

**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Estatística espacial e sensoriamento remoto para a predição
volumétrica em florestas de *Eucalyptus* spp.**

Esthevan Augusto Goes Gasparoto

Dissertação apresentada para obtenção do título de
Mestre em Ciências. Programa de Recursos
Florestais. Opção em: Silvicultura e Manejo Florestal

**Piracicaba
2016**

Esthevan Augusto Goes Gasparoto
Engenheiro Florestal

**Estatística espacial e sensoriamento remoto para a predição volumétrica
em florestas de *Eucalyptus* spp.**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:
Prof. Dr. **JOÃO LUIS FERREIRA BATISTA**

Dissertação apresentada para obtenção do título de
Mestre Ciências. Programa de Recursos Florestais.
Opção em: Silvicultura e Manejo Florestal

**Piracicaba
2016**

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP**

Gasparoto, Esthevan Augusto Goes

Estatística espacial e sensoriamento remoto para a predição volumétrica em florestas de *Eucalyptus* spp. / Esthevan Augusto Goes Gasparoto. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2016.

118 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz".

1. Krigagem com deriva externa
2. Krigagem ordinária
3. Imagens de satélite
4. Regressão linear múltipla I. Título

CDD 634.9734
G249e

“Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor”

AGRADECIMENTOS

Ao Prof. Dr. João L. F. Batista, da Escola Superior de Agricultura “Luiz de Queiroz” / USP, meu orientador, pela confiança, e profundos ensinamentos transmitidos, nas mais diversas áreas do conhecimento, ao longo de dois anos e meio de convívio acadêmico.

Ao Prof. Dr. Peter L. Marshall, da *University of British Columbia* no Canadá, por me receber como pesquisador visitante junto ao seu grupo de trabalho em *Forest Measurements and Biometrics*, transmitindo-me conhecimentos com paciência e ampliando meus horizontes acadêmicos e pessoais.

Ao Prof. Dr. Cláudio R. Thiersch, da Universidade Federal de São Carlos, *campus* Sorocaba, por ainda na graduação, me apresentar ao mundo do Inventário e da Biometria Florestal, instigando-me na busca de soluções inovadoras para os problemas reais da mensuração florestal.

A empresa de celulose e papel International Paper do Brasil S.A, nas pessoas de Cristiane Lemos, Adriano Almeida, Luis Fernando, Thalita Faria, Thiago Freitas, Karla Weber, Guilherme Pontes e João Morato, pela confiança e interesse no desenvolvimento do projeto, fornecendo recursos logísticos e base de dados para que a realização deste se concretizasse.

Aos professores da Escola Superior de Agricultura “Luiz de Queiroz” / USP, Dr. Luiz Estraviz, Dr. Edson Vidal, Dr. Silvio Ferraz, Dra. Luciana Duque, Dr. Paulo Justiniano e Dr. Gener Pereira, pelos preciosos conhecimentos transmitidos ao longo das disciplinas cursadas nesta instituição.

Aos amigos do Centro de Métodos Quantitativos (CMQ), por proporcionarem um ambiente saudável de amizade, de debates técnicos e trocas de experiências acadêmicas e florestais.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio à pesquisa concedido na forma de bolsa de mestrado.

Aos meus pais Alaor e Sueli, pelo amparo, suporte e confiança sempre demonstrados para comigo. Ao meu irmão Fillipe pelo companheirismo e trocas de ideias. Aos meus avós e demais familiares pelo apoio nos momentos de necessidade.

A Emily Shinzato, pela cumplicidade, companheirismo e incentivo, por acompanhar de perto as minhas angústias, e estar sempre disposta a dar bons conselhos para que eu pudesse superar os diversos desafios encontrados.

A todos, que, de maneira geral, me apoiam e acreditam nas minhas boas intenções, os meus mais honestos agradecimentos. Muito Obrigado!

*A Deus pelas oportunidades
Aos meus pais Alaor, Sueli e meu irmão Fillipe, por todo o suporte
A todos aqueles dedicados a promover a inovação científica e tecnológica no Brasil
À minha amada Emily, pelo amor e companheirismo
Dedico.*

SUMÁRIO

RESUMO	9
ABSTRACT	11
LISTA DE TABELAS.....	17
LISTA DE ABREVIATURAS E SIGLAS	19
1 INTRODUÇÃO	21
2 REVISÃO BIBLIOGRÁFICA.....	23
2.1 Conceitos de inventário florestal	23
2.1.1 Amostragem aleatória simples	24
2.1.2 Estimador de regressão	25
2.2 Conceitos de sensoriamento remoto (SR)	26
2.2.1 Resolução espacial	27
2.2.2 Resolução espectral	28
2.2.3 Resolução radiométrica	29
2.2.4 Resolução temporal	30
2.2.5 Aplicações do SR no contexto florestal.....	30
2.3 Conceitos de estatística espacial.....	31
2.3.1 Krigagem ordinária (KO).....	32
2.3.2 Krigagem com deriva externa (KDE).....	33
3 MATERIAL E MÉTODOS	35
3.1 Caracterização da área de estudo.....	35
3.2 Caracterização das imagens e dos sensores remotos.....	36
3.2.1 Satélite LANDSAT-8	36
3.2.2 Satélite RAPIDEYE.....	39
3.2.3 Satélite GEOEYE-1	42
3.2.4 Imagens aerotransportadas	45
3.2.5 Processamento das imagens - Padronização	47
3.3 Caracterização da base de dados.....	47
3.4 Técnicas preditivas	48
3.4.1 Seleção de variáveis e regressão linear múltipla (RLM).....	48
3.4.2 Verificação de pressuposições	50
3.4.3 Krigagem ordinária e Krigagem com deriva externa.....	50
3.4.4 Estimativa da média e construção de intervalos de confiança.....	52
3.4.5 Estimativa de redução da intensidade amostral.....	53

3.5 Fluxograma geral do método	53
4 RESULTADOS E DISCUSSÃO	54
4.1 Padronização das imagens e geração de covariáveis.....	55
4.2 Análise exploratória da variável resposta	56
4.3 Regressão linear múltipla	57
4.3.1 Seleção de variáveis, ajuste e desempenho dos modelos de RLM	57
4.3.2 Estimativas volumétricas.....	60
4.4 Estimativas geoestatísticas.....	66
4.4.1 Krigagem ordinária (KO).....	66
4.4.1.1 Ajuste e desempenho da KO	66
4.4.1.2 Estimativa volumétrica.....	67
4.4.2 Krigagem com deriva externa (KDE).....	70
4.4.2.1 Ajuste e desempenho da KDE	70
4.4.2.2 Estimativa volumétrica.....	74
4.5 Análise comparativa dos estimadores espaciais	79
4.6 Análise comparativa dos estimadores clássicos.....	82
5 CONCLUSÕES.....	83
REFERÊNCIAS.....	85
ANEXO	93
APÊNDICE.....	103

RESUMO

Estatística espacial e sensoriamento remoto para a predição volumétrica em florestas de *Eucalyptus* spp.

O inventário florestal é uma das principais ferramentas na gestão dos recursos florestais, uma vez que as informações geradas por ele são utilizadas ao longo de toda a cadeia produtiva do setor. Desta forma, erros nas estimativas volumétricas dos inventários florestais devem ser controlados. Inúmeras informações podem ser obtidas a partir de imagens orbitais ou aerotransportadas, uma vez que podem cobrir facilmente toda a área de interesse, e estão comumente disponíveis em empresas florestais ou ao usuário final. A utilização de preditores derivados das imagens pode trazer benefícios para as estimativas do inventário florestal. Desta forma, a aplicação de técnicas de regressão linear múltipla (RLM) ganhou espaço no setor devido a sua facilidade de aplicação. Porém, a RLM não leva em consideração a dependência espacial entre as unidades amostrais, sendo que a geoestatística pode ser utilizada para prever a distribuição espacial do estoque de madeira (VTCC) para uma dada região. A modelagem geoestatística mais simples como a krigagem ordinária (KO), por considerar apenas a dependência espacial entre os pontos não amostrados, pode apresentar erros de predição nestes locais. Tais erros podem ser reduzidos com a aplicação de técnicas mais robustas como a Krigagem com Deriva Externa (KDE), pois esta agrega as informações obtidas das imagens com a distribuição espacial do volume. Buscando-se avaliar as vantagens da integração do Sensoriamento Remoto (SR) ao inventário florestal foram testados 4 tipos diferentes de imagens; as oriundas dos satélites LANDSAT8, RAPIDEYE e GEOEYE, e as provenientes de aeronaves (Imagens Aerotransportadas). Avaliou-se também diferentes tipos de estimativas para a predição volumétrica sendo estas RLM, KDE e KO. A melhor estimativa serviu de variável auxiliar para o estimador de regressão (ER), sendo os resultados comparados com a abordagem tradicional da amostragem aleatória simples (AAS). Os resultados demonstraram por meio da validação cruzada que as estimativas da KDE foram mais eficientes que as estimativas da KO e da RLM. Os melhores preditores (variáveis auxiliares) foram aqueles derivados do satélite LANDSAT8 e do satélite RAPIDEYE. Obteve-se como produto das estimativas de KDE e RLM mapas capazes de detectar áreas com mortalidade ou anomalias em meio a formação florestal. A utilização de uma estimativa de KDE utilizando imagens LANDSAT8 como medida auxiliar para o ER permitiu reduzir o erro amostral da AAS de 3,87% para 2,34%. Da maneira tradicional, tal redução de erro apenas seria possível com um aumento de mais 99 unidades amostrais.

Palavras-chave: Krigagem com Deriva Externa; Krigagem Ordinária; Imagens de satélite; Regressão Linear Múltipla

ABSTRACT

Spatial Statistics and Remote Sensing applied to estimating volume in *Eucalyptus* spp. forests

Forest Inventory (FI) is one of the main tools for managing forest resources, once the information derived from FI is used along the sector production chain. When estimating volume, errors resulting from FI are common, therefore these errors must be controlled. Once orbital or airborne imaging data are easily acquired for an entire area, and are commonly available in forest companies or for the end user, much information can be obtained from these products. The use of predictor derived from images can be of significant benefits to forest inventory estimates. For that reason, the application of linear multiple regression (LMR) techniques have taken place in the forest sector, due to the facilities of its application. However, the LMR technique does not take the spatial dependence among sample units in consideration, the geostatistics utilized to predict the spatial distribution of the wood stock (VTCC) for a specific region. Simpler geostatistical modeling as the ordinary kriging (OK), just takes in consideration the spatial dependence among non-sampled points, because of that, prediction errors can be found. Such errors can be reduced when techniques that are more robust are applied, such as the kriging with external drift (KED) approach. This technique aggregates the information obtained from the images with the spatial distribution of the volume. In order to evaluate the advantages of Remote Sensing and Forest Inventory integration, we considered 4 different types of images, from the satellites LANSAT 8, RAPIDEYE, GEOEYE and from airborne images. When predicting volume, three different approaches were evaluated: LMR, EDK, OK. The best model among those evaluated, served as auxiliary variable for the regression estimator (RE). The result were then compared to the traditional approach, simple random sampling (SRS). This approach showed, through a cross-validation, that the KDE estimates were more efficiently than the OK and the LMR. The best predictor model (auxiliary variables) were derived from LADNSAT 8 and RAPIDEYE satellites. There is a significant advantage to using the KDE and LMR approaches, as it allows for a spatial representation of areas with mortality or anomalies, in a forest environment. The combination of KDE approach and LANDSAT 8 images as an auxiliary method for the RE, abled the decrease of the sampling error of SRS from 3.87% to 2.34%. The traditional approaches to conduct plantation inventories would allow for this error reduction, only if there were an increase of 99 more sampling units.

Keywords: Kriging with External Drift; Ordinary Kriging; Satellite images; Multiple Linear Regression

LISTA DE FIGURAS

- Figura 1 - Processo de captação da energia eletromagnética, refletida por alvos terrenos, em um sensor orbital. Fonte: Adaptado de Meneses e Almeida (2012) e Pics About Spaces (2015).....27
- Figura 2–Exemplo do impacto de diferentes resoluções espaciais no detalhamento de uma mapa temático na Europa Central. Fonte: GCS (2015).....28
- Figura 3 – Número e largura das bandas do espectro eletromagnético, bem como a interação destas bandas com diferentes de alvos terrestres para os sensores dos Satélites RAPIDEYE. Fonte: RAPIDEYE (2015)28
- Figura 4 – Exemplos de imagens com diferentes níveis de resolução radiométrica em Brasília-DF Fonte: (MENESES; ALMEIDA, 2012).29
- Figura 5 – Semivariograma experimental típico para dados sem tendência e com ruído (efeito pepita). Cada ponto pode ser obtido por meio da aplicação da função variograma na distância h . Ajuste teórico de um modelo Esférico32
- Figura 6 – Distribuição dos talhões de *Eucalyptus* spp. e das parcelas amostrais que compõem a área de estudo, Horto Paineiras, Espírito Santo do Pinhal, SP35
- Figura 7 – Ocorrência de eventos importantes ao longo do tempo, Horto Paineiras, Espírito Santo do Pinhal, SP.....36
- Figura 8 – Composição colorida (R7-G5-B3) das imagens do satélite LANDSAT-8, geradas em janeiro de 2014, Horto Paineiras, Espírito Santo do Pinhal, SP38
- Figura 9 – Composição colorida (R3-G2-B1) das imagens do satélite RAPIDEYE, geradas em março de 2014, Horto Paineiras, Espírito Santo do Pinhal, SP41
- Figura 10 – Composição colorida (R1-G2-B3) das imagens do satélite GEOEYE-1, geradas em maio de 2013, Horto Paineiras, Espírito Santo do Pinhal, SP 44
- Figura 11 – Composição colorida (R1-G2-B3) das imagens aerotransportadas, obtidas em setembro de 2013, Horto Paineiras, Espírito Santo do Pinhal, SP46
- Figura 12 – Forma padrão das funções geoestatísticas implementadas no pacote “*gstat*”. Fonte: Pebesma (2014).....52
- Figura 14 – Fluxograma explicativo das etapas de processamento realizadas.....54
- Figura 15 – Análise exploratória em VTCC. A-Gráfico boxplot; B-Histograma; C-Gráfico Quantil-Quantil57

- Figura 16 – Valor de LOOCV RMSE (m^3/ha) em função do número de variáveis preditoras. A: imagens LS8; B: imagens RAPIDEYE; C: imagens GEOEYE e D: imagens Aerotransportadas 58
- Figura 17 – Análise gráfica dos ajustes. 1-Preditores LANDSAT8; 2-Preditores RAPIDEYE; 3-Preditores GEOEYE; 4-Preditores Aerotransportado. A-Gráfico Observado x Estimado; B-Dispersão dos resíduos e C-Histograma dos resíduos 60
- Figura 18 – Mapa de distribuição de VTCC (m^3/ha) para toda a fazenda – RLM utilizando preditores derivados das imagens LS8 62
- Figura 19 – Mapa de distribuição de VTCC (m^3/ha) para toda a fazenda – RLM utilizando preditores derivados das imagens RAPIDEYE 63
- Figura 20 – Mapa de distribuição de VTCC (m^3/ha) para toda a fazenda – RLM utilizando preditores derivados das imagens GEOEYE 64
- Figura 21 – Mapa de distribuição de VTCC (m^3/ha) para toda a fazenda – RLM utilizando preditores derivados das imagens Aerotransportadas 65
- Figura 22 – Semivariograma experimental para Krigagem Ordinária da variável VTCC 66
- Figura 23 – Ajuste do Semivariograma experimental para Krigagem Ordinária da variável VTCC 67
- Figura 24 – Análise gráfica do ajuste de KO. A-Gráfico Observado x Estimado; B-Dispersão dos resíduos e C-Histograma dos resíduos 67
- Figura 25 – Mapa de distribuição de VTCC (m^3/ha) para toda a fazenda, realizado por meio da KO 69
- Figura 26 – Comparação dos semivariogramas experimentais de KO e KDE para da variável VTCC. A: LS8; B: RAPIDEYE; C: GEOEYE e D: Aerotransportada 71
- Figura 27 – Ajuste do Semivariograma residual da Krigagem com Deriva Externa. A: LS8; B: RAPIDEYE; C: GEOEYE e D: Aerotransportada 72
- Figura 28 – Análise gráfica dos ajustes. 1-Preditores LANDSAT8; 2-Preditores RAPIDEYE; 3-Preditores GEOEYE; 4-Preditores Aerotransportado. A-Gráfico Observado x Estimado; B-Dispersão dos resíduos e C-Histograma dos resíduos 73
- Figura 29 – Mapa de distribuição de VTCC (m^3/ha) para toda a fazenda – KDE utilizando variáveis auxiliares derivados das imagens LANDSAT-8 75
- Figura 30 – Mapa de distribuição de VTCC (m^3/ha) para toda a fazenda – KDE utilizando variáveis auxiliares derivados das imagens RAPIDEYE 76

- Figura 31 – Mapa de distribuição de VTCC (m^3/ha) para toda a fazenda – KDE utilizando variáveis auxiliares derivados das imagens GEOEYE77
- Figura 32 – Mapa de distribuição de VTCC (m^3/ha) para toda a fazenda – KDE utilizando variáveis auxiliares derivados das imagens Aerotransportadas .78
- Figura 33 – Comparativo visual da capacidade de detecção de mortalidade no plantio, das predições de RLM, KDE e KO utilizando diferentes variáveis auxiliares. Pontos Verdes: Unidades Amostrais81
- Figura 34 –Matriz de correlação das variáveis preditoras oriundas das imagens LANDSAT8 e submetidas ao processo de *Best Subset Selection*. Varx=VTCC95
- Figura 35 –Matriz de correlação das variáveis preditoras oriundas das imagens RAPIDEYE e submetidas ao processo de *Best Subset Selection*. Varx=VTCC96
- Figura 36 –Matriz de correlação das variáveis preditoras oriundas das imagens GEOEYE e submetidas ao processo de *Best Subset Selection*. Varx=VTCC97
- Figura 37 –Matriz de correlação das variáveis preditoras oriundas das imagens Aerotransportadas e submetidas ao processo de *Best Subset Selection*. Varx=VTCC98

LISTA DE TABELAS

Tabela 1 – Variáveis derivadas do satélite LS8, e utilizadas como preditoras no processo de seleção. SRij: i/j com i e j = B1, B2, ..., B11	37
Tabela 2 – Variáveis derivadas do satélite RapidEye, e utilizadas como preditoras no processo de seleção. SRij: i/j com i e j = B1, B2, ..., B5	40
Tabela 3 – Variáveis derivadas do satélite GeoEye, e utilizadas como preditoras no processo de seleção. SRij: i/j com i e j = B1, B2, ..., B5	43
Tabela 4 – Variáveis derivadas do satélite Aerotransportada, e utilizadas como preditoras no processo de seleção. SRij: i/j com i e j = B1, B2, ..., B5	45
Tabela 5 – Estatística descritiva das 110 parcelas utilizadas. Volume Total com Casca (VTCC), Altura total média (HT), Diâmetro a altura do peito (DAP) e Área Basal (AB)	48
Tabela 6 – Modelos geoestatísticos implementados no pacote “gstat”. Fonte: Pebesma (2014)	51
Tabela 7 – Resumo das etapas iniciais do processamento das imagens	56
Tabela 8 – Resumo das estatísticas de ajuste RLM. *Valores de R ² Ajustado e RMSE e RMSE% obtidos a partir da validação cruzada LOOCV	58
Tabela 9 – Resumo das estatísticas de ajuste KDE. *Valores de RMSE% obtidos a partir da validação cruzada LOOCV	72
Tabela 10 – Resumo das estatísticas de ajuste RLM, KO e KDE. *Valores de RMSE% obtidos a partir da validação cruzada LOOCV	79
Tabela 11 – Resumo dos resultados da estimativa de VTCC obtidos pelas estimativas de AAS e ER.	82
Tabela 12 - Estatísticas do ajuste de RLM para o conjunto de dados LANDSAT8. *Valores de R ² Ajustado e RMSE% obtidos a partir da validação cruzada LOOCV	99
Tabela 13 - Estatísticas do ajuste de RLM para o conjunto de dados RAPIDEYE. *Valores de R ² Ajustado e RMSE% obtidos a partir da validação cruzada LOOCV	99
Tabela 14 - Estatísticas do ajuste de RLM para o conjunto de dados GEOEYE. *Valores de R ² Ajustado e RMSE% obtidos a partir da validação cruzada LOOCV	100
Tabela 15 - Estatísticas do ajuste de RLM para o conjunto de dados Aerotransportados. *Valores de R ² Ajustado e RMSE% obtidos a partir da validação cruzada LOOCV	100

- Tabela 16 – Seleção de modelos geostísticos candidatos da KO. RMSE%: Raíz do erro quadrático médio em porcentagem; ME: Erro médio; COR1: Correlação de Pearson (Observado x Estimado) e COR2: Correlação de Pearson (Estimado x Resíduo) 101
- Tabela 17 – Seleção de modelos geostísticos candidatos da KDE (LANDSAT8). RMSE%: Raíz do erro quadrático médio em porcentagem; ME: Erro médio; COR1: Correlação de Pearson (Observado x Estimado) e COR2: Correlação de Pearson (Estimado x Resíduo)..... 101
- Tabela 18 – Seleção de modelos geostísticos candidatos da KDE (RAPIDEYE). RMSE%: Raíz do erro quadrático médio em porcentagem; ME: Erro médio; COR1: Correlação de Pearson (Observado x Estimado) e COR2: Correlação de Pearson (Estimado x Resíduo)..... 101
- Tabela 19 – Seleção de modelos geostísticos candidatos da KDE (GEOEYE). RMSE%: Raíz do erro quadrático médio em porcentagem; ME: Erro médio; COR1: Correlação de Pearson (Observado x Estimado) e COR2: Correlação de Pearson (Estimado x Resíduo)..... 102
- Tabela 20 – Seleção de modelos geostísticos candidatos da KDE (Aerotransportada). RMSE%: Raíz do erro quadrático médio em porcentagem; ME: Erro médio; COR1: Correlação de Pearson (Observado x Estimado) e COR2: Correlação de Pearson (Estimado x Resíduo) 102

LISTA DE ABREVIATURAS E SIGLAS

AAS	-Amostragem Aleatória Simples
AB	-Área Basal
AGE	-Idade do Plantio
Bes	-Modelo Geostatístico Bessel
Cir	-Modelo Geostatístico Circular
CV%	-Coeficiente de Variação em Porcentagem
DAP	-Diâmetro a altura do peito
ER	-Estimador de Regressão
Exp	-Modelo Geostatístico Exponencial
Gau	-Modelo Geostatístico Gaussiano
HT	- Altura total média
IBÁ	-Indústria Brasileira de Árvores
KDE	-Krigagem com Deriva Externa
KO	-Krigagem Ordinária
KU	-Krigagem Universal
LiDAR	-Light Detection and Ranging
Lin	-Modelo Geostatístico Linear
LOOCV	-Validação Cruzada Leave-One-Out
LS5	-Satélite LANDSAT-5
LS7	-Satélite LANDSAT-7
LS8	-Satélite LANDSAT-8
Matgen	-Material Genético
NASA	-National Aeronautics and Space Administration
NDVI	-Normalized Difference Vegetation Index
NDVI _{RE}	-Normalized Difference Vegetation Index RedEdge
OLI	-Operational Land Imager Sensor
Pen	-Modelo Geostatístico Pentaesférico
Per	-Modelo Geostatístico Periódico
Pow	-Modelo Geostatístico de Potência
RLM	-Regressão Linear Múltipla
RMSE	-Raíz do Erro Quadrático Médio
RS	-Estado do Rio Grande do Sul
Shp	-Modelo Geostatístico Esférico
SP	-Estado de São Paulo
SR	-Sensoriamento Remoto
SR _{ij}	-Simple Ratio Index
TIRS	-Thermal Infrared Sensor
USGS	-United States Geological Survey
VANT	-Veículo aéreo não tripulado
VTCC	-Volume Total com Casca

1 INTRODUÇÃO

As florestas plantadas desempenham papel fundamental no setor florestal brasileiro e mundial, em especial pelo fornecimento de matérias-primas para diversas finalidades, tais como: a produção de celulose, papel, bioenergia, madeira, painéis reconstituídos, fármacos, entre outros (IBÁ, 2015). Sendo assim, as plantações florestais contribuem na redução das pressões antrópicas sobre as florestas naturais (GARLIPP; FOELKEL, 2009; PANY et al., 2015).

A área total ocupada por florestas plantadas vem crescendo a cada ano (IBÁ, 2015; PANY et al. 2015). Com base no último levantamento realizado em 2014 pela Indústria Brasileira de Árvores (IBÁ), o Brasil possui 7,7 milhões de hectares de florestas plantadas, destacando-se em 1º lugar entre os maiores produtores de carvão vegetal, 4º lugar entre os produtores de celulose, 7º lugar entre os produtores de painéis reconstituídos e em 9º lugar entre os países produtores de papel (IBÁ, 2015).

