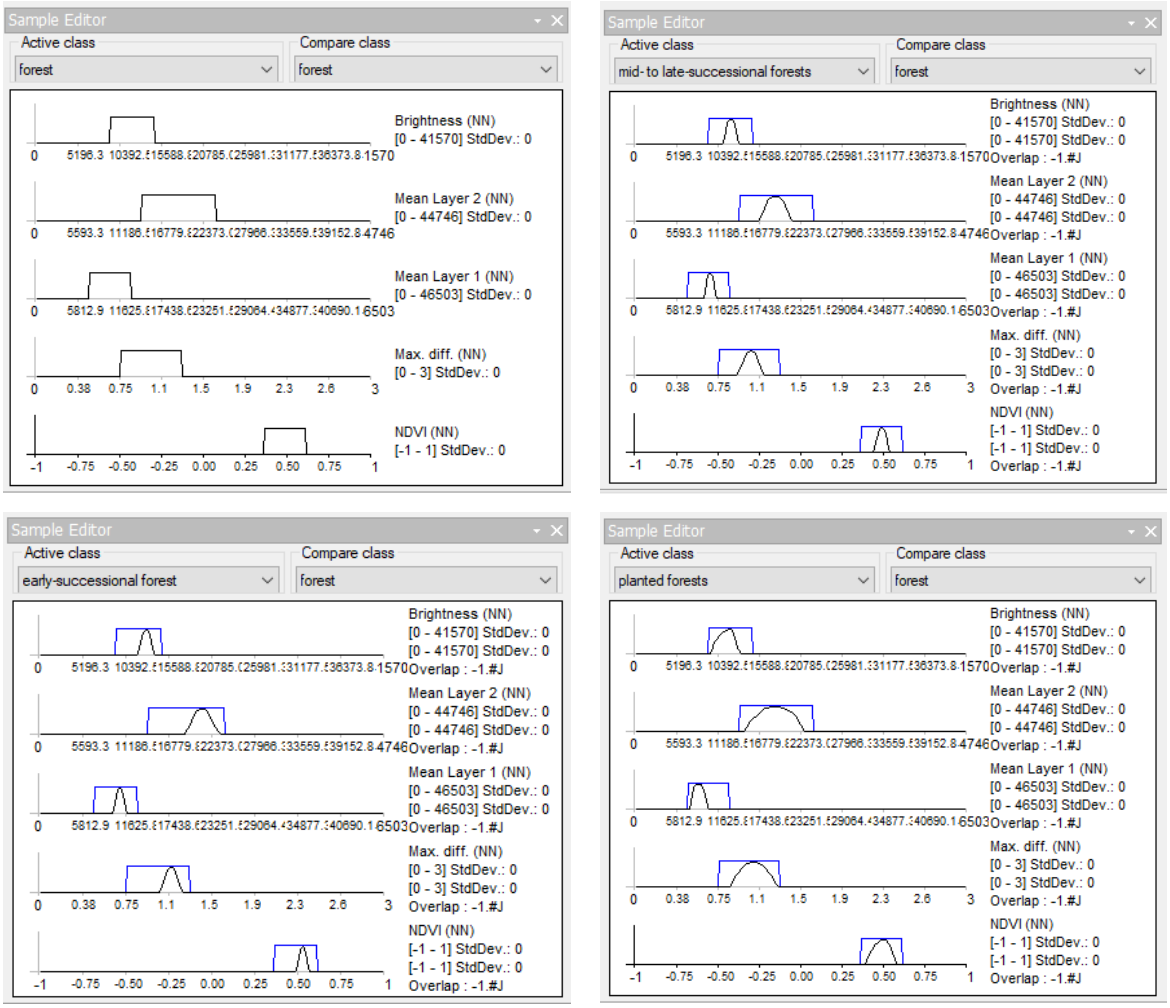


FIG. 9. ECOGNITION DEVELOPER 8.7 WINDOW DISPLAYING MEMBERSHIP FUNCTIONS OF VARIABLES USED FOR CLASSIFYING LANDSAT-8 SCENES.

Fig. 9 shows pfds manually drawn, since the software allows users easily handling pdfs according to the frequency distribution of the collected samples. The amount of collected samples by class varied from ~2,000 to 5,000, which was sufficiently large to represent each class. Since the pdfs are drawn based on the underlying samples, it is very important that the samples be sufficiently and carefully collected along the whole image.

By following the described hierarchical structure, the classification was split into two phases: parent and child. The parent class (I) forest is composed by rectangular pdfs (Fig. 9) and it gives the maximum probability (100%) of classification to an image object located between the pdf limits. Likewise, the minimal probability (0%) is given to an object outside of the limits, being automatically classified as ‘non-forest’. The child classes, in turn, are composed by pdfs adherent

to the observed frequency distributions (Fig. 9). It is in this phase that image objects were then classified into a certain forest (child) class, basing on fuzzy logic.

Subsequently, a post-classification visual analysis was accomplished to manually edit misclassified objects, allowing reaching a more assertive classification. The salt-and-pepper effect is unexpressive in object-oriented classifications (Blaschke, 2010) and then it had not to be corrected. Lastly, the accuracy of the classifications was assessed by means of the Kappa index of agreement, as explained by Congalton and Green (2008). The kappa value is one of the existent indices used for classifying the degree or strength of agreement (Landis and Koch, 1977) of classified imagery, as shown in Table 5. The samples of ground truth necessary to calculate the Kappa index were analyzed with assistance from Google Earth imagery.

TABLE 5. AGREEMENT CLASSES IN RELATION TO KAPPA VALUES.

Kappa statistic	Strength of agreement
< 0.00	Poor
0.00 - 0.20	Slight
0.21 - 0.40	Fair
0.41 - 0.60	Moderate
0.61 - 0.80	Substantial
0.81 - 1.00	Almost perfect

In fewer words, the image classification is summarized into: (i) image segmentation into objects, (ii) choosing of classes (parent and child ones) and their membership functions – definition of the class hierarchy, (iii) collection of object samples by class, (iv) handling of the membership functions (or pdfs), (v) image object classification, and (vi) post-classification analysis.

3.3 PIXEL-LEVEL MODELING

The next step of the procedure (Fig. 7, step 6) consisted of creating two pixel-level linear models to be compared between them. These models were used to estimate and map wood volume ($m^3\ ha^{-1}$) through data from satellite image pixel. The first model is based on surface reflectance values. The second model is based on latitude and longitude, fitted to be applied in the technique called the regression-kriging (RK). Both models were regressed with two randomly selected sub-datasets: (i) fitting, with ~90% of data participating in the modeling, and (ii) validation, with ~10% of data with participation. The following sections describe the fitting procedures in detail.

3.3.1 Surface reflectance model (SRM)

This model employs wood volume ($m^3\ ha^{-1}$) of clusters as output variable, and surface reflectance values from Landsat-8 OLI bands as input variables. Fig. 10 describes the analytical procedure necessary for fitting the surface reflectance model (SRM).

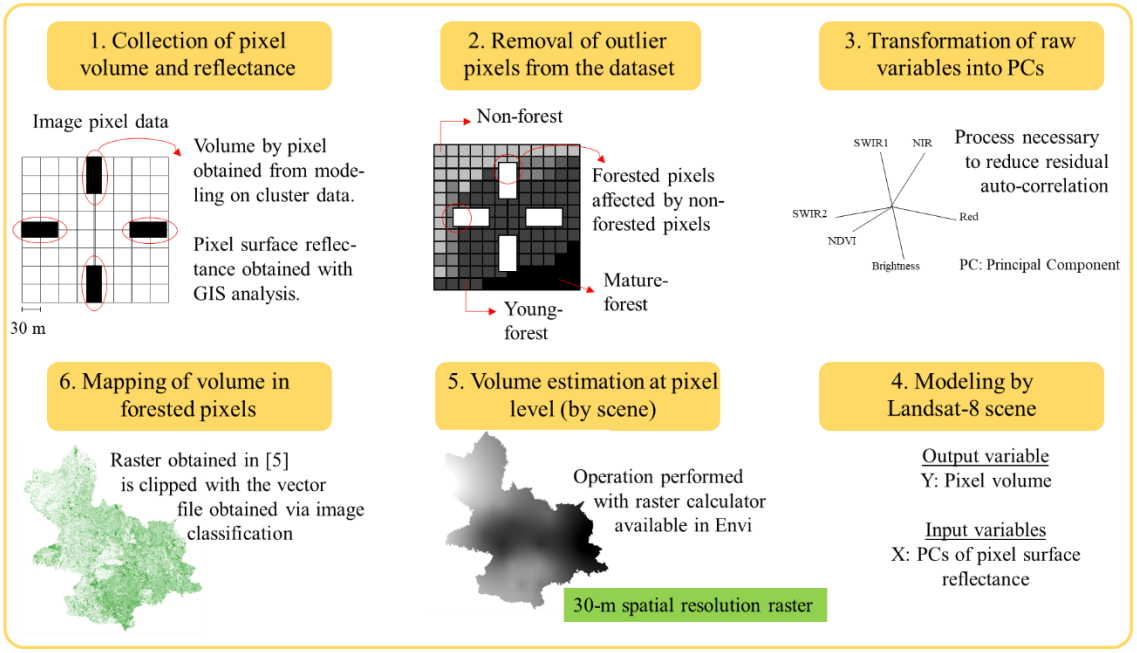


FIG. 10. ANALYTICAL PROCEDURE FOR MODELING WOOD VOLUME AT PIXEL LEVEL.

To fit the SRM, a data consistency was necessary to eliminate outlier pixels (Fig. 10, step 2) that do not represent the variable of interest, as discussed by Fazakas et al. (1999). This occurs because the localization of the clusters in field is not exactly the same in digital maps, due to GPS errors. In addition, the surface reflectance of pixels suffers influence of neighbor pixels (Maack et al., 2016). In this study, such influence occurs when pixel reflectance values of clusters allocated in the border of forests are affected by non-forested pixels (Hall et al., 2006). In this sense, upper and lower outlier pixels were removed from the database by using the rules [19] and [20], respectively.

$$L_u = Q_3 + 1.5ir \qquad [19]$$

$$L_l = Q_1 - 1.5ir \qquad [20] \qquad \text{being } ir = Q_3 - Q_2$$

Where,

L_u and L_l: upper and lower limit of the data range, respectively. Qi: i-th quartile.

The input and output variables were extracted at pixel level through GIS analysis on imagery (Fig. 10, step 1). The chosen OLI bands were: Red, NIR, SWIR1, SWIR2, Brightness, and NDVI. The stepwise method was used to select the most important input variables, leaving variables statistically significant ($\alpha \leq 5\%$) and removing those non-significant ones. The Shapiro-Wilk statistic for testing the normality indicated that the input variables are normally distributed at 5% probability level. Fig. 11 shows frequency distributions of the input variables. App. 2 shows further samples of surface reflectance of observed natural vegetation.

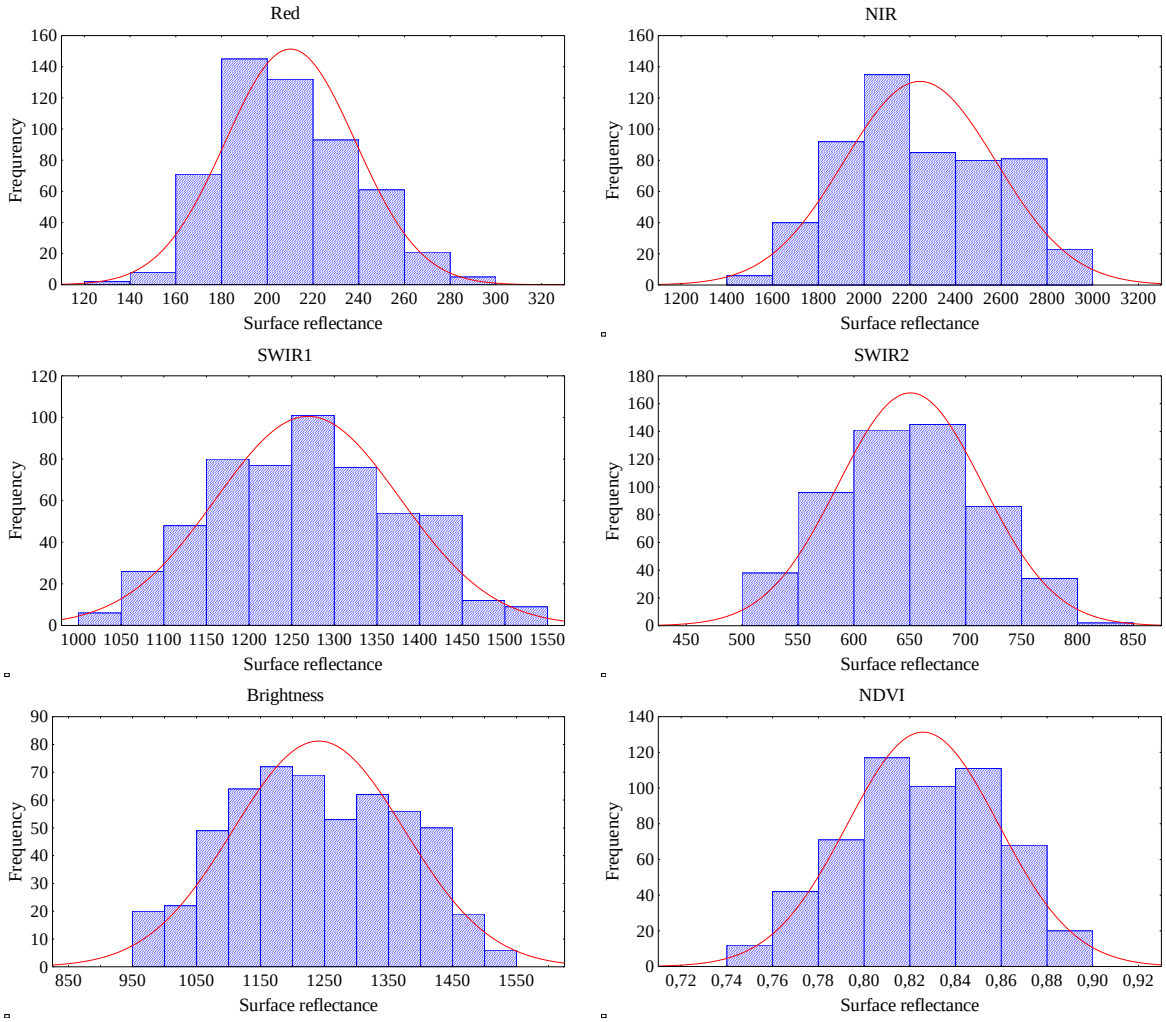


FIG. 11. DISTRIBUTION OF CLASSES OF SURFACE REFLECTANCE IN FORESTED PIXEL SAMPLES USED FOR MODELING THE TOTAL TREE VOLUME.

Despite the explanatory variables follow the normal distribution (Fig. 11), satellite sensor data are commonly correlated (Hayashi et al., 2015). As expected, correlations of up to ~50% were

found among some input variables, meaning that residuals from the models may be autocorrelated if raw variables are used. Such condition is of especial importance because non-autocorrelation between residuals is an assumption of the classical linear regression model (Gujarati and Porter, 2011). To fix that problem, the raw input variables (i.e., surface reflectance) were transformed into principal components (PCs), which is one of the most used transformations for satellite data (Meng et al., 2009) (Fig. 10, step 3). In spite of the fact that the addition of PCs enlarges the model length, they do not increase the statistical complexity of the models.

In the principal components analysis (PCA), PCs with significant eigenvalues capture the largest amount of variance of raw data (Johnson and Wichern, 2007), therefore, only these PCs were used in the regression analysis. The SRM was fitted through the method of ordinary least squares and can be expressed by [21]:

$$V_i = \alpha_0 + \alpha_1 PC_{i1} + \alpha_2 PC_{i2} + \cdots + +\alpha_j PC_{ij} \quad [21]$$

in which PC is given by [22]:

$$PC_{ij} = \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + +\beta_p X_{ip} \quad [22]$$

Where,

V: wood volume, $m^3\ ha^{-1}$. PC: principal component. X: raw input variable. j: number of PCs with $\lambda \geq 1$. p: number of variables selected by stepwise. α_i: coefficients of the volume model. β_i: coefficients of the PC.

As it was used seven Landsat scenes to cover the whole study area, the model [21] was fitted by scene (Fig. 10, step 4), using the specific set of clusters related to each scene, totaling thus seven SRMs. The separation by scene was necessary due to the variation of reflectance values among them, caused by solar illumination, haze and atmospheric conditions, among others. Fig. 12 shows the portion covered by each Landsat scene and the portion of data for the modeling.

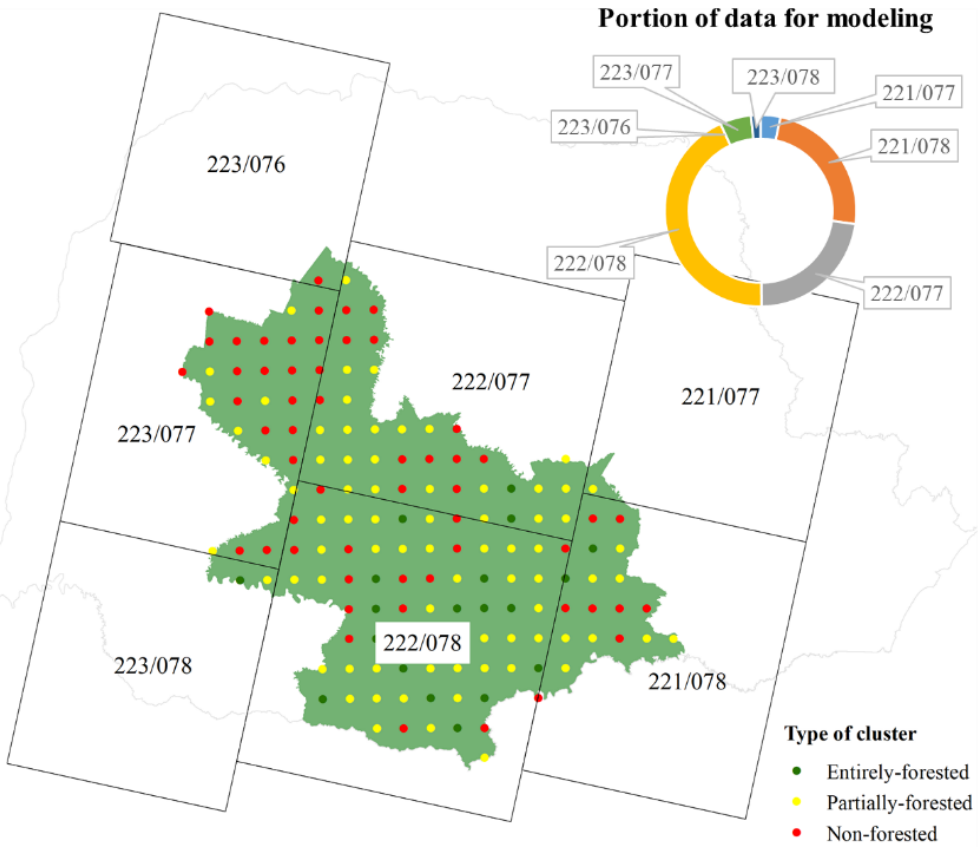


FIG. 12. COVERAGE OF LANDSAT-8 SCENES OVER THE STUDY AREA AND SPATIAL DISTRIBUTION OF CLUSTERS.

The equations from Landsat scenes 223/076 and 223/078 were replaced by similar equations from bordering scenes, due to the lack of or low number of forested pixels, respectively (Fig. 12). With the raster files of surface reflectance data, these equations were used to estimate the volume at pixel level by using the raster calculator available in Envi 4.7, creating thus a raster file of estimated wood volume. Finally, the last step (Fig. 10, step 6) was to clip such raster files by using vector files stemmed from the image classifications. As these vector files delimit forest patches from those non-forest ones, the clipping removes all pixels out of forest.

3.3.2 Regression-kriging (RK) and the RK model

The regression-kriging (RK) was an alternative method tested to estimate and map wood volume. RK is a geostatistical, hybrid method that combines kriging with a regression-kriging model (RKM) (Odeh et al., 1995). The role of the RKM is to introduce regression residuals into

the spatial analysis (Meng et al., 2009), i.e., into the regressed and spatialized variable (i.e., wood volume). Here, instead of using reflectance data, volume was modeled as a function of latitude and longitude, which are much easier to obtain than image data. These two variables are possibly few correlated with wood volume, but the main idea is to take advantage of the capability of the RK method in fixing estimated values with addition of residuals. Latitude and longitude were used as input variables in the RKM, as shown in [23]. Mello et al. (2015), Scolforo et al. (2015), and Angulo-Martínez et al. (2009) also employed geographical variables in RKMs.

$$V_i = \alpha_0 + \alpha_1 X_i + \alpha_2 Y_i \qquad [23]$$

Where,

V: wood volume of a pixel, in $m^3\ ha^{-1}$. X: latitude of a pixel. Y: longitude of a pixel. α_i: coefficients of the RKM.

The procedures to fit the RKM and execute the RK are described in the sections that follow and are summarized into six steps in Fig. 13.

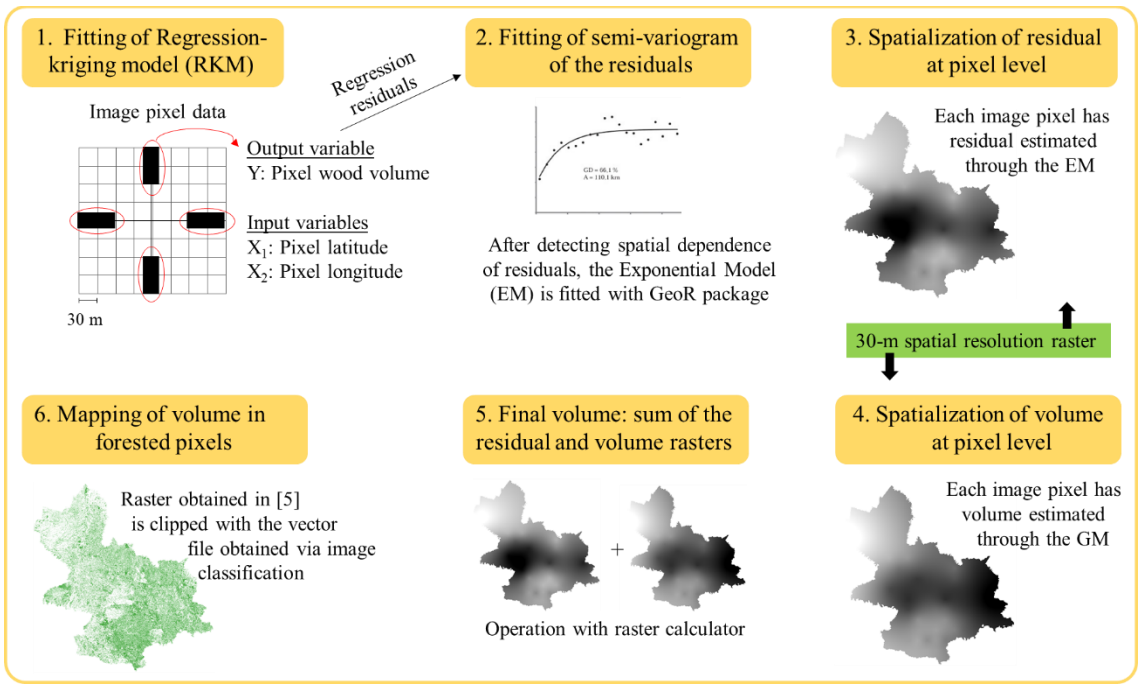


FIG. 13. ANALYTICAL PROCEDURE FOR APPLYING REGRESSION-KRIGING TO WOOD VOLUME AT PIXEL LEVEL.