A fim de gerir adequadamente todos esses recursos florestais, a principal ferramenta utilizada por empresas e gestores da área, é o inventário florestal. O inventário, é capaz de fornecer informações quantitativas e qualitativas dos povoamentos de interesse, permitindo assim, um maior detalhamento dos recursos florestais e fornecendo subsídios para as tomadas de decisões (SOARES et al., 2006).

Com o objetivo de melhorar a assertividade do inventário florestal, diversos métodos e preditores vêm sendo globalmente pesquisados, destacando-se dois grupos de iniciativas: o primeiro envolve a instalação de parcelas em campo e a subsequente modelagem da dependência espacial entre as parcelas, por meio da geoestatística univariada (MELLO, 2004; MORELI et al., 2006; KANEGAE JUNIOR et al., 2007; MELLO et al., 2009, RÉ et al., 2009; SANTANA, 2011; LEAL et al., 2011; GUEDES et al., 2012; SILVA, 2014; RIBEIRO, 2014; PELISSARI et al., 2015); já o segundo grupo busca associar as parcelas de campo com imagens de satélite, obtidas por sensoriamento remoto, utilizando técnicas de regressão (CANAVESI et al., 2010; JI et al., 2012; BERRA et al., 2012; ZHANG et al., 2014; ZHU; LIU, 2015; BARROS et al., 2015; SOUSA et al., 2015).

Ambas as abordagens se mostram interessantes, por buscarem soluções para aumentar a confiabilidade das estimativas produzidas pelo inventário florestal. Porém, quando utilizadas de forma isolada, podem apresentar limitações em sua aplicação.

A geoestatística, no entanto, pode ser aplicada de forma multivariada, quando além da variável principal, espacializada, existir uma ou mais variáveis secundárias capazes de explicar parte da variação observada nos dados. A Krigagem com Deriva Externa (KDE) (WACKERNAGEL, 2003; HENGL et al., 2007), também chamada por alguns autores de Krigagem Universal (KU) (BIVAND et al., 2011; HENGL et al., 2009), caracteriza-se como um destes métodos, e pode ser aplicada sempre quando as variáveis secundárias estiverem presentes em toda a área de interesse, condicionante facilmente alcançada ao se trabalhar com imagens de satélite ou aerotransportadas.

Estas imagens apresentam valores de refletância correlacionados com o volume de madeira nas florestas (mensurado em campo) e podem fornecer informações importantes para a modelagem geoestatística multivariada *pixel a pixel* ao longo de toda a área de interesse. Utilizando esta abordagem, seria possível a detecção de regiões com mortalidade, além de regiões com alta ou baixa produtividade, e que ocorram em locais não contemplados por unidades amostrais.

Dessa forma, o presente trabalho teve por objetivo geral avaliar a eficiência da utilização do interpolador geoestatístico multivariado KDE, frente ao interpolador univariado Krigagem Ordinária (KO), à regressão linear múltipla (RLM) e ao estimador da amostragem aleatória simples (AAS), com foco na estimativa volumétrica em florestas comerciais de *Eucalyptus* spp.. Além disso, foram avaliados os efeitos na qualidade das estimativas multivariadas ao se utilizar covariáveis oriundas de quatro sensores imageadores diferentes (LANDSAT 8, RAPIDEYE, GEOEYE e Aerotransportado).

2 REVISÃO BIBLIOGRÁFICA

A seguir listam-se os tópicos da revisão de literatura diretamente relacionados com o escopo da dissertação. No primeiro item são apresentados alguns conceitos de inventário florestal bem como a descrição matemática dos estimadores da Amostragem Aleatória Simples (AAS) e do Estimador de Regressão (ER), ambos utilizados na análise dos dados. Em seguida são introduzidos os principais princípios de aquisição de dados por sensores imaginadores, bem como as características intrínsecas das imagens, que podem influenciar na capacidade explicativa destas nas aplicações florestais. Por fim, aborda-se alguns conceitos básicos para a entendimento da modelagem da dependência espacial (Geoestatística) bem como a formulação matemática dos estimadores de Krigagem Ordinária (KO) e Krigagem com Deriva Externa (KDE).

2.1 Conceitos de Inventário Florestal

Toda a tomada de decisão requer informação (KANGAS; MALTAMO, 2006). No contexto florestal, a obtenção de informações sobre a floresta como um todo denomina-se inventário florestal (KANGAS; MALTAMO, 2006; QUEIROZ, 2012). Conforme Queiroz (2012), o inventário florestal pode ser definido como o ramo da ciência florestal que visa avaliar as variáveis qualitativas e quantitativas da floresta e suas inter-relações, assim como as dinâmicas de crescimento e a sucessão florestal, servindo de base para a formulação de planos de utilização dos produtos florestais, manejo integrado da floresta, bem como para alicerçar propostas de planos de desenvolvimento e política florestal de caráter local, regional ou nacional.

Para a obtenção das informações em campo, o inventário florestal tradicional baseia-se na teoria da amostragem (BATISTA et al., 2014). Desta forma, utilizam-se amostras mensuradas em campo a fim de inferir sobre a população como um todo. Isso ocorre, pois um censo florestal (mensuração de todas as árvores), é na maioria dos casos impossível ou impraticável, devido ao elevado custo operacional e a grande extensão das florestas (KÖHL et al., 2006; KANGAS; MALTAMO, 2006)

A estimativa da média populacional de uma variável de interesse em um inventário florestal, dependerá da média amostral mais um intervalo de confiança, que por sua vez, é derivado da variância amostral observada entre as parcelas (KANGAS; GOVE; SCOTT, 2006). Além disso, a formulação do estimador do intervalo de confiança também varia conforme o procedimento amostral adotado (KANGAS, 2006).

Nesse contexto, a incerteza amostral máxima admitida em inventários florestais pode variar entre empresas, ou conforme a legislação específica. No caso das plantações comerciais equiâneas, é comum que a incerteza amostral máxima tolerada esteja abaixo de 10% (CAMPOS; LEITE, 2013). Desta forma, a redução dos intervalos de confiança em torno da média, em um inventário florestal é, na maioria das vezes, associada ao aumento da intensidade amostral.

Porém, sabe-se que a operação de coleta de dados em campo compõem a maior parte dos custos em um inventário florestal (CAMPOS; LEITE, 2013). Neste sentido, um aumento na intensidade amostral pode ser custoso, comprometendo o orçamento destinado para execução da atividade.

A fim de solucionar esse dilema, isto é, o aumento da precisão em inventários florestais, sem que para isso se eleve os custos com o aumento da intensidade amostral, o uso do sensoriamento remoto surge como um grande aliado na obtenção e na geração das informações de interesse em diferentes escalas (KÖHL et al., 2006; SHINZATO, 2015).

2.1.1 Amostragem Aleatória Simples

A amostragem aleatória simples (AAS), também conhecida como amostragem casual simples (ACS), baseia-se no princípio de que cada amostra de mesmo tamanho (n) tem a mesma probabilidade de ser obtida, ou seja, selecionam-se aleatoriamente as unidades amostrais. Este delineamento é mais indicado em populações homogêneas, sendo o seu estimador um dos mais aplicados comercialmente na indústria devido a sua simplicidade (QUEIROZ, 2012; CAMPOS; LEITE, 2013). De acordo com Kangas (2006), os estimadores da AAS são descritos conforme as Equações 1, 2, 3 e 4, abaixo:

-Média populacional

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad (1)$$

-Variância populacional

$$S_y^2 = \sum_{i=1}^n \frac{(y_i - \bar{y})^2}{n-1} \quad (2)$$

-Erro padrão da média com correção para populações finitas

$$S_{\bar{y}} = \sqrt{\frac{S_y^2}{n} \left(1 - \frac{n}{N}\right)} \quad (3)$$

-Erro amostral em porcentagem ($\alpha=0,05$)

$$E\% = \frac{t_{(n-1)}^{0,05} * S_{\bar{y}}}{\bar{y}} * 100 \quad (4)$$

Onde: i é o índice de indica as amostras (parcelas), n é o tamanho da amostra, y_i é a medida de interesse na amostra, N é o tamanho populacional e t é valor da distribuição t de Student para $\alpha=0,05$.

2.1.2 Estimador de Regressão

Segundo Batista et al. (2014), o objetivo do estimador de regressão é sempre aumentar a precisão da estimativa da medida de interesse. Para isso, o estimador de regressão sempre requer o conhecimento da média populacional da medida auxiliar. Nos inventários florestais, o estimador de regressão pode ser utilizado para aumentar a precisão da medida de produção de volume de madeira com base em atributos de simples medição em florestas.

O estimador de regressão é considerado apropriado quando há uma relação linear entre a medida de interesse e a medida auxiliar, contudo, a reta que representa essa relação não obrigatoriamente passa pela origem (BATISTA et al., 2014).

Conforme Kangas (2006) e Schreuder et al. (2004), o estimador de regressão para média segue a expressão da reta (Equação 5):

$$\bar{y}_R = \bar{y} + \hat{\beta}(\bar{X} - \bar{x}) \quad (5)$$

Onde: \bar{y}_R é o valor médio da variável de interesse estimado com base na variável auxiliar; \bar{y} é a média amostral da variável de interesse; \bar{x} é a média amostral da variável auxiliar; \bar{X} é a média populacional da variável auxiliar e $\hat{\beta}$ é o coeficiente de regressão que representa a inclinação da relação linear, sendo obtido pelo método dos quadrados mínimos da regressão linear (Equação 6):

$$\hat{\beta} = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (6)$$

Conforme Batista et al. (2014), como $\hat{\beta}$ é a estimativa do coeficiente de inclinação da relação linear, a relação pode ser tanto positiva ($\beta > 0$), quanto negativa ($\beta < 0$). O estimador de regressão será igualmente eficiente em ambos os casos, desde que haja uma forte relação linear entre a medida de interesse (Y) e a medida auxiliar (X).

Por fim, o erro padrão de \bar{y}_R é dado pela Equação 7.

$$S_{\bar{y}_R} = S_{y,x} \sqrt{\left(\frac{1}{n} + \frac{(X-\bar{x})^2}{SS_x}\right) \frac{N-n}{N}} \quad (7)$$

Onde: $S_{y,x}$ é a interação y e x calculada pela Equação 8:

$$S_{y,x} = \sqrt{\frac{SS_y - \frac{(SP_{xy})^2}{SS_x}}{n-2}} \quad (8)$$

Onde: $SS_y = \sum_{i=1}^n (y_i - \bar{y})^2$; $SS_x = \sum_{i=1}^n (x_i - \bar{x})^2$ e $SP_{xy} = \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})$

2.2 Conceitos de Sensoriamento remoto (SR)

Segundo Elachi (1987), o Sensoriamento Remoto (SR) caracteriza-se pela capacidade de aquisição de informações sobre um objeto/alvo sem que para isso se entre em contato físico com ele. Para tanto, utiliza-se de forma conjunta sensores remotos, equipamentos para processamento de dados e equipamentos de transmissão de dados instalados a bordo de aeronaves, espaçonaves, ou outras plataformas, capazes de registrar e analisar as interações entre a radiação eletromagnética e as substâncias que compõem o objeto de estudo. Permitindo assim, a análise de fenômenos e processos que ocorrem na superfície do planeta Terra (NOVO, 2010).

A Figura 1 demonstra esquematicamente o processo da aquisição da imagens por sensores remotos passivos. Segundo Jensen (2009), esse processo resume aos seguintes passos: uma fonte de energia, como a radiação solar, se propaga através da atmosfera até atingir a superfície da terra, interagindo com a mesma. A partir desta

interação, essa energia se propaga na atmosfera sendo captada pelos sensores remotos, gerando informação digital.

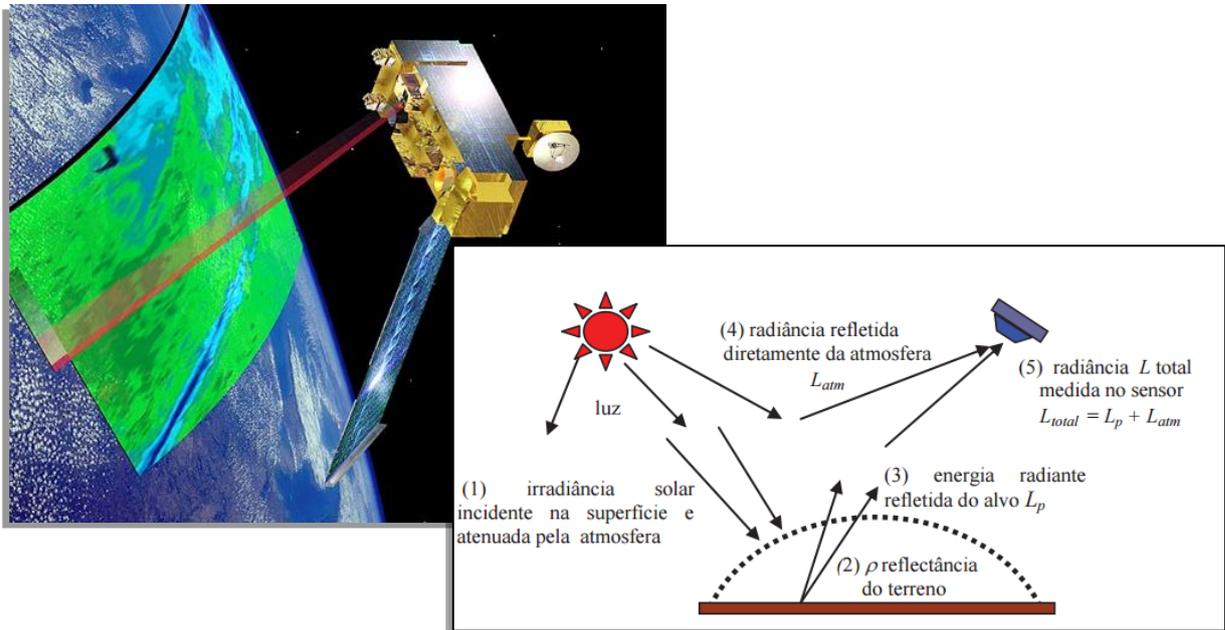


Figura 1 - Processo de captação da energia eletromagnética, refletida por alvos terrenos, em um sensor orbital. Fonte: Adaptado de Meneses e Almeida (2012) e Pics About Spaces (2015)

2.2.1 Resolução Espacial

O tamanho individual do elemento de área imageada no terreno representa em qualquer tipo de sensor uma propriedade importante da imagem: a resolução espacial. Por definição, um objeto somente pode ser detectado, quando o tamanho deste é, no mínimo, igual ou maior do que a resolução espacial da imagem (MENESES; ALMEIDA, 2012).

Em terrenos naturais as feições da paisagem apresentam uma alta correlação espacial mostrando pouca variabilidade ao longo de extensas áreas, nestas situações sensores com média ou baixa resolução espacial são indicados. Já, para uma área com alta variabilidade de tipos de objetos, como numa área urbana, a utilização de sensores de alta resolução espacial é necessária para que se possa diferenciar as casas, as ruas, os prédios, etc (MENESES; ALMEIDA, 2012).

Para que o sensoriamento remoto em ambientes florestais seja operacional, a resolução espacial do sensor imageador deve ser apropriada para a aplicação demandada. Desta forma, a resolução espacial escolhida determinará o conteúdo da informação e o erro de mensuração (KÖHL et al., 2006).

A Figura 2 exemplifica os efeitos que diferentes resoluções espaciais podem apresentar sobre a qualidade de detalhamento de um mapa temático.

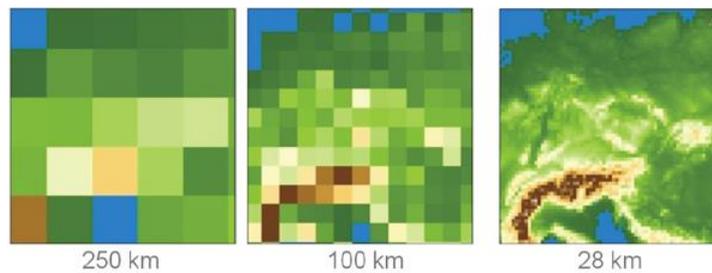


Figura 2—Exemplo do impacto de diferentes resoluções espaciais no detalhamento de uma mapa temático na Europa Central. Fonte: GCS (2015)

2.2.2 Resolução Espectral

A resolução espectral é uma medida da largura das faixas espectrais e da sensibilidade do sistema sensor em distinguir entre dois níveis de intensidade do sinal de retorno (resolução radiométrica) (NOVO, 2010). Comparativamente, um sensor tem melhor resolução espectral se ele possui maior número de bandas situadas em diferentes regiões espectrais e com larguras estreitas de comprimentos de onda (MENESES; ALMEIDA, 2012).

No âmbito florestal, quando se deseja distinguir diferentes tipos de espécies e detectar características de sanidade dos plantios, a utilização de sensores com maior resolução espectral é indicada (KÖHL et al., 2006).

Para exemplificar esse conceito, a Figura 3 demonstra a quantidade e a largura das bandas (resolução espectral) dos sensores remotos da constelação de satélites RAPIDEYE, bem como cada uma das bandas se relaciona com diferentes alvos terrestres (vegetação verde, vegetação seca e solo exposto).

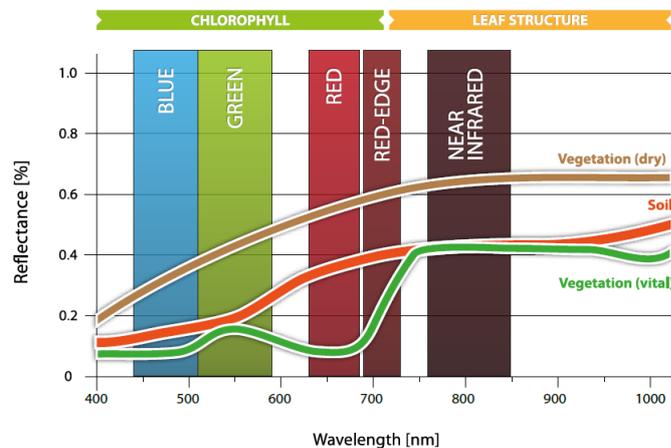


Figura 3 – Número e largura das bandas do espectro eletromagnético, bem como a interação destas bandas com diferentes de alvos terrestres para os sensores dos Satélites RAPIDEYE. Fonte: RAPIDEYE (2015)

2.2.3 Resolução Radiométrica

A resolução radiométrica de um sensor influenciará no reconhecimento dos padrões espectrais das feições, atividade fundamental no mapeamento de uso e cobertura do solo e delimitação dos talhões florestais (VALENTE, 2001). Segundo Novo (2010), a resolução radiométrica descreve a habilidade de um sensor distinguir variações no nível de energia refletida, emitida ou retroespalhada que deixa a superfície do alvo.

Esta energia apresenta diferenças de intensidade contínuas, as quais precisam ser detectadas, registradas e reproduzidas pelo sensor (NOVO, 2010). Em função da intensidade da radiação de entrada no sensor, a resposta de saída dos detectores é convertida eletronicamente em um número digital discreto (MENESES; ALMEIDA, 2012).

A variação na resolução radiométrica corresponde ao número de *bits* utilizados para codificar os números no formato binário. Cada *bit* grava uma exponencial elevado a 2, sendo que o valor máximo de brilho disponível depende do número de bits utilizado (KÖHL et al., 2006).

A Figura 4 demonstra a gradual perda de informação do alvo imageado a medida que a capacidade de detecção radiométrica (resolução radiométrica) diminui.

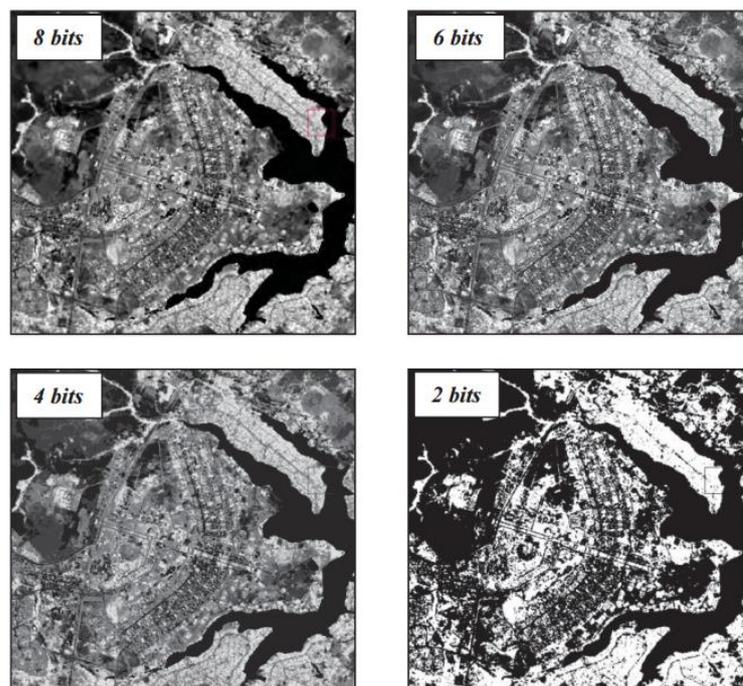


Figura 4 – Exemplos de imagens com diferentes níveis de resolução radiométrica em Brasília-DF Fonte: (MENESES; ALMEIDA, 2012).

2.2.4 Resolução Temporal

Tendo em vista que muitas das informações de sensoriamento remoto utilizadas em empresas florestais provêm de imagens de satélite, como o LANDSAT8, a resolução temporal é muito importante, principalmente quando o objetivo é o monitoramento sequencial e temporal de uma dada floresta. Esta resolução, segundo Köhl et al. (2006), refere-se ao período de revisita e duração de tempo que o satélite leva para completar um ciclo orbital inteiro, expresso em dias. O satélite LANDSAT8, por exemplo, leva 16 dias para completar um ciclo orbital.

2.2.5 Aplicações do SR no contexto florestal

Informações de sensoriamento remoto obtidas por meio de sensores orbitais ou aerotransportados são utilizadas há décadas no contexto florestal, principalmente, no fornecimento de informações cartográficas das áreas de interesse, necessárias para um correto planejamento das atividades de inventário florestal (HOLMGREN; THURESSON, 1998; COUTO, 1993).

Entretanto, com o advento dos sensores imageadores e a gradual redução dos custos para aquisição, tratamento e a utilização destas imagens pelo usuário final, as possibilidades de aplicação do sensoriamento remoto ao inventário florestal superam a simples aplicação para fins cartográficos.

Pesquisadores estudando o comportamento da correlação entre o volume de povoamentos florestais e imagens orbitais, constataram que este, de modo geral, se correlaciona negativamente com os valores de refletância obtidos das imagens, sendo observados valores de correlação entre -0,26 a -0,82, a depender do sensor digital utilizado, do comprimento de onda avaliado, das resoluções espacial e radiométrica e da densidade do povoamento florestal (LOPES, 2005; FIORELLA; RIPPLE, 1993; TROTTER et al., 1997; MAGNUSSON, 2006; RIBEIRO; 2007)

Desta forma, após a constatação de que os valores de refletância obtidos por sensores remotos, nos mais diversos comprimentos de onda do espectro radiométrico, apresentam correlação com as características dendrométricas dos povoamentos florestais, e são capazes de varrer amplas áreas com relativa facilidade e baixos custos, novas abordagens surgiram a fim de aproveitar tais características, para a estimativa de parâmetros florestais por meio de imagens.

De maneira geral, para a estimativa volumétrica por meio de sensores remotos, utilizam-se parcelas de inventário florestal, alocadas em campo, para

calibração de modelos preditivos. Nestes modelos, as diversas bandas espectrais dos sensores remotos, constituem as variáveis predictoras, enquanto que os parâmetros florestais de interesse, as variáveis resposta a serem estimadas *pixel a pixel*, sobre todo o domínio da imagem (CANAVESI et al., 2010; JI et al., 2012; BERRA et al., 2012; ZHANG et al., 2014; ZHU; LIU, 2015; BARROS et al., 2015; SOUSA et al., 2015).

2.3 Conceitos de Estatística Espacial

Conforme Cressie (1993), o termo estatística espacial é utilizado para descrever uma gama de modelos e métodos estatísticos desenvolvidos para a análise de dados espacialmente referenciados. Segundo o mesmo autor, a estatística espacial pode ser dividida em três principais ramos, sendo esses: a variação discreta do espaço (*discrete spatial variation*), o processo de pontos espacializados (*spatial point processes*) e a variação contínua do espaço ou a geoestatística (*geoestatistics*).

No caso de formações florestais e das variáveis ambientais de maneira geral, sabe-se que estas apresentam com frequência dependência espacial, ou seja, a variação observada nas variáveis estudadas, tais como o volume, a altura média, o diâmetro, entre outras, não é aleatória. Porém, apresentam um padrão de distribuição contínuo capaz de ser descrito e modelado com o auxílio de técnicas geoestatísticas.

Diversos autores já estudaram a aplicação das técnicas geoestatísticas, na formação de estratos mais homogêneos, na predição volumétrica de madeira e no estudo do comportamento das variáveis dendrométricas em relação ao espaço (MELLO, 2004; MORELI et al., 2006; KANEGAE JUNIOR et al., 2007; MELLO et al., 2009, RÉ et al., 2009; SANTANA, 2011; GUEDES et al., 2012).

Conforme Yamamoto e Landim (2013), uma modelagem geoestatística pode ser dividida basicamente em três etapas. A primeira consiste na construção de um semivariograma experimental (Figura 5), com base na função variograma (Equação 9).

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \quad (9)$$

Onde: $\hat{\gamma}(h)$ é a semivariância estimada entre pares de pontos na mesma distância h ; $N(h)$ é o número de pares de pontos separados por um vetor h ; $Z(x_i)$ é o valor da variável regionalizada no ponto x_i (coordenada geográfica) e $Z(x_i + h)$ é o valor no ponto $x_i + h$.

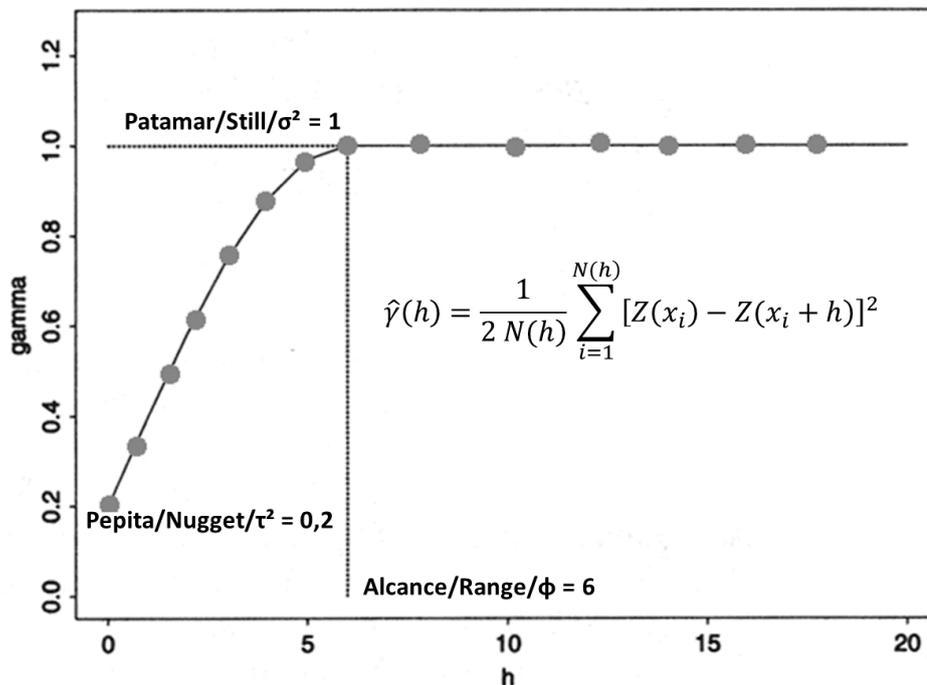


Figura 5 – Semivariograma experimental típico para dados sem tendência e com ruído (efeito pepita). Cada ponto pode ser obtido por meio da aplicação da função variograma na distância h . Ajuste teórico de um modelo Esférico

A segunda etapa, após a construção do semivariograma consiste na escolha e ajuste de modelos matemáticos teóricos que sejam capazes de explicar o padrão do comportamento espacial da variável em estudo (Figura 5), diversos modelos estão disponíveis na literatura (HENGL, 2009; CHILE'S; DELFINER, 2012; YAMAMOTO; LANDIM, 2013). Dependendo dos objetivos do estudo, a terceira etapa poderá ser desenvolvida ou não, para tanto, aplica-se o modelo teórico ajustado na etapa anterior em cada local x_i onde a estimativa da variável regionalizada é de interesse.

2.3.1 Krigagem Ordinária (KO)

O estimador espacial denominado krigagem ordinária (KO), é o estimador mais difundido e conhecido da estatística espacial, tem como base os dados amostrados da variável regionalizada e as propriedades estruturais do semivariograma obtido a partir destes dados, ou seja, o valor estimado da variável de interesse em uma dada localidade irá depender apenas da distância e do peso de seus vizinhos mais próximos, conforme a modelagem da dependência espacial no semivariograma (JOURNEL; HUIJBREGTS, 1978). Conforme Yamamoto e Landim (2013), o estimador da krigagem ordinária é dado pela Equação 10.

$$\hat{Z}_{KO}(x_o) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (10)$$

Onde: \hat{Z}_{KO} é o estimador do valor na posição x_o (média ponderada dos dados); n o número de pontos amostrais vizinhos utilizados para a predição do valor não amostrado Z_{KO}^* (vizinhança da krigagem); λ_i é o i -ésimo peso atribuído a cada i -ésima observação da variável de interesse na posição x_i ; $Z(x_i)$ definido pelo semivariograma.

2.3.2 Krigagem com Deriva Externa (KDE)

A geoestatística, no entanto, também pode ser aplicada de forma multivariada, quando além da variável espacial, existir uma ou mais variáveis secundárias capazes de explicar parte da variação observada nos dados. Para tanto, as variáveis secundárias devem apresentar alguma correlação com a variável principal, sendo que o aumento da correlação entre estas variáveis, resultará em uma melhoria na eficiência da predição (WATANABE et al., 2009).

A utilização da krigagem com deriva externa (KDE) também requer que os dados sejam multicolocalizados. De acordo com Wackernagel (2003), o estimador da krigagem com deriva externa resulta da somatória dos valores da variável regionalizada Z nos pontos x_i multiplicados por um peso λ_i (Equação 11).

$$\hat{Z}_{KDE}(x_o) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (11)$$

Os pesos λ_i provêm da resolução do sistema de Equações 12 (WACKERNAGEL, 2003):

$$\begin{cases} \sum_{j=1}^n \lambda_j C_r(x_i - x_j) - \mu_1 - \mu_2 s(x_i) = C_r(x_i - x_o) & \text{para } i = 1, n \\ \sum_{j=1}^n \lambda_j = 1 \\ \sum_{j=1}^n \lambda_j s(x_j) = s(x_o) \end{cases} \quad (12)$$

Onde: \hat{Z}_{KDE} é o estimador do valor na posição x_o (média ponderada dos dados); n o número de pontos amostrais vizinhos utilizados para a predição do valor não amostrado \hat{Z}_{KDE} (vizinhança da krigagem); λ_i é o i -ésimo peso atribuído a cada i -

ésima observação da variável de interesse na posição x_i ; $s(x_i)$ definido pelo semivariograma; μ_1 e μ_2 são multiplicadores de Lagrange e $C_r(x_i - x_j)$ é a covariância dos resíduos entre os pontos x_i e x_j .

Segundo Watanabe et al. (2009), a aplicação deste método considera a função variograma residual como modelo de correlação espacial. O cálculo do variograma residual exige que as variáveis primária e secundárias sejam conhecidas em todos os pontos amostrais e correlacionadas entre si.

3 MATERIAL E MÉTODOS

3.1 Caracterização da área de estudo

A pesquisa foi desenvolvida em povoamentos clonais de *Eucalyptus* spp., pertencentes a empresa International Paper do Brasil S/A., localizados na região oeste do Município de Espírito Santo do Pinhal, SP (entre coordenadas 22°11'0" e 22°13'0" de Latitude Sul e 46°53'30" e 46°48'30" de Longitude Oeste de Greenwich). De acordo com a classificação climática de Köeppen, o município apresenta clima do tipo "Cwa", ou seja, clima temperado úmido caracterizado por invernos secos e verões quentes, sendo que as médias anuais de precipitação na região, giram em torno de 1.540mm (CEPAGRI, 2015).

A área de estudo (Figura 6) localiza-se a um raio de 30km da Unidade Fabril consumidora e apresenta altitude e declividade média de 731m e 8,5%, respectivamente. Foram estudados 44 talhões adjacentes entre si, que totalizam 707,7ha de florestas de *Eucalyptus* spp. distribuídas em 9 tipos de materiais genéticos diferentes. A área foi manejada sob o regime de reforma com espaçamento médio entre plantas de 3x3m. Além disso, os plantios possuíam idades variando de 4,2 a 6,7 anos no momento da coleta dos dados em campo.

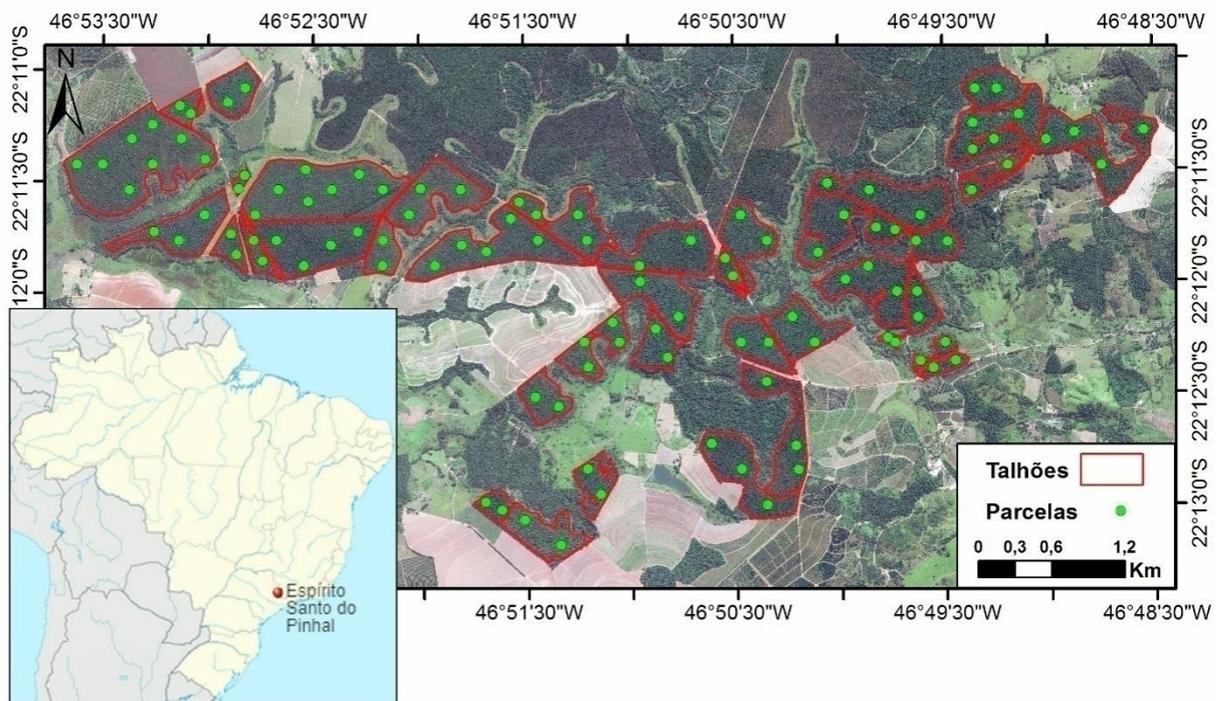


Figura 6 – Distribuição dos talhões de *Eucalyptus* spp. e das parcelas amostrais que compõem a área de estudo, Horto Paineiras, Espírito Santo do Pinhal, SP

3.2 Caracterização das imagens e dos sensores remotos

As imagens utilizadas na pesquisa, foram obtidas por diferentes sensores remotos, sendo três sensores orbitais e um sensor aéreo. A diferença máxima de tempo entre as diferentes imagens não foi superior a 7 meses. Da mesma forma, a diferença máxima de tempo entre as imagens e os levantamentos em campo também não ultrapassou 7 meses. A Figura 7 demonstra todos os eventos importantes relacionados à pesquisa em uma linha do tempo.

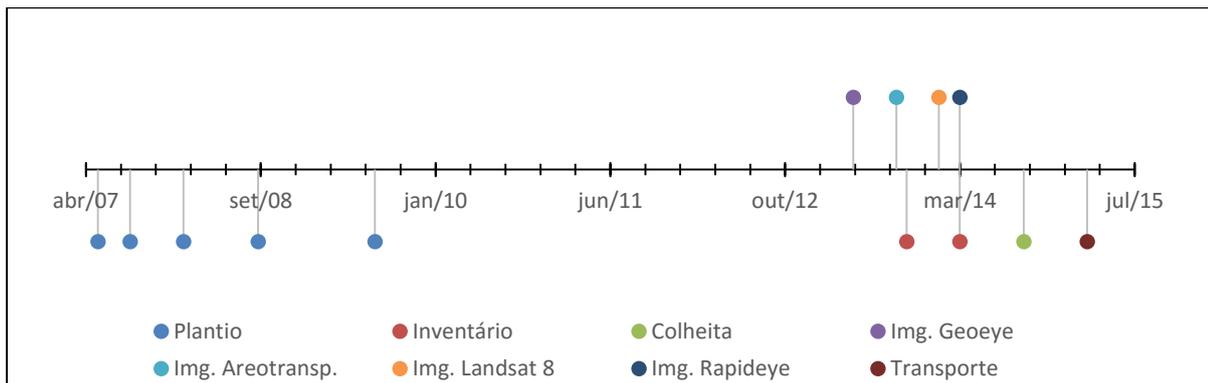


Figura 7 – Ocorrência de eventos importantes ao longo do tempo, Horto Paineiras, Espírito Santo do Pinhal, SP

As características de cada sensor, bem como as imagens derivadas deles, serão detalhadas a seguir.

3.2.1 Satélite LANDSAT-8

O satélite LANDSAT8 (LS8) é o oitavo satélite da série de satélites LANDSAT, dos quais apenas dois estão em operação (LS7 e LS8). O satélite LS8 é uma iniciativa conjunta da *National Aeronautics and Space Administration* (NASA) com outras empresas e órgãos governamentais americanos e é operado atualmente pela *U.S Geological Survey* (USGS), a qual disponibiliza as imagens provenientes do satélite de forma gratuita (USGS, 2013). O satélite apresenta resolução temporal de 16 dias e é equipado com dois sensores imageadores, OLI e o TIRS, que produzem imagens em diferentes comprimentos de onda e com características complementares.

As imagens provenientes do sensor *Operational Land Imager* (OLI), apresentam resolução espacial de 15 e 30m, resolução radiométrica de 16bits e resolução espectral que vai de 0,43 μ m até 1,38 μ m, distribuídas em 9 bandas. Tais imagens são indicadas para diversas finalidades, que vão desde o mapeamento e monitoramento da vegetação e territórios até estudos relacionados a água costeira e aerossóis (Banda 1), além da detecção precisa de nuvens (Banda 9) (USGS, 2013).

Já as imagens provenientes do sensor *Thermal Infrared Sensor* (TIRS), apresentam resolução espacial de 100m, resolução radiométrica de 16bits e resolução espectral que vai do 10,6 μ m ao 12,5 μ m, distribuídas em 2 bandas. Tais imagens são indicadas principalmente para a modelagem da evapotranspiração e da temperatura superficial em grandes áreas (USGS, 2013).

Desta forma, a fim de avaliar a eficiência da utilização de imagens LS8 na predição volumétrica, além da utilização destas 11 bandas puras, foram gerados também outros índices de vegetação, o *Normalized Difference Vegetation Index* (NDVI) e todas as combinações de *Simple Ratio* (SR_{ij}) entre as bandas, conforme formulação matemática disponível em Ponzoni et al. (2012). Na Tabela 1 é apresentado a lista completa de preditores ou covariáveis, derivadas a partir das imagens LS8, geradas em janeiro de 2014, e utilizadas nas análises subsequentes.

Tabela 1 – Variáveis derivadas do satélite LS8, e utilizadas como predictoras no processo de seleção.
SR_{ij}: i/j com i e j = B1, B2, ..., B11

Nome	Sigla	Frequência (μm)	Fonte
Banda 1 – <i>Coastal aerosol</i>	B1	0,43 – 0,45	
Banda 2 – <i>Blue</i>	B2	0,45 – 0,51	
Banda 3 – <i>Green</i>	B3	0,53 – 0,59	
Banda 4 – <i>Red</i>	B4	0,64 – 0,67	
Banda 5 – <i>Near Infrared (NIR)</i>	B5	0,85 – 0,88	
Banda 6 – <i>SWIR 1</i>	B6	1,57 – 1,65	USGS (2013)
Banda 7 – <i>SWIR 2</i>	B7	2,11 – 2,29	
Banda 8 – <i>Panchromatic</i>	B8	0,50 – 0,68	
Banda 9 – <i>Cirrus</i>	B9	1,36 – 1,38	
Banda 10 – <i>Thermal Infrared 1</i>	B10	10,60 – 11,19	
Banda 11 – <i>Thermal Infrared 2</i>	B11	11,50 – 12,51	
<i>Normalized Difference Vegetation Index</i>	NDVI	-----	Ponzoni et al. (2012)
<i>Simple Ratio</i>	SR _{ij}	-----	

A Figura 8 demonstra uma composição colorida das imagens LS8 da área de interesse e utilizadas neste estudo.

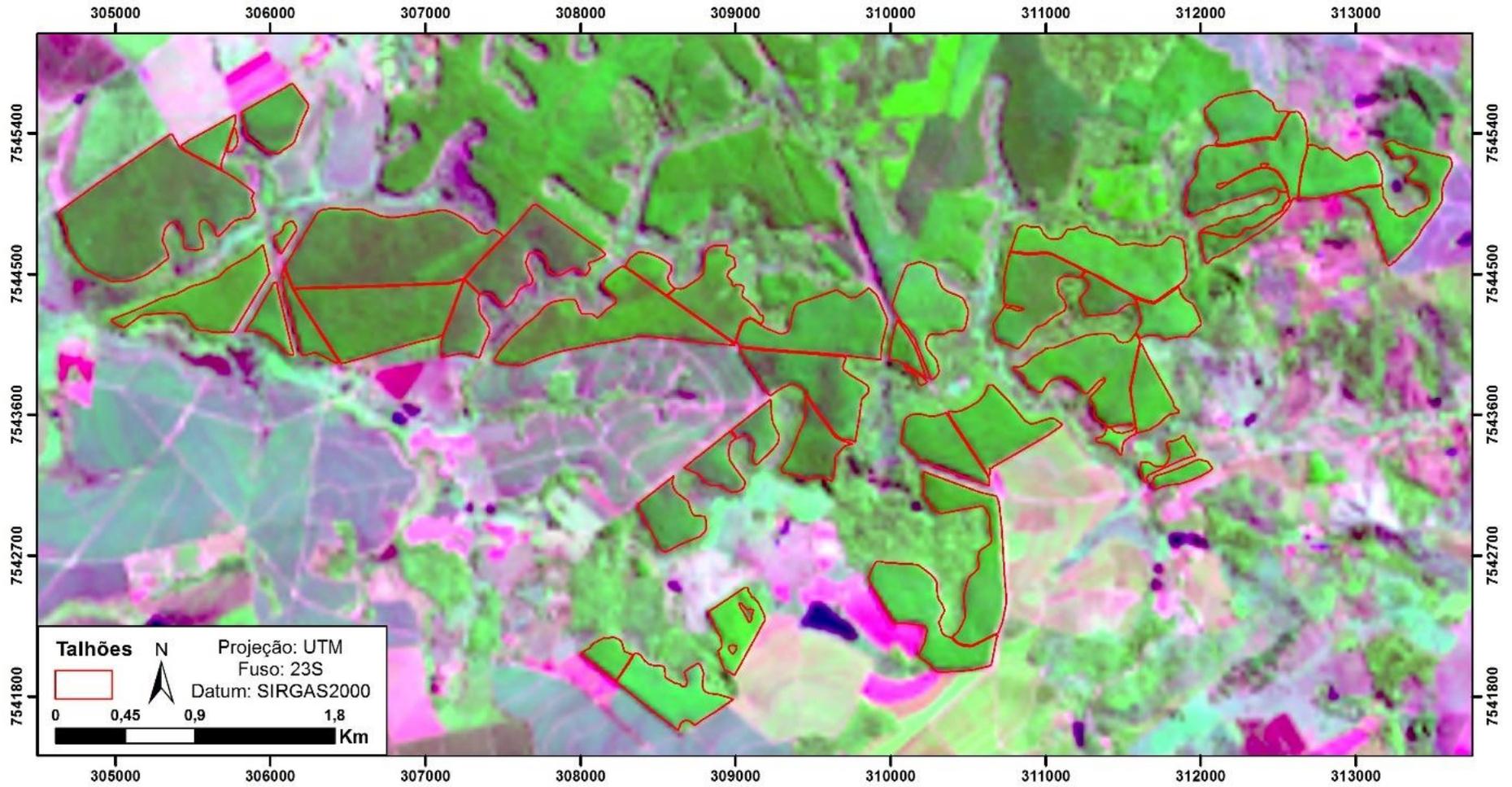


Figura 8 – Composição colorida (R7-G5-B3) das imagens do satélite LANDSAT-8, geradas em janeiro de 2014, Horto Paineiras, Espírito Santo do Pinhal, SP

3.2.2 Satélite RAPIDEYE

A empresa *Surrey Satellite Technology*, na Inglaterra, foi responsável pela fabricação e lançamento da constelação de cinco satélites idênticos, colocados em órbita heliossíncrona no ano de 2008. Denominados RAPIDEYE, os satélites são atualmente operados pela empresa alemã BlackBridge (BLACKBRIDGE, 2015). As imagens orbitais pós-processadas da constelação de satélites RAPIDEYE, apresentam uma resolução espacial de 5 metros, resolução radiométrica de 12bits e resolução espectral que abrange o Azul até o Infravermelho Próximo, distribuídas em cinco intervalos de onda no espectro de luz (RAPIDEYE, 2015). Características estas, que segundo Tyc et al. (2005), são ideais para o monitoramento agrícola e florestal, pois associam médias resoluções, espacial e espectral, à uma capacidade diária de coleta de imagens.

Os satélites foram os primeiros satélites comerciais a apresentarem a banda *RedEdge* (0,69 – 0,73 μ m), tal comprimento de onda é sensível a mudanças no conteúdo de clorofila dos vegetais. Sendo que estudos já demonstraram o sucesso da utilização desta banda no monitoramento da sanidade da vegetação, na classificação de diferentes espécies de vegetação e na mensuração dos níveis de proteína e nitrogênio na biomassa vegetal (EITEL et al., 2011; RECIO et al., 2011; EITEL et al., 2007).

Com objetivo de avaliar a eficiência da utilização de imagens RAPIDEYE na predição volumétrica, além a utilização das 5 bandas puras, foram gerados também outros índices de vegetação, o *Normalized Difference Vegetation Index* (NDVI), o *Normalized Difference Vegetation Index RedEdge*(NDVI_{RE}) e todas as combinações de *Simple Ratio*(SR_{ij}) entre as bandas, conforme formulação matemática disponível em Ponzoni et al. (2012) e Wu et al. (2009). Na Tabela 2 é apresentado a lista completa de preditores ou covariáveis, derivadas a partir das imagens RapidEye utilizadas nas análises subsequentes e geradas em março de 2014.

Tabela 2 – Variáveis derivadas do satélite RapidEye, e utilizadas como preditoras no processo de seleção. SR_{ij}: i/j com i e j = B1, B2, ..., B5

Nome	Sigla	Frequência (µm)	Fonte
Banda 1 – <i>Blue</i>	B1	0,44 – 0,51	BlackBridge (2015)
Banda 2 – <i>Green</i>	B2	0,52 – 0,59	
Banda 3 – <i>Red</i>	B3	0,63 – 0,69	
Banda 4 – <i>RedEdge</i>	B4	0,69 – 0,73	
Banda 5 – <i>NearInfrared (NIR)</i>	B5	0,76 – 0,85	
<i>Normalized Difference Vegetation Index</i>	NDVI	-----	Ponzoni et al. (2012)
<i>Simple Ratio</i>	SR _{ij}	-----	Wu et al. (2009)
<i>Normalized Difference Vegetation Index RedEdge</i>	NDVI _{RE}	-----	

Na Figura 9 pode-se observar uma composição colorida das imagens RAPIDEYE da área de interesse e utilizadas neste estudo.