The first step (Fig. 13, step 1) consists of fitting the RKM [23] with geographical variables as input variables, and wood volume ($m^3\ ha^{-1}$) from pixels as output variable. The formula $r_i = y_i - \hat{y}_i$ was used to calculate the residuals (r_i) obtained from the RKM, where y_i is the observed wood volume and $\hat{y}_i$ is the wood volume estimated by the RKM. To fit the RKM, three assumptions were considered: $E(r_i) = 0$, $Cov(r_i\ y_i) = 0$, and $Cov(r_i\ x_i) = 0$ (Odeh et al., 1995).

After that, an Exponential Model (EM) [24] was fitted to the semivariogram (SV) of the residuals (r_i) (Fig. 13, step 2). The EM was chosen following experiences from Scolforo et al. (2015). The function *variofit* from the GeoR package was used for fitting the semivariogram (Ribeiro Junior and Diggle, 2001).

$$\gamma(h) = \tau^2 + \sigma^2 \left[1 - e^{-\left(\frac{h}{\varphi}\right)}\right] \qquad \text{for } h < \varphi \qquad [24]$$

$$\gamma(h) = \tau^2 + \sigma^2 \qquad \text{for } h \geq \varphi$$

The spatial dependence of the regression residuals was quantified by means of the ratio [$r(h)$] [25] between nugget effect and sill. The value of $r(h)$ was then classified into three classes of spatial dependence (Table 6), as proposed by Cambardella et al. (1994).

$$r(h) = \frac{\tau^2}{\tau^2 + \sigma^2} 100 \qquad [25]$$

Where,

$\gamma(h)$: semivariance. τ^2: nugget effect. σ^2: random variance. h: distance between clusters, in km. φ: range; e: exponential. $\tau^2 + \sigma^2$: sill of the semivariogram.

TABLE 6. CLASSES OF SPATIAL DEPENDENCE OF THE VARIABLE OF INTEREST.

Ratio - $r(h)$	Class
≤ 25%	Strongly spatially dependent
25% to 75%	Moderately spatially dependent
> 75%	Weakly spatially dependent

Once fitted, the EM was applied to spatialize the regression residuals at pixel level (Fig. 13, step 3), in which the final product was a 30-m spatial resolution raster. Subsequently, the wood volume was also spatialized at pixel level (Fig. 13, step 4), but now using the RKM. To that end, a

vector format file containing points with latitude and longitude from each image pixel is required for the wood volume estimation. This operation also produces a 30-m spatial resolution raster. The software ArcGis (v. 10.3) was the main auxiliary tool in these described steps.

The two raster files created in the steps 3 and 4 were added to each other aiming to sum, at pixel level, regression residuals and estimated wood volumes (Fig. 13, step 5). This operation was performed with the raster calculator available in ArcGis. The final step (Fig. 13, step 6) is similar to the step 6 developed in the last model (Fig. 10).

3.4 SRM FOR ESTIMATING FOREST BIOMASS AND CARBON

After estimating tree biomass and carbon through equations shown in Table 2, the same procedures performed for the wood volume (section [3.3.1]) were applied for the forest biomass and carbon. Thus, surface reflectance models (SRMs) for estimating forest biomass [26] were fitted by Landsat scene, with pixel reflectance values transformed into PCs.

$$Y_i = \alpha_0 + \alpha_1 PC_{i1} + \alpha_2 PC_{i2} + \cdots + \alpha_j PC_{ij} \quad [26]$$

in which PC is given by:

$$PC_{ij} = \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}$$

Where,

Y: forest biomass / forest carbon, Mg ha^{-1}. PC: principal component. X: raw input variable. j: number of PCs with $\lambda \geq 1$. p: number of variables selected by stepwise. α_i: coefficients of the model. β_i: coefficients of the PC.

3.5 ASSESSMENT OF GOODNESS OF FIT

Goodness-of-fit statistics were applied to the height-diameter models [11], and pixel-level models [21], [23], and [26], in order to examine their accuracy. As the SRMs were fitted by scene, to evaluate the accuracy of the global volume estimation, such statistics were proportionally

weighed in relation to the portion which each Landsat scene covers on the study area (Fig. 12). The accuracy statistics were: adjusted coefficient of determination [27], root mean square error [28], relative RMSE [29], mean error [30], and absolute mean error [31]:

$$R^2_{adj.} = 1 - \left[\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}\left(\frac{n-1}{n-p}\right)\right] \quad [27]$$

$$RMSE = \frac{1}{n-p}\left[\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\right]^{\frac{1}{2}} \quad [28]$$

$$RMSE(\%) = RMSE\,\frac{100}{\bar{y}} \quad [29]$$

$$ME = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) \quad [30]$$

$$AME = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad [31]$$

Where:

$R^2_{adj.}$: adjusted coefficient of determination. RMSE: root mean square error. ME: mean error. AME: absolute mean error. y: observed volume. $\hat{y}$: predicted volume. $\bar{y}$: mean observed volume. n: number of observations. p: number of parameters.

3.6 EXTRAPOLATION OF WOOD VOLUME

3.6.1 Extrapolation at a coarse scale

After estimating volume at tree level and then volume by unit area ($m^3\ ha^{-1}$), wood volume was extrapolated to non-sampled areas at a coarse scale (Fig. 7, step 7a). This step consists of directly extrapolating wood volume estimated in the clusters, to the area that each cluster represents. For entirely forested clusters, such represented area is 400 km^2. For partially forested

clusters, such represented area is the proportional to the cluster's forested area; for example, if a cluster is 50% into forest, its represented area is 200 km². For non-forested clusters, its represented area is 0 km² and consequently the extrapolation is set to 0.

3.6.2 Extrapolation at a coarse scale supported by image classification

The procedure of directly extrapolating wood volume estimated in clusters to their represented areas (section 3.6.1) is the simplest way to obtain totals for the study area. Here is proposed embedding image classification [section 3.2.2 (ii)] to that method of direct extrapolation. The map of forested and non-forested areas stemmed from the image classification is required. The forested area by cluster was quantified by clipping such map into 152 ~400-km² pieces, which are the represented area of each cluster.

This alternative supported by image classification then consists of extrapolating the wood volume ($m^3\ ha^{-1}$) estimated in the clusters to the classified forested area (ha) corresponding to each cluster. Volume equal to zero was assigned to represented areas from non-forested clusters. The total volume is the sum of the extrapolated wood volume by cluster.

3.6.3 Extrapolation at a fine scale

To extrapolate wood volume at fine scale (Fig. 7, step 7b), modeling and remotely sensed data were combined. Instead of directly extrapolating the variables to non-sampled areas, as in the coarse-scale case, methods of image interpretation were employed, in order to both delimit forested patches along the study area, as to estimate such variables at pixel level. The wood volume was firstly estimated at pixel level by means of the SRM and RK, described in section [3.3]. Each of these methods allowed me to obtain wood volume for the whole study area. Capabilities to extrapolating the wood volume at a coarse scale was compared with the results of a finer scale (section [3.6.1]). Such comparison consisted of verifying if wood volume, estimated by the methods that use Landsat image data, fits into the confidence interval estimated by the method without imagery applications (i.e., coarse scale). The procedure for estimating the confidence interval for the NFI data was explained in the section [3.2.1 (iii)].

4. RESULTS

4.1 ESTIMATION OF TREE VARIABLES

Table 7 shows the height-diameter equations fitted for each stratum of h/d. Table 8 gives goodness-of-fit statistics obtained from the height-diameter models fitted to each stratum of height/diameter ratio (h/d). The values of RMSE for the total (1.91 m, or 17.1%) fit into the expected errors found in the literature. The values of ME and AME indicate a great accuracy; ME for the total was close to zero, indicating overall unbiased estimates, and AME reveals an error of 1.49 m for the total database.

TABLE 7. EQUATIONS APPLIED TO ESTIMATE TREE HEIGHT IN FORESTED CLUSTERS.

Height-diameter equations	**Specificity**
$h - 1.3 = \frac{Dbh^2}{(1.3289 + 0.19069Dbh)^2}$	Stratum 1: $0.2 \leq h/d < 0.7$
$h - 1.3 = \frac{Dbh^2}{(1.2764 + 0.21369Dbh)^2}$	Stratum 2: $0.7 \leq h/d < 1.2$
$h - 1.3 = \frac{Dbh^2}{(1.3049 + 0.2000Dbh)^2}$	Stratum 3: $1.2 \leq h/d < 1.7$
$h - 1.3 = \frac{Dbh^2}{(1.2529 + 0.22881Dbh)^2}$	Stratum 4: $h/d \geq 1.7$

Dbh: diameter at breast height (over-bark), in cm. h: tree total height, in m. h/d: height/ diameter ratio.

TABLE 8. STATISTICS OF GOODNESS OF FIT FOR THE FITTED HEIGHT-DIAMETER MODELS.

Stratum	$R^2_{adj.}$	RMSE*	RMSE(%)	ME*	AME*	DF
$0.2 \leq h/d < 0.7$	0.68	2.66	26.0%	-0.04	2.05	580
$0.7 \leq h/d < 1.2$	0.76	1.87	14.5%	0.01	1.50	176
$1.2 \leq h/d < 1.7$	0.93	1.12	8.5%	0.01	0.89	510
$h/d \geq 1.7$	0.94	0.88	7.0%	0.01	0.72	28
Overall	0.80	1.91	17.1%	-0.01	1.49	1,294

* Value given in m. RMSE: Root mean square error. ME: Mean error. AME: Absolute mean error. DF: Degrees of freedom.

Fig. 15 shows scatter plots of the residuals by stratum of h/d. As previously noted with the values of ME (Table 8), the plots show that, in general, the models provided unbiased residuals along the input variable Dbh for all strata.

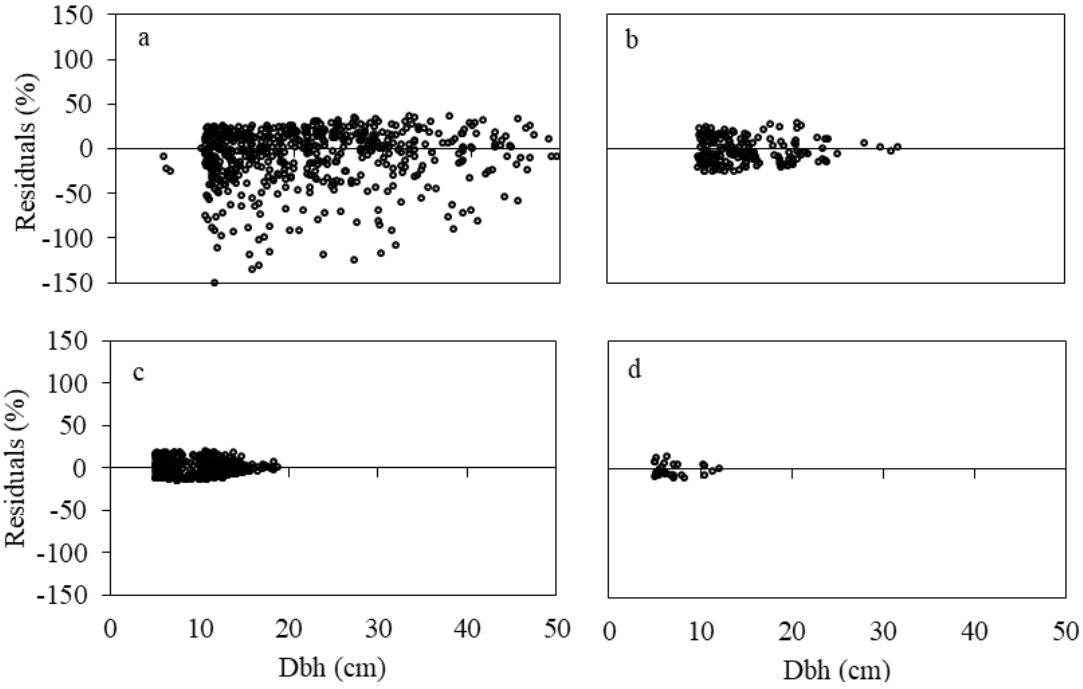


FIG. 14. PLOT OF RESIDUALS FROM THE HEIGHT-DIAMETER MODELING APPLIED TO THE STRATA (A) STRATUM 1, (B) STRATUM 2, (C) STRATUM 3, (D) STRATUM 4. DBH: DIAMETER AT BREAST HEIGHT. SEE STRATUM SPECIFICATIONS IN TABLE 7.

Unlike the heights of trees, the variables volume of stem, tree biomass and carbon were estimated by equations taken from literature. Fig. 16 shows estimates of these three variables, related to the diameter at breast height (Dbh).

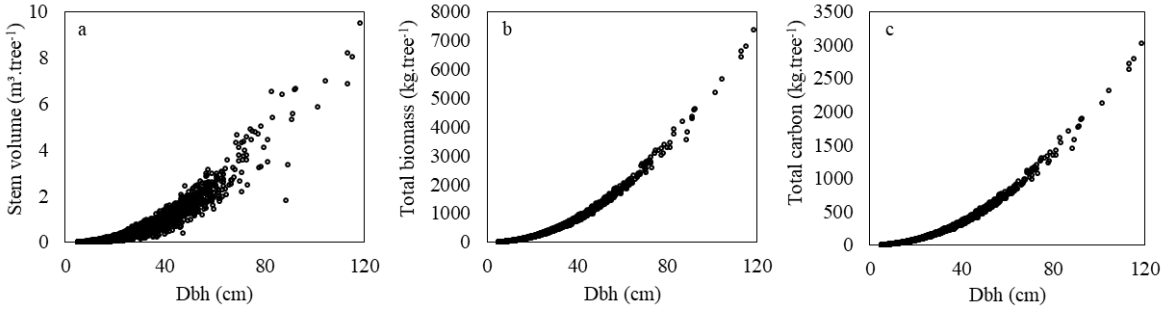


FIG. 15. ESTIMATION OF STEM VOLUME (A), TOTAL TREE BIOMASS (B) AND TOTAL TREE CARBON (C).

As the NFI's database does not provide observed volume of the stems, neither biomass and carbon of trees, unfortunately it is not possible evaluate the accuracy of these three variables. However, Fig. 16 shows that the estimated values by tree were coherent, with normal factor form ranging from 0.4579 to 0.7284, with average of 0.5536.

4.2 PIXEL-LEVEL ESTIMATES OF WOOD VOLUME

After estimating stem volumes at tree level (m^3) and wood volume at stand level ($m^3\ ha^{-1}$), two pixel-level methods for estimating wood volume were tested. The first of them, the SRM, is a linear model that uses, as input variables, surface reflectance values from Landsat-8 OLI bands, whose reflectances were transformed into Principal Components (PCs). Table 9 shows regression coefficients from this linear model fitted to each satellite scene, as well as their associated PCs.

The second pixel-level model for estimating wood volume, the RKM, is a preliminary step for the RK technique that spatialize the regression residuals of the variable of interest (i.e., wood volume). Regression coefficients from the RKM are presented at the bottom of Table 9. The assumptions $E(r_i) = 0$, $Cov(r_i\ y_i) = 0$, and $Cov(r_i\ x_i) = 0$ were met (Odeh et al., 1995).

TABLE 9. REGRESSION COEFFICIENTS OF PIXEL-LEVEL MODELS OF WOOD VOLUME AND ASSOCIATED PRINCIPAL COMPONENTS OF LANDSAT-8 OLI BANDS.

Surface reflectance model			
Scene	Final equation		
221/077	V = 638.15341 - 0.27073 PC1 + 0.49191 PC2		
221/078 and 223/076	V = 167.8204 - 0.70919 PC1		
222/077 and 223/078	V = 175.51159 - 0.09704 PC1 - 0.65708 PC2		
222/078	V = 94.30008 + 0.04800 PC1 - 0.12708 PC2 + 3.87062 PC3		
223/077	V = 44.18151 - 2.77950 PC1 + 4.33302 PC2		
Scene	Principal Component		
221/077	PC1: 0.99703 NIR + 0.07707 SWIR1		
	PC2: -0.07707 NIR + 0.99703 SWIR1		
221/078 and 223/076	PC1: 0.98484 Red - 0.13166 SWIR1 + 0.11294 SWIR2		
222/077 and 223/078	PC1: 0.91761 Brightness + 0.39749 SWIR2		
	PC2: -0.39749 Brightness + 0.917607 SWIR2		
222/078	PC1: 0.36164 Brightness + 0.91678 NIR + 0.16950 SWIR1		
	PC2: 0.28390 Brightness - 0.28146 NIR + 0.91662 SWIR1		
	PC3: 0.88804 Brightness - 0.28336 NIR - 0.36206. SWIR1		
223/077	PC1: -0.04918 Brightness - 0.00011 NDVI + 0.01766 NIR - 0.3610 SWIR1 + 0.9311 SWIR2		
	PC2: 0.90197 Brightness - 0.00187 NDVI - 0.30130 NIR - 0.30261 SWIR1 - 0.06398 SWIR2		
Scene	Principal Component	Eigenvalue	Cumulative variance (%)
221/077	PC1	55,710.9	89.3
	PC2	6,698.2	100.0
221/078 and 223/076	PC1	47,453.0	97.9
222/077 and 223/078	PC1	37,204.1	95.8
	PC2	1,635.6	4.2
222/078	PC1	131,862.0	92.7
	PC2	10,398.2	7.2

	PC3	40.0	0.1
223/077	PC1	6,8714.0	95.9
	PC2	2,817.8	99.8
Regression-kriging model			
Total area	V = 1051.51885 - 0.00031 Lat - 0.00011 Long		

V: wood volume, $m^3\ ha^{-1}$. PC: principal component. NIR: near infra-red. SWIR: short-wave infra-red. NDVI: normalized difference vegetation index. Lat: latitude. Long: longitude.

Table 10 gives goodness-of-fit statistics from the SRMs and RKM. With the low variation of the statistics between the fitting and validation datasets, the results indicate that the regression coefficients were accurately estimated and can be applied without loss of accuracy. However, the equations associated with the scenes 221/077 and 223/077 were fitted with a small database and therefore they have limited applications to other studies. Particularly to this research, such scenes have a small coverage over the study area (Fig. 12); therefore, the equations associated to them have low contribution (3.4% and 5.4%, respectively, Table 10) in the average accuracy.

TABLE 10. STATISTICS OF GOODNESS OF FIT FOR PIXEL-LEVEL MODELS OF VOLUME, AND WEIGHTS ASSOCIATED TO COVERAGE OF LANDSAT-8 SCENES.

Surface reflectance model								
Scene	Dataset	$R^2_{adj.}$	RMSE*	RMSE%	ME*	AME*	DF	w (%)
221/077	Fitting	0.61	59.22	30.7%	0.00	46.33	41	3.4%
	Validation	0.59	67.16	31.5%	18.78	37.64	4	
221/078 and 223/076	Fitting	0.11	28.53	27.4%	0.00	20.73	299	24.3%
	Validation	0.10	32.45	33.0%	-5.64	23.86	43	
222/077 and 223/078	Fitting	0.33	45.82	39.3%	0.00	36.56	283	23.0%
	Validation	0.32	53.22	41.8%	12.35	41.10	36	
222/078	Fitting	0.39	40.10	28.9%	0.00	33.08	542	44.0%
	Validation	0.39	42.63	29.0%	9.86	34.59	77	
223/077	Fitting	0.22	59.16	41.9%	0.00	49.53	66	5.4%
	Validation	0.18	55.31	51.0%	8.73	36.08	7	
Total area	Fitting	0.43	41.94	32.5%	-1.57	33.05	1,231	100%
	Validation	0.46	42.37	32.3%	-6.84	33.53	167	
Regression-kriging model								
Total area	Fitting	0.11	53.36	49.0%	0.00	41.55	60	100%
	Validation	0.11	41.78	47.2%	-13.23	26.57	10	

* Value given in $m^3\ ha^{-1}$. RMSE: Root mean square error. ME: Mean error. AME: Absolute mean error. DF: degrees of freedom. w: weight of coverage.

To the SRM, DF indicate the number of pixels taken from the areas occupied by forested clusters, minus the number of parameters of the model. To the RKM, DF was reduced to 60 (and 10, for validation) because each cluster have only one geographical location, which is its midpoint (Table 10).

The RKM had accuracy worse than the SRM (Table 10), however, it is important to highlight that this model is not used to directly estimate wood volume. Instead, estimated volumes and their respective regression residuals are spatialized and added each other in order to obtain the final estimated wood volume. Fig. 17a and 17c shows plotting of estimated and observed volumes for the total database, for both tested models. Fig. 17b and 17d shows the histogram of the regression residuals along with the normal curve.

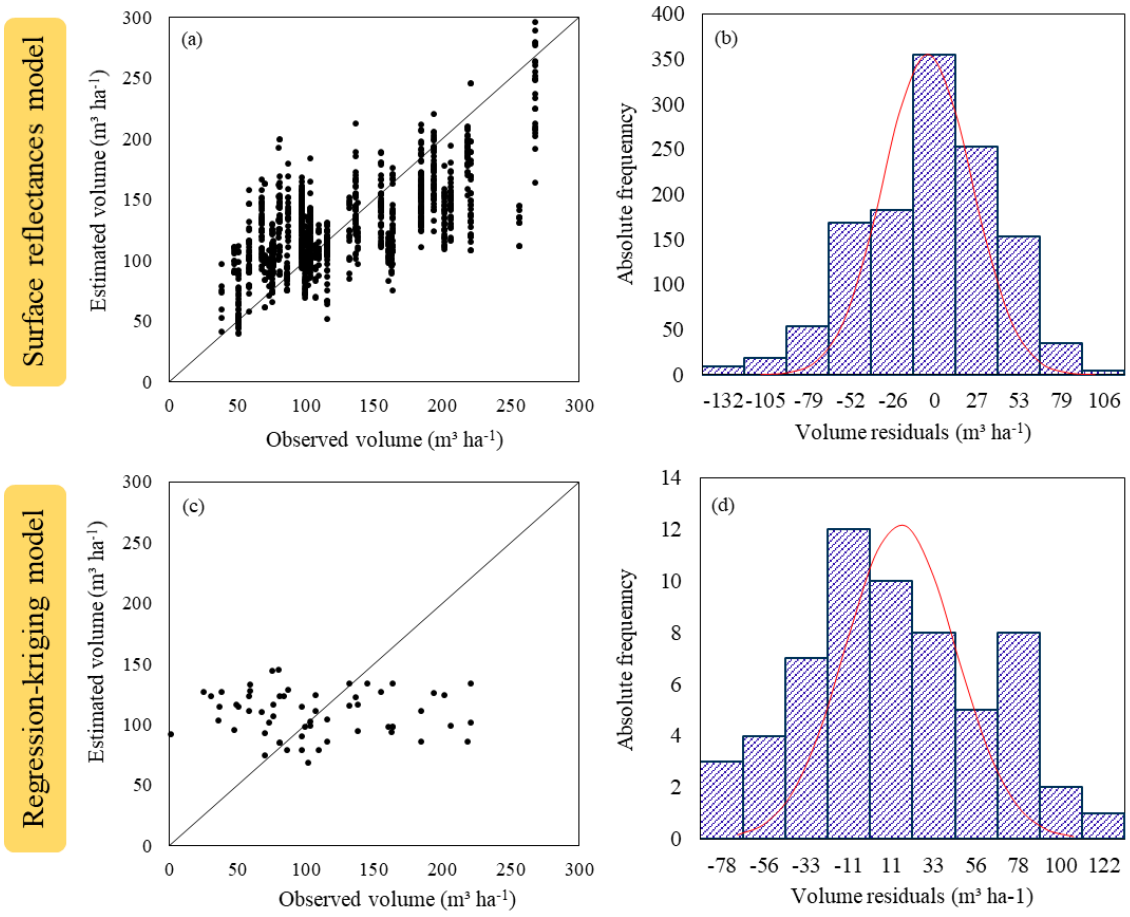


FIG. 16. PLOTTING OF ESTIMATED OVER OBSERVED VOLUMES (A AND C) AND HISTOGRAM OF RESIDUALS OF VOLUME ESTIMATES (B AND D). RED LINE: NORMAL DISTRIBUTION CURVE.