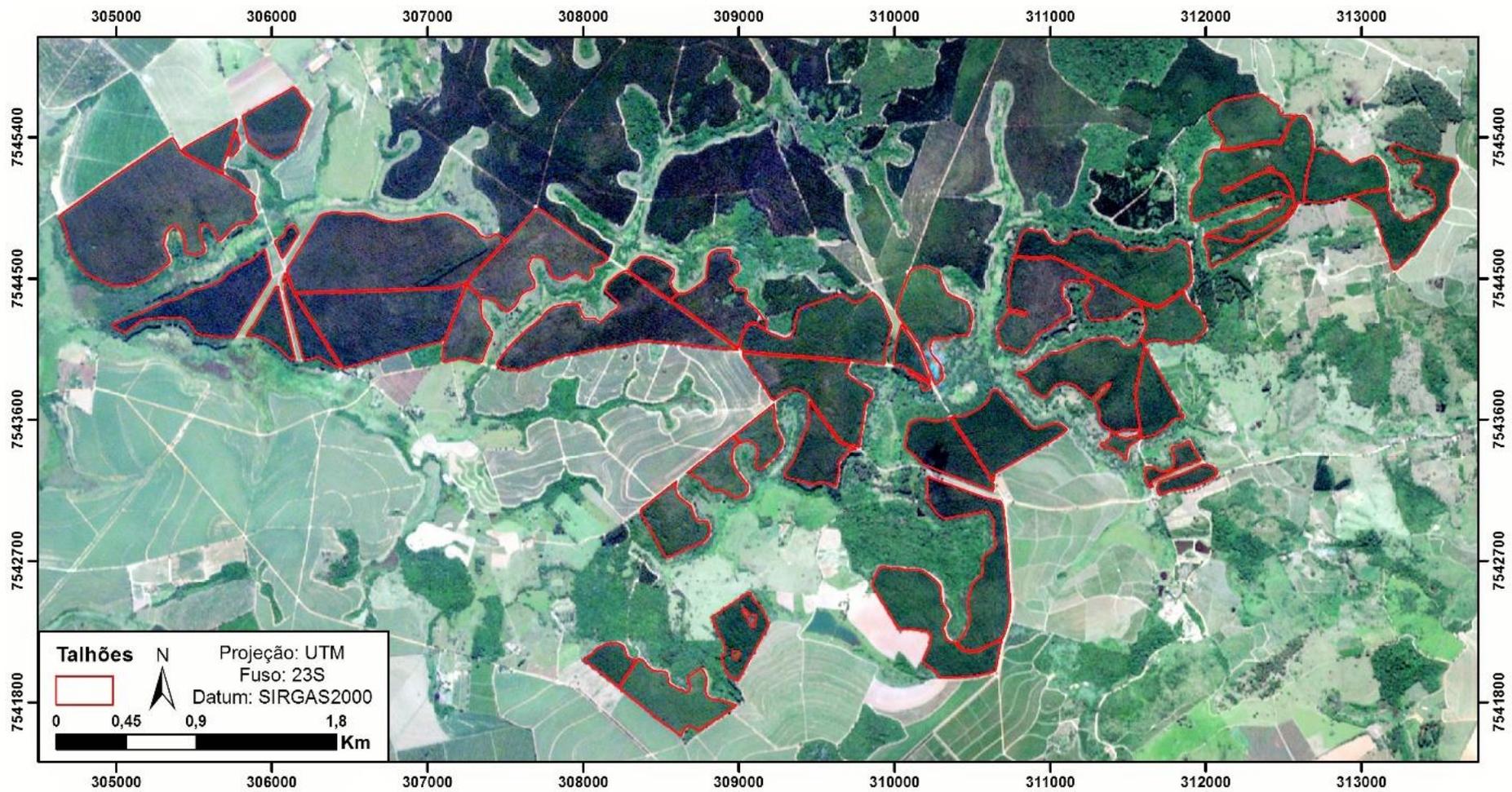


Figura 9 – Composição colorida (R3-G2-B1) das imagens do satélite RAPIDEYE, geradas em março de 2014, Horto Paineiras, Espírito Santo do Pinhal, SP

3.2.3 Satélite GEOEYE-1

O satélite GEOEYE-1 foi fabricado pela empresa *General Dynamics* nos Estados Unidos da América, e posto em órbita heliossíncrona em setembro de 2008 (GEOEYE, 2008). Atualmente o satélite é operado pela empresa *Digital Globe*, uma das principais empresas globais em imageamento por satélite e principal fornecedora de imagens para aplicações como *Google Earth* e *Google Maps*.

As imagens orbitais pós-processadas do satélite, apresentam uma resolução espacial de 0,5 metro, resolução radiométrica de 11bits e resolução espectral variando do Azul ao Infravermelho Próximo, distribuídas em cinco intervalos de onda no espectro de luz (GEOEYE, 2008). Tais características fizeram do satélite GEOEYE-1 o mais avançado satélite de observação terrestre na época em que foi lançado, estando ainda hoje entre os mais modernos satélites do mundo (DIGITALGLOBE, 2015).

Segundo Embrapa (2015), as imagens do satélite são utilizadas principalmente para: mapeamentos urbanos e rurais que exijam alta precisão dos dados (cadastramento, redes, planejamento, telecomunicações, saneamento, transportes), mapeamentos básicos e aplicações gerais em sistemas de informação geográfica; levantamento e monitoramento do uso e cobertura das terras (com ênfase em áreas urbanas), estudo de áreas verdes urbanas; agricultura de precisão, cadastros rurais, índices de vegetação e laudos periciais em questões ambientais.

As imagens GEOEYE-1, utilizadas nesse estudo, foram obtidas por meio da plataforma *Garmin Birds Eye Satellite Imagery*, que disponibiliza imagens de diferentes satélites, no comprimento de luz visível, para clientes de dispositivos *Garmin*. Dessa forma, apenas as bandas Azul, Verde e Vermelha foram utilizadas.

A eficiência da utilização das imagens GEOEYE-1, na predição volumétrica, foi avaliada utilizando 3 bandas (RGB), geradas pelo satélite, e todas as combinações de *Simple Ratio* (SR_{ij}) entre estas bandas. Devido a ausência de bandas no comprimento do infravermelho próximo, não foi possível calcular o índice de vegetação NDVI. Na Tabela 3 é apresentado a lista completa de preditores ou covariáveis, derivadas a partir das imagens aerotransportadas, obtidas em maio de 2013, e utilizadas nas análises subsequentes.

Tabela 3 – Variáveis derivadas do satélite GeoEye, e utilizadas como preditoras no processo de seleção. SR_{ij}: i/j com i e j = B1, B2, ..., B5

Nome	Sigla	Frequência (µm)	Fonte
Banda 1 – <i>Blue</i>	B1	0,45 – 0,51	GeoEye (2008)
Banda 2 – <i>Green</i>	B2	0,51 – 0,58	
Banda 3 – <i>Red</i>	B3	0,65 – 0,69	
<i>Simple Ratio</i>	SR _{ij}	-----	Ponzoni et al. (2012)

Na Figura10 é apresentado uma composição colorida das imagens GEOEYE-1 da área de interesse, utilizadas neste estudo.

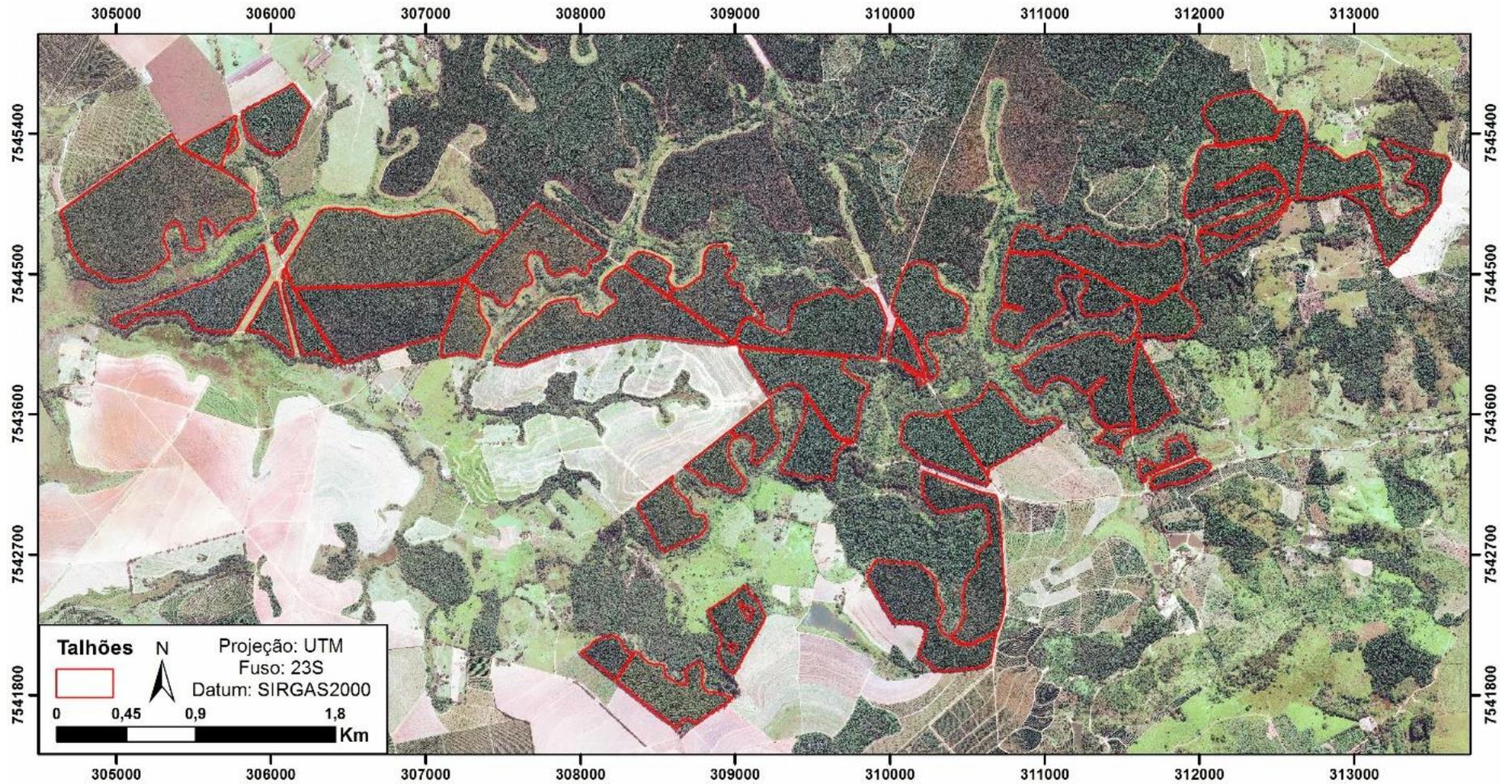


Figura 10 – Composição colorida (R1-G2-B3) das imagens do satélite GEOEYE-1, geradas em maio de 2013, Horto Paineiras, Espírito Santo do Pinhal, SP

3.2.4 Imagens Aerotransportadas

O Grupo WZ foi responsável pela aquisição das imagens aéreas da área de estudo, por meio de um veículo aéreo tripulado, modelo Cessna e munido de uma câmera Sony Alpha 7R de 36,4 megapixels, equipada com lentes de 35 mm, estando o equipamento integrado ao sistema de imageamento *Kitphoto*.

O plano de voo foi calculado para que as imagens apresentassem uma taxa de sobreposição de 60% nas linhas e 40% nas entrelinhas. As imagens foram posteriormente processadas no software *Agisoft PhotoScan*. Desta forma, o produto final do sobrevoo, foi uma imagem mosaicada da fazenda, com uma resolução espacial de 0,26m e resolução espectral no comprimento de onda da luz visível, indo do Azul ao Vermelho, distribuídas em três intervalos de onda no espectro de luz.

A eficiência da utilização das imagens aerotransportadas, para a predição volumétrica, foi avaliada utilizando apenas as 3 bandas puras geradas pelo equipamento e todas as combinações de *Simple Ratio* (SR_{ij}) entre estas bandas. Devido à ausência de bandas no comprimento do infravermelho próximo, não foi possível calcular o índice de vegetação NDVI. Na Tabela 4 é apresentado a lista completa de preditores ou covariáveis, derivadas das imagens aerotransportadas e obtidas em setembro de 2013.

Tabela 4 – Variáveis derivadas do satélite Aerotransportada, e utilizadas como preditoras no processo de seleção. SR_{ij} : i/j com i e $j = B1, B2, \dots, B5$

Nome	Sigla	Frequência (μm)	Fonte
Banda 1 – <i>Blue</i>	B1	-----	
Banda 2 – <i>Green</i>	B2	-----	-----
Banda 3 – <i>Red</i>	B3	-----	
<i>Simple Ratio</i>	SR_{ij}	-----	Ponzoni et al. (2012)

Na Figura 11 pode-se observar o mosaico final de imagens aerotransportadas da área de interesse utilizadas neste estudo.

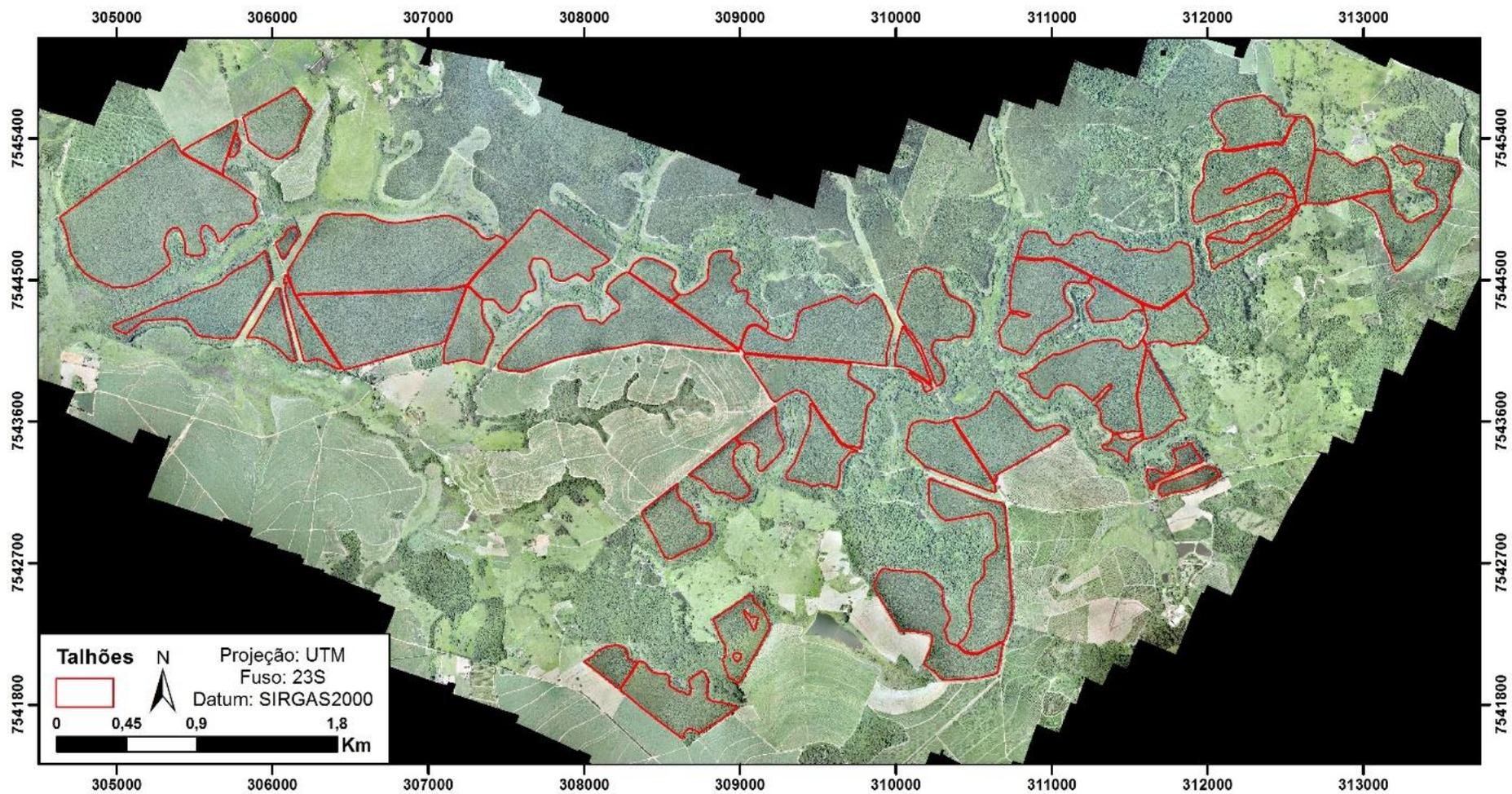


Figura 11 – Composição colorida (R1-G2-B3) das imagens aerotransportadas, obtidas em setembro de 2013, Horto Paineiras, Espírito Santo do Pinhal, SP

3.2.5 Processamento das Imagens – Padronização

Todas as imagens e índices de vegetação foram processadas em ambiente computacional R (R CORE TEAM, 2015), com o auxílio dos pacotes para processamento de imagens e manipulação de dados espacializados “*raster*”, “*sp*” e “*mapproj*” (PEBESMA, 2004; HIJMANS; ET TEN, 2012; PEBESMA; BIVAND, 2005; LEWIN-KOH; BIVAND, 2012). Nessa etapa, as imagens foram processadas, para que apresentassem um mesmo *datum* (SIRGAS 2000, 23S), sistema de coordenadas (UTM) e resolução espacial, sendo a resolução espacial padrão definida como 10 metros para todas as análises.

3.3 Caracterização da base de dados

Os dados de campo, utilizados nesse estudo, foram disponibilizados pela empresa proprietária da área e fazem parte da base operacional de parcelas de inventário contínuo da companhia. As parcelas foram mensuradas entre outubro de 2013 e março de 2014. O procedimento de amostragem utilizado foi o sistemático com dois estágios, com no mínimo 2 parcelas por talhão e com uma intensidade amostral média de 1 parcela para cada 6,4 hectares, totalizando 110 parcelas na área total.

A metodologia de medição consistiu-se na adoção de parcelas circulares de 400 m². Sendo que para cada uma delas, mediu-se a circunferência a 1,30 m do solo (CAP) de todas as árvores, a altura total das 15 primeiras árvores centrais, mais a altura das árvores dominantes, conforme o conceito de Assmann (1970). As alturas das demais árvores, não medidas nas parcelas, foram estimadas, por meio de modelos hipsométricos genéricos disponíveis em Bartoszeck et al. (2002).

Com o intuito de conhecer a distribuição espacial das parcelas (exigência geoestatística) e a sua conseqüente localização em relação às imagens orbitais, todas as parcelas tiveram seu centro georreferenciado com a utilização de GPS de navegação, modelo GARMIN 62SC. Como ponto central da parcela foi feita a média de 500 pontos coletados durante 8 minutos no centro de cada parcela. A estimativa volumétrica das árvores foi realizada utilizando-se o modelo volumétrico de Schumacher e Hall (1933), ajustado com base em cubagens da região. A estatística descritiva de algumas variáveis do inventário para as 110 parcelas amostradas é observada na Tabela 5.

Tabela 5 – Estatística descritiva das 110 parcelas utilizadas. Volume Total com Casca (VTCC), Altura total média (HT), Diâmetro a altura do peito (DAP) e Área Basal (AB)

	DAP (cm)	HT (m)	AB (m ²)	VTCC/ha
Mínimo	11,5	17,7	13,8	122,9
Média	16,1	25,1	23,2	247,6
Máximo	19,7	29,9	33,3	388,4
CV%	9,8	8,1	15,4	20,5

3.4 Técnicas Preditivas

Cada um dos conjuntos de dados, oriundos das imagens, foram tratados de forma independente, isto é, os preditores (imagens) oriundos do satélite LS8 não foram misturados àqueles provenientes dos satélites RAPIDEYE e GEOEYE-1, nem aos preditores derivados das imagens aerotransportadas.

Além da avaliação e utilização dos diversos preditores obtidos por meio das imagens (Bandas puras e Índices de Vegetação), avaliou-se também o efeito da utilização de preditores cadastrais, isto é, variáveis disponíveis na base cadastral das parcelas e talhões. Neste sentido foram disponibilizados também para o método de seleção de variáveis, descrito a seguir, a variável contínua Idade do Plantio (AGE) e a variável categórica Material Genético (MATGEN), sendo estas as únicas variáveis coincidentes em cada uma das análises.

O método de seleção de variáveis e predição, descrito a seguir, foi realizado uma vez para cada conjunto de imagens no ambiente computacional R 3.1.2. (R CORE TEAM, 2015).

3.4.1 Seleção de Variáveis e Regressão Linear Múltipla (RLM)

Inicialmente para reduzir os efeitos da colinearidade, comum entre preditores derivados de imagens orbitais ou aéreas, foram calculados os coeficientes de correlação linear de Pearson (ρ) entre todas variáveis, sendo eliminada da análise aquelas que apresentassem correlação (ρ) superiores a 0,9 (DORMAN et. al., 2013; SILVA, 2014). Para tanto, utilizou-se nesta etapa o pacote estatístico “*usdm*” no ambiente R (NAIMI, 2015; R CORE TEAM, 2015).

Em seguida, aplicou-se o método de seleção de variáveis *Best Subset Selection*, implementado computacionalmente no pacote “*leaps*” em ambiente R (LUMLEY; 2014; R CORE TEAM, 2015). Este método, executa uma busca exaustiva de modelos, e é conhecido na literatura pela sua robustez. Ao contrário de outros

métodos, este é capaz de testar todas as combinações possíveis de modelos, de forma computacionalmente eficiente para o conjunto de preditores fornecidos (JAMES et al., 2013). Com isto, pode-se eliminar variáveis redundantes e simplificar o modelo resultante.

Para a aplicação do método ajustam-se inicialmente todos os modelos com apenas uma variável preditora, escolhendo-se o melhor entre eles. Em seguida, ajustam-se todos os possíveis modelos com duas variáveis preditoras, escolhendo-se o melhor entre eles. Este processo ocorre até que o número de variáveis no modelo seja igual ao número de variáveis preditoras. Ao final do processo o número total de combinações avaliadas é próximo à 2^p , sendo p o número de variáveis preditoras (HASTIE et al., 2008; JAMES et al., 2013).

Nesta fase, obteve-se como *output*, o melhor modelo linear para cada tamanho de modelo possível ($n=0, n=1, \dots, n=p$). A etapa seguinte consistiu na escolha do melhor modelo dentre os modelos candidatos de diferentes tamanhos.

Para tanto, a escolha do melhor modelo baseou-se na raiz do erro quadrático médio (RMSE) obtida pela validação cruzada *leave-one-out* (LOOCV). Conforme James et al. (2013), a aplicação da validação cruzada permite um melhor entendimento do comportamento do modelo frente a uma base de dados não utilizada no ajuste do mesmo. O desenvolvimento algébrico para o ajuste de modelos lineares múltiplos pelo método dos mínimos quadrados possui amplo referencial teórico e pode ser consultado em Hastie et al. (2008) e James et al. (2013).

Além disso, conforme Kangas e Maltamo (2006), o valor de RMSE pode ser obtido aplicando-se as seguintes formulações (Formulas 13 e 14).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (13)$$

$$RMSE\% = 100 * \frac{RMSE}{\bar{y}} \quad (14)$$

Onde: y_i é o valor da variável y mensurada na posição i e \hat{y}_i é o valor da variável y estimada na posição i ; \bar{y} é a média dos valores observados e n é o número de observações.

3.4.2 Verificação de pressuposições

De acordo com Peña e Slate (2006), o ajuste de modelos lineares, simples ou múltiplos, depende de quatro pressuposições básicas: (1) existência de linearidade entre variáveis preditoras e variável resposta, (2) normalidade dos resíduos, (3) homocedasticidade (ou variância constante dos resíduos) e (4) independência entre resíduos. Sendo assim, após a detecção dos melhores modelos, por meio dos métodos *Best Subset Selection* e LOOCV, para cada tipo de imagem (LS8, RAPIDEYE, GEOEYE-1 e Aerotransportada), verificou-se o atendimento de tais pressuposições por meio de análises gráficas e utilizando-se o pacote “gvlma” no software R, o qual realiza um conjunto de testes estatísticos para avaliar atendimento ou não destas pressuposições (PEÑA; SLATE, 2014; R CORE TEAM, 2015).

3.4.3 Krigagem Ordinária e Krigagem com Deriva Externa

Conforme Yamamoto e Landim (2013), a krigagem ordinária (KO) é um método local de estimativa. Sendo assim, a estimativa em um ponto não amostrado resultará da combinação linear dos valores encontrados na vizinhança próxima, considerando-se o ajuste da função variograma. Já a krigagem com deriva externa (KDE) é uma técnica que permite fazer a estimativa de uma variável aleatória em um ponto não amostrado com base no padrão espacial descrito por uma variável secundária (YAMAMOTO, 2009).

O desenvolvimento algébrico, de ambos os métodos, foi descrito brevemente nos Itens 2.3.1 e 2.3.2 (Revisão Bibliográfica), porém, podem ser consultados com maior profundidade em diversas literaturas específicas da área, tais como Journel e Huijbregts (1978), Wackernagel (2003), Diggle e Ribeiro Jr. (2007), Hengl (2009), além da publicação nacional de Yamamoto e Landim (2013).

As interpolações geoestatísticas utilizando a KO e a KDE, foram realizadas seguindo as metodologias e procedimentos demonstrados por Hengl (2009), Bivand et al. (2011) e Rossiter (2015). Para tanto, utilizou-se unicamente o software R (R CORE TEAM, 2015), juntamente com os pacotes: geoestatístico “gstat”, e de manipulação de dados espaciais “raster”, “sp” e “maptools” (PEBESMA, 2004; HIJMANS; ETEN, 2012; PEBESMA; BIVAND, 2005; LEWIN-KOH; BIVAND, 2012).

Para a aplicação das técnicas, primeiramente verificou-se o grau de dependência espacial das amostras por intermédio do semivariograma (residual ou não), que avalia a variância entre pares de pontos separados por uma distância, e

permite identificar a presença de correlação entre as unidades amostradas (YAMAMOTO; LANDIM, 2013).

Em seguida ajustou-se os semivariogramas por meio do método dos mínimos quadrados ordinários, utilizando os seguintes modelos geoestatísticos: Esférico (Sph), Exponencial (Exp), Linear (Lin), Circular (Cir), Pentaesférico (Pen), Gaussiano (Gau), Bessel (Bes), Logaritimico (Log), Potência (Pow) e Periódico (Per).

Na Tabela 6 é possível verificar a formulação matemática destes modelos, enquanto na Figura 12 observa-se o padrão de forma que cada função pode assumir. Todos esses modelos estão implementados no pacote “gstat” no ambiente R (PEBESMA, 2004; R CORE TEAM, 2015).

Tabela 6 – Modelos geoestatísticos implementados no pacote “gstat”. Fonte: Pebesma (2014)

Modelo	Sintaxe	Formulação $\hat{\gamma}(h) =$	Alcance (h)
Esférico	Sph	$\frac{3h}{2a} - \frac{1}{2} \left(\frac{h}{a}\right)^3$	$0 \leq h \leq a$ $h > a$
Exponencial	Exp	$1 - \exp\left(\frac{-h}{a}\right)$	$h \geq 0$
Linear	Lin	h	$h \geq 0$
Circular	Cir	$\frac{2h}{\pi a} \sqrt{1 - \left(\frac{h}{a}\right)^2} + \frac{2}{\pi} \arcsin \frac{h}{a}$	$0 \leq h \leq a$ $h > a$
Pentaesférico	Pen	$\frac{15h}{8a} - \frac{5}{4} \left(\frac{h}{a}\right)^3 + \frac{3}{8} \left(\frac{h}{a}\right)^5$	$0 \leq h \leq a$ $h > a$
Gaussiano	Gau	$1 - \exp\left(-\left(\frac{h}{a}\right)^2\right)$	$h \geq 0$
Bessel	Bes	$1 - \frac{h}{a} K_1\left(\frac{h}{a}\right)$	$h \geq 0$
Logaritimico	Log	0 $\log(h + a)$	$h = 0$ $h > 0$
Potência	Pow	h^a	$h \geq 0$ $0 < a \leq 2$
Periódico	Per	$1 - \cos\left(\frac{2\pi h}{a}\right)$	$h \geq 0$

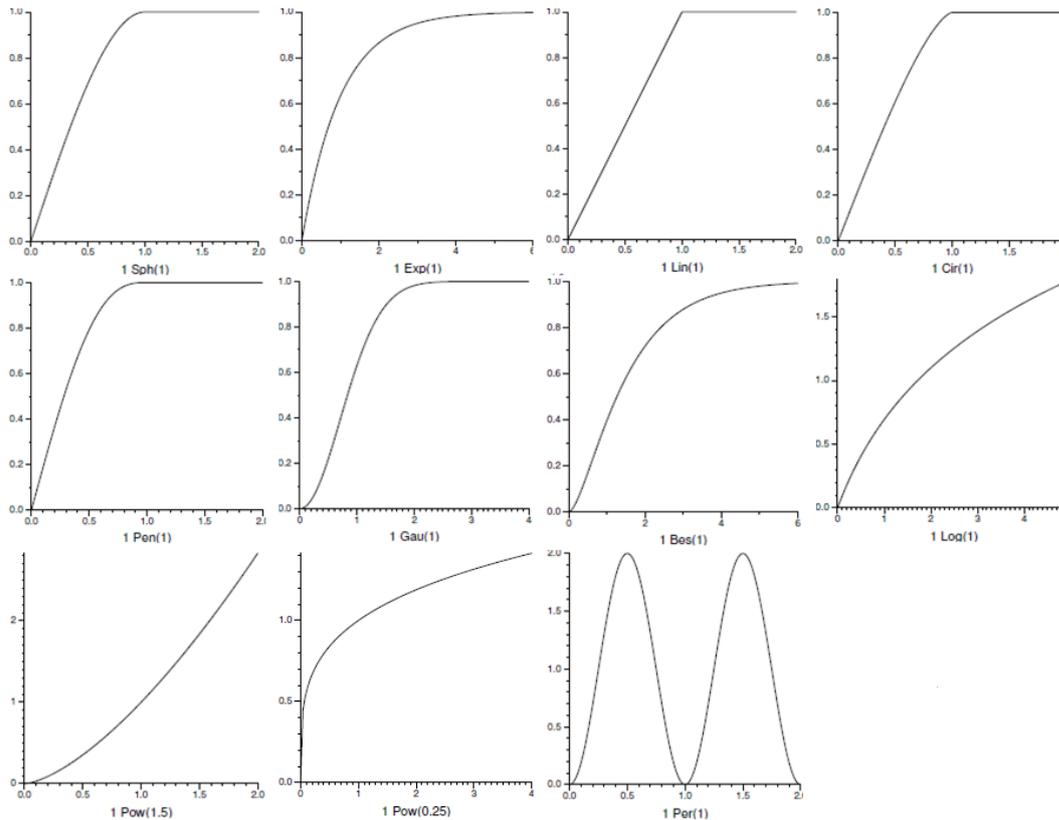


Figura 12 – Forma padrão das funções geoestatísticas implementadas no pacote “*gstat*”. Fonte: Pebesma (2014)

A escolha dos melhores modelos geoestatísticos de KO e KDE foi realizada com base na Raíz do Erro Quadrático Médio (RMSE) obtida pela validação cruzada *leave-one-out* (LOOCV). Para os ajustes de KDE, em cada grupo de preditores (LS8, RAPIDEYE, GEOEYE-1 e Aerotransportado), utilizou-se como variáveis auxiliares as mesmas variáveis selecionadas, no procedimento de seleção de variáveis realizado nas estimativas de Regressão Linear Múltipla (RLM).

Um fluxograma explicativo de todas as etapas de processamento pode ser consultado no Item 3.6..

3.4.4 Estimativa da Média e Construção de Intervalos de Confiança

Com o objetivo de quantificar os benefícios, ou não, da utilização das técnicas multivariadas para a predição volumétrica frente a abordagem tradicional do inventário florestal, foram estimados a média populacional e os intervalos de confiança da média para variável volume total com casca (VTCC).

Primeiramente calculou-se tais estimativas utilizando o estimador tradicional da amostragem aleatória simples (AAS), cuja formulação pode ser consultada no Item 2.2.1 e em Batista et al. (2014).

Em seguida as mesmas estimativas de média populacional e intervalo de confiança foram calculadas empregando-se o estimador de regressão (ER) (amostragem com variável auxiliar), cuja formulação pode ser consultada no Item 2.2.2 e em Batista et al. (2014). Nesta etapa utilizou-se como variável auxiliar o mapa de estimativas de VTCC produzido pela combinação da melhor técnica multivariada (RLM ou KDE) e pelo melhor conjunto de variáveis auxiliares (LS8, RAPIDEYE, GEOEYE-1 ou Aerotransportadas), identificado por meio dos valores de RMSE% gerados durante a validação cruzada LOOCV de todos os métodos.

A aplicação do estimador de regressão possui a característica de promover mecanismos para a correção de possíveis erros sistemáticos na estimativa da medida auxiliar, além de fornecer um conjunto de formulações válidas para o cálculo dos intervalos de confiança da média (MCROBERT et al., 2013)

3.4.5 Estimativa de redução da intensidade amostral

Após a quantificação do Erro Amostral em porcentagem (E%) de cada um dos métodos (AAS e ER), pode-se identificar ou não a possível redução dos intervalos de confiança gerados a partir das técnicas multivariadas. Uma vez identificado este ganho, calculou-se o número de parcelas que deveriam ser instaladas no procedimento de AAS para que fosse gerado o mesmo E% observado no ER.

Conforme Batista et al. (2014), na AAS o número de parcelas necessárias para que se obtenha um valor de erro amostral previamente fixado pode ser gerado aplicando-se a seguinte formulação (Formula 15):

$$n = \frac{N(t_{\alpha;n-1}.CV\%)^2}{N.E\%+(t_{\alpha;n-1}.CV\%)^2} \quad (15)$$

Onde: n é o número de amostras necessárias para que E% seja contemplado; E% refere-se ao erro amostral almejado; N é o tamanho populacional; CV% é o coeficiente de variação da população e t é o valor da distribuição t de *Student*.

3.5 Fluxograma Geral do Método

A Figura 14 demonstra na forma de um fluxograma todas as etapas descritas acima e desenvolvidas durante as atividades do projeto. Além disso, um estudo dirigido exemplificando de forma simplificada a aplicação da metodologia no ambiente R, pode ser consultada no item Apêndice.

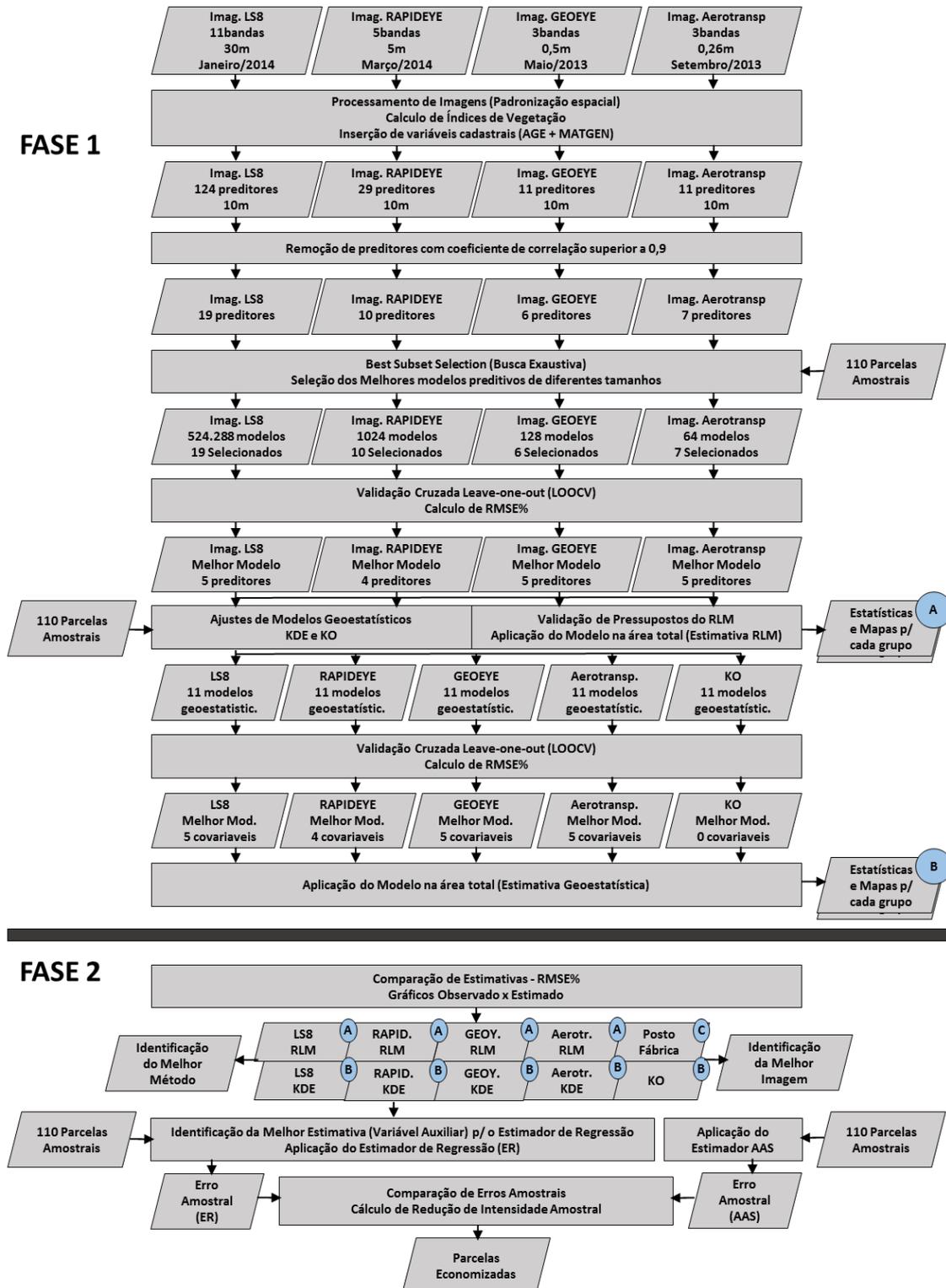


Figura 13 – Fluxograma explicativo das etapas de processamento realizadas.

4 RESULTADOS E DISCUSSÃO

A seguir apresentam-se os principais resultados e contribuições do presente trabalho no âmbito da mensuração florestal. Inicialmente, são abordados os resultados do pré-processamento das imagens de satélite e das imagens aéreas, referentes às etapas de geração de covariáveis e eliminação de variáveis redundantes. Em seguida, são apresentados os resultados da modelagem de RLM, da KO e da KDE, e é feita uma comparação visual e numérica das estimativas. E ao final, são expostos os resultados de uma comparação da AAS e do estimador de regressão (ER), com o volume posto-fábrica.

4.1 Padronização das imagens e geração de covariáveis.

Após o tratamento inicial das imagens, visando a padronização de resolução espacial, sistema de coordenadas e *datum*, as variáveis cadastrais (AGE e MATGEN) foram inseridas ao conjunto de variáveis também de forma espacializada. Em seguida foram geradas todas as combinações possíveis do índice *Simple Ratio* e criados os índices de vegetação NDVI (LS8 e RAPIDEYE) e NDVI_{RE} (RAPIDEYE). A Tabela 7 demonstra um resumo dos resultados obtidos durante as etapas iniciais de pré-processamento e padronização dos dados, sendo que as imagens LANDSAT8, RAPIDEYE, GEOEYE e aerotransportadas tiveram, respectivamente, 124, 29, 11 e 11 preditores gerados.

Em seguida, cada conjunto de preditores foi submetido a uma matriz de correlação, buscando-se a remoção de preditores com elevada colinearidade, isto é, coeficiente de correção de Pearson maior que 0.9 ou menor que -0.9. Conforme James et al. (2013), a presença de colinearidade entre variáveis predictoras pode causar problemas no contexto da regressão linear, pois torna-se difícil a separação dos efeitos de cada preditor sobre a variável resposta. A colinearidade reduz ainda a acurácia das estimativas dos coeficientes da regressão e aumenta o erro padrão dos parâmetros.

Ao final do processo de remoção de variáveis com elevada colinearidade, restaram-se: 19 variáveis derivadas do satélite LS8, 10 oriundas do satélite RAPIDEYE, 6 do satélite GEOEYE e 7 das imagens Aerotransportadas. Essas variáveis foram, em seguida, submetidas ao processo de busca exaustiva de modelos, denominado *Best Subset Selection*. A matriz de correção dos preditores selecionados, após a remoção daqueles com elevada colinearidade, pode ser observada no Item Anexo para cada grupo de imagens, nas Figuras 34, 35, 36 e 37.

Tabela 7 – Resumo das etapas iniciais do processamento das imagens

Imagem	Dados de Entrada		Geração de Índices		Remoção de colinearidade	
	Número de Bandas	Resolução espacial (m)	Número de Bandas	Resolução espacial (m)	Número de Bandas	Resolução espacial (m)
LANDSAT-8	11	15/30/100	124	10	19	10
RAPIDEYE	5	5	29	10	10	10
GEOEYE	3	0,5	11	10	6	10
Aerotransportada	3	0,26	11	10	7	10

4.2 Análise exploratória da variável resposta

Antes da continuidade das análises, ou seja antes da aplicação do algoritmo *Best Subset Selection*, foi realizada uma análise exploratória da variável resposta Volume Total Com Casca (VTCC), expresso em m³/ha.

De acordo com Diggle e Ribeiro Jr (2007), a análise exploratória dos dados é parte integral da estatística moderna e deve também ser realizada, quando o objetivo é uma investigação geoestatística. Yamamoto e Landim (2013), demonstram que a presença de observações discrepantes (*outliers globais*), pode influenciar negativamente todos os pontos na etapa de construção dos semivariogramas. Neste contexto, a Figura 15A demonstra por meio de uma análise *boxplot*, que as 110 observações de VTCC não apresentaram dados discrepantes que demonstrassem a presença de *outliers globais*.

A aplicação da krigagem exige que a variável de interesse apresente normalidade, fazendo-se necessários mecanismos de transformação e normalização, antes mesmo, da aplicação do método (YAMAMOTO; LANDIM, 2013). As Figuras 15B e 15C, demonstram por meio de um histograma e um gráfico quantil-quantil, que a variável VTCC apresentou distribuição normal, descartando-se, assim, a necessidade de transformação.

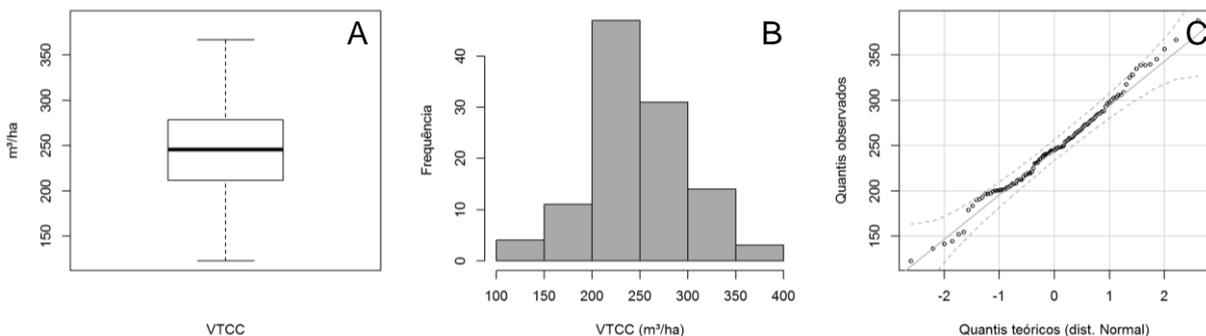


Figura 14 – Análise exploratória em VTCC. A-Gráfico boxplot; B-Histograma; C-Gráfico Quantil-Quantil

4.3 Regressão Linear Múltipla

4.3.1 Seleção de variáveis, ajuste e desempenho dos Modelos de RLM

Foram disponibilizados para o algoritmo de *Best Subset Selection* as variáveis preditoras de cada grupo de imagem (Item 4.1), juntamente com a variável resposta (Item 4.2), sendo limitado em 5, o número máximo de preditores no modelo resultante.

Todas as combinações possíveis de modelos foram analisadas, sendo escolhido aquela que apresentou menor RMSE%, obtido pelo método de validação cruzada LOOCV.

Na Figura 16, é apresentado graficamente, os valores mínimos de RMSE (m^3/ha) obtidos em função do número de preditores, para cada grupo de variáveis preditoras. Conforme demonstram as Figuras 16A e 16B, observa-se que os preditores de média resolução espacial e de maior resolução espectral (LS8 e RAPIDEYE), apresentaram um melhor desempenho em relação àqueles de alta resolução espacial e baixa resolução espectral (GEOEYE e Aerotransportado) (Figuras 16C e 16D).

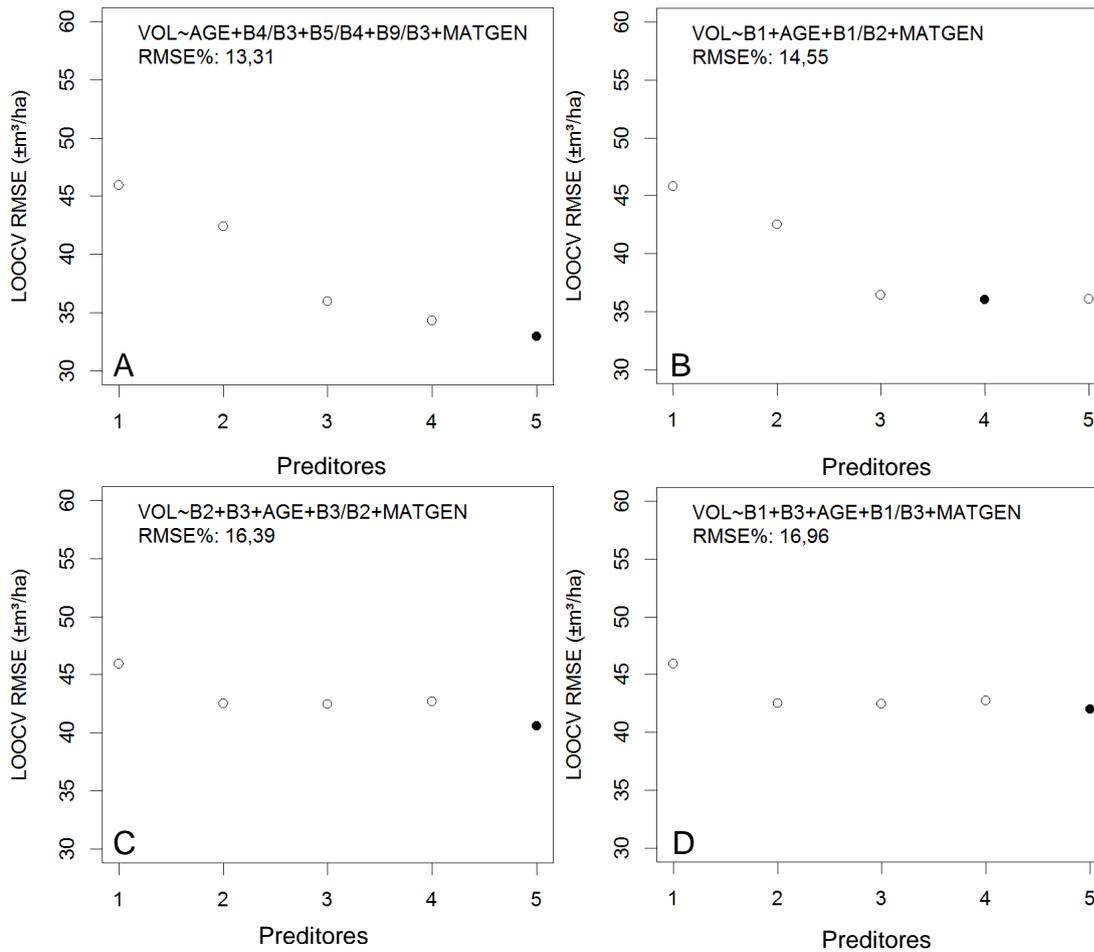


Figura 15 – Valor de LOOCV RMSE (m³/ha) em função do número de variáveis preditoras. A: imagens LS8; B: imagens RAPIDEYE; C: imagens GEOEYE e D: imagens Aerotransportadas

A Tabela 8 demonstra um resumo dos resultados obtidos nesta etapa. As estimativas de todos os parâmetros, juntamente com as estatísticas de erro padrão dos parâmetros, podem ser consultadas nos Anexo, nas Tabelas 12, 13, 14 e 15.

Tabela 8 – Resumo das estatísticas de ajuste RLM. *Valores de R² Ajustado e RMSE e RMSE% obtidos a partir da validação cruzada LOOCV

Conjunto de Dados	Modelo	*R² Ajust.	*RMSE (m³/ha)	*RMSE%
LANDSAT8	$VTCC \sim f\left(\frac{B4}{B3}, \frac{B5}{B4}, \frac{B9}{B3}, AGE, MATGEN\right)$	0,58	32,96	13,31
RAPIDEYE	$VTCC \sim f\left(B1, \frac{B1}{B2}, AGE, MATGEN\right)$	0,50	36,03	14,55
GEOEYE	$VTCC \sim f\left(B2, B3, \frac{B3}{B2}, AGE, MATGEN\right)$	0,37	40,59	16,39
Aerotransp.	$VTCC \sim f\left(B1, B3, \frac{B1}{B3}, AGE, MATGEN\right)$	0,32	41,99	16,96

Canavesi et al. (2010) utilizaram variáveis de imagens hiperespectrais do sensor Hyperion EO-1 (30m e 242 bandas + 9 índices de vegetação) para a predição volumétrica em plantios eucalipto na região de Capão Bonito (SP) utilizando RLM. Entretanto, os autores não obtiveram resultados satisfatórios sendo encontrados erros de subestimativa na ordem de 209 m³/ha ao se comparar com o volume das parcelas.

Em estudo semelhante, Berra et al. (2012) utilizando preditores derivados do satélite LANDSAT5 (30m, 4 bandas + 2 índices de vegetação), estimaram o volume de madeira em formações florestais de *Eucalyptus* spp. na região leste do Estado do Rio Grande do Sul (RS). O melhor modelo, identificado pelos autores, para a predição do volume de madeira, utilizou como variáveis preditoras a Banda 2 (verde) do satélite LS5 juntamente com uma combinação da *Simple Ratio* (SR) entre as Bandas 3 (vermelho) e 4 (infravermelho próximo). Como resultado, os autores obtiveram o volume de madeira com erro de 106 m³/ha ou RMSE de 22,8%.

Shinzato (2015), avaliando o uso da integração de imagens aerotransportadas de alta resolução espacial com preditores oriundos de escaneamentos a laser, para predição volumétrica em *Eucalyptus* spp. observou que as estimativas utilizando apenas as imagens aerotransportadas, apresentaram os piores resultados. A autora obteve valor de R² Ajustado de 0,48 e RMSE de 8,45%.