Fig. 17a reveals that the SRM had a moderate tendency in overestimating the lower values of wood volume and underestimating the higher ones. Estimates from the RKM, in turn, presented estimates seemingly dependent from at most one input variable, which was latitude (*p*-value > 0.05). However, a good accuracy of this analysis depends mainly on residuals (of the RKM) normally distributed, once normality is a condition necessary to apply the RK technique (Odeh et al., 1995).

In a visual analysis, Fig. 17b and 17d reveal a distribution of the regression residuals very close to the normality and it aligns with the values of ME (Table 10), in which positive and negative residuals nullify themselves, so that the overall ME is a tiny value around zero. The Shapiro-Wilk test confirms that the residuals from both models are normally distributed at 95% probability level.

In addition to the normality, another issue of greater importance for the application of the RKM is that the regression residuals have spatial dependence further than ~20 km, which is the minimum distance among clusters. The practical asymptotic range (95% of the sill) of the EM, which is equal to three times the value of range (φ) (Wackernagel, 2003), was of 269.6 km, meaning that there is spatial dependence until such distance. With this result, the residuals can be properly spatialized with the NFI grid distance. By following the Formula [25] and classes from Table 6, residuals were classified as moderately spatially dependent ($r(h)$ = 55.32%). The equation of the semivariogram is expressed in [32].

$$\gamma(h) = 16.3804 + 13.2294\left[1 - e^{-\left(\frac{h}{90{,}000}\right)}\right] \quad [32]$$

Where,

γ(h): semivariance. h: distance between clusters. *e*: exponential.

4.3 PIXEL-LEVEL ESTIMATES OF FOREST BIOMASS AND CARBON

After estimating total biomass and carbon at tree and stand levels (Mg ha^{-1}), the SRM was applied to estimate these variables at pixel level, as accomplished for wood volume (section [4.2]). As proposed in the methodology, carbon was directly estimated by a ratio of biomass-carbon = 0.41 (Table 2). The RK method was not tested here, as for volume (section [4.2]). Regression coefficients from the reflectance model and from the PCs are presented in Table 11.

TABLE 11. REGRESSION COEFFICIENTS OF PIXEL-LEVEL MODELS FOR BIOMASS AND CARBON, AND ASSOCIATED PRINCIPAL COMPONENTS OF LANDSAT-8 OLI BANDS.

Scene	Final equation
221/077	Bm = 512,194.85908 - 217.12670 PC1 + 404.56031 PC2 / C = 0.41 Bm
221/078 and 223/076	Bm = 138,297.46773 - 8.79914 PC1 + 37.16360 PC2 / C = 0.41 Bm
222/077 and 223/078	Bm = 190,136.4365 - 107.03014 PC1 - 600.65401 PC2 / C = 0.41 Bm
222/078	Bm = 114,958.7886 + 30.73977 PC1 - 83.48692 PC2 / C = 0.41 Bm

223/077	Bm = 95,020.79487 + 38.47558 PC1 - 111.54384 PC2 - 735.56314 PC3 / C = 0.41 Bm

Scene	Principal Component
221/077	PC1: 0.99703 NIR + 0.07707 SWIR1
	PC2: - 0.07707 NIR + 0.99703 SWIR1
221/078 and 223/076	PC1: 0.05796 Red + 0.72880 SWIR1 + 0.38012 SWIR2 + 4.93E-07 NDVI + 0.56657 Brightness
	PC2: -0.05638 Red + 0.37664 SWIR1 + 0.47177 SWIR2 - 0.00006 NDVI - 0.79524 Brightness
222/077 and 223/078	PC1: 0.91691 Brightness + 0.39909 SWIR2
	PC2: - 0.39909 Brightness + 0.91691 SWIR2
222/078	PC1: - 0.00430 Red + 0.93232 NIR + 0.01194 SWIR2 + 0.00007 NDVI + 0.36143 Brightness
	PC2: 0.31315 Red - 0.17400 NIR + 0.83123 SWIR2 - 0.00023 NDVI + 0.42510 Brightness
223/077	PC1: 0.05386 Red + 0.49415 SWIR1 + 0.19831 SWIR2 + 0.00003 NDVI + 0.84474 Brightness
	PC2: 0.05070 Red + 0.77847 SWIR1 + 0.32462 SWIR2 - 0.00010 NDVI - 0.53482 Brightness
	PC3: 0.73922 Red - 0.30507 SWIR1 + 0.60033 SWIR2 - 0.00042 NDVI - 0.00960 Brightness

Scene	Principal Component	Eigenvalue	Cumulative variance (%)
221/077	PC1	55,710.9	89.3
	PC2	6,698.2	100.0
221/078 and 223/076	PC1	66,541.2	89.2
	PC2	7,500.1	99.3
222/077 and 223/078	PC1	37,204.1	95.8
	PC2	1635,62	100.0
222/078	PC1	128,533.0	94.8
	PC2	6,597.4	99.7
223/077	PC1	12,443.1	88.2
	PC2	1,435.2	98.3
	PC3	143.2	99.4

Bm: forest biomass, Mg ha^{-1}. C: forest carbon, Mg ha^{-1}. PC: principal component. NIR: near infra-red. SWIR: short-wave infra-red. NDVI: normalized difference vegetation index.

Table 12 and Table 13 give, respectively, goodness-of-fit statistics from the SRMs fitted to biomass and carbon, to each Landsat scene.

TABLE 12. STATISTICS OF GOODNESS OF FIT FOR PIXEL-LEVEL MODELS FOR BIOMASS, AND WEIGHTS ASSOCIATED TO COVERAGE OF LANDSAT-8 SCENES.

Scene	Dataset	$R^2_{adj.}$	RMSE*	RMSE%	ME*	AME*	DF	w (%)
221/077	Fitting	0.63	45,776.67	27.3%	0.00	35,713.43	41	3.4%
	Validation	0.62	40,372.97	26.8%	12,253.03	33,454.62	4	
221/078 and 223/076	Fitting	0.02	26,217.43	26.6%	0.00	20,143.75	299	24.3%
	Validation	0.01	26,210.96	27.3%	-3,059.96	21,261.50	43	
222/077 and 223/078	Fitting	0.42	37,260.77	35.6%	0.00	31,131.66	283	23.0%
	Validation	0.41	38,746.82	38.8%	-4,376.26	32,412.92	36	
222/078	Fitting	0.09	41,019.99	31.3%	0.00	34,864.16	542	44.0%
	Validation	0.08	38,063.67	30.3%	-5,244.76	33,435.83	77	
223/077	Fitting	0.05	48,902.19	39.3%	0.00	43,771.85	66	5.4%
	Validation	0.05	42,379.61	35.2%	-3,998.16	38,509.64	7	
Total area	Fitting	0.35	37,500.13	31.7%	0.00	30,979.30	1,231	100%

	Validation	0.34	35,767.86	31.7%	-4,017.89	30,239.97	167	

* Value given in Mg ha^{-1}. RMSE: Root mean square error. ME: Mean error. AME: Absolute mean error. DF: degrees of freedom. w: weight of coverage.

TABLE 13. STATISTICS OF GOODNESS OF FIT FOR PIXEL-LEVEL MODELS FOR CARBON, AND WEIGHTS ASSOCIATED TO COVERAGE OF LANDSAT-8 SCENES.

Scene	Dataset	$R^2_{adj.}$	RMSE*	RMSE%	ME*	AME*	DF	w (%)
221/077	Fitting	0.63	18,768.43	27.3%	0.00	14,642.51	41	3.4%
	Validation	0.62	16,552.92	26.8%	5,023.74	13,716.39	4	
221/078 and 223/076	Fitting	0.02	10,749.15	26.6%	0.00	8,258.94	299	24.3%
	Validation	0.01	10,746.49	27.3%	-1,254.58	8,717.22	43	
222/077 and 223/078	Fitting	0.42	15,276.92	35.6%	0.00	12,763.98	283	23.0%
	Validation	0.41	15,886.20	38.8%	-1,794.27	13,289.30	36	
222/078	Fitting	0.09	16,818.20	31.3%	0.00	14,294.31	542	44.0%
	Validation	0.08	15,606.10	30.3%	-2,150.35	13,708.69	77	
223/077	Fitting	0.05	20,049.90	39.3%	0.00	17,946.46	66	5.4%
	Validation	0.05	17,375.64	35.2%	-1,639.25	15,788.95	7	
Total area	Fitting	0.35	15,375.05	31.7%	0.00	12,701.51	1,231	100%
	Validation	0.34	14,664.82	31.7%	-1,647.33	12,398.39	167	

* Value given in Mg ha^{-1}. RMSE: Root mean square error. ME: Mean error. AME: Absolute mean error. DF: degrees of freedom. w: weight of coverage.

In general, the SRMs fitted for biomass (Table 12) and carbon (Table 13) also had an acceptable performance, aligned with biomass and carbon surveys. As in the volume models, the validation analysis also confirms the great performances of the biomass and carbon models.

Fig. 18a shows plot of estimated and observed forest biomass for the total database. Fig. 18b shows the histogram of residuals of the estimates along with the normal curve. Plots for the variable carbon are not presented because they have the same behavior as biomass.

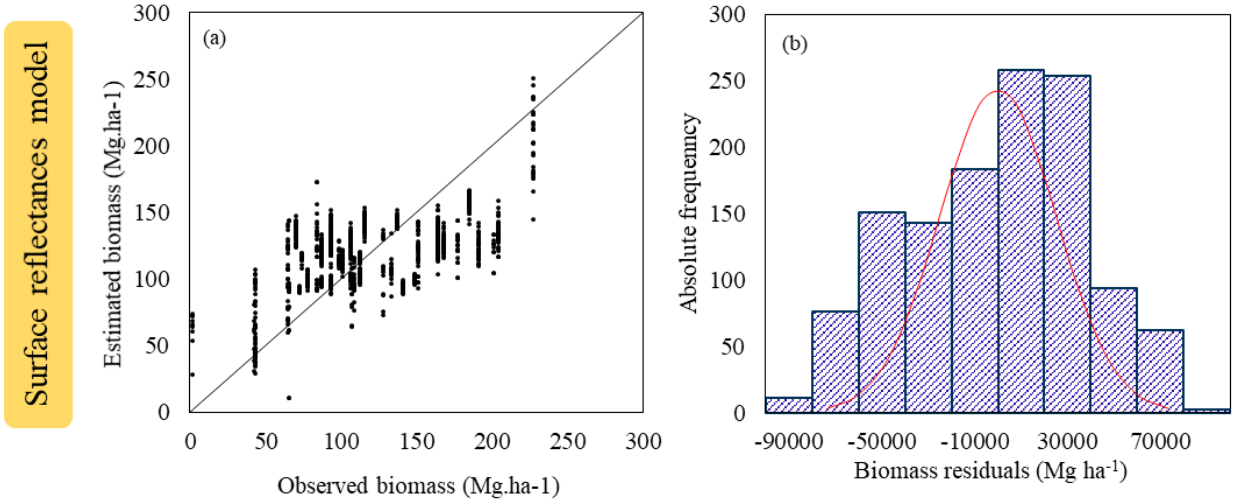


FIG. 17. PLOTTING OF ESTIMATED OVER OBSERVED BIOMASS (A) AND HISTOGRAM OF RESIDUALS OF BIOMASS ESTIMATES (B). RED LINE: NORMAL DISTRIBUTION CURVE.

4.4 IMAGE SEGMENTATION AND CLASSIFICATION

The image segmentation process provided by eCognition Developer 8.7 resulted in so homogeneous objects, in some cases, that the smallest objects reached size of up to one pixel. Fig. 19 shows samples of results of the image segmentation and classification in three common land-uses in the State of Parana.

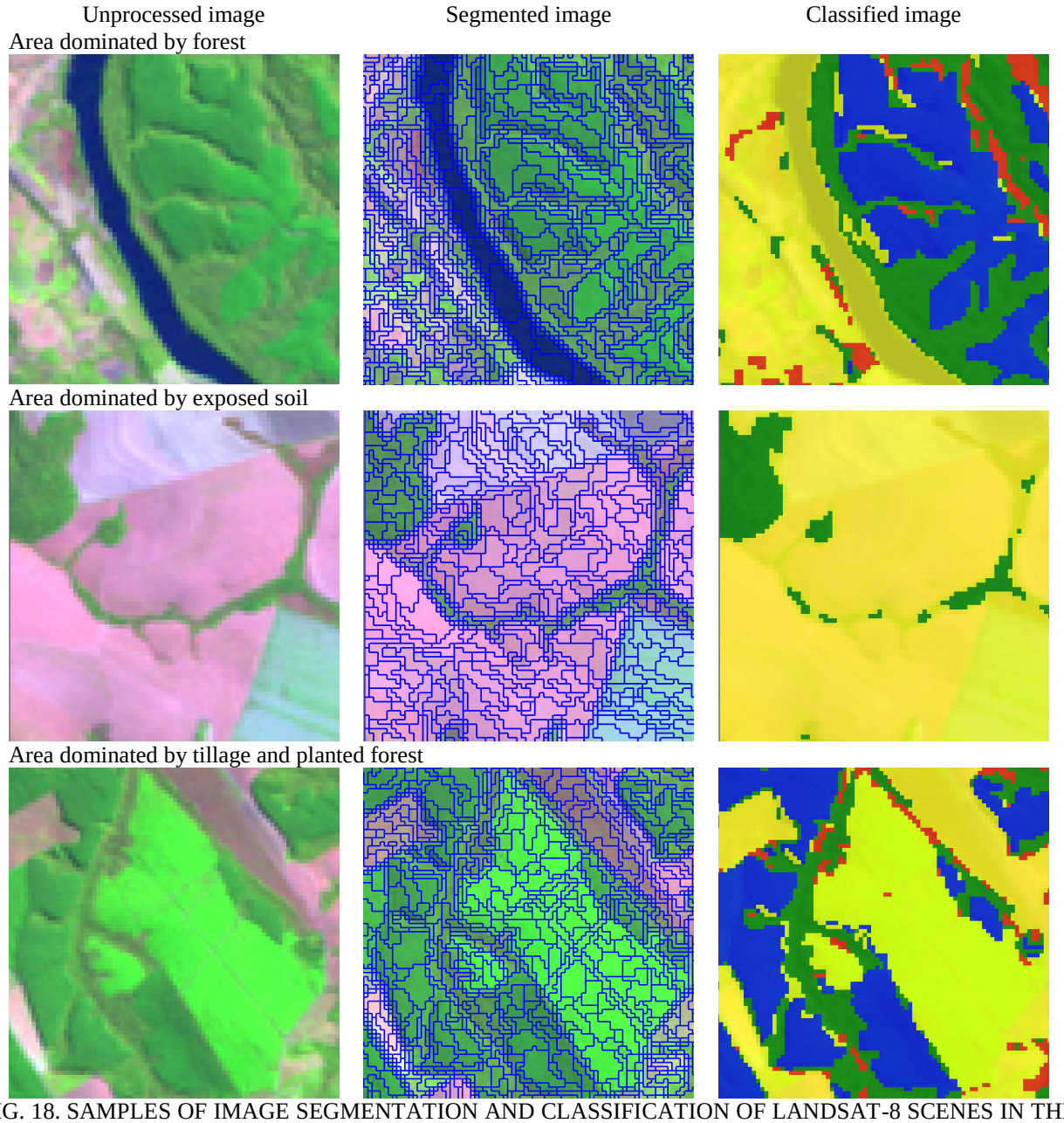


FIG. 18. SAMPLES OF IMAGE SEGMENTATION AND CLASSIFICATION OF LANDSAT-8 SCENES IN THE STATE OF PARANA. GREEN PATCHES: MID- TO LATE-SUCCESSIONAL FORESTS. RED PATCHES: EARLY-SUCCESSIONAL FORESTS. BLUE PATCHES: PLANTED FORESTS. YELLOW PATCHES: NON-FORESTED AREAS. BLUE LINES: DELIMITATION OF OBJECTS (SEGMENTATION).

The results obtained in the image segmentation indicate that the adopted parameters and weighs of image layers satisfied a high-quality pattern of segmentation. Accurate results also were obtained in the image classification; as shown in Table 14, kappa values ranged from 0.81 to 0.90, which correspond to almost perfect agreement (Table 5).

TABLE 14. CONFUSION MATRIX OF THE IMAGE CLASSIFICATIONS BY LANDSAT SCENE.

221/077 (kappa = 0.86)					
Class	PF	ESF	MLSF	NF	Total
PF	32	3	2	1	38
ESF	1	24	2		27
MLSF	2	1	26		29
NF		2		34	36
Total	35	30	30	35	130

221/078 (kappa = 0.81)					
Class	PF	ESF	MLSF	NF	Total
PF	27		1	1	29
ESF		23	1	4	28
MLSF			28		28
NF	3	7		25	35
Total	30	30	30	30	120

222/077 (kappa = 0.86)					
Class	PF	ESF	MLSF	NF	Total
PF	28		1		29
ESF		23			23
MLSF	1		26		27
NF	1	7	3	30	41
Total	30	30	30	30	120

222/078 (kappa = 0.87)					
Class	PF	ESF	MLSF	NF	Total
PF	29		2	1	32
ESF		24	1		25
MLSF	1		26		27
NF		6	1	29	36
Total	30	30	30	30	120

223/076 (kappa = 0.90)					
Class	PF	ESF	MLSF	NF	Total
PF	27				27
ESF	3	28	2	1	34
MLSF		2	27		29
NF			1	29	30
Total	30	30	30	30	120

223/077 (kappa = 0.90)					
Class	PF	ESF	MLSF	NF	Total
PF	26				26
ESF		28	2		30
MLSF	4		27		31
NF		2	1	30	33
Total	30	30	30	30	120

223/078 (kappa = 0.86)					
Class	PF	ESF	MLSF	NF	Total
PF	28				28
ESF		22		1	23
MLSF	2		28		30
NF		8	2	29	39
Total	30	30	30	30	120

PF: Planted forests. ESF: Early-successional forests. MLSF: Mid- to late-successional forests. NF: Non-forest.

Fig. 20 and appendices 3 to 6 show further samples of results of the image classifications. As the image segmentation attained a high level of refinement (Fig. 19), it is possible to recognize fine details along the classified imagery, such as tracks of main roads, electrical transmission lines, narrow patches of riparian forests, and waterways crossing forests (Fig. 20, appendices 3 to 6).

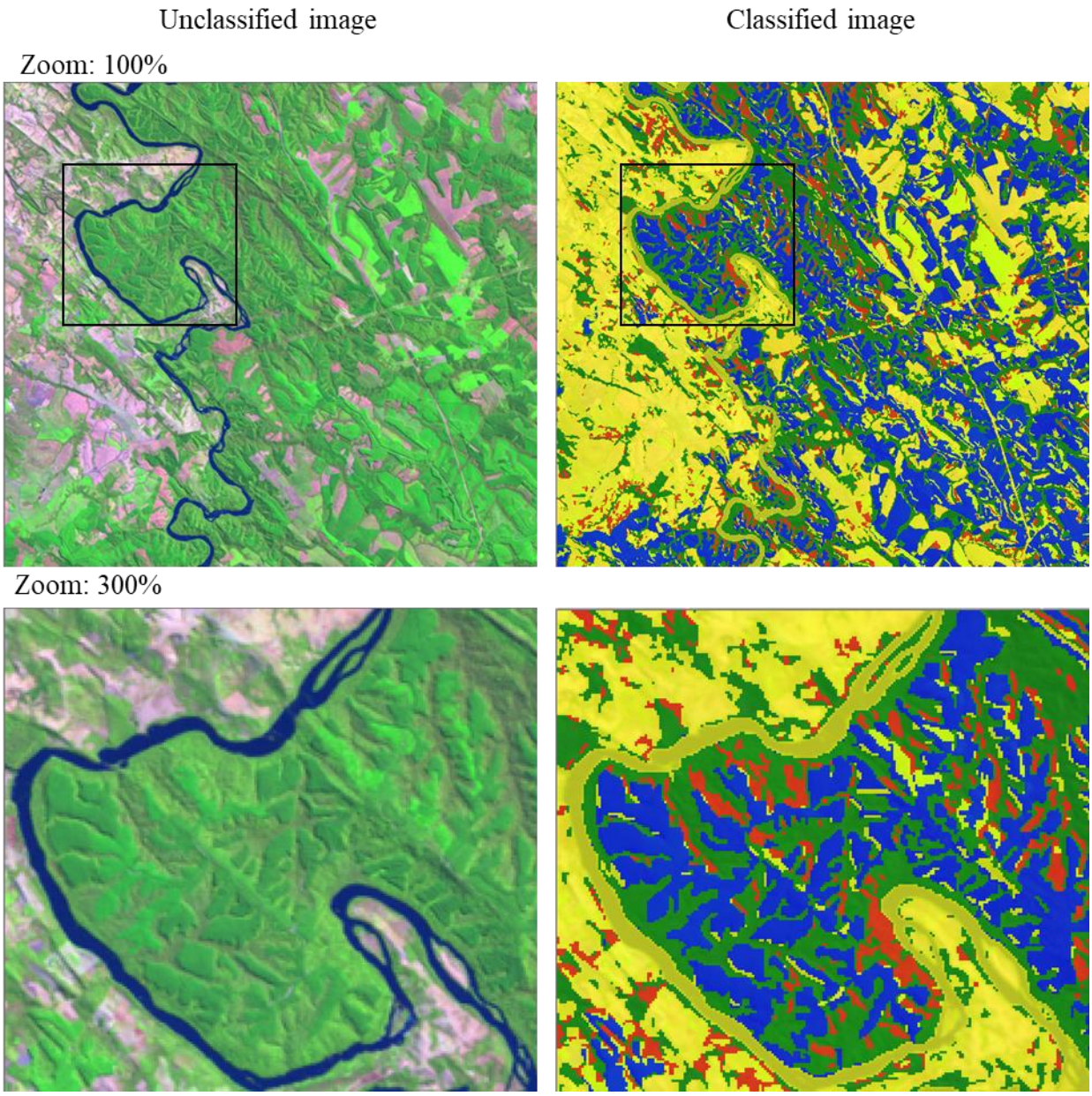


FIG. 19. SAMPLES OF IMAGE CLASSIFICATION OF LANDSAT-8 SCENE (221/77) IN THE STATE OF PARANA. GREEN PATCHES: MID- TO LATE-SUCCESSIONAL FORESTS. RED PATCHES: EARLY-SUCCESSIONAL FORESTS. BLUE PATCHES: PLANTED FORESTS. YELLOW PATCHES: NON-FORESTED AREAS.

Misclassifications occurred mainly due to early-successional forests be erroneously classified as non-forest (e.g., tillage, pasture), and vice versa, followed by shadow effects, and planted forest be misclassified as mid- to late-successional forest, and vice-versa. Most of the misclassifications due to shadow effects occurred in the south of the study area, where is the most mountainous areas. Such regions had more manual edits on classification.

Fig. 21 provides the result of image classification for the study area, as well as presents a ground slope map originated from SRTM (30-m spatial resolution) available on the internet.

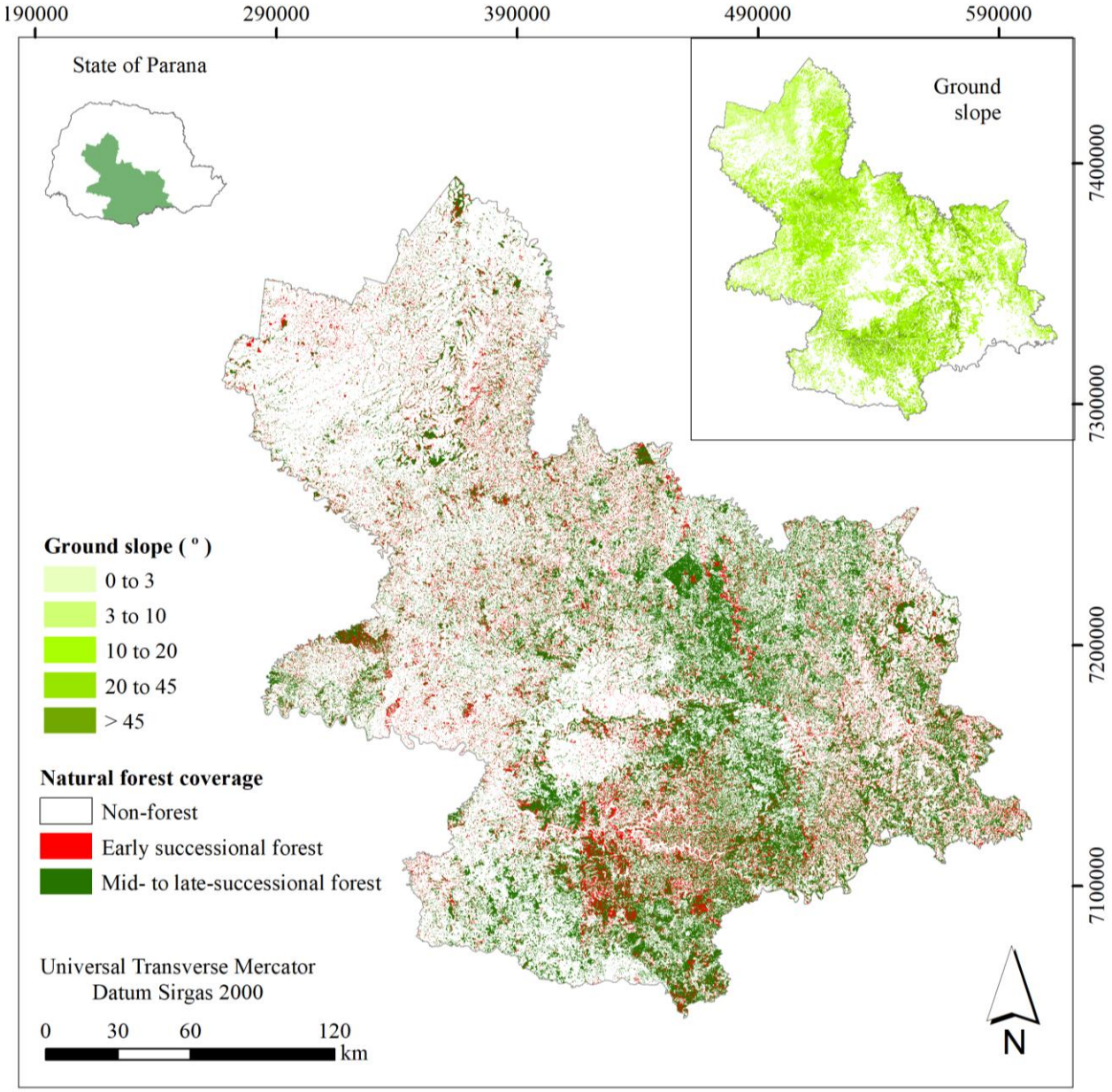


FIG. 20. SPATIAL DISTRIBUTION OF EARLY-SUCCESSIONAL AND MID- TO LATE-SUCCESSIONAL FORESTS AND GROUND SLOPE IN PART OF THE STATE OF PARANA.

The total area classified as ‘early-successional forest’ and ‘mid- to late-successional forest’ (Fig. 21) were 432,528 ha and 1,330,041 ha, respectively. These two classes combined totalize 1,762,569 ha of forested area. Estimates from the current decade indicate that the Atlantic Forest biome shelters only 11.3% of its original forest area, i.e., 88.7% has been lost to other land uses (Ribeiro et al., 2009). The findings indicate that 31.7% of the study area is still covered by natural

forest; almost three times the average of the biome Atlantic Forest. As shown in Fig. 21, the most deforested regions are in the North, Northwest, and Mid-west of the study area. Fig. 22 approximates the scale for better visualizing the classifications in the North/Northwest, and Fig. 23 approximates for the mid-west region.

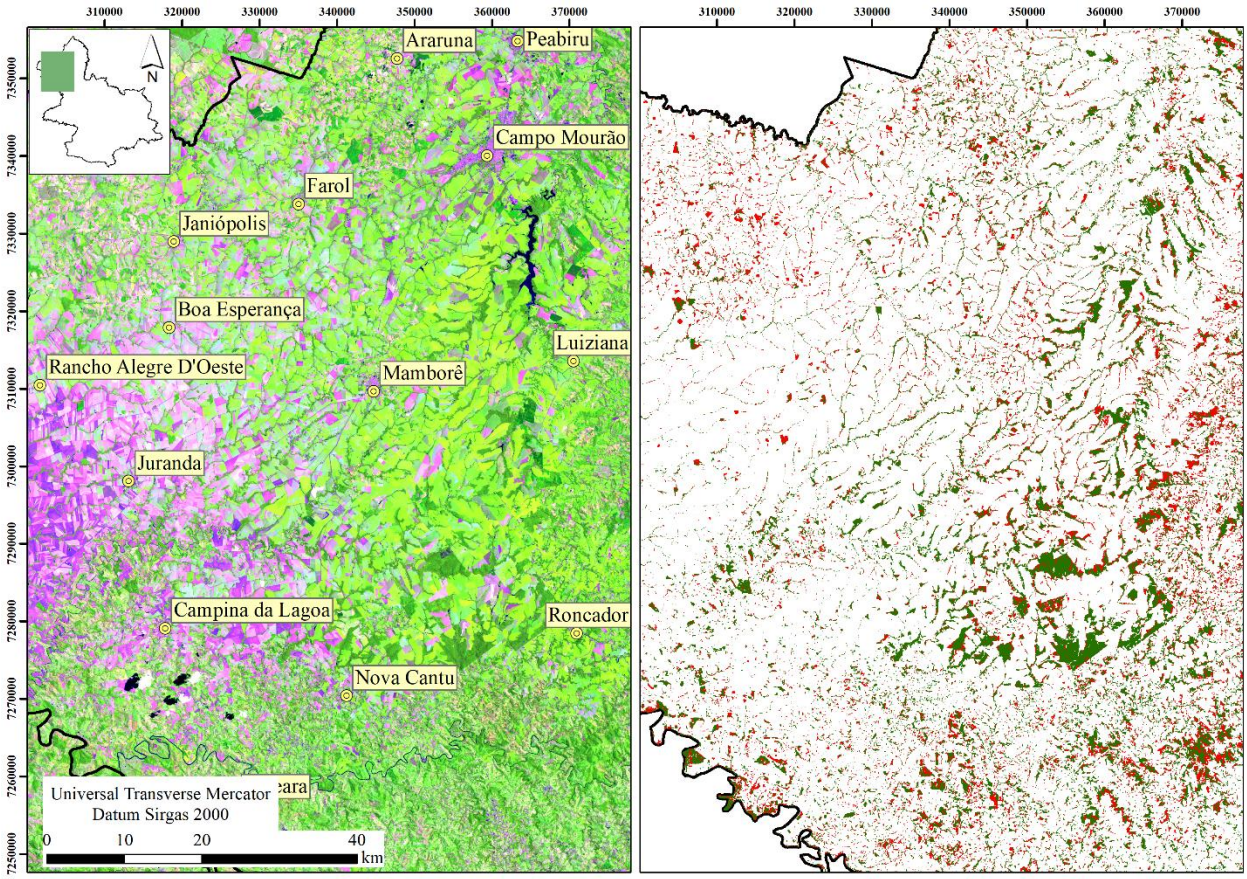


FIG. 21. LAND-USE AND CLASSIFICATION IN THE NORTHERN REGION OF THE STUDY AREA. GREEN PATCHES: MID- TO LATE-SUCCESSIONAL FORESTS. RED PATCHES: EARLY-SUCCESSIONAL FORESTS.

Reports from Agriculture and Supply Department (SEAB, 2017) point out that these regions (North, Northwest, and Mid-west) had large portions of lands devoted to crops as soybeans, corn, sugarcane, among others. As example, over the 2014-2015 period, the whole State of Parana designated more than 10.6 million ha to agricultural crops, from which ~1.1 million ha came from 21 counties highlighted in Fig. 22 and Fig. 23. In addition, six of the 25 most productive (in terms of agriculture) counties belong to these highlighted regions, being: Mamborê (9th most productive), Campo Mourão (16th), Luiziana (18th), and Rocandor (24th), in the Northern region (Fig. 22), besides Guarapuava (5th), and Candói (15th), in the Mid-west of the study area (Fig. 23).

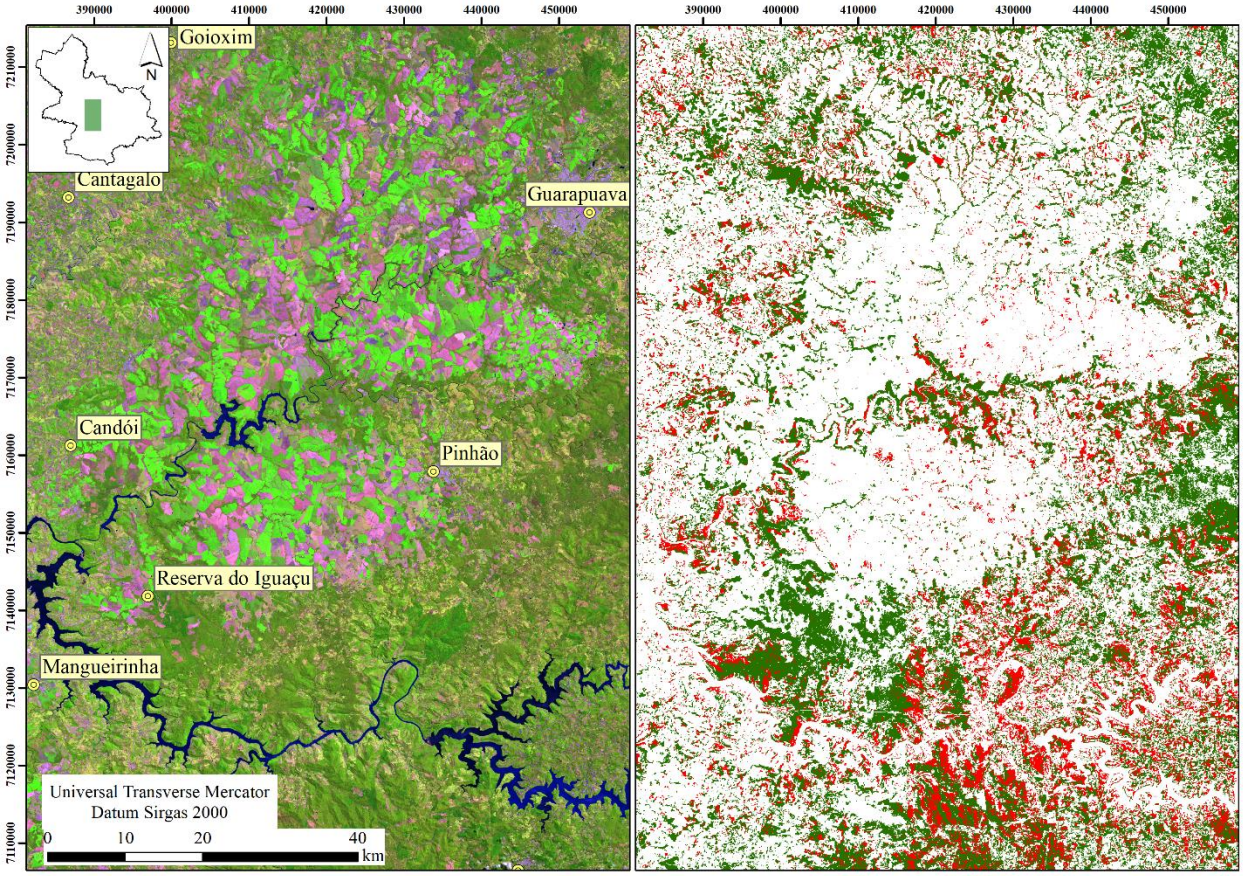


FIG. 22. LAND-USE AND CLASSIFICATION IN THE MID-WEST REGION OF THE STUDY AREA. GREEN PATCHES: MID- TO LATE-SUCCESSIONAL FORESTS. RED PATCHES: EARLY-SUCCESSIONAL FORESTS.

Another reason for these regions be the most deforested is by being in flat grounds. The maps shown in Fig. 21 indicate a strong relation between natural forest coverage and ground slope. This is specially evidenced in the regions pointed in Fig. 22 and Fig. 23, whose ground slopes are slighter. On the other hand, the South and the East appear as the most forested regions and are across the sharper reliefs (Fig. 24). Therefore, most of the manual edits were made in these declined regions, in order to fix misclassifications of very shadowed and darker image objects.

Two of the most important Parana's conservation units (CUs) are located in these regions: 'Área de Proteção Ambiental da Serra da Esperança', occupying an area of ~206,555 ha, and 'Reserva Biológica das Araucárias', with 14,930 ha. Fig. 24 approximates the scale for better visualizing the land use and classification in the southern region.

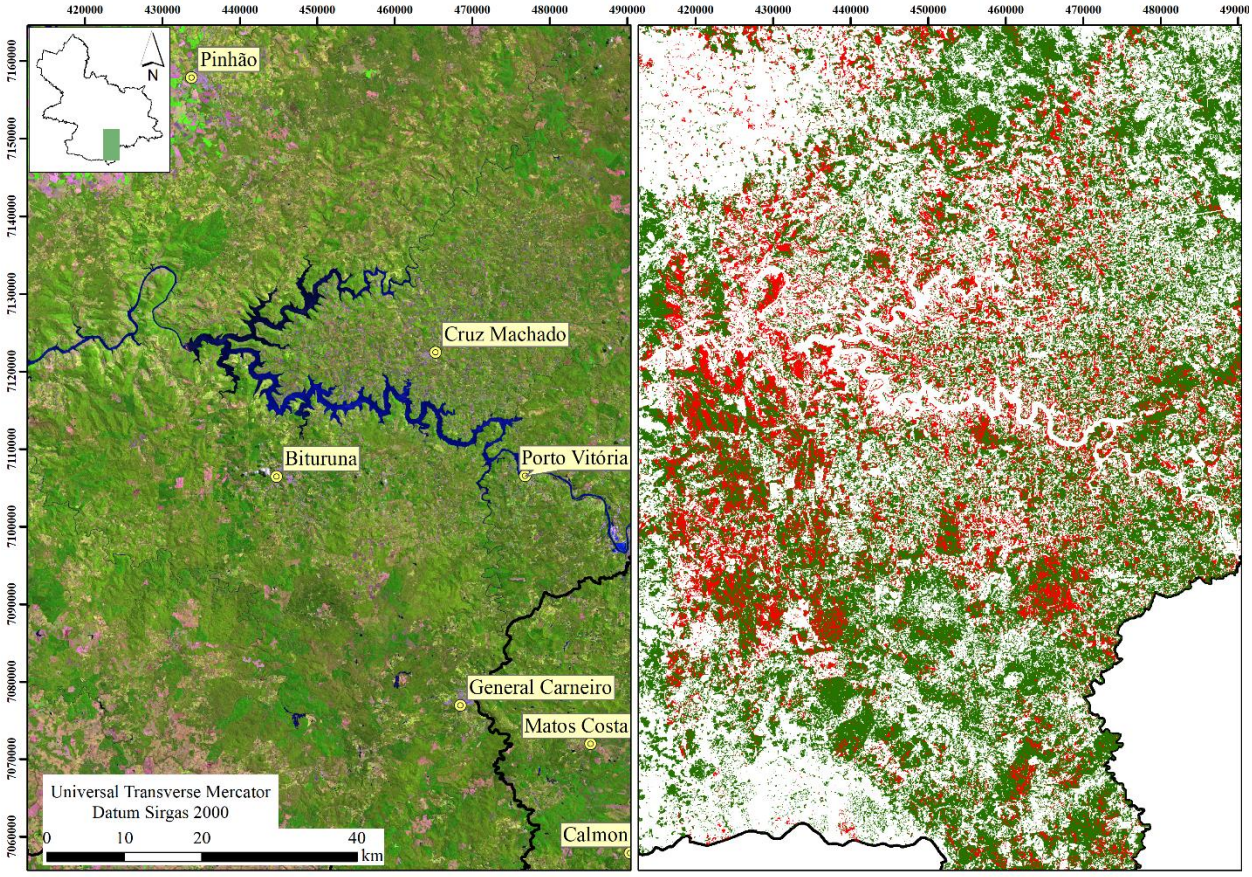


FIG. 23. LAND-USE AND CLASSIFICATION IN THE SOUTHERN REGION OF THE STUDY AREA. GREEN PATCHES: MID- TO LATE-SUCCESSIONAL FORESTS. RED PATCHES: EARLY-SUCCESSIONAL FORESTS.

4.5 COARSE-SCALE ESTIMATION OF WOOD VOLUME

4.5.1 Extrapolation unsupported by image classification

As proposed, the coarse-scale estimation consisted of directly extrapolating wood volume ($m^3\ ha^{-1}$) observed in clusters, to the area that each cluster represents, i.e., ~400 km^2. Table 15 gives sampling error (obtained through MFD, adapted from Chacko, 1965), totals, and other main results of the wood volume. Table 15 also shows frequency by type of cluster and their represented area, i.e., 400 km^2 each. The extrapolated forested area (EFA) means the proportion of forest found in the cluster extrapolated to its represented area. Total volume was obtained by multiplying mean volume and EFA.

TABLE 15. MAIN SAMPLING STATISTICS FOR THE VARIABLE WOOD VOLUME, WITH EXTRAPOLATION OF TOTALS UNSUPPORTED BY IMAGE CLASSIFICATION.

Statistics		Unit	Type of cluster			
			Entirely forested (A)	Partially forested (B)	(C) Non-Forested	Total (A+B)
Frequency*		absolute	7	89	56	96
		%	4.6%	58.6%	36.8%	63.2%
Represented area		ha	280,000	3,560,000	2,240,000	3,840,000
		%	4.6%	58.6%	36.8%	63.2%
Extrapolated forested area		ha	280,000	1,707,000	0.0	1,987,000
		%	14.1%	85.9%	0.0%	100.0%
Sampling error		$m^3\ ha^{-1}$	18.0	15.8	-	13.6
		%	8.3%	19.8%	-	14.7%
Confidence interval (5%)	Lower limit	$m^3\ ha^{-1}$	198.6	64.3	-	84.8
	Mean	$m^3\ ha^{-1}$	216.6	80.1	-	99.3
	Upper limit	$m^3\ ha^{-1}$	234.6	96.0	-	113.9
	Lower limit	m^3	55,603,212	109,703,290	-	168,468,681
	Total	m^3	60,639,475	136,757,044	-	197,396,518
	Upper limit	m^3	65,675,737	163,810,798	-	226,324,356

*Total frequency of forested and non-forested clusters (A+B+C) = 152 (100%).

4.5.2 Extrapolation supported by image classification

As a second alternative of extrapolating wood volume at coarse scale, the extrapolation supported by image classification is more complex than the first alternative, by requiring remote sensing applications. Total volume is obtained by multiplying mean volume and forested area delimited by image classification (FADIC), instead of multiplying by the EFA, as in the first case. Table 16 shows the FADIC and total volume by type of cluster.

TABLE 16. MAIN SAMPLING STATISTICS FOR THE VARIABLE WOOD VOLUME, WITH EXTRAPOLATION OF TOTALS SUPPORTED BY IMAGE CLASSIFICATION.

Statistics	Unit	Type of cluster			
		Entirely forested (A)	Partially forested (B)	(C) Non-Forested	Total (A+B+C)
Frequency	absolute	7	89	56	152
	%	4.6%	58.6%	36.8%	100.0%
Forested area delimited by image classification	ha	97,987	1,186,667	477,916	1,762,569
	%	5.6%	67.3%	27.1%	100.0%
Total	m^3	21,060,437	89,952,970	-	111,013,407

Fig. 25 illustrates the spatial distribution of wood volume by cluster, and the area occupied by each one.