Na Figura 17 pode-se observar a análise gráfica dos resíduos para cada um dos ajustes de RLM. Conforme a imagem, todos os quatro modelos avaliados (Tabela 8), respeitaram as suposições de ajuste para modelos lineares descritas em James et al. (2013), como homocedasticidade e normalidade de resíduos. Sendo assim, conforme a análise gráfica dos ajustes (Figura 17) e as estatísticas de bondade de ajuste demonstradas na Tabela 8, evidencia-se a superioridade dos preditores oriundos do satélite LS8 para a estimativa volumétrica na formação florestal de *Eucalyptus* spp., objeto desse estudo.

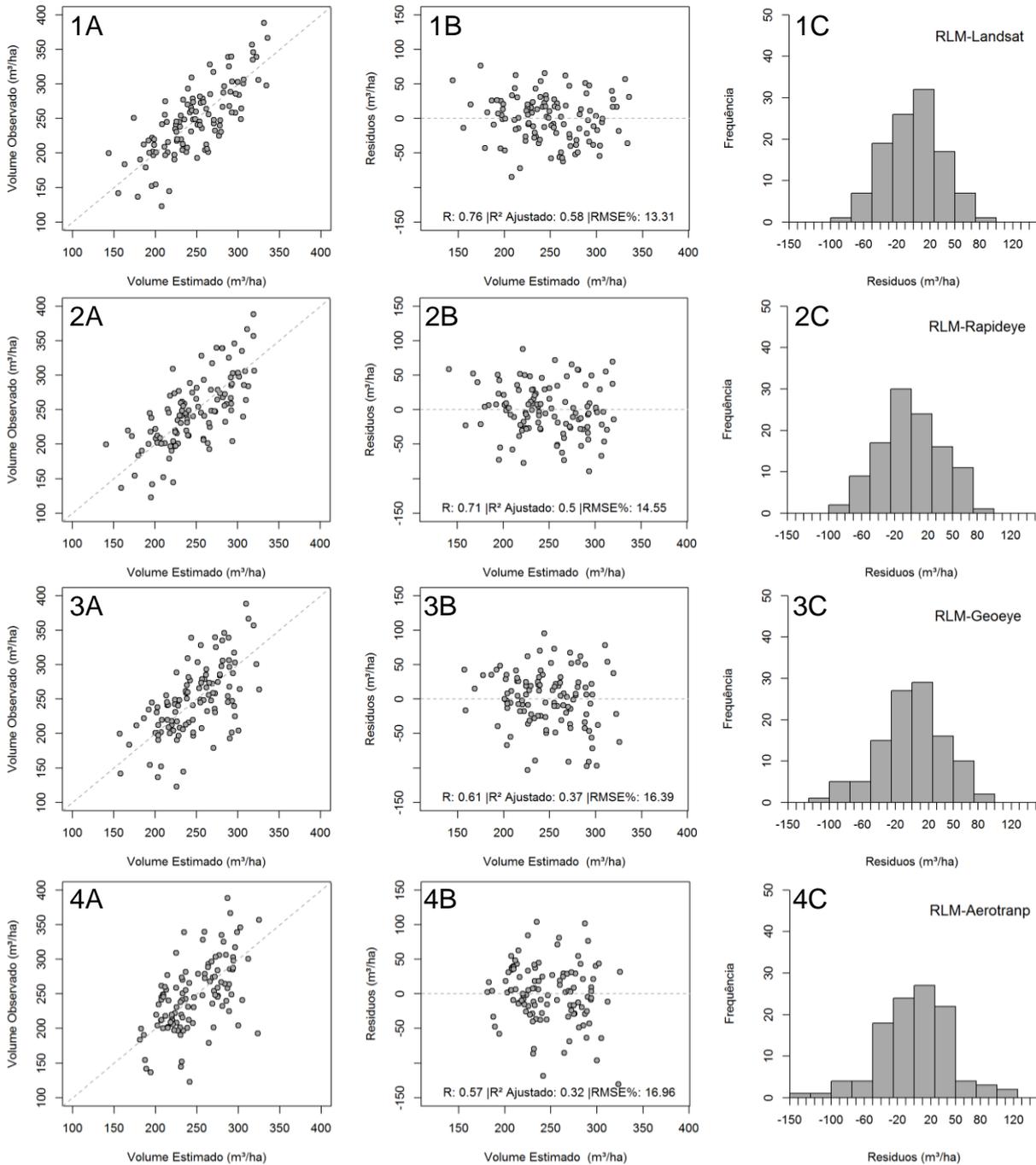


Figura 16 – Análise gráfica dos ajustes. 1-Preditores LANDSAT8; 2-Preditores RAPIDEYE; 3-Preditores GEOEYE; 4-Preditores Aerotransportado. A-Gráfico Observado x Estimado; B- Dispersão dos resíduos e C-Histograma dos resíduos

4.3.2 Estimativas volumétricas

Uma vez ajustado o melhor modelo de RLM para cada conjunto de preditores, e em posse das variáveis predictoras em área total, pôde-se aplicar o modelo e gerar a estimativa volumétrica para toda a área de interesse. Esse procedimento permite a confecção de um mapa de distribuição da variável de interesse em área total, além da extração de estatísticas de interesse a nível de talhão e/ou em área total.

Nesse sentido, as estimativas médias de VTCC/ha foram muito próximas entre as predições de LS8 (245,98m³/ha) e RAPIDEYE (245,61m³/ha). Além disso, o conjunto de imagens Aerotransportadas produziu a maior estimativa (246,35m³/ha), enquanto as imagens GEOEYE produziram a menor estimativa (240,13m³/ha).

As Figuras 18, 19, 20 e 21 demonstram os resultados da aplicação dos modelos de RLM para os conjuntos de dados LS8, RAPIDEYE, GEOEYE e Aerotransportados, respectivamente.

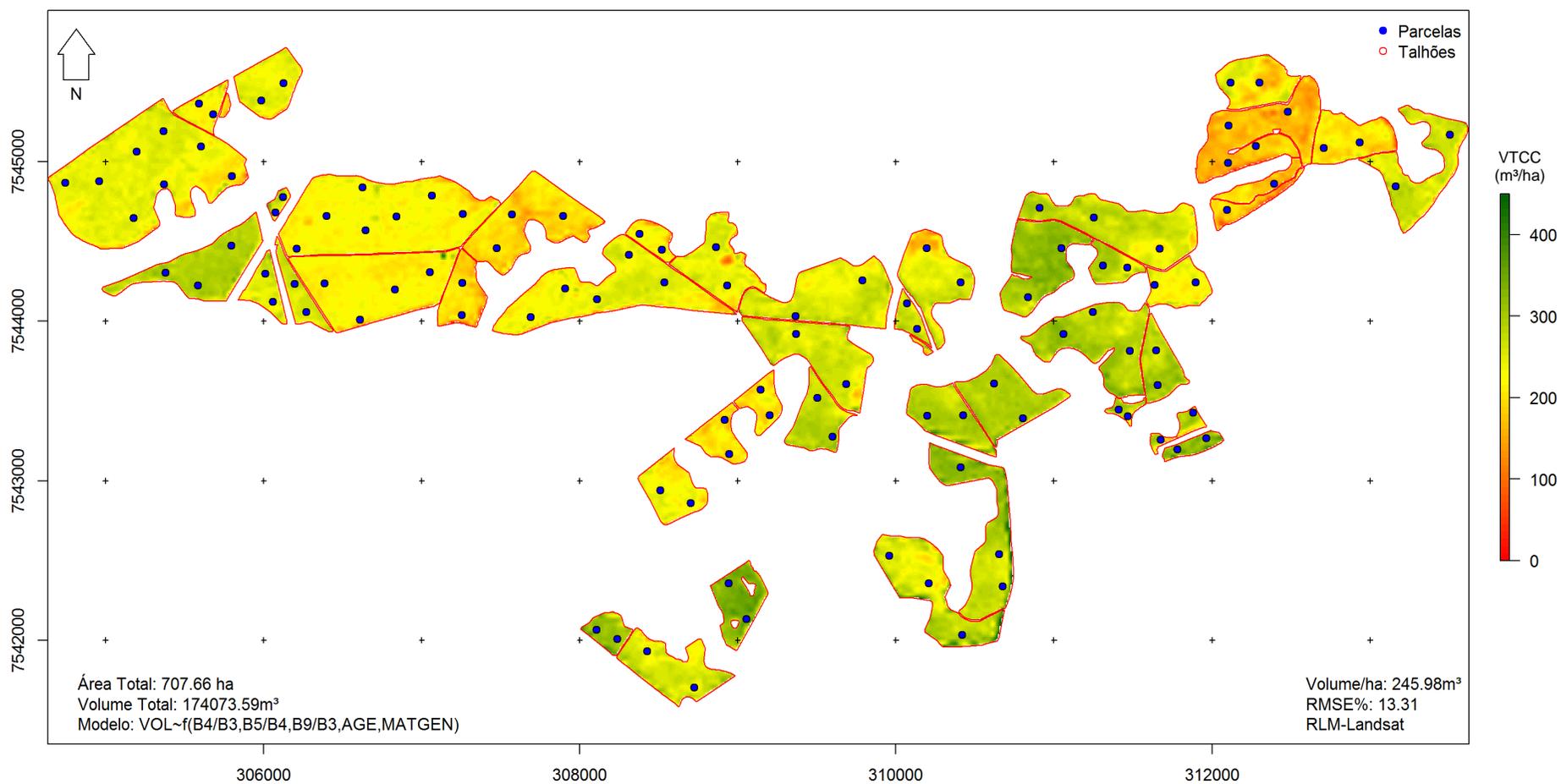


Figura 17 – Mapa de distribuição de VTCC (m³/ha) para toda a fazenda – RLM utilizando preditores derivados das imagens LS8

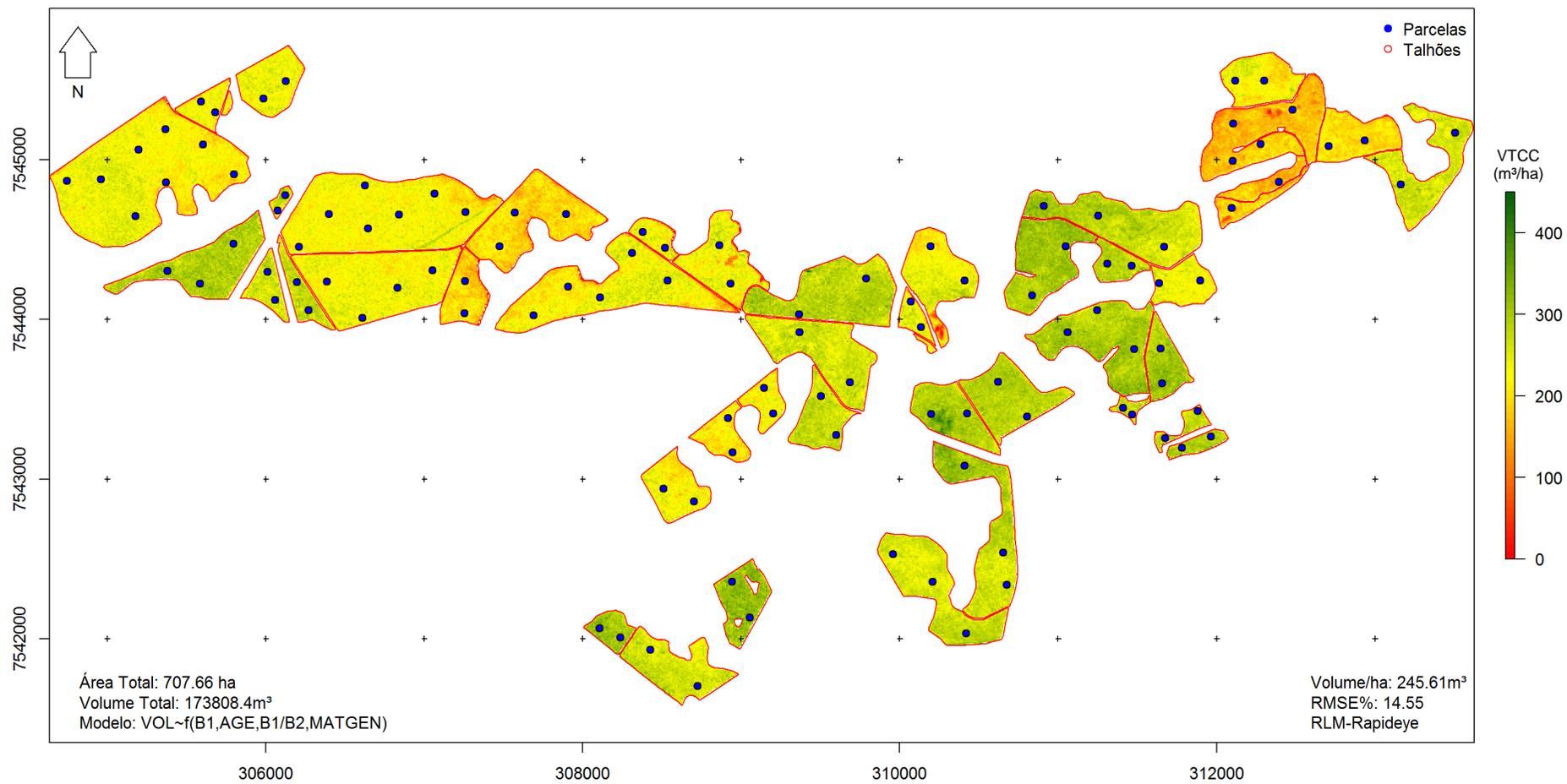


Figura 18 – Mapa de distribuição de VTCC (m³/ha) para toda a fazenda – RLM utilizando preditores derivados das imagens RAPIDEYE

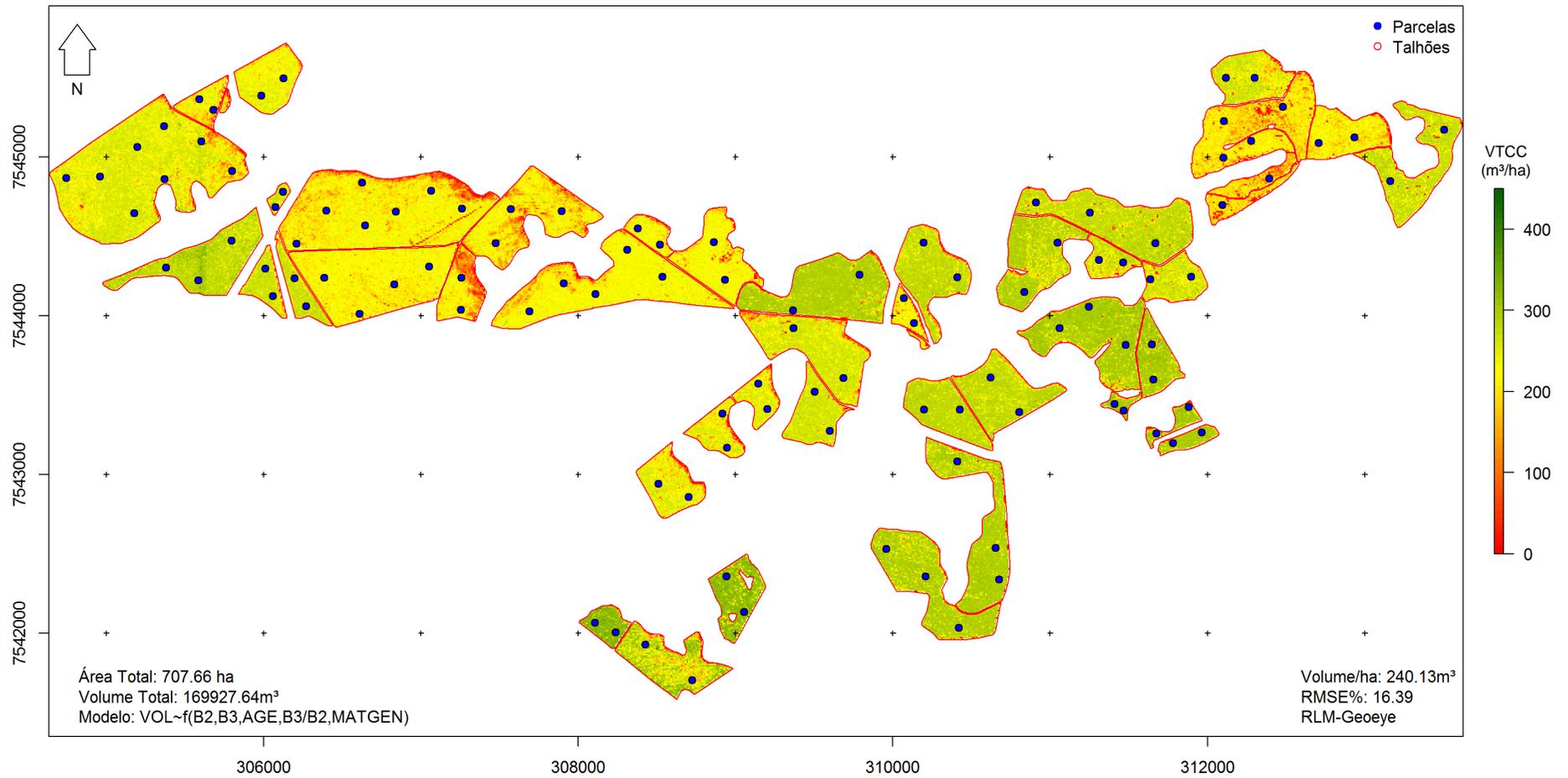


Figura 19 – Mapa de distribuição de VTCC (m³/ha) para toda a fazenda – RLM utilizando preditores derivados das imagens GEOEYE

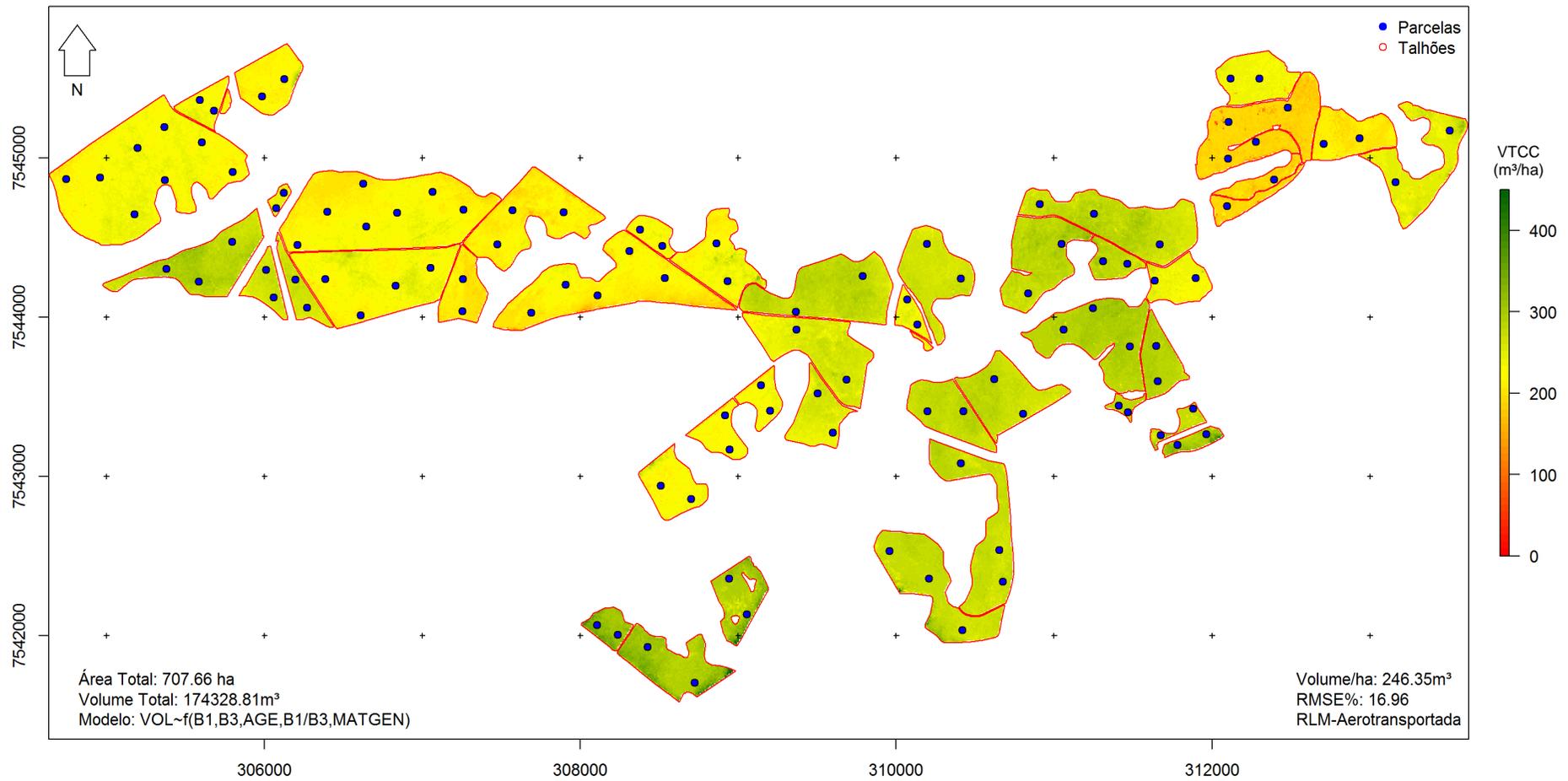


Figura 20 – Mapa de distribuição de VTCC (m³/ha) para toda a fazenda – RLM utilizando preditores derivados das imagens Aerotransportadas

4.4 Estimativas Geoestatísticas

A seguir, são apresentados os principais resultados das estimativas geoestatísticas de KO e KDE.

4.4.1 Krigagem Ordinária (KO)

4.4.1.1 Ajuste e desempenho da KO

Conforme Cressie (1993), para as estimativas de KO, utiliza-se apenas as próprias amostras espacializadas, descartando-se a necessidade de variáveis auxiliares. Desta forma, obteve-se como resultados iniciais a construção do semivariograma experimental demonstrado na Figura 22.

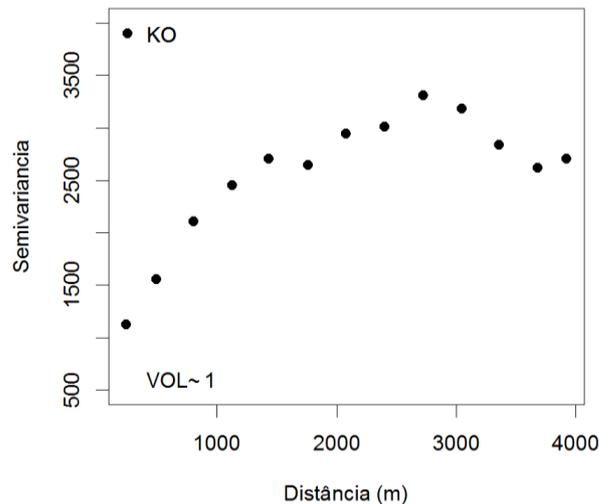


Figura 21 –Semivariograma experimental para Krigagem Ordinária da variável VTCC

Após a avaliação de todos os modelos geoestatísticos candidatos, descritos no Item 3.4.3., foi escolhido àquele que apresentou o menor RMSE% obtido pela validação cruzada LOOCV. O modelo selecionado nessa etapa foi modelo Exponencial com os parâmetros Pepita (τ^2) = 100,7, Patamar (σ^2) = 2896,5 e Alcance (ϕ) = 664,5. O ajuste do modelo ao semivariograma pode ser observado na Figura 23, enquanto as estatísticas envolvendo a seleção entre os modelos candidatos podem ser encontradas em Anexo na Tabela 16.

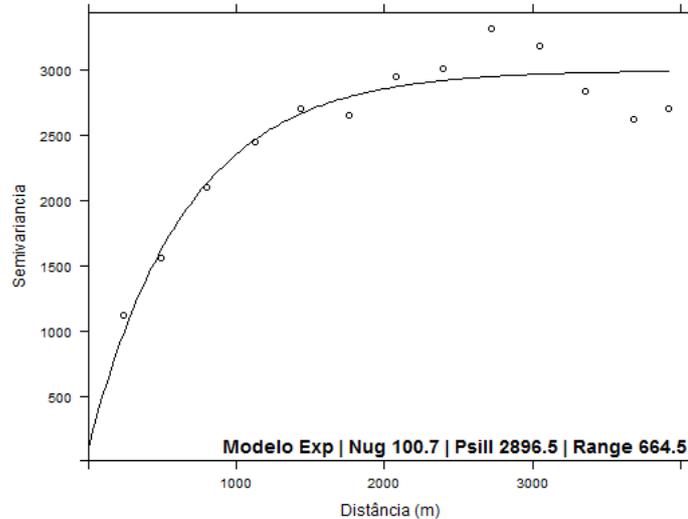


Figura 22 – Ajuste do Semivariograma experimental para Krigagem Ordinária da variável VTCC

A Figura 24 demonstra os resultados obtidos pela validação cruzada LOOCV do modelo selecionado. Assim como observado por Souza (2007), as estimativas da KO apresentaram leve tendência de suavização das predições (Figura 24A), ou seja, ocorreu a subestimativa de valores altos e a superestimava de valores baixos. Entretanto, ainda assim, os resíduos apresentaram homocedasticidade e normalidade (Figuras 24B e 24C).