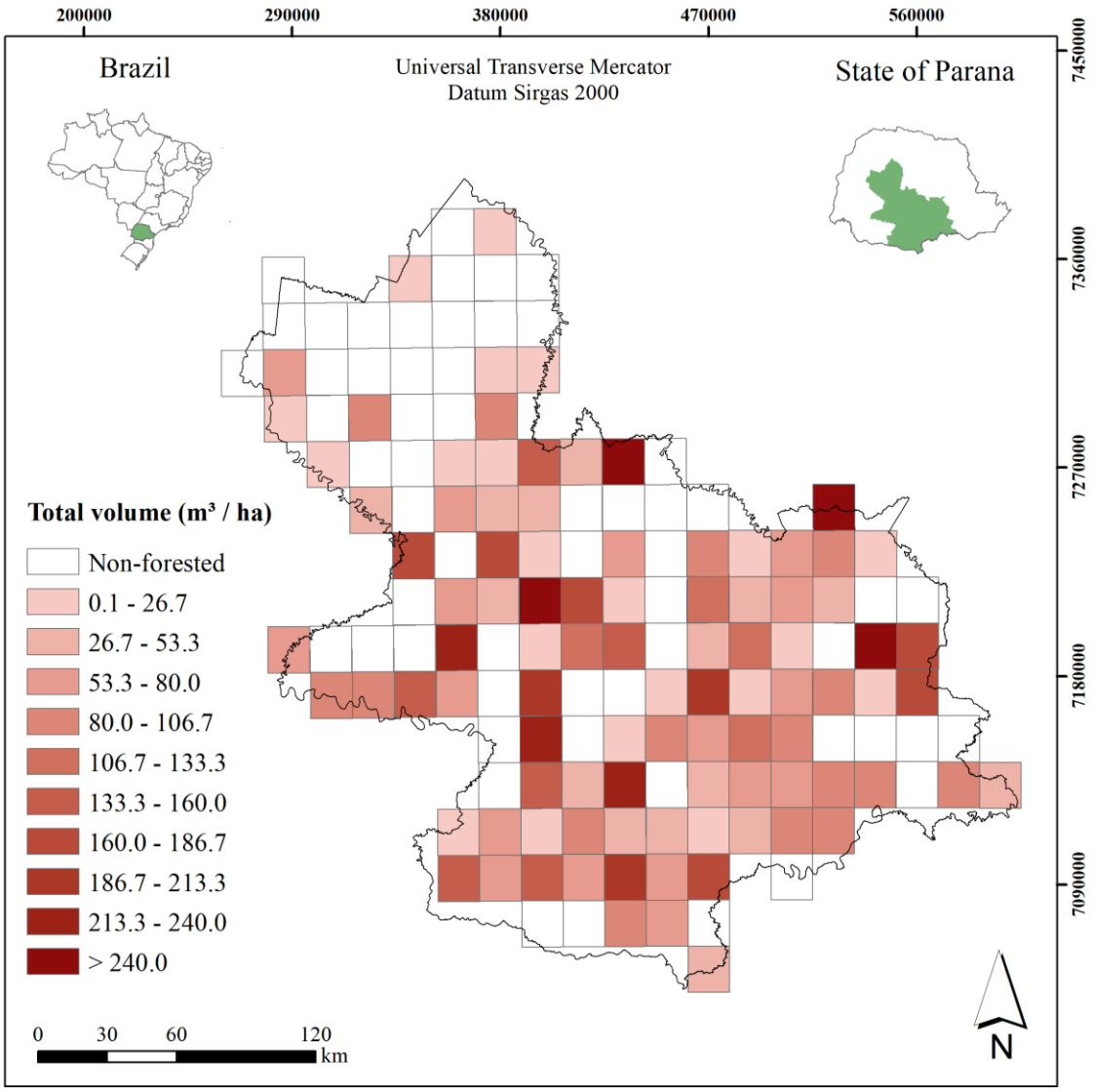


FIG. 24. SPATIAL DISTRIBUTION OF WOOD VOLUME BY EACH CLUSTER.

4.6 FINE-SCALE ESTIMATION OF WOOD VOLUME

By means of the SRMs, estimation of mean wood volume was of ~123.7849 $m^3\ ha^{-1}$ (or 10.8099 $m^3\ pixel^{-1}$, where 1 pixel = 900 m^2); considering the total forested area of 1,762,569 ha observed for the whole study area (Fig. 26), the total wood volume is of 210,961,589 m^3. With the RK geostatistical method, in turn, the mean volume was of ~115.7 $m^3\ ha^{-1}$ (or 10.4109 $m^3\ pixel^{-1}$) and 203,326,674 m^3 for the study area. As each of these methods (linear regression and regression-kriging) estimated wood volume at pixel level, the frequency distribution of pixels is presented in classes of volume in Fig. 26.

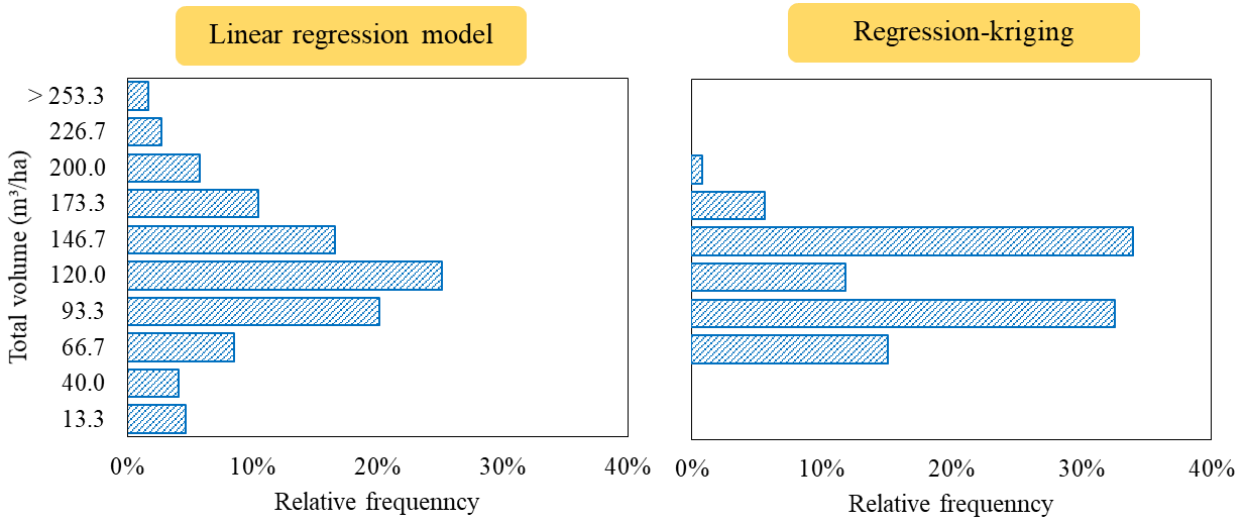


FIG. 25. FREQUENCY OF PIXELS *VERSUS* CLASSES OF WOOD VOLUME ESTIMATED BY PIXEL-LEVEL METHODS OF MAPPING.

The method that uses surface reflectances as input variables produced a frequency of volumes much closer to a normal distribution than the regression-kriging method. In addition, this geostatistical method failed in estimating volume of pixels in extreme classes. The spatial distribution of the wood volume ($m^3\ ha^{-1}$) estimated through the SRM and RK, are shown in Fig. 27 and Fig. 28, respectively.

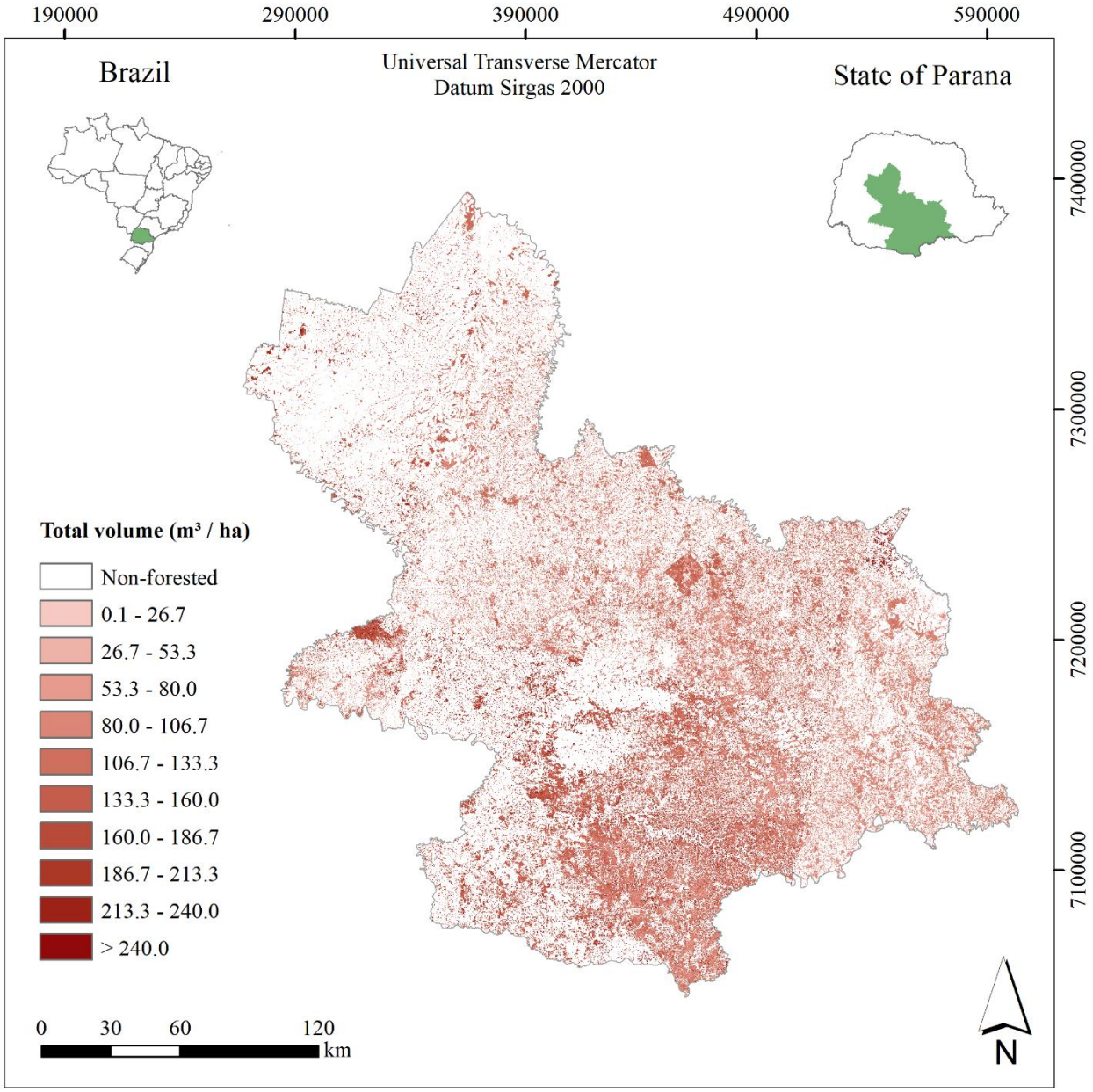


FIG. 26. SPATIAL DISTRIBUTION OF TOTAL TREE VOLUME ESTIMATED IN FORESTED PATCHES FROM MODELING WITH SURFACE REFLECTANCE AND IMAGERY CLASSIFICATION.

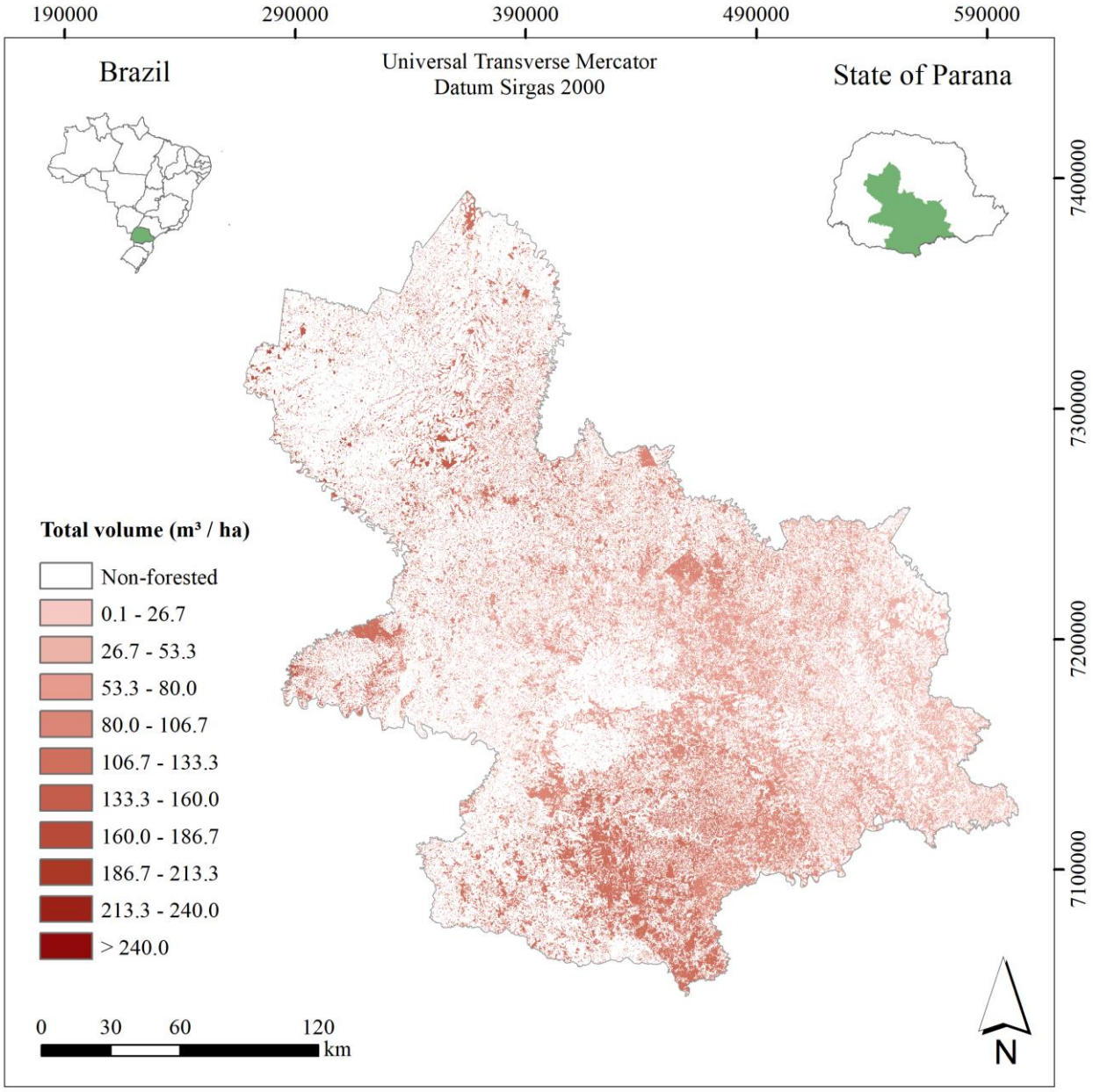


FIG. 27. SPATIAL DISTRIBUTION OF TOTAL TREE VOLUME ESTIMATED IN FORESTED PATCHES FROM REGRESSION KRIGING METHOD.

4.7 FINE-SCALE ESTIMATION OF FOREST BIOMASS AND CARBON

In relation to the biomass estimates, the mean biomass was of ~120.1156 Mg ha^{-1} (or 10.8104 Mg $pixel^{-1}$) and a total biomass of 217,736,862 Mg stored in the study area. Considering the ratio 1 unit of biomass to 0.41 of carbon (Table 2), the mean carbon stock results in ~49.24738 MgC ha^{-1} (or 4.4323 MgC $pixel^{-1}$), meaning that the whole study area stores 89,272,113 MgC, i.e.,

approximately 327,331,082 Mg of CO_2-equivalent. Fig. 29 illustrates the spatial distribution by classes of biomass and carbon (Mg ha^{-1}) across the study area.

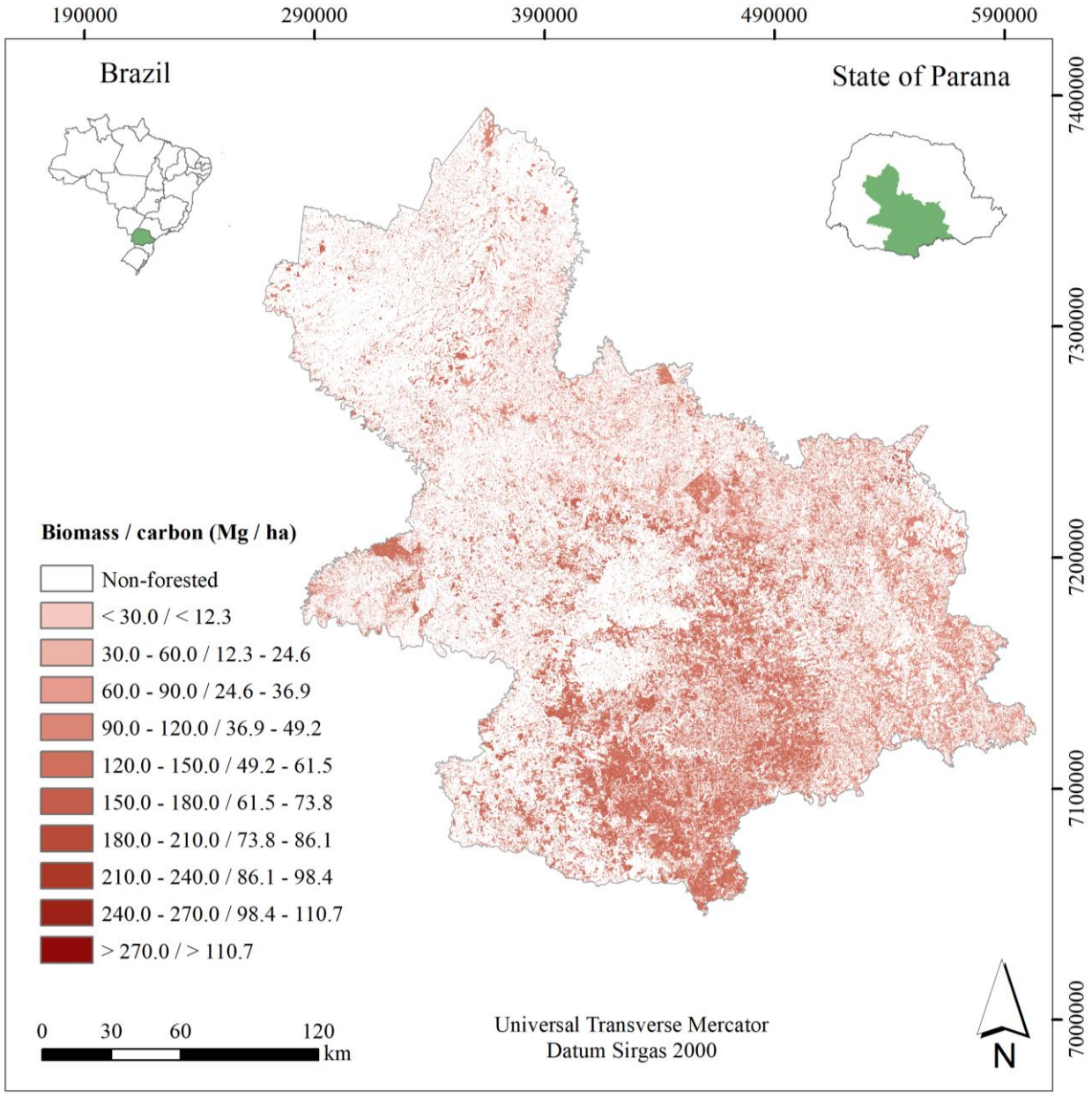


FIG. 28. SPATIAL DISTRIBUTION OF TOTAL FOREST BIOMASS AND CARBON ESTIMATED IN FORESTED PATCHES FROM SURFACE REFLECTANCE MODELING AND IMAGERY CLASSIFICATION.

5. DISCUSSION

5.1 PERFORMANCE OF THE RK FOR ESTIMATING WOOD VOLUME

The findings reveal that the RK method was unable to estimate wood volume from pixels with lower and upper stocks, precisely the ones with volume < 50 $m^3\ ha^{-1}$ and > 210 $m^3\ ha^{-1}$ (Fig. 26). This limitation had already been noted for the RKM, in Fig. 17c, where wood volume estimates ranged from ~50 to ~150 $m^3\ ha^{-1}$, while observed volumes ranged from ~25 to ~225 $m^3\ ha^{-1}$. This fact indicates us that the analytical procedure of adding rasters of spatialized residuals and volumes was limited in fixing volume estimates from less and more stocked pixels. Despite its worse performance, the RMSE provided by RK fits into the results from surveys combining remotely sensed data and geostatistical methods (Yadav and Nandy, 2015; Scolforo et al., 2015; Scolforo et al., 2016).

One of the main reasons why the RK has been less reliable than SRM concerns the input variables. In the tested RKM, wood volume was limited to the variation of latitude and longitude. If one of these variables is not significant, the volume then becomes dependent on only one geographical variable, as occurred in this study (where only latitude was significant at 95% probability level). Therefore, the RKM fitted to this study can attain extreme volumes under extreme latitudes only, but this last is limited to the study area. Another underlying deadlock of the tested RKM is that forest resources may not be stocked in a logical spatial gradient, what would hamper the performance of the models based on geographical location. The SRMs, in turn, were more efficient possibly because they estimate volume as function of spectral responses of vegetation, regardless of pixel geographical locations (as the RKMs), being therefore able to estimate upper and lower volume. In addition, forest variables commonly can be spatialized via remote sensing data (Viana et al., 2012), though a moderate tendency was noted in the SRMs (Fig. 17a).

As commented by Powell et al. (2010), the fact that no real parameter is known makes it difficult to state outright which method is superior; however, the findings suggest that the SRM was an alternative better than the RK method. As main highlight, the maps derived from the SRMs present extreme classes of volume, therefore they should better reproduce spatially ground realities. Although it is pointed this superiority, both methods provided close estimates for mean and total

volume, with relative difference of only 3.7%. In addition, the RK method provided an overall result considerably close to the SRMs.

In forest surveys, RK has been successfully applied along with variables from various natures, inclusive remotely sensed data. Meng et al. (2009) tested four geostatistical methods aiming at estimating forest basal area in a region of ~35,000 km²; RK provided the smallest errors among the other geostatistical methods. The main difference between this study and that one from Meng et al. is that they used Landsat ETM+ data as input variables of the linear model, rather than geographical variables. Meng et al. had no limitation in estimating lower or upper values and obtained values of R^2 around 0.90. On the other hand, Viana et al. (2012) used remotely sensed data combined with RK for estimating AGB; the authors obtained low performance of the RK method, but they associated this result with a low spatial dependence of the variable.

Aiming to spatialize aboveground carbon (AGC) in tropical forests, Scolforo et al. (2016) also used RK combined with longitude, besides another variable related to the biome. The authors obtained a mean error of ~ 58% and a residual distribution notably biased, but relatively better than the residuals showed in Fig. 17c. The commented limitation in estimating volume of extreme classes also was noted in Scolforo et al., indicating that their variable of interest (AGC) was limited to longitude such as in this study, where wood volume was limited to latitude.

In a survey also with AGC, but in a larger study area (State of Minas Gerais, Brazil), Scolforo et al. (2015) selected the variables latitude and altitude in the model used to the RK. In this case, their residuals were much less biased than the residuals showed in Fig. 17c. The inclusion of altitude in the model possibly contributed to a better residual distribution, once elevation is commonly related to AGB and carbon, being thus constantly used to estimate such variables (Alves, et al., 2010; Zhang et al. 2009; Houghton et al., 2001).