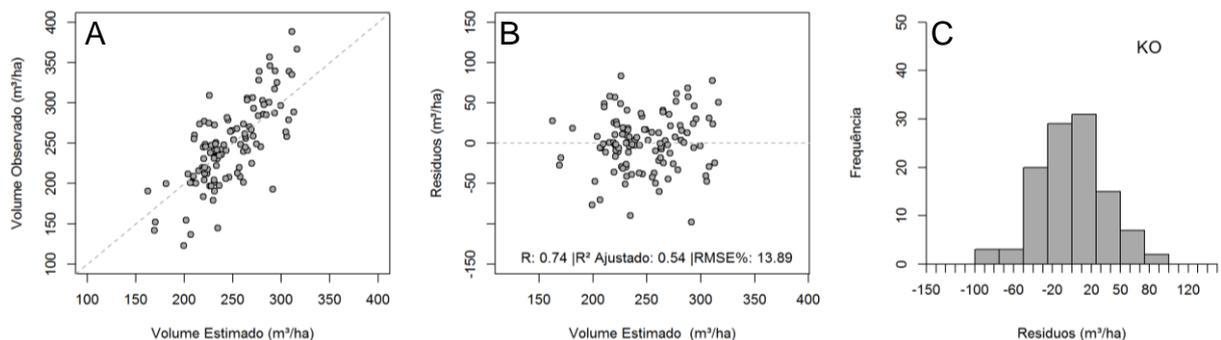


Figura 23 – Análise gráfica do ajuste de KO. A-Gráfico Observado x Estimado; B-Dispersão dos resíduos e C-Histograma dos resíduos

4.4.1.2 Estimativa volumétrica

Na Figura 25 observa-se o mapa de distribuição da variável de interesse em área total, produzido por meio da aplicação da KO. Uma vez que, para esta estimativa, utiliza-se apenas as unidades amostrais conforme o ajuste no semivariograma, o mapa resultante não apresenta capacidade de detectar clareiras provocadas por mortalidade ou furto, ou qualquer outro evento de baixa ou alta produtividade que não

tenha sido captado pelas unidades amostrais. Ainda assim, diversos autores utilizam a KO devido a sua facilidade de aplicação e capacidade de espacialização do volume.

Guedes et al. (2012), utilizaram a KO para definir estratos mais homogêneos em plantações de *Eucalyptus* spp.. Isso permitiu a utilização de estimadores da amostragem casual estratificada (ACE), reduzindo os intervalos de confiança das estimativas. Já Ribeiro (2014), utilizou a KO para produzir mapas de classe de sítio em plantios de *Pinus taeda* L. no estado do Paraná. Com os mapas resultantes o autor pode propor um retalhamento (mais homogêneo) dos plantios, facilitando o planejamento estratégico.

No presente estudo a estimativa do VTCC da KO foi de 245,61m³/ha. Além disso, o valor de RMSE% gerado pela validação cruzada LOOCV foi de 13,89%, tal valor é inferior a incerteza associada as predições da RLM com as variáveis do RAPIDEYE (14,55%), GEOEYE (16,39%) e Aerotransportada (16,96%).

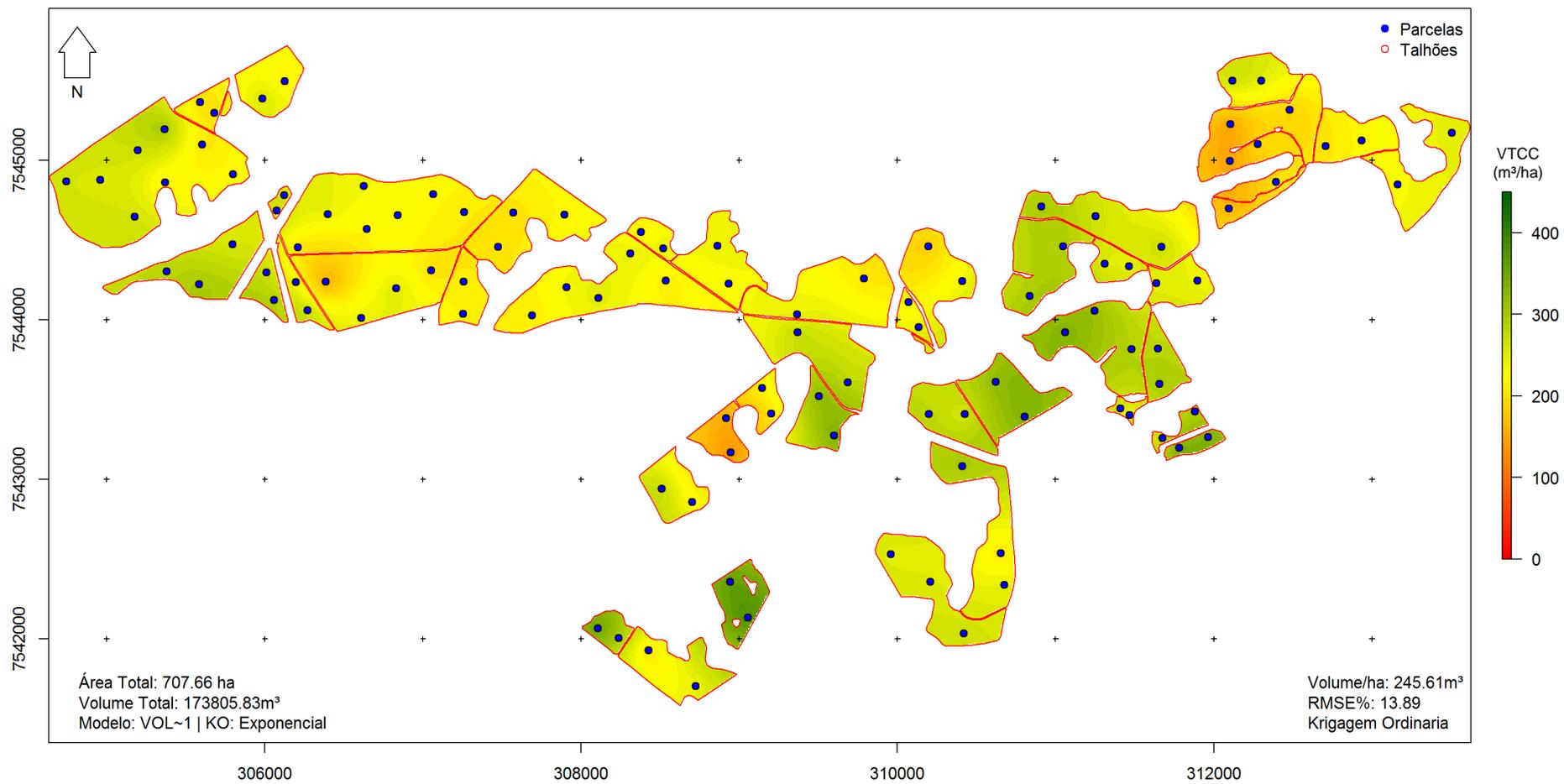


Figura 24 – Mapa de distribuição de VTCC (m³/ha) para toda a fazenda, realizado por meio da KO

4.4.2 Krigagem com Deriva Externa (KDE)

4.4.2.1 Ajuste e desempenho da KDE

Assim como na KO, a abordagem da KDE inicia-se com a construção do semivariograma experimental. Porém nesta etapa, insere-se variáveis auxiliares que possuem a capacidade de explicar parte da variação da variável principal, isto é, o novo semivariograma possui sua variabilidade total reduzida e recebe o nome de semivariograma residual (YAMOMOTO; LANDIM, 2013). Na Figura 26 é apresentada uma comparação do semivariograma original da KO com os semivariogramas residuais da KDE, utilizando as variáveis auxiliares derivadas das imagens LS8, RAPIDEYE, GEOEYE e Aerotransportadas.

Desta forma, conforme demonstra a Figura 26A, a utilização de variáveis auxiliares do LS8 explicou 67,1% da semivariância média inicial dos dados, enquanto que a utilização de variáveis auxiliares do RAPIDEYE explicou 60,5% (Figura 26B). A utilização de covariáveis derivadas do satélite GEOEYE e das imagens Aerotransportadas tiveram pior desempenho, reduzindo a semivariância média inicial em 47,9% e 43,6%, respectivamente (Figuras 26C e 26D).

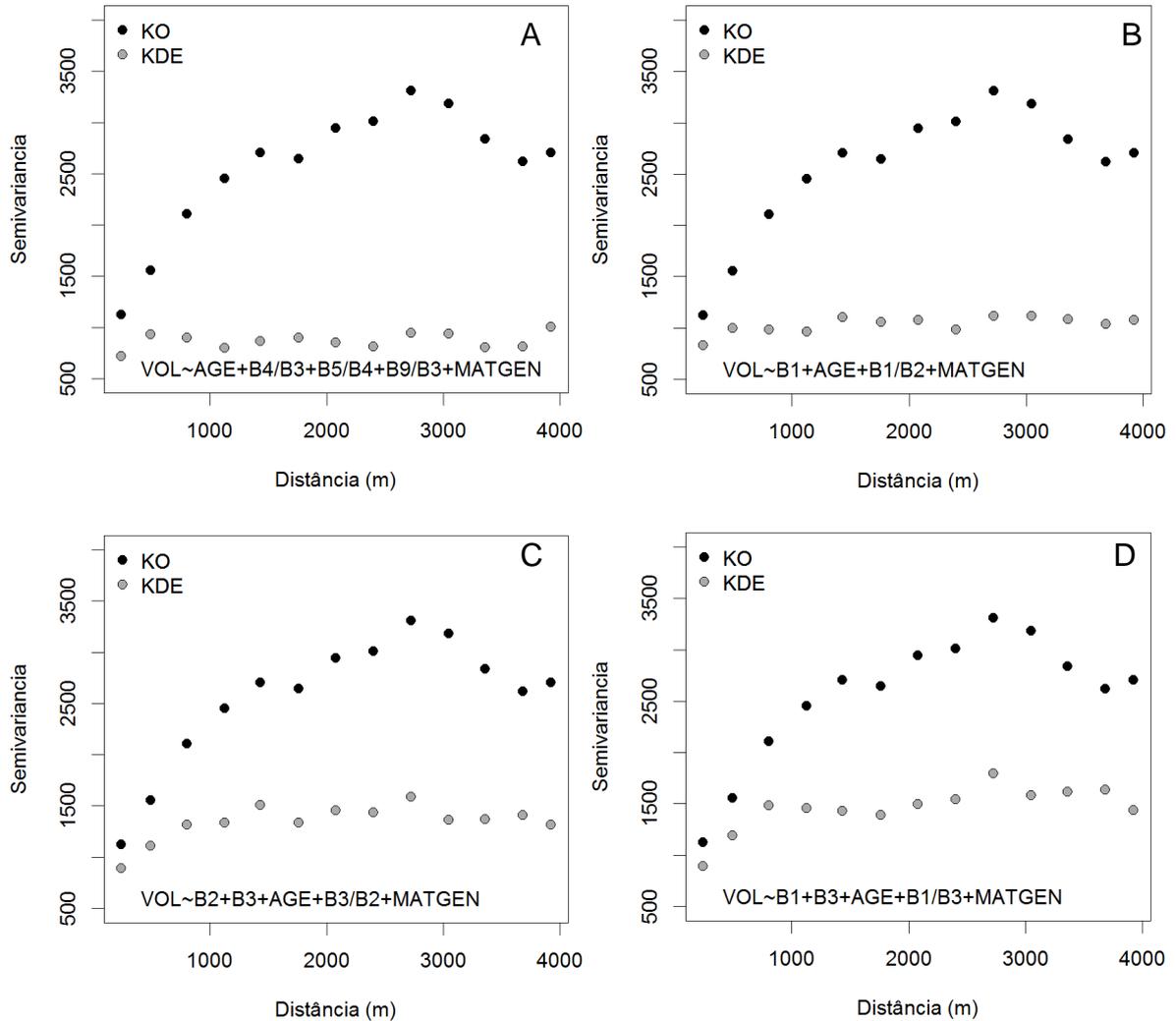


Figura 25 – Comparação dos semivariogramas experimentais de KO e KDE para da variável VTCC. A: LS8; B: RAPIDEYE; C: GEOEYE e D: Aerotransportada

A modelagem de cada um dos semivariogramas residuais se deu após a avaliação de todos os modelos geoestatísticos candidatos, descritos no Item 3.4.3., Desta forma, foi escolhido aquele que apresentou o menor RMSE% obtido pela validação cruzada LOOCV.

A Tabela 9 demonstra um resumo dos modelos selecionados. Além disso, os ajustes de tais modelos ao semivariograma podem ser observados na Figura 27, enquanto as estatísticas envolvendo a seleção entre os modelos candidatos pode ser encontrada no Item Anexo nas Tabelas 17, 18, 19 e 20.

Tabela 9 – Resumo das estatísticas de ajuste KDE. *Valores de RMSE% obtidos a partir da validação cruzada LOOCV

Conjunto de Dados	Modelo	Pepita (τ^2)	Patamar (σ^2)	Alcance (ϕ)	*RMSE%
LANDSAT8	Bes	0	896,8	98,3	12,30
RAPIDEYE	Bes	209	811	109,1	13,41
GEOEYE	Pow	373,5	1055,6	539,9	14,34
Aerotransp.	Gau	685,4	812,7	401,6	14,12

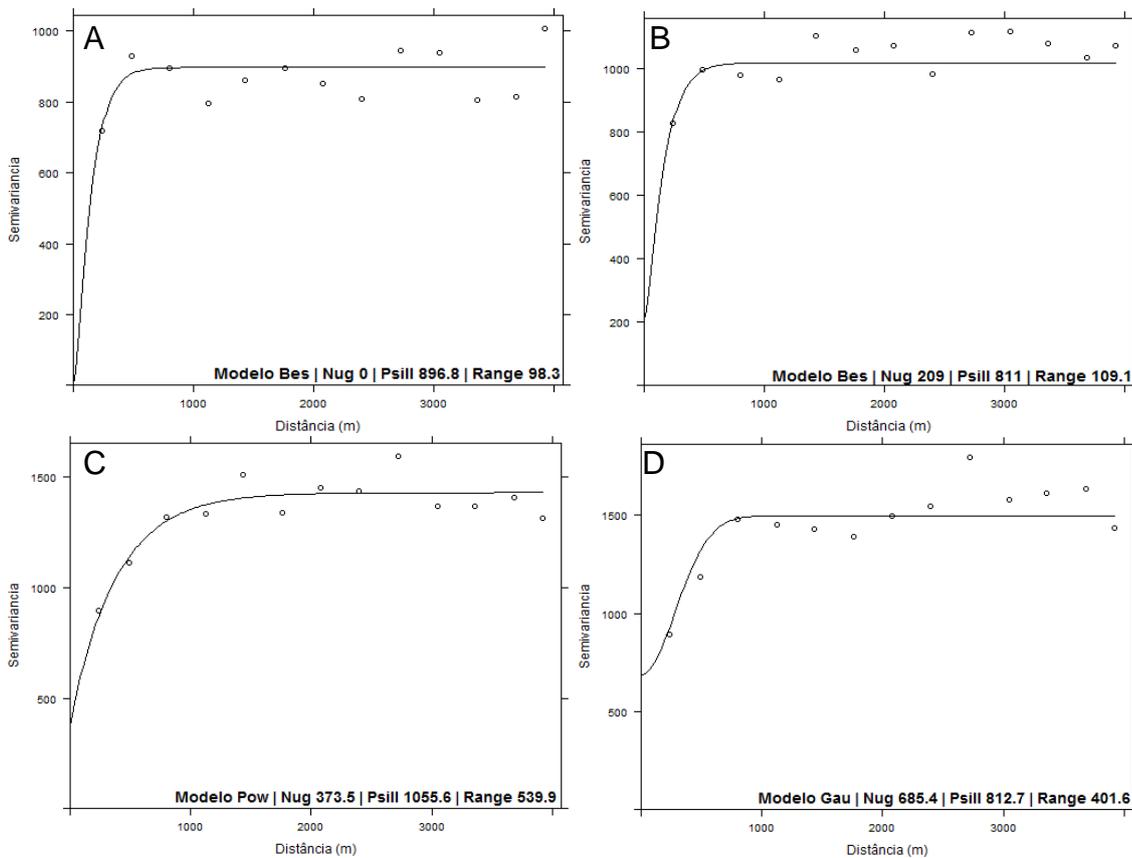


Figura 26 – Ajuste do Semivariograma residual da Krigagem com Deriva Externa. A: LS8; B: RAPIDEYE; C: GEOEYE e D: Aerotransportada

A Figura 28, demonstra a análise gráfica dos ajustes da KDE, sendo observada uma melhora no desempenho de todos os grupos, quando comparado à performance da RLM (Figura 17).

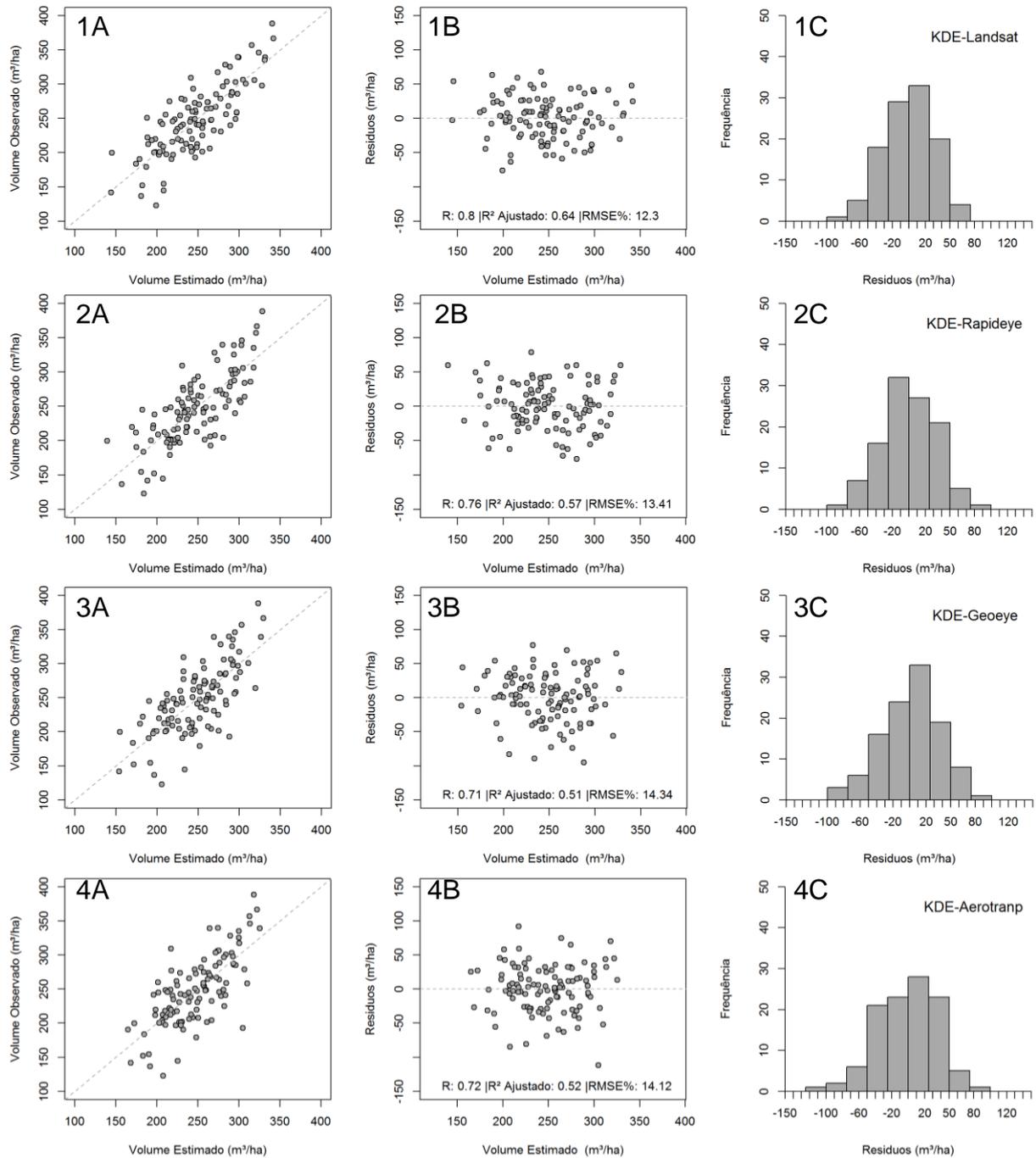


Figura 27 – Análise gráfica dos ajustes. 1-Preditores LANDSAT8; 2-Preditores RAPIDEYE; 3-Preditores GEOEYE; 4-Preditores Aertransportado. A-Gráfico Observado x Estimado; B-Dispersão dos resíduos e C-Histograma dos resíduos

Tsui et al. (2013) avaliaram a utilização de três técnicas geoestatísticas multivariadas (Cokrigagem Ordinária, Krigagem com Deriva Externa e Cokrigagem da Regressão) para estimar a biomassa de florestas temperadas no Canadá, utilizando a integração de dados LiDAR e de radar. Os autores observaram que para diferentes intensidades amostrais a KDE produziu os melhores resultados com valores de RMSE variando de 22,7 a 28,7%, a depender da intensidade amostral.

Sales et al. (2007) utilizaram a KDE para prever a biomassa acima do solo para todo o estado de Rondônia. Os autores utilizaram como medidas auxiliares as variáveis categóricas de fisionomia florestal e tipo de solo, além da variável de altitude do terreno. O estudo demonstrou que a utilização da KDE possui um potencial de reduzir em até 27% a incerteza da predição, se comparada com a utilização de uma média simples. Ainda segundo os autores, a KDE possui potencial para ser aplicada sobre toda Amazônia brasileira para promover estimativas mais acuradas de biomassa.

4.4.2.2 Estimativa volumétrica

Após o ajuste de cada um dos semivariogramas residuais e de posse das variáveis auxiliares em área total, pôde-se aplicar os modelos geoestatísticos e gerar a estimativa volumétrica para toda a área de interesse.

Esse procedimento apresenta características da KO a medida que permite a confecção de um mapa de distribuição da variável de interesse em área total, mas também apresenta características da RLM, uma vez que a estimativa de cada um dos *pixels* é influenciada pelas variáveis auxiliares.

Dessa forma, caso o conjunto de variáveis auxiliares apresente elevada correlação com a variável resposta o mapa possuirá um aspecto mais errático semelhante às predições da RLM, caso contrário, o produto da predição apresentará um aspecto mais suavizado típico da KO.

As estimativas de média global de VTCC/ha foram muito próximas entre as predições de LS8 (246,42m³/ha), RAPIDEYE (245,54m³/ha) e Aerotransportadas (246,04m³/ha), seguindo a mesma tendência observada na RLM a estimativa utilizando o conjunto de variáveis auxiliares do satélite GEOEYE produziu a menor média (240,1m³/ha).

As Figuras 29, 30, 31 e 32 demonstram os resultados da aplicação dos modelos de KDE para os conjuntos de dados LS8, RAPIDEYE, GEOEYE e Aerotransportados, respectivamente.