5.2 PERFORMANCE OF THE SRMS FOR ESTIMATING WOOD VOLUME, BIOMASS AND CARBON

In the last section, it was addressed reasons why the SRM was superior to RK in this study. Here, some sources of error to fit the SRMs are highlighted. Firstly, the large variation of stand-level volume observed in NFI plots can be cited as one of the main causes of loss of accuracy. In this study, it was observed many entirely forested clusters (i.e., the 40 plots classified into a forest

class) with low wood stock, or even without any stock. This occurs because the NFI inclusion criteria (of Dbh ≥ 10 cm) may exclude, for example, most of the trees belonging to an early-successional forest. If a young vegetation like that is compared to a mature forest, they are likely to have less spectral discrepancy than volume discrepancy, given their greenness and wetness similarities.

Secondly, the cluster's sub-units are too small (20 x 50 m = 1,000 m², see Fig. 6) to be considered as single plots, due to the 30-m spatial resolution. Fassnacht et al. (2014) address problems related to size of sample units in remote sensing-derived estimation, including loss of accuracy. Such condition obligates us to expand the collection of pixels across the total area covered by the cluster, i.e. 40,000 m² (or 4 ha). In this case, total area is the area formed by the distances between sub-units 1 to 3, and 2 to 4 (which is 200 m each). Despite this total area is a reasonably sufficient sample size, it contemplates pixels from both sampled as non-sampled forest patches, hoping that they have similar stocks. For this reason, only entirely forested clusters had their total area expanded, since the other clusters have non-forested pixels over their total area. By following such needed strategy of expansion, a possible source of error concerns the incompatibility between pixels from sampled and non-sampled forest patches. This undesirable situation would entail in calibrating the models with data of stocks (derived from sub-unit's data) incompatible to ground truth existing in the pixels from non-sampled patches. In addition to these two limitations, the complex composition and structure from the Brazilian tropical forests (Malhi et al., 1999) also are cited as negative influences to fit satellite-derived models (Mallinis et al., 2004).

As positive results from this study, it is cited the improvement in accuracy given when the SRM is modeled by Landsat scene. Although it is not presented here, a single SRM for the whole study area was tested and had performance much worse than the specific-by-scene SRMs used in this research. Aiming at estimating AGB in different sites, Foody et al. (2003) found that the contribution of Landsat bands to predict AGB differed between sites, so that bands that were important at one site were unimportant at other sites. As consequence of this, the vegetation indices that are strongly correlated with biomass also differed between sites, as well as the magnitude and direction of regression parameters. The results from Foody et al. are somehow similar to the results of regression analysis of this study, in which the spectral bands selected by the stepwise changed among the Landsat scenes, both for wood volume (Table 9), as for forest biomass and carbon (Table

11). As the Brazilian NFI plots cover a huge variability of forest types at a national scale, these findings suggest that it is of special importance to stratify the NFI plots into succession classes or sites. Unfortunately, the small number of observations allowed me to stratify data only by Landsat scene. Gains provided by stratification has been obtained even in planted forest surveys (Yadav and Nandy, 2015).

The forestry literature is full of surveys proposing satellite-derived methods for estimating AGB and AGC, however, little effort has been devoted to the variable wood volume. Surveys that estimate forest variables as function of remote sensing data have generally showed weaker relations with wood volume than AGB. Fig. 30 shows Pearson correlations of forest variables (wood volume, biomass and carbon) along the spectral range. Indeed, Fig. 30 reveal that the biomass and carbon estimated in this study have stronger relations with spectral bands than wood volume.

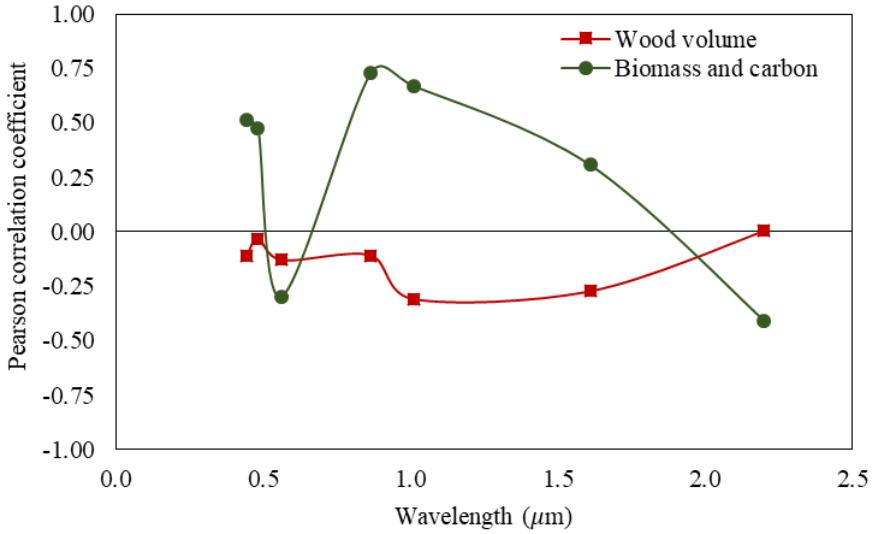


FIG. 29. CORRELATION BETWEEN LANDSAT-8 OLI BANDS AND FOREST VARIABLES FROM NFI PLOTS.

Mäkelä and Pekkarinen (2004) and Mohammadi et al. (2010) used Landsat imagery to estimate wood volume in homogeneous forests, obtaining RMSE ranging from 48% to 100%. In deciduous forest, Tomppo et al. (2002) found up to 180% of error in volume estimates. On the other hand, all the AGB studies cited below had better performances. Lu et al. (2004) cite many authors who found strong relations between the visible bands and biomass, as well as negative relations between the middle infrared bands and wood volume. Despite the authors cited by Lu et al. studied coniferous and mixed forests typically from temperate regions, the relation between forest variable and spectral bands were overall similar to this study (Fig. 30).

Kim et al. (2013) also cite many other authors that used models analogous to what was here named SRMs, in which significant relationships between biomass and spectral bands have been detected, especially the visible and middle infrared bands. As found by these authors, the principal components (PCs) of NIR, NDVI, and Brightness had strong positive correlation with biomass and carbon, indicating an important role of them for biomass and carbon estimations. To the wood volume, the relationships were weaker, in which the two strongest correlations came from Brightness and SWIR1 (Fig. 30). Correlations between NDVI and wood volume was ~ -0.16 (third strongest), whereas with biomass and carbon was ~0.70 (second strongest).

In forestry, satellite-derived estimates of forest resources have been widely tested by means of various methods, including combination with multivariate analysis, and with artificial neural networks (ANN). Fazakas et al. (1999) used the *k* nearest-neighbor (kNN) algorithm to estimate wood volume and biomass on NFI sample plot data. After calibrating the algorithm with ground values and Landsat TM bands, the method employed by the authors assigns values of wood volume and biomass to forested pixels, basing on spectral responses from the *k=5* nearest-neighbor pixels. On plot level, Fazakas et al. found RMSE higher than (~66% for both forest variables) the ones observed here. However, they noted that the RMSE decreased at levels larger than at the plot level. Such as in this study, the authors detected that the kNN method overestimated low wood volumes. Labrecque et al. (2006) evaluated four methods (DRR, kNN, LCC, and BioCLUST – see meaning of each abbreviation in their article) to map biomass, also obtaining RMSE values very close to results of this study.

One difference between this study and the one from Fazakas et al. (1999) is that here NDVI was considered as input variable, which was selected in some SRMs by the stepwise selection method (Table 9). This inclusion tends to improve the capability of the models because NDVI reflects green vegetation density and is directly related to forest biomass (Deo et al., 2017; Fassnacht et al., 2014). NDVI and other vegetation indices combined were used by Gu et al. (2006) to model wood volume from Landsat TM sensor data. With the same estimation method kNN, Gu et al. found RMSE smaller (~44%) than the values obtained by Fazakas et al. (1999). Likewise, Dube and Mutanga (2015) modeled AGB and noted that, when vegetation indices (including NDVI) were used, the models yielded better accuracy statistics than the models that disregarded such indices. The same was found by Zhu and Liu (2015).

A different method of predicting forest attributes from satellite data is presented in Hall et al. (2006): called the BioSTRUCT method. These authors developed a two-steps method for estimating volume and AGB, in which: first, stand height and crown closure are estimated as function of Landsat ETM+ spectral responses; and second, wood volume and AGB are estimated as function of height and crown closure estimated in the first step. Hall et al. found RMSE better than the traditional direct methods involving modeling. Despite these authors propose an interesting alternative, their method is applied to (Canadian) boreal forests, where crown closure remarkably differs from tropical forests, as the Brazilian ones. In spite of the fact that tropical forests have overall canopies closer than the boreal ones, they shelter taller trees, with more vertical variation, as well as a number of main tree species > 200 ha^{-1}, whereas boreal forests have typically few main tree species (Malhi et al., 1999). These factors portray the spectral response of height and crown closure, increasing thus the difficult in estimating such variables in mixed forests (Kim et al., 2013). The BioSTRUCT method used by Hall et al., therefore, is more likely to work better in more homogenous forest types, what may explain the more accurate results obtained by the authors.

5.3 WOOD VOLUME ESTIMATIONS AT COARSE AND FINE SCALES

The wood volume observed in NFI plots (clusters) was directly extrapolated by means of two coarse-scale methods. The first alternative consists of obtaining total volume considering the EFA, which is the proportion of forest of each cluster multiplied by 400 km². At this coarse scale, estimate for the study area indicates 197,396,518 m³ in total, with confidence intervals (CI) ranging from 168,468,681 m³ to 226,324,356 m³ (Table 15), at 95% probability level.

In the second alternative, total volume was obtained considering delimitation of the forested area by image classification. A total volume of 111,013,407 m³ was estimated in this case, i.e., value 43.8% smaller in relation to the total estimated in the first case, which was of 197,396,518 m³. Despite unexpected, such large difference is due to some reasons. One of them is because, in the first alternative, EFA is of 400 km² for entirely forested clusters (Table 15), whereas in the second alternative, the FADIC was of ~140 km² per cluster. For partially forested clusters, EFA (~192 km² per cluster) also was smaller than FADIC (~133 km² per cluster).

As EFA and FADIC are the extrapolation factors, respectively, for the first and second alternatives, the variation of them directly affects the estimated totals. The main issue to consider

in comparing both alternatives is that, the larger the number of entirely forested clusters, the larger the difference between the totals, in which extrapolations without assistance of image classification will always produce totals larger than extrapolations with assistance of image classification. In this case, a smaller difference between totals would be obtained with compensations from larger FADICs of partially forested clusters.

In relation to the SRM and RK methods, surprisingly, both the fine methods fitted into the CI obtained at coarse scale. The total volume estimated through SRM was of 210,961,589 m^3, and through RK 203,326,674 m^3, a difference of, respectively, 6.9% and 3.0% in relation to the coarse-scale estimation.

The fact that the fine-scale estimation fitted into the CI provided by the coarse-scale one was also found by Gobakken et al. (2012). In a similar analysis with AGB stock, Gobakken et al. used plot data from the Norwegian NFI to investigate the performance of two sampling strategies based on Airbone Laser Scanner (ALS) data. The authors compared AGB estimates with assistance of ALS and estimates solely derived from field measurements. As results, CIs of the two ALS-based methods fitted into the CI of the direct method (without assistance of ALS), having a relative difference for the mean AGB of ~8% and ~2%, in relation to the total AGB directly estimated from field data. Gobakken et al. found smaller differences in relation to this study, which may be associated with a more accurate data source (based on LIDAR), a more homogeneous forest type, as well as the finer sampling intensity employed in the Norwegian NFI (Gobakken et al., 2012).

Finally, the fitting of the fine-scale estimates into the CI provided by the coarse-scale estimate is a good indicator that Brazil's NFI sampling intensity produces reliable estimates of wood volume. These findings confirm what many authors have concluded about capabilities provided by satellite data-derived estimates (Bohlin et al., 2017; Maack et al., 2016; Kim et al., 2013; Labrecque et al., 2006; Krankina et al., 2004; Fazakas et al., 1999). The comparison test accomplished with the Landsat and RapidEye images indicates that the impact of the image spatial resolution is tiny (at least comparing the 30-m and 5-m resolutions) in the overall estimate of wood volume. However, it is important to highlight that such analysis does not quantify the effect of the spatial resolution on the classification and delimitation of forest patches. Finer spatial resolution images will probably provide more forest patches in the classifications (due to capability of detecting smaller patches), increasing thus the forest area in relation to coarser scale images.

6. CONCLUSION

In this study, wood volume from NFI plots was estimated and mapped at pixel level through two methods: (i) linear regression model (SRM) based on surface reflectance data, and (ii) regression kriging (RK) combined with a linear model based on geographical variables (latitude and longitude). These fine-scale methods provided close estimates for the total volume (difference of ~8%), confirming the first raised hypothesis. However, the SRM had the best capability to mapping the spatial distribution of the wood volume. The SRM also provided accurate estimates and mappings of forest biomass and carbon stocks. The low performance of the RK method in spatially mapping the studied variables is due to the fact that the geographical variables is limited to how big is the study area.

Estimates of the total wood volume derived from SRM and RK fit into the confidence interval obtained at a coarse scale, what contrasts the second raised hypothesis. This indicates that Brazil's NFI sampling intensity produces reliable estimates of wood volume. Even with a broad sampling intensity of 1:100,000, the results reveal that both the tested methods (SRM and RK) of estimating wood volume can be applied to the Brazilian NFI plot data. To reinforce such statement, further surveys should be developed in the other Brazilian biomes, mainly in regions where there is larger variation of wood volume and biomass stocks.

The coarse-scale extrapolation supported by image classification underestimates the total wood volume, in relation to that unsupported one. This occurs due to the lesser forested area found in the method with image classification, in relation to the represented forested area associated to each cluster. Further researches should be conducted with other NFI database to confirm that conclusion.

7. RECOMMENDATIONS

Some lessons were learned about pixel-level modeling by using Brazil's NFI plot data and satellite imagery. This study adopted Landsat image pixels and field measurements from Brazil's NFI plots as data source for modeling wood volume, biomass and carbon. To whom will develop novel works, it should be kept in mind that the cluster's sub-units are too small (1,000 m²) to compose the pairs of data together with the pixels' reflectance values, since the image pixels are of 900 m². An alternative is to perform an image fusion, resizing the pixel size of Landsat imagery from 30 m to 15 m. Thus, a certain variable taken from the sub-unit would compose the database instead of it taken from the entire cluster, as occurred in this study. With this alternative, the minimum unit is no longer the cluster, but rather the cluster's sub-unit.

There are two ways of improving the goodness of fit statistics of the SRMs. One of them is reducing the minimum unit of pairs of data, as cited above. Beyond the cited alternative, the minimum unit could be reduced until plots' sub-units of 100 m². Another way is to fit models by Landsat scene, as done in this study, rather than when by an entire mosaic of images. The use of PCs did not improve RKM accuracy, but they provide more consistent estimators. Although it was not used in this study, visible-bands could improve the SRMs of biomass and carbon.

Further studies should be conducted employing RK with other variables besides longitude and latitude, such as altitude, soil attributes, climate variables, etc. In addition, the RK method is likely to be more suitable for larger areas, as the whole state of Parana.

As the Brazilian NFI does not provide tree-level equations (for tree height, wood volume, biomass and carbon), it is very important to find robust equations. This ensures accurate tree-level estimates, providing in turn accurate pixel-level estimates. Otherwise, large errors and inconsistent estimates may be obtained, as negative values.

In relation to the image classifications, the use of eCognition and object-oriented classification are recommended for Landsat imagery. To attain a good segmentation and classification, the user must pay attention in the parameters described in the section 3.2.2 ii. There is no fixed parameter for each situation. The user should test different parameters of shape and compactness and then choose the most proper for each case. As Landsat imageries are too large, the user can split them into smaller scenes, reducing thus the number of objects by scene. This automatically reduces the computational burden, as reported in the section 3.2.2 ii.

REFERENCES

ALVES, L. F.; VIEIRA, S. A.; SCARANELLO, M. A.; CAMARGO, P. B.; SANTOS, F. A. M.; JOLY, C. A.; MARTINELLI, L. A. Forest structure and live aboveground biomass variation along an elevational gradient of tropical Atlantic moist forest (Brazil). **Forest Ecology and Management**, v. 260, p. 679-691, 2010.

ANGULO-MARTÍNEZ, M.; LÓPEZ-VICENTE, M.; VICENTE-SERRANO, S. M.; BEGUERIA, S. 2009. Mapping rainfall erosivity at a regional scale: a comparison of interpolation methods in the Ebro Basin (NE Spain). **Hydrol. Earth Syst. Sci.**, v. 13, p. 1907-1920, 2009.

BLACKARD, J. A.; FINCO, M. V.; HELMER, E. H.; HOLDEN, G. R.; HOPPUS, M. L.; JACOBS, D. M.; LISTER, A. J.; MOISEN, G. G.; NELSON, M. D.; RIEMANN, R.; ET AL. Mapping U.S. forest biomass using nationwide forest inventory data and moderate resolution information. **Remote Sensing of Environment**, v. 112, p. 1658-1677, 2008.

BLASCHKE, T. Object based image analysis for remote sensing. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 65, p. 2-16, 2010.

BOHLIN, J.; BOHLIN, I.; JONZÉN, J.; NILSSON, M. Mapping forest attributes using data from stereophotogrammetry of aerial images and field data from the national forest inventory. **Silva Fennica**, v. 51, n. 2, p. 1-18, 2017.

BRITES, R. S.; BIAS, E. S.; ROSA, A. N. C. S. **Classificação por regiões**. In: Introdução ao processamento de imagens de sensoriamento remoto. MENESES, P. R.; ALMEIDA, T. Universidade de Brasília, Brasília, DF, 266p. 2012.

BROOKS, E. B.; COULSTON, J. R.; WYNNE, R. H.; THOMAS, V. A. Improving the precision of dynamic forest parameter estimates using Landsat. **Remote Sensing of Environment**, v. 179, p. 162-169, 2016.

CAMBARDELLA, C. A.; MOORMAN, T. B.; NOVAK, J. M.; PARKIN, T. B.; KARLEN, D. L.; TURCO, R. F.; KONOPKA, A. E. Field-Scale Variability of Soil Properties in Central Iowa Soils. **Soil Sci. Soc. Am. J.** v. 58, p. 1501-1511, 1994.

CANADA'S NATIONAL FOREST INVENTORY ESTIMATION PROCEDURES, v.1.13, 2004. Available in: https://nfi.nfis.org/en/estimation.

CHACKO, V. J. **A manual on sampling techniques for forest surveys.** Dehra Dun: India, P.Z.O. 1965, 172p.

CIENCIALA, E.; TOMPPO, E.; SNORRASON, A.; BROADMEADOW, M.; COLIN, A.; DUNGER, K.; EXNEROVA, Z.; LASSERRE, B.; PETERSSON, H.; PRIWITZER, T.; ET AL. Preparing Emission Reporting from Forests: Use of National Forest Inventories in European Countries. **Silva Fennica**, v. 42, n. 1, p. 73-88, 2008.

COHEN, W. C.; GOWARD, S. N. Landsat's Role in Ecological Applications of Remote Sensing. **BioScience**, v. 54, n. 6, p. 535-545, 2004.

BRASIL. **CONAMA - Conselho Nacional Do Meio Ambiente.** Resolução nº 2, de 18 de Março de 1994. Diário Oficial da República Federativa do Brasil, Brasília, DF, 18 mar. 1994.

CONGALTON, R. G., GREEN, K. **Assessing the Accuracy of Remotely Sensed Data: Principles and Practices**. 2nd edition. CRC Press. 2008.

DAVID, H. C.; ARAÚJO, E. J. G.; MORAIS, V. A.; SCOLFORO, J. R. S.; MARQUES, J. M.; PÉLLICO NETTO, S.; MACFARLANE, D. W. Carbon stock classification for tropical forests in Brazil: Understanding the effect of stand and climate variables. **Forest Ecology and Management**, v. 404, p. 241-250, 2017.

DEO, R. K.; RUSSELL, M. B.; DOMKE, G. M.; WOODALL, C. W.; FALKOWSKI, M. J.; COHEN, W. B. Using Landsat Time-Series and LiDAR to Inform Aboveground Forest Biomass Baselines in Northern Minnesota, USA. **Canadian Journal of Remote Sensing**, v. 43, n. 1, p. 28-47, 2017.

DUBE, T.; MUTANGA, O. Evaluating the utility of the medium-spatial resolution Landsat 8 multispectral sensor in quantifying aboveground biomass in uMgeni catchment, South Africa. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 101, p. 36-46, 2015.

ENE, L. T.; NAESSET, E.; GOBAKKEN, T.; MAUYA, E. W.; BOLLANDSAS, O. M.; GREGOIRE, T. G.; STAHL, G.; ZAHABU, E. Large-scale estimation of aboveground biomass in miombo woodlands using airborne laser scanning and national forest inventory data. **Remote Sensing of Environment**, v. 186, p. 626-636, 2016.

FASSNACHT, F. E.; HARTIG, F.; LATIFI, H.; BERGER, C.; HERNÁNDERZ, J.; CORVALÁN, P.; KOCH, B. Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass. **Remote Sensing of Environment**, v. 154, p. 102-114, 2014.

FAZAKAS, Z.; NILSSON, M.; OLSSON, H. Regional forest biomass and wood volume estimation using satellite data and ancillary data. **Agricultural and Forest Meteorology**, v. 98-99, p. 417-425, 1999.

FOODY, G. M.; BOYD, D. S.; CUTLER, M. E. J. Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. **Remote Sensing of Environment**, v. 85, p. 463-474, 2003.

FRANKLIN, S. E. **Remote sensing for sustainable forest management**. Lewis Publishers, 407p. 2001.

FRAZIER, R. J.; COOPS, N. C.; WULDER, M. A.; KENNEDY, R. Characterization of aboveground biomass in an unmanaged boreal forest using Landsat temporal segmentation metrics. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 92, p. 137-146, 2014.

GARA, T.; MURWIRA, A.; DUBE, T.; SIBANDA, M.; RWASOKA, D. T.; NDAIMANI, H.; CHIVHENGE, E.; HATENDI, C. M. Estimating forest carbon stocks in tropical dry forests of Zimbabwe: exploring the performance of high and medium spatial-resolution multispectral sensors. **Southern Forests**, v. 2016, p. 1–10, 2006.

GLEASON, C. J.; IM, J. A Review of Remote Sensing of Forest Biomass and Biofuel: Options for Small-Area Applications. **GIScience & Remote Sensing**, v. 48, n. 2, p. 141-170, 2011.

GOBAKKEN, T.; NAESSET, E.; NELSON, R.; BOLLANDSAS, O. M.; GREGOIRE, T. G.; STAHL, G.; HOLM, S.; ORKA, H. O.; ASTRUP, R. Estimating biomass in Hedmark County, Norway using national forest inventory field plots and airborne laser scanning. **Remote Sensing of Environment**, v. 123, p. 443–456, 2012.

GOEKING, S. A. Disentangling Forest Change from Forest Inventory Change: A Case Study from the US Interior West. **Journal of Forestry**, v. 113, n. 5, p. 475-483, 2015.

GOOGLE EARTH-MAPS. http://mapas.google.com. Consulted between 08/01/2017 to 10/30/2017.

GU, H.; DAI, L.; WU, G.; XU, D.; WANG, S.; WANG, H. Estimation of wood volumes by integrating Landsat TM imagery and forest inventory data. **Science in China: Series E Technological Sciences**, v. 49, p. 54-62, 2006.

GUJARATI, D. N.; PORTER, D. C. **Basic Econometrics.** McGraw HillBookman, 6th ed., 2011
HALL, R. J.; SKAKUN, R. S.; ARSENAULT, R. J.; CASE, B. S. Modeling forest stand structure attributes using Landsat ETM+ data: Application to mapping of aboveground biomass and stand volume. **Forest Ecology and Management**, v. 225, p. 378-390, 2006.

HAYASHI, R.; KERSHAW JR., J. A.; WEISKITTEL, A. Evaluation of alternative methods for using LIDAR to predict aboveground biomass in mixed species and structurally complex forests in Northeastern North America. **Math. Comput. For. Nat.-Res. Sci.**, v. 7, n. 2, p. 49-65, 2015.

HENGL, T.; HEUVELINK, G. B. M.; ROSSITER, D. G. About regression-kriging: From equations to case studies. **Computers & Geosciences**, v. 33, p. 1301-1315, 2007.

HOUGHTON, R. A.; LAWRENCE, K. T.; HACKLER, J. L.; BROWN, S. The spatial distribution of forest biomass in the Brazilian Amazon: a comparison of estimates. **Global Change Biology**, v. 7, p. 731-746, 2001.

IBDF - INSTITUTO BRASILEIRO DE DESENVOLVIMENTO FLORESTAL. **Inventário Florestal do Pinheiro no Sul do Brasil**. Fundação de Pesquisas Florestais do Paraná. Curitiba, 1978, 327p.

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**, 6th ed., Pearson Prentice Hall, New Jersey, 2007.

KIM, H.; BETTINGER, P.; CIESZEWSKI, C. Reflections on the Estimation of Stand-Level Forest Characteristics Using Landsat Satellite Imagery. **Applied Remote Sensing Journal**, v. 2, n. 2, p. 45 56, 2013.

KRANKINA, O. N.; HARMON, M. E.; COHEN, W. B.; OETTER, D. R.; ZYRINA, O.; DUANE, M. V. Carbon stores, sinks, and sources in forests of Northwestern Russia: Can we reconcile forest inventories with remote sensing results? **Climatic Change**, v. 67, p. 257-472, 2004.

LABRECQUE, S.; FOURNIER, R. A.; LUTHER, J. E.; PIERCEY, D. A comparison of four methods to map biomass from Landsat-TM and inventory data in western Newfoundland. **Forest Ecology and Management** v. 226, p. 129-144, 2006.

LANDIS, J., KOCH, G. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, p. 159-174, 1977.

LILLESAND, T.; KIEFER, R. W.; CHIPMAN, J. **Remote Sensing and Image Interpretation**. John Wiley & Sons. 704p. 2014

LOETSCH, F.; HALLER, K. E.; ZOHRER, F. **Forest inventory**. 2nd ed. BLV Verlagsgesellschaft, 905p. 1973.

LU, D. The potential and challenge of remote sensing-based biomass estimation. **International Journal of Remote Sensing**, v. 27, n. 7, p.1297-1328, 2006.

LU, D.; CHEN, Q.; WANG, G.; MORAN, E.; BATISTELLA, M.; ZHANG, M.; LAURIN, G. V.; SAAH, D. Aboveground Forest Biomass Estimation with Landsat and LiDAR Data and Uncertainty Analysis of the Estimates. **International Journal of Forestry Research**, v. 2012, p. 1-16, 2012.

LU, D.; MAUSEL, P.; BRONDÍZIO, E.; MORAN, E. Relationships between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon Basin. **Forest Ecology and Management**, v. 198, p. 149-167, 2004.

MAACK, J.; LINGENFELDER, M.; WEINACKER, H.; KOCH, B. Modelling the standing timber volume of Baden-Württemberg – A large-scale approach using a fusion of Landsat, airborne LiDAR andNational Forest Inventory data. **International Journal of Applied Earth Observation and Geoinformation**, v. 49, p. 107-116, 2016.

MAKELA, H.; PEKKARINEN, A. Estimation of forest stand volumes by Landsat TM imagery and stand-level field-inventory data. **Forest Ecology and Management**, v. 196, p. 245-255, 2004.
MALHI, Y.; BALDOCCHI, D. D.; JARVIS, P. G. The carbon balance of tropical, temperate and boreal forests. **Plant, Cell and Environment**, v. 22, p. 715-740, 1999.

MALLINIS, G.; KOUTSIAS, N.; MAKRAS, A.; KARTERIS, M. Forest Parameters Estimation in a European Mediterranean Landscape Using Remotely Sensed Data. **Forest Science**, v. 50, n. 4, p. 450-460, 2004.

MALTAMO, M.; PACKALÉN, P.; SUVANTO, A.; KORHONEN, K. T.; MEHTATALO, L.; HYVONEN, P. Combining ALS and NFI training data for forest management planning: a case study in Kuortane, Western Finland. **Eur. J. Forest Res.**, v. 128, p. 305-317, 2009.

MARGOLIS, H. A.; NELSON, R. F.; MONTESANO, P. M.; BEAUDOIN, A.; SUN, G.; ANDERSEN, H-E.; WULDER, M. A. Combining satellite Lidar, airborne Lidar, and ground plots to estimate the amount and distribution of aboveground biomass in the boreal forest of North America. **Can. J. For. Res**., v. 45, p. 838-855, 2015.

MCROBERTS, R. E.; TOMPPO, E.; SCHADAUER, K.; VIDAL, C.; STAHL, G.; CHIRICI, G.; LANZ, A.; CIENCIALA, E.; WINTER, S.; SMITH, W. B. Harmonizing national forest inventories. **Journal of Forestry**, v. 107, n. 4, p.179-187, 2009.

MELLO, C. R.; VIOLA, M. R.; OWENS, P. R.; MELLO, J. M.; BESKOW, S. Interpolation methods for improving the RUSLE R-factor mapping in Brazil. **Journal of Soil and Water Conservation**, v. 70, n. 3, p. 182-197, 2015.

MENESES, P. R.; SANO, E. E. **Classificação pixel a pixel de imagens**. In: Introdução ao processamento de imagens de sensoriamento remoto. MENESES, P. R.; ALMEIDA, T. Universidade de Brasília, Brasília, DF, 266p. 2012.

MENG Q.; CIESZEWSKI, C.; MADDEN, M. Large area forest inventory using Landsat ETM+: A geostatistical approach. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 64, p. 27-36, 2009.

MOHAMMADI, J.; JOIBARY, S. S.; YAGHMAEE, F.; MAHINY, A. S. Modelling forest stand volume and tree density using Landsat ETM+ data. **International Journal of Remote Sensing**, v. 31, n. 11, p. 2959–2975, 2010.

NAZEER, M.; NICHOL, J. E.; YUNG, Y-K. Evaluation of atmospheric correction models and Landsat surface reflectance product in an urban coastal environment. I**nternational Journal of Remote Sensing**, v. 35, n. 6, p. 1-21, 2014.

ODEH, I. O. A.; MCBRATNEY, A. B.; CHITTLEBOROUGH, D. J. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. **Geoderma**, v. 67, p. 215-226, 1995.

PÉLLICO NETTO, S.; BRENA, D. A. **Inventário Florestal**. 1st ed. Curitiba, Edited by the authors, 1997, 316p.

PÉLLICO NETTO, S.; PELISSARI, A. L.; CYSNEIROS, V. C.; BONAZZA, S.; SANQUETTA, C. R. Sampling procedures for inventory of commercial volume tree species in Amazon Forest. **Anais da Academia Brasileira de Ciências**, v. 89, n. 3, p. 1829-1840, 2017.

PÉLLICO NETTO, S.; SANQUETTA, C. R.; CARON, B. O.; BEHLING, A.; SIMON, A. A.; CORTE, A. P. D.; BAMBERG, R. Ground level photosynthetically active radiation dynamics in stands of Acacia mearnsii De Wild. **Anais da Academia Brasileira de Ciências**, v. 87, n. 3, p. 1833-1845, 2015.

PÉLLICO NETTO, S.; STEFANELLO, F. R.; PELISSARI, A. L.; DAVID, H. C. Mapping of sites in forest stands. **Anais da Academia Brasileira de Ciências**, v. 86, n. 4, p. 2025-2037, 2014.
PONZONI, F. J.; SHIMABUKURO, Y. E.; KUPLICH, T. M. **Sensoriamento Remoto da Vegetação**. 2nd ed. São Paulo, Oficina de textos. 160p. 2012.

POWELL, S. L.; COHEN, W. B.; HEALEY, S. P.; KENNEDY, R. E.; MOISEN, G. G.; PIERCE, K. B.; OHMANN, J. L. Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches. **Remote Sensing of Environment**, v. 114, p. 1053-1068, 2010.

QUEIROZ, W. T. **Amostragem em inventário florestal**. Belém – Universidade Federal Rural da Amazônia, 2012, 441p.

RATUCHNE, L. C. **Equações alométricas para a estimativa de biomassa, carbono e nutrientes em uma Floresta Ombrófila Mista**. Guarapuava, 101 p. Dissertação (Mestrado em Agronomia) – Universidade Estadual do Centro-Oeste, Guarapuava, 2010.

RIBEIRO, M .C.; METZGER, J. P.; The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. **Biological Conservation**, v. 142, p. 1141-1153, 2009.

RIBEIRO JÚNIOR, P. J.; DIGGLE, P. J. geoR: **A package for geostatistical analysis**. R-NEWS, London, v. 1, n. 2, p. 15-18, June 2001.

SCHROEDER, T. A.; GRAY, A.; HARMON, M. E.; WALLIN, D. O.; COHEN, W. B. Estimating live forest carbon dynamics with a Landsat-based curve-fitting approach. **Journal of Applied Remote Sensing**, v. 2, p. 1-20, 2008.

SCOLFORO, H. F.; SCOLFORO, J. R. S.; MELLO, C. R.; MELLO, J. M.; FERRAZ FILHO, A. C. Spatial Distribution of Aboveground Carbon Stock of the Arboreal Vegetation in Brazilian Biomes of Savanna, Atlantic Forest and Semi-Arid Woodland. **Plos One**, v. 10, n. 6, p. 1-20, 2015.

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

SECRETARY OF AGRICULTURE AND SUPPLYING (SEAB). Rural Economy Department. Tabela de Produção Agrícola por Município. 2017. Available at: **http://www.agricultura.pr.gov.br/modules/conteudo/conteudo.php?conteudo=137**

TEMPFLI, K.; KERLE, N.; HUURNEMAN, G. C.; JANSSEN, L. L. F. **Principles of remote sensing: an introductory textbook**. (ITC Educational Textbook Series; Vol. 2). Enschede: ITC. 2009.

THENKABAIL, P. S.; HALL, J.; LIN, T.; ASHTON, M. S.; HARRIS, D.; ENCLONA, E. A. Detecting floristic structure and pattern across topographic and moisture gradients in a mixed species Central African forest using IKONOS and Landsat-7 ETM+ images. **International Journal of Applied Earth Observation and Geoinformation**, v. 4, p. 255-270, 2003.

TOMPPO, E.; GSCHWANTNER, T.; LAWRENCE, M.; MCROBERTS, R. E. **National Forest Inventories: Pathways for common reporting**. Springer. 612p. 2010.

TOMPPO, E.; NILSSON, M.; ROSENGREN, M.; AALTO, P.; KENNEDY, P. Simultaneous use of Landsat-TM and IRS-1C WiFS data in estimating large area tree stem volume and aboveground biomass. **Remote Sensing of Environment**, v. 82, p. 156-171, 2002.

TORRES-SÁNCHEZ, J.; LÓPEZ-GRANADOS, F.; PEÑA, J. M. An automatic object-based method for optimal thresholding in UAV images: Application for vegetation detection in herbaceous crops. **Computers and Electronics in Agriculture**, v. 114, p. 43–52, 2015.

VIANA, H.; ARANHA, J.; LOPES, D.; COHEN, W. B. Estimation of crown biomass of *Pinus pinaster* stands and shrubland above-ground biomass using forest inventory data, remotely sensed imagery and spatial prediction models. **Ecological Modelling**, v. 226, p. 22-35, 2012.

VIBRANS, A. C.; SEVEGNANI, S.; GASPER, A. L.; LINGER, D. V. **Inventário Florístico Florestal de Santa Catarina: Floresta Ombrófila Mista**. Blumenau: Edifurb, v. 3, 2013, 440 p.

VIDAL, C.; ALBERDI, I.; REDMOND, J.; VESTMAN, M.; LANZ, A.; SCHADAUER, K. The role of European National Forest Inventories for international forestry reporting. **Annals of Forest Science**, v. 73, p. 793-806, 2016.

WACKERNAGEL, H. **Multivariate Geostatistics: An Introduction with Applications**. Springer-Verlag Berlin Heidelberg, 3rd ed. 388p, 2003.

WANG, G.; WENG, Q. **Remote Sensing of Natural Resources**. 1st ed. Taylor & Francis Series in Remote Sensing Applications. 580p. 2013.

WILSON, B. T.; WOODALL, C. W.; GRIFFITH, D. M. Imputing forest carbon stock estimates from inventory plots to a nationally continuous coverage. **Carbon Balance and Management**, v. 8, n. 1, p. 1-15, 2013.

YADAV, K. V.; NANDY, S. Mapping aboveground woody biomass using forest inventory, remote sensing and geostatistical techniques. **Environ. Monit. Assess**., v. 187, n. 5, p. 1-12, 2015.

ZHANG, X.; WANG, M.; LIANG, X. Quantitative classification and carbon density of the forest vegetation in Lüliang Mountains of China. **Plant Ecol.**, v. 201, p. 1-9, 2009.

ZHENG, G.; CHEN, J. M.; TIAN, Q. J.; JU, W. M.; XIA, X. Q. Combining remote sensing imagery and forest age inventory for biomass mapping. **Journal of Environmental Management**, v. 85, p. 616-623, 2007.

ZHENG, G.; RADEMACHER, J.; CHEN, J.; CROW, T.; BRESEE, M.; MOINE, J. L.; RYU, S-R. Estimating aboveground biomass using Landsat 7 ETM+ data across a managed landscape in northern Wisconsin, USA. **Remote Sensing of Environment**, v. 93, p. 402-411, 2004.

ZHU, X.; LIU, D. Improving forest aboveground biomass estimation using seasonal Landsat NDVI time-series. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 102, p. 222-231, 2015.

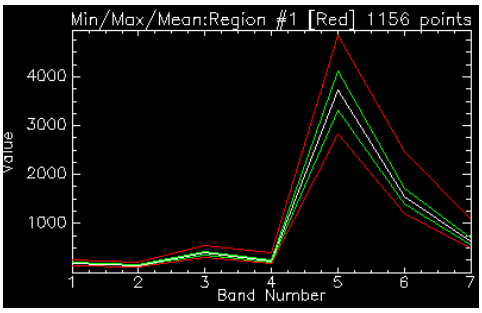
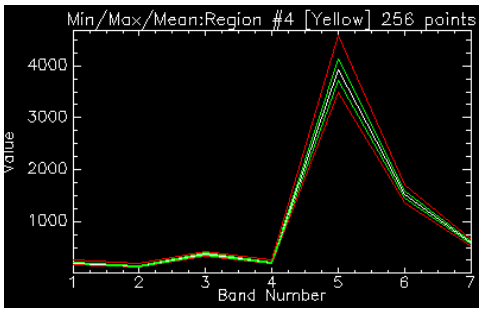
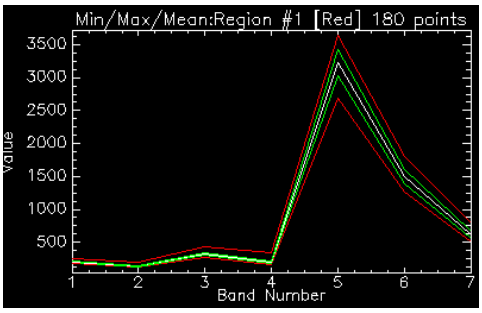
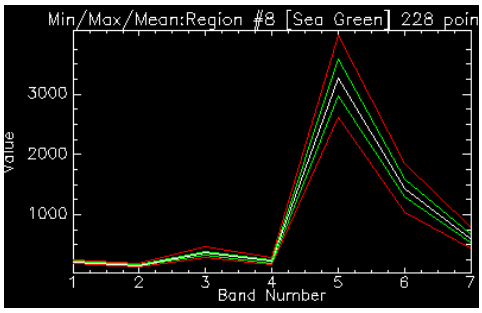
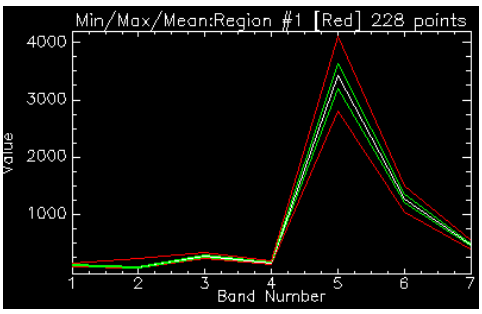
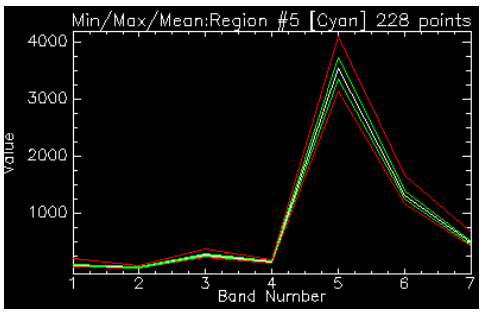
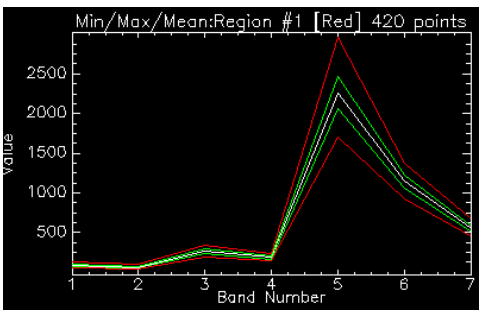
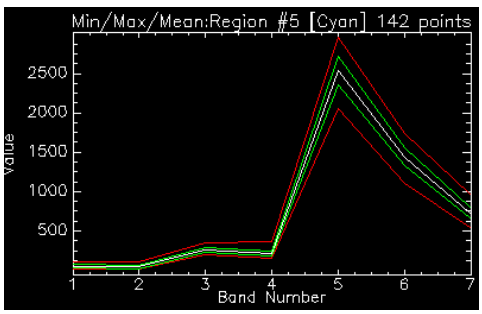
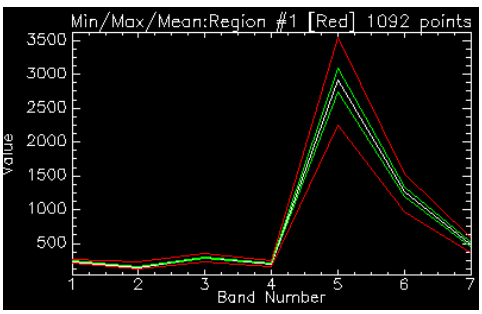
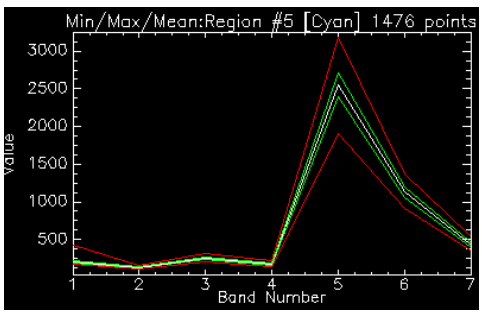
APPENDIX

App. 1. Main sampling statistics for the variable wood volume, with standard error obtained through the ratio estimator for SS with cluster.

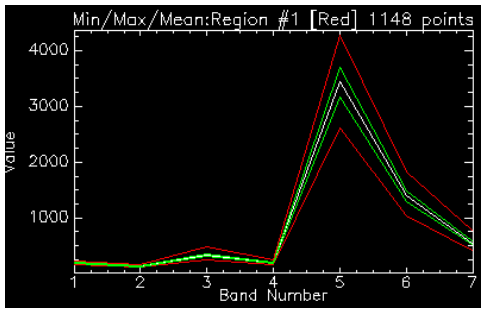
Statistics		Unit	Type of cluster			
			Entirely forested (A)	Partially forested (B)	(C) Non-Forested	Total (A+B)
Frequency*		absolute	7	89	56	96
		%	4.6%	58.6%	36.8%	63.2%
Represented area		ha	280,000	3,560,000	2,240,000	3,840,000
		%	4.6%	58.6%	36.8%	63.2%
Extrapolated forested area		ha	280,000	1,707,000	0.0	1,987,000
		%	14.1%	85.9%	0.0%	100.0%
Sampling error		$m^3\ ha^{-1}$	65.1	61.8	-	66.3
		%	30.1%	77.1%	-	82.7%
Confidence interval (5%)	Lower limit	$m^3\ ha^{-1}$	151.5	18.3	-	13.8
	Mean	$m^3\ ha^{-1}$	216.6	80.1	-	99.3
	Upper limit	$m^3\ ha^{-1}$	281.7	141.9	-	146.4
	Lower limit	m^3	42,402,738	31,280,236	-	23,612,896
	Total	m^3	60,639,475	136,757,044	-	197,396,518
	Upper limit	m^3	78,876,212	242,233,852	-	249,901,192

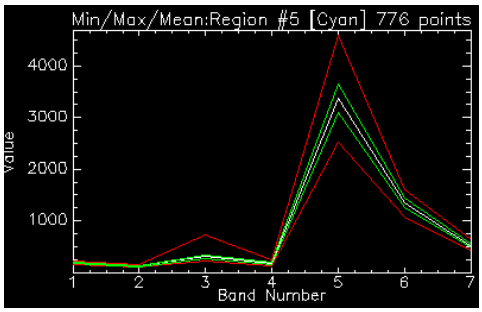
*Total frequency of forested and non-forested clusters (A+B+C) = 152 (100%).

App. 2. Samples of surface reflectance of natural vegetation for Landsat-8 OLI bands.

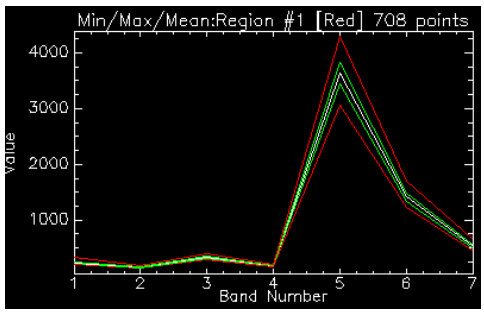
	Sample 1	Sample 2
221/077		
221/078		
222/077		
222/078		
223/076		

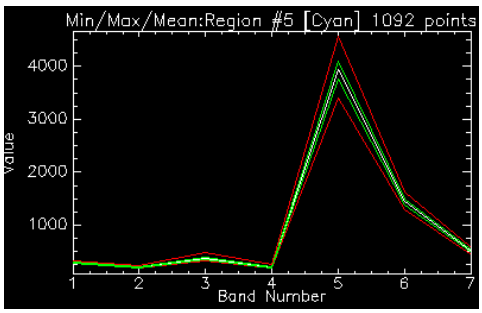
223/077





223/078

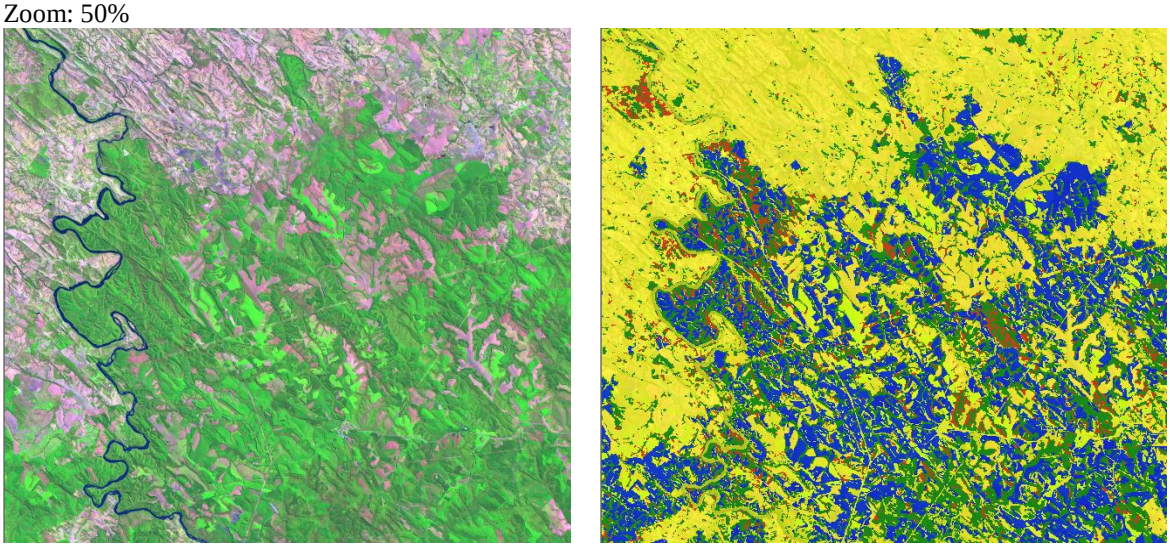




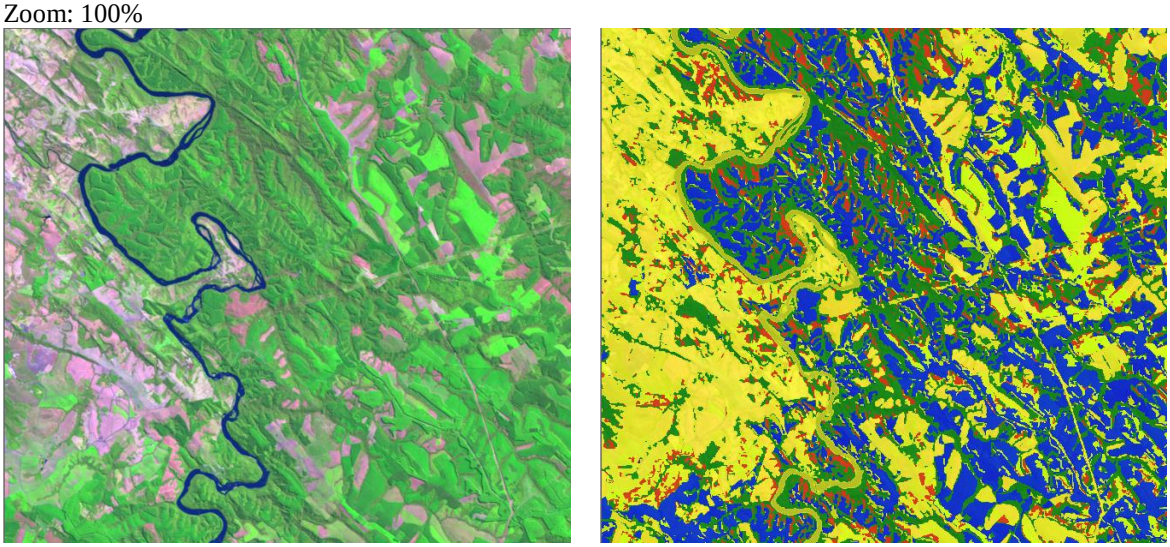
App. 3. Samples of image classification of Landsat-8 scene (221/77) in the State of Parana.

Unclassified image | Classified image

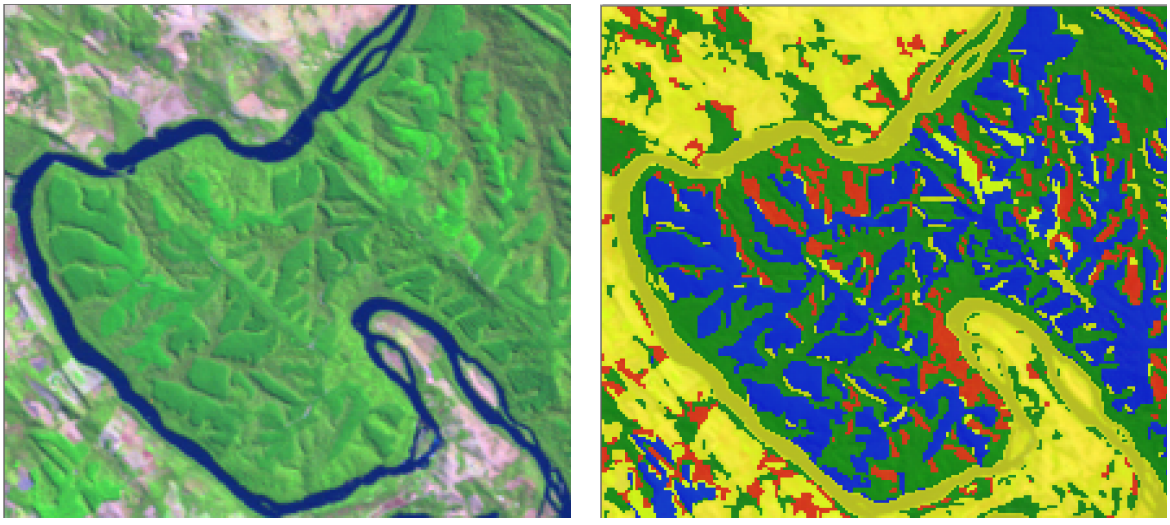
Zoom: 50%



Zoom: 100%



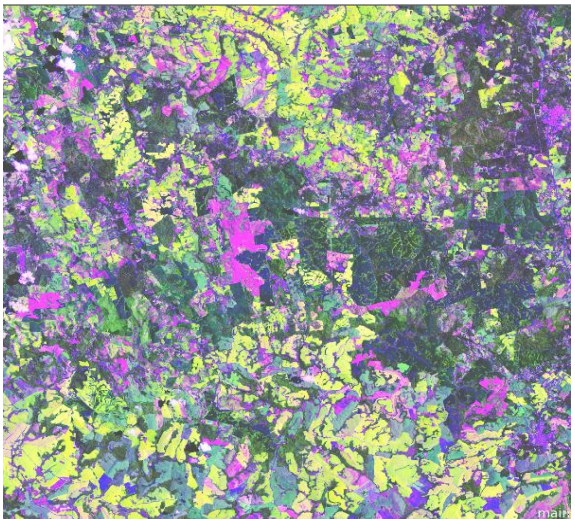
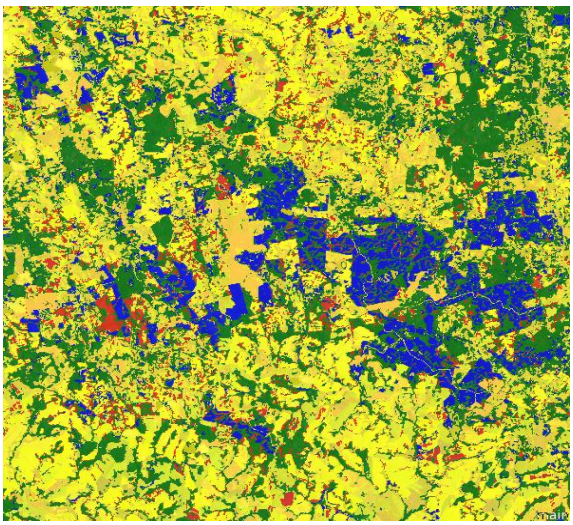
Zoom: 300%



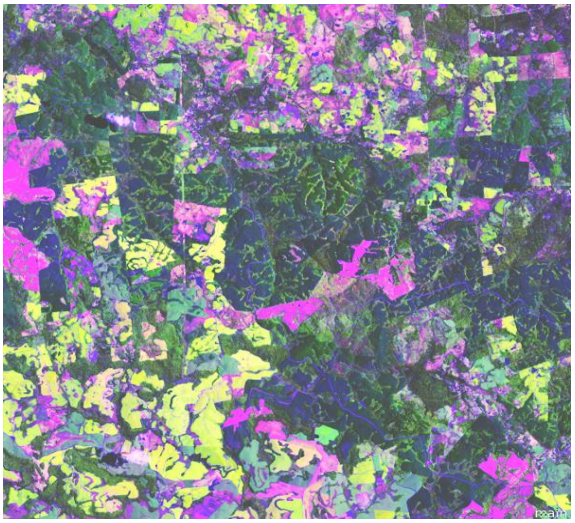
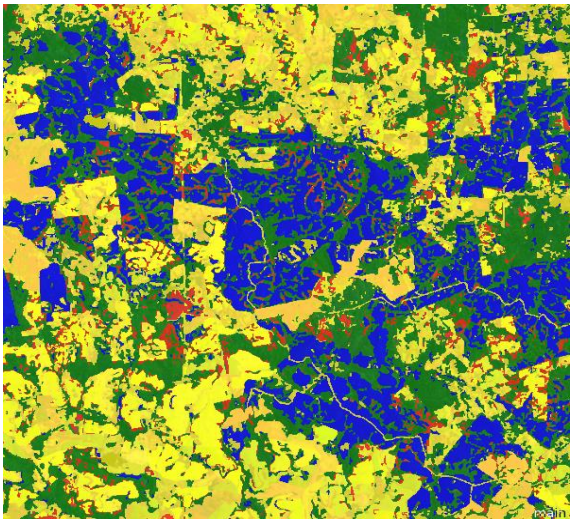
App. 4. Samples of image classification of Landsat-8 scene (222/77) in the State of Parana.

Unclassified image | Classified image

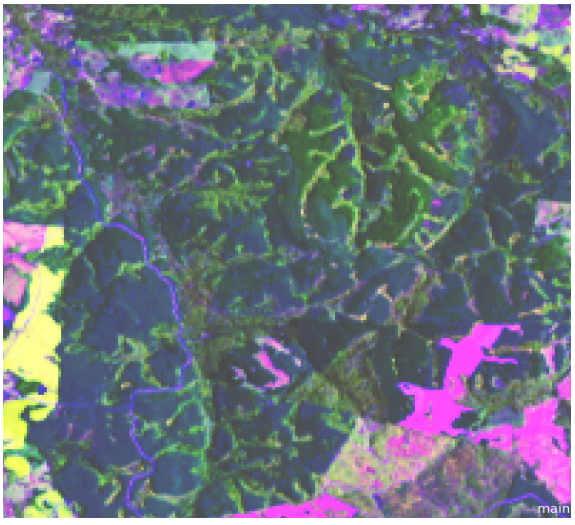
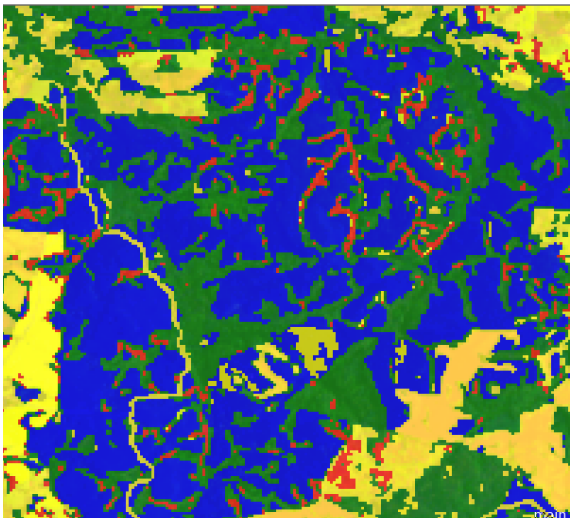
Zoom: 50%

Zoom: 100%

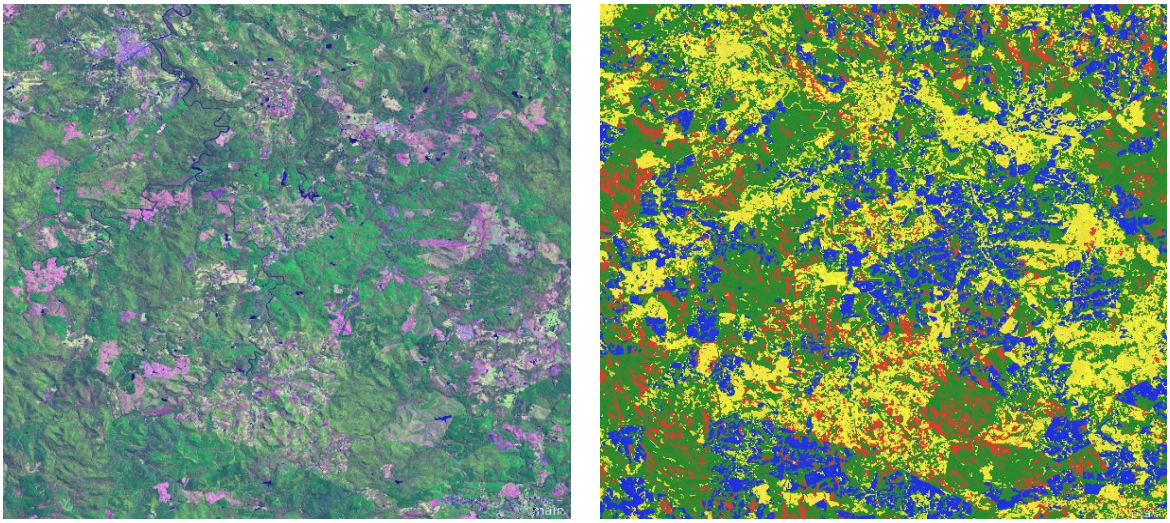
Zoom: 300%

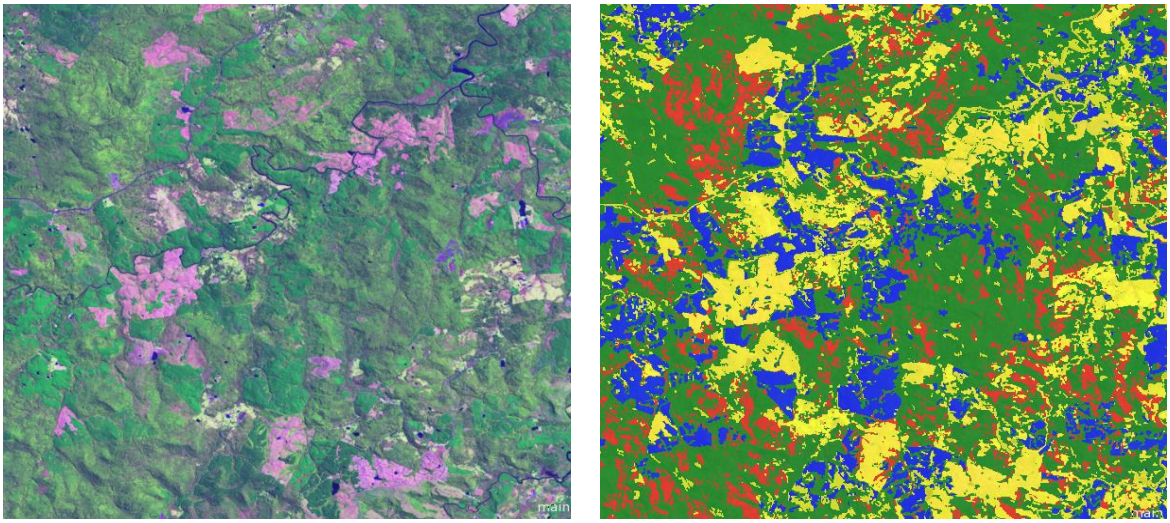
App. 5. Samples of image classification of Landsat scene (222/78) in the State of Parana.

Unclassified image | Classified image

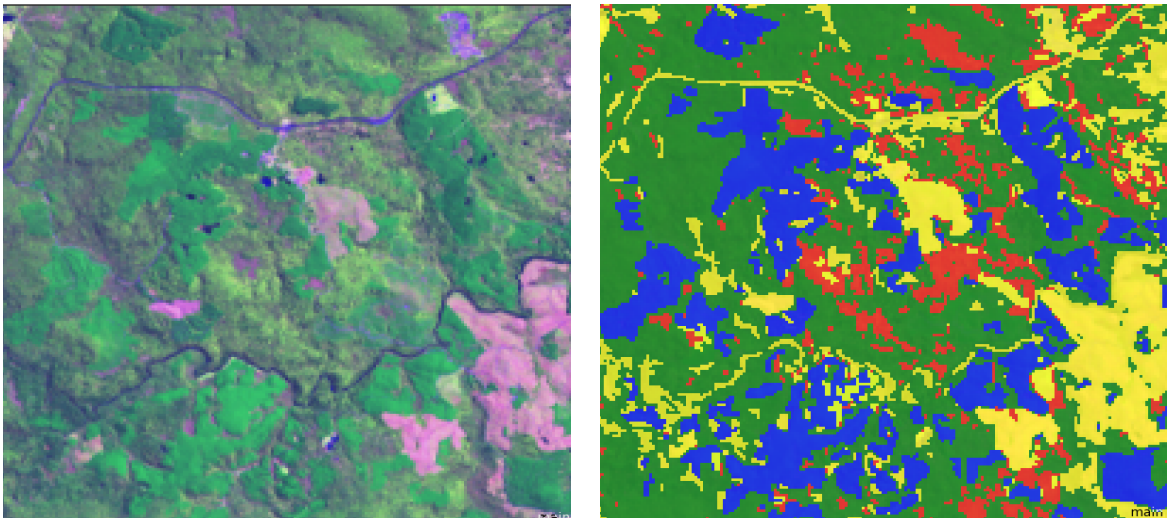
Zoom: 50%



Zoom: 100%



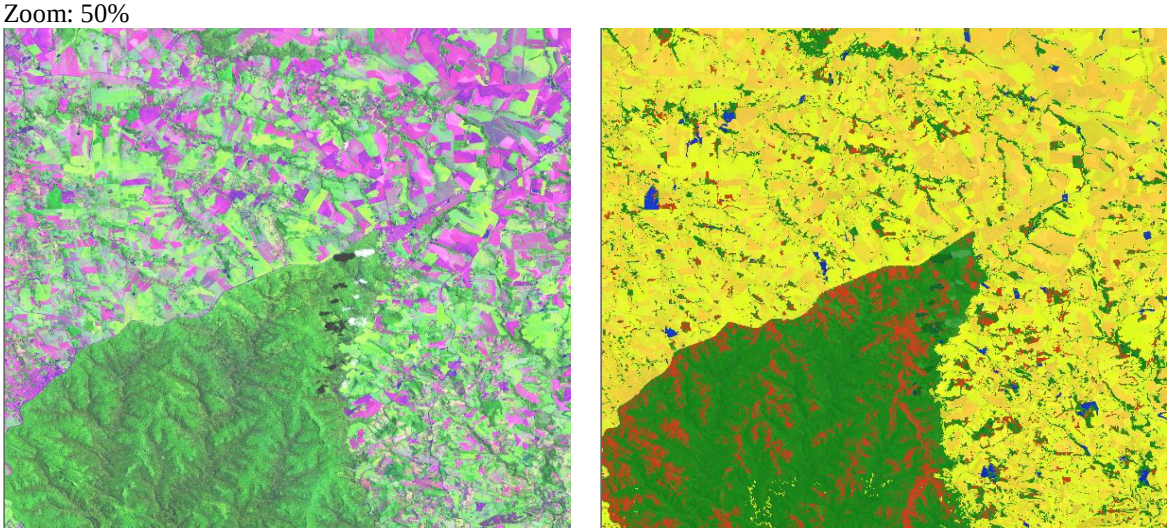
Zoom: 300%



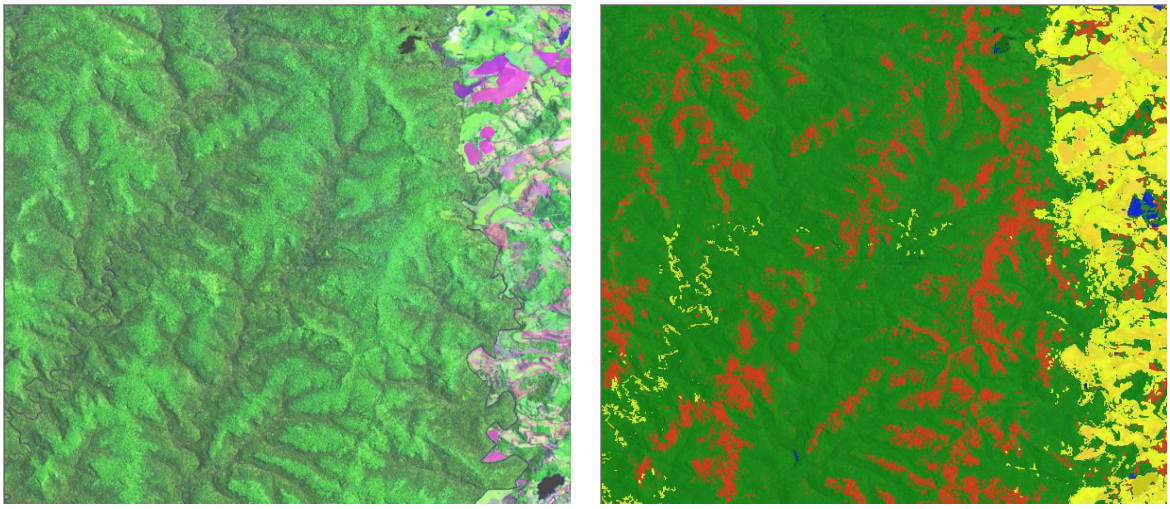
App. 6. Samples of image classification of Landsat scene (223/77) in the State of Parana.

Unclassified image | Classified image

Zoom: 50%



Zoom: 100%



Zoom: 300%