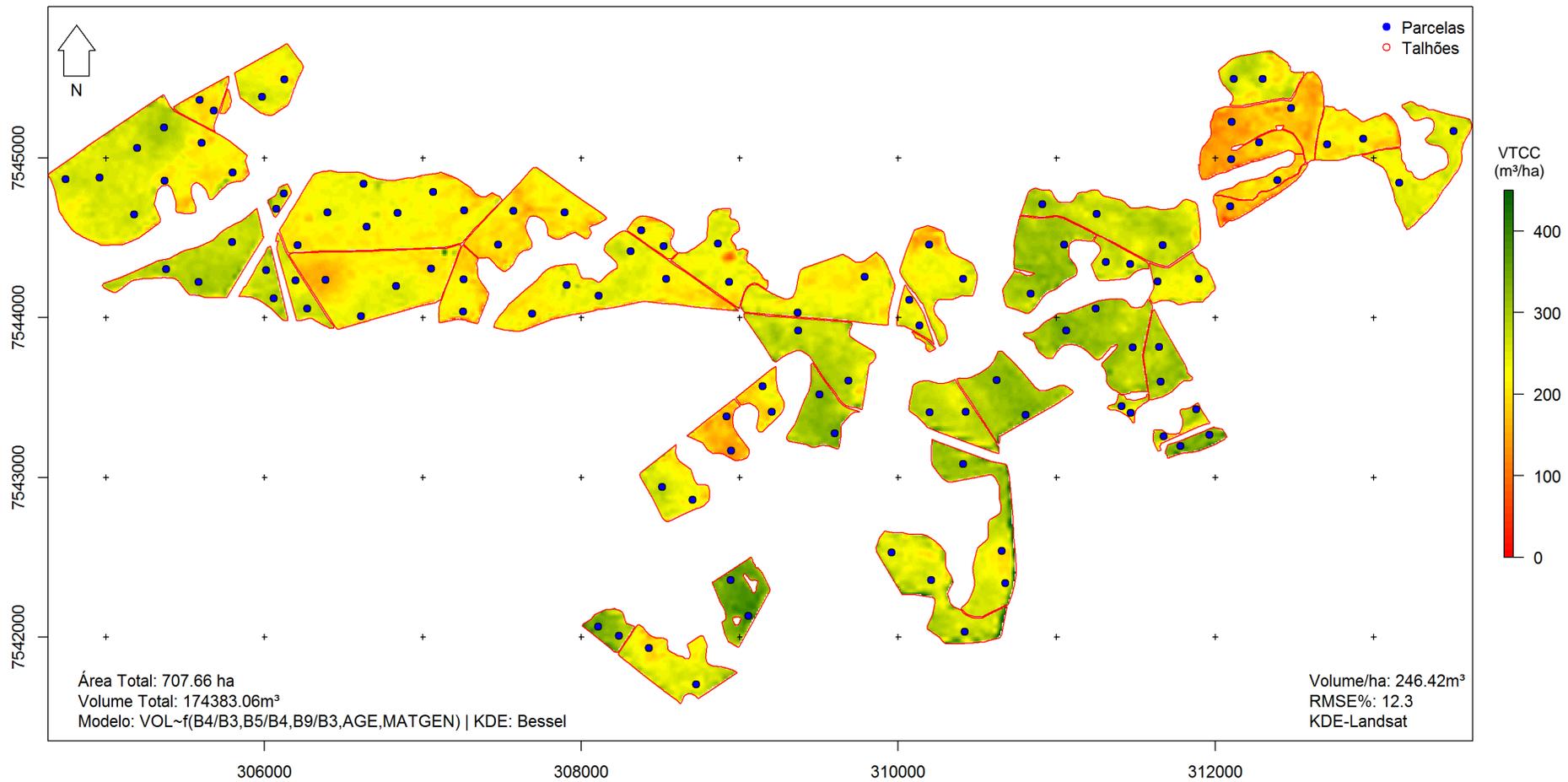


Figura 28 – Mapa de distribuição de VTCC (m³/ha) para toda a fazenda – KDE utilizando variáveis auxiliares derivados das imagens LANDSAT-8

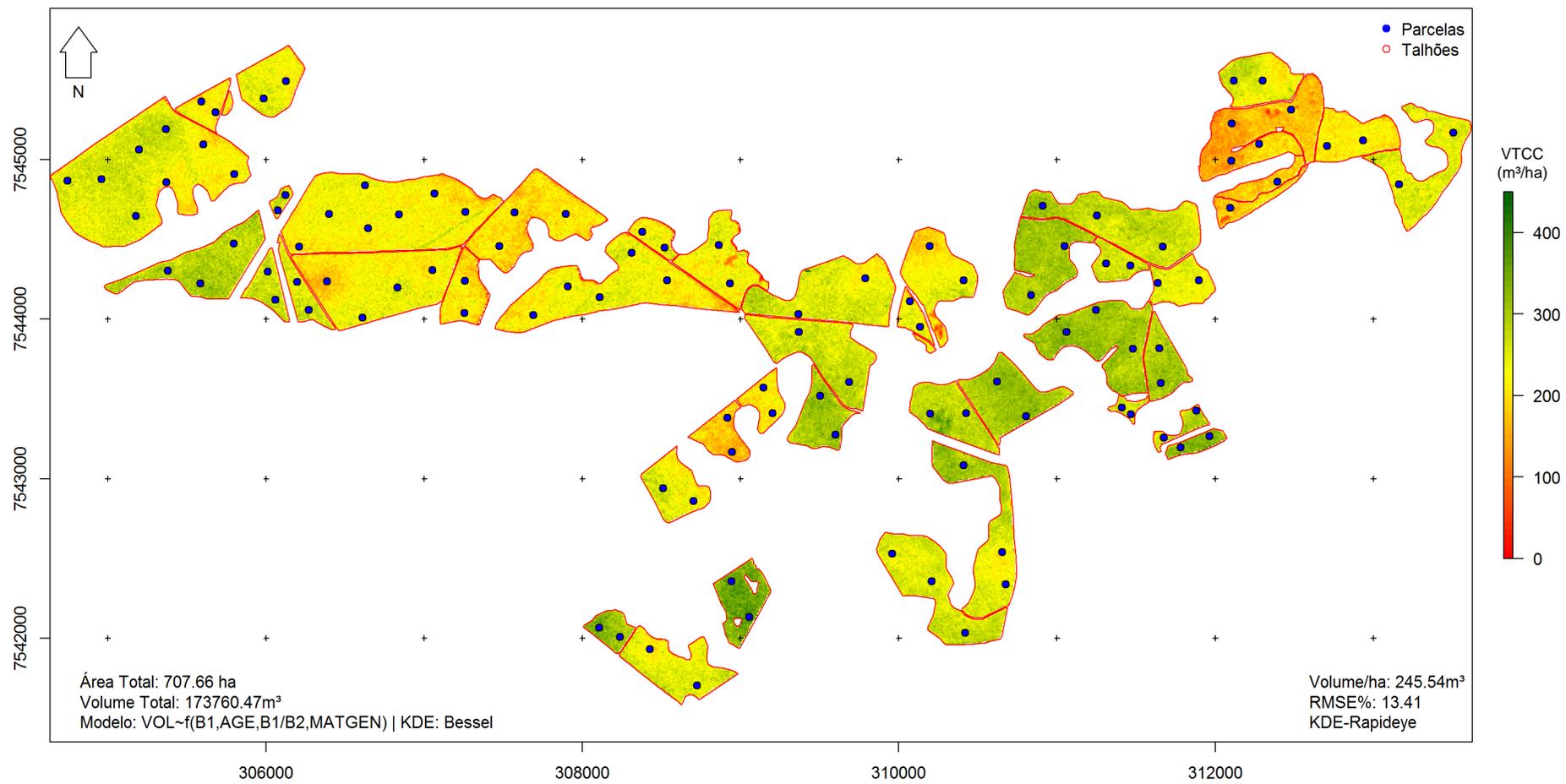


Figura 29 – Mapa de distribuição de VTCC (m³/ha) para toda a fazenda – KDE utilizando variáveis auxiliares derivados das imagens RAPIDEYE

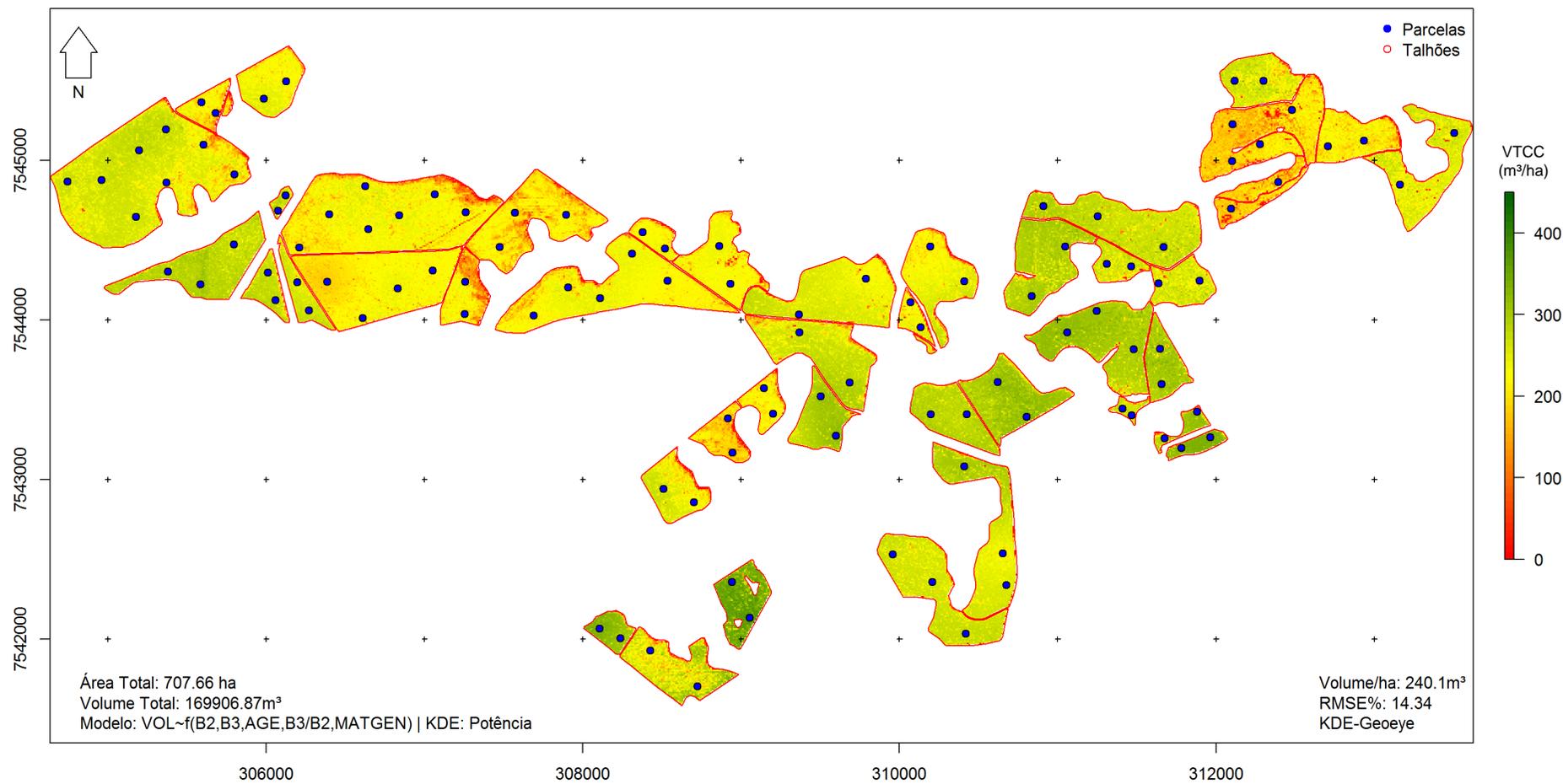


Figura 30 – Mapa de distribuição de VTCC (m³/ha) para toda a fazenda – KDE utilizando variáveis auxiliares derivados das imagens GEOEYE

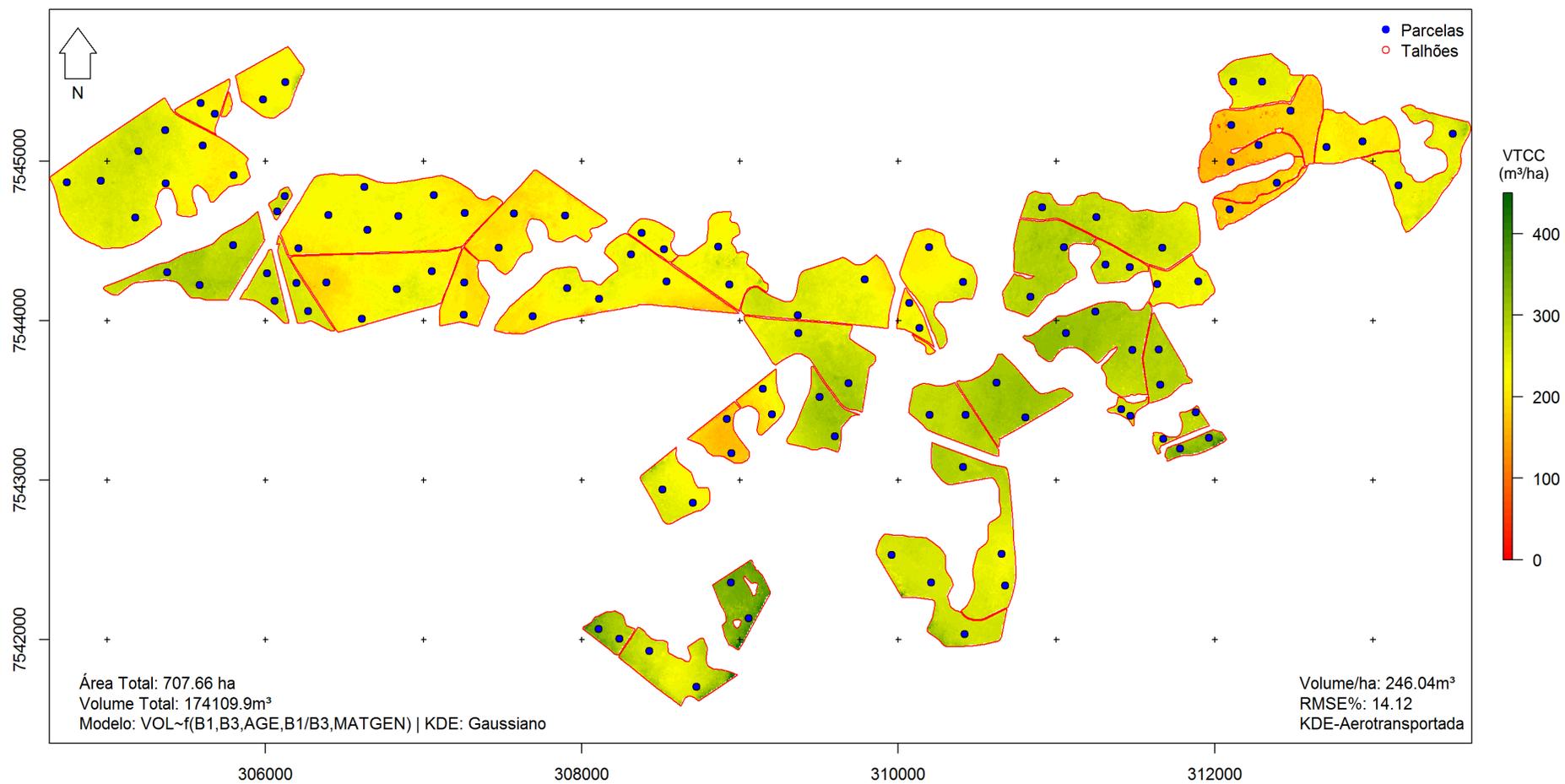


Figura 31 – Mapa de distribuição de VTCC (m³/ha) para toda a fazenda – KDE utilizando variáveis auxiliares derivados das imagens Aerotransportadas

4.5 Análise comparativa dos estimadores espaciais

A Tabela 10 apresenta um compilado de todos os resultados obtidos durante as etapas preditivas descritas anteriormente.

Tabela 10 – Resumo das estatísticas de ajuste RLM, KO e KDE. *Valores de RMSE% obtidos a partir da validação cruzada LOOCV

Método	Preditores / Auxiliares	R ² Ajustado	RMSE (m ³ /ha)	RMSE%	VTCC/ha	VTCC Fazenda
RLM	LS8	0,58	32,96	13,31	245,98	174.073,59
	RAPIDEYE	0,50	36,03	14,55	245,61	173.808,40
	GEOEYE	0,37	40,59	16,39	240,13	169.927,64
	Aerotr.ansp.	0,32	41,99	16,96	246,35	174.328,81
KO	-----	0,54	34,39	13,89	245,61	173.805,83
KDE	LS8	0,64	30,45	12,30	246,42	174.383,06
	RAPIDEYE	0,57	33,19	13,41	245,54	173.760,47
	GEOEYE	0,51	35,50	14,34	240,10	169.906,87
	Aerotr.ansp.	0,52	34,95	14,12	246,04	174.109,90

A Figura 33 permite realizar uma análise visual e comparativa da qualidade das predições utilizando-se diferentes métodos preditivos, e diferentes variáveis predictoras em um talhão com mortalidade (ocupada com sub-bosque de vegetação nativa e braquiárias).

Como já se esperava a KO, devido à ausência de variáveis auxiliares, se mostrou incapaz de detectar a anomalia na plantação. Desta forma o método interpolou volume no local da mortalidade, entre as duas parcelas amostrais, superestimando, desta forma, o volume naqueles pontos (Figura 33).

As estimativas utilizando os preditores GEOEYE e Aerotransportados, se mostraram pouco eficientes na detecção da mortalidade do plantio, isso pode ser atribuído ao número restrito de bandas espectrais disponibilizadas para o processo de seleção de variáveis, nestes dois conjuntos de preditores. Ainda assim, as estimativas com variáveis GEOEYE se mostraram mais coerentes que aquelas que utilizaram imagens Aerotransportadas, uma vez que a imagem de satélite é gerada em uma única cena, contendo as mesmas condições de angulação, espalhamento atmosférico e incidência solar (NOVO, 2010).

O mosaico de imagens aerotransportadas utilizado neste estudo, apresentou algumas fragilidades relacionadas às distorções, ao efeito *blur*, além de diferenças na angulação e na exposição solar entre as diferentes fotos que compuseram o mosaico.

Tais fatores inseriram ruído nos valores de refletância das variáveis preditoras e prejudicaram o desempenho delas na modelagem de características da floresta.

Nesse contexto, Puliti et al. (2015) estudaram a utilização de imagens de veículos aéreos não tripulados (VANT) com elevada resolução espacial para a predição de diversos atributos florestais em uma floresta boreal na Noruega. Os autores avaliaram as imagens aéreas não apenas como fornecedoras de preditores espectrais, mas também como fornecedoras de métricas estruturais da floresta, obtidas por meio de visão computacional (menos susceptível as fragilidades expostas acima). Dessa forma, foram obtidas boas estimativas volumétricas de RLM com valores de R^2 Ajustado de 0,85 e RMSE de 14,95%.

As estimativas que utilizaram preditores oriundos das imagens LS8 e RAPIDEYE, apresentaram melhores desempenhos estatísticos e visuais, independentemente do estimador utilizado (RLM ou KDE). Conforme demonstra a Figura 33, ambos os conjuntos de dados foram capazes de detectar a região de mortalidade no talhão, entretanto, foram verificadas diferenças na capacidade de definição e delimitação da clareira.

Nesse sentido, quando o objetivo do inventário florestal é um levantamento de maior escala (nível de fazenda ou um conjunto delas), recomenda-se o uso de imagens LS8. Isso permitirá a detecção de média a grandes clareiras ou anomalias, que possuam área superior ao tamanho do *pixel* padrão do satélite, ou seja, a resolução espacial (30x30m ou 900m²). Já as imagens RAPIDEYE apresentam capacidade de detecção de anomalias 36x vezes superior (25m²) se comparadas com a capacidade do LS8. Dessa forma, estas imagens apresentam grande vantagem em inventários de menor escala (nível de talhão - fazenda), porém, ao contrário das imagens LS8, não são de distribuição gratuita e apresentam custo aproximado de U\$1,28/km² (EOTEC, 2012).

Com base nos resultados apresentados na Tabela 10, nos gráficos da qualidade dos ajustes demonstrados nas Figuras 17, 24 e 28, e no quadro comparativo da Figura 33, selecionou-se como a melhor estimativa, aquela gerada pelo conjunto de variáveis auxiliares LS8, utilizando-se o estimador da KDE. Sendo assim, tal predição foi utilizada como variável auxiliar durante a estimativa gerada pelo Estimador de Regressão (ER).

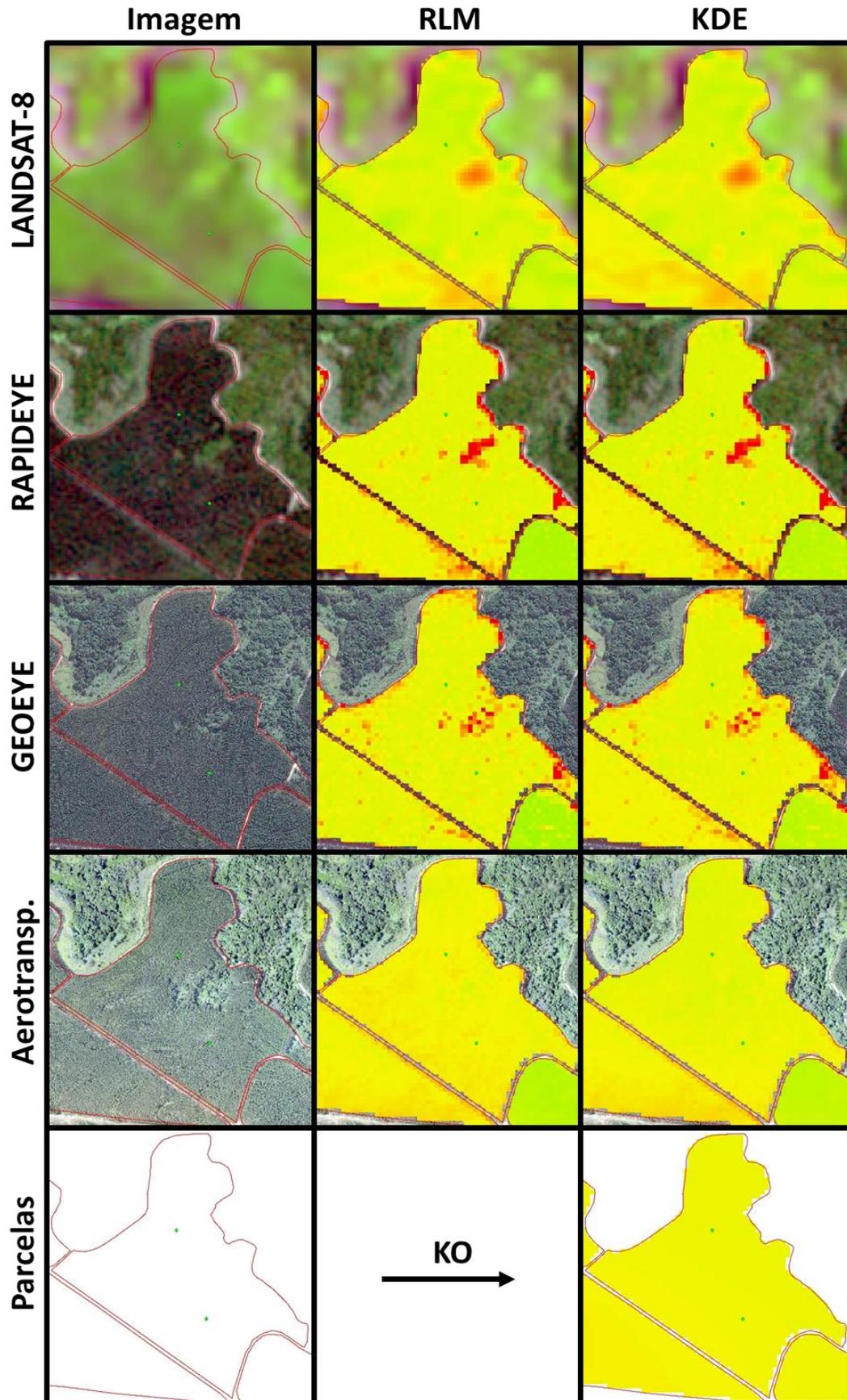


Figura 32 – Comparativo visual da capacidade de detecção de mortalidade no plantio, das predições de RLM, KDE e KO utilizando diferentes variáveis auxiliares. Pontos Verdes: Unidades Amostrais

4.6 Análise comparativa dos estimadores clássicos

A Tabela 11 demonstra os resultados da estimativa da média populacional juntamente com outras métricas, com base nos estimadores da estatística clássica, AAS e ER.

Tabela 11 – Resumo dos resultados da estimativa de VTCC obtidos pelas estimativas de AAS e ER.

Estimadores	VTCC (m³/ha)	IC (m³/ha)	VTCC Fazenda	IC Fazenda	IC (%)
AAS	247,59	±9,58	175.212,40	±6.777,73	±3,87
ER	246,65	±5,78	174.544,20	±4.093,25	±2,34

Conforme o cálculo de suficiência amostral, descrita no Item 3.4.5. (Equação 15), caso se desejasse reduzir o intervalo de confiança da amostragem AAS para o mesmo nível proporcionado pelo ER, seriam necessárias além das 110 parcelas iniciais, a instalação de mais 99 parcelas, totalizando 209 unidades amostrais.

Ressalta-se que o método permitiu uma redução no intervalo de confiança da média, sem elevar os custos com aumento da intensidade amostral.

5 CONCLUSÕES

Os resultados apresentados neste trabalho demonstram as vantagens da integração de dados de Sensoriamento Remoto (SR) ao inventário florestal. A aplicação de variáveis auxiliares provenientes de sensores aerotransportados ou orbitais no contexto da geoestatística multivariada (KDE), produziu resultados mais promissores que os estimadores univariados (KO) ou que não considerem os efeitos da estrutura de dependência espacial (RLM).

A disponibilidade de uma variabilidade maior de bandas espectrais demonstrou benefícios em detrimento de elevadas resoluções espaciais. O produto das estimativas da KDE pode ser utilizado como uma medida auxiliar no contexto da estatística clássica.

Tais resultados evidenciam a importância geoestatística multivariada para a predição volumétrica em florestas, se comparada a outros métodos de inferência. Com isto aproveita-se, um recurso de baixíssimo custo e muitas vezes já disponível nas empresas florestais (Imagens Orbitais), para aumentar a confiabilidade e a precisão do inventário florestal, sem para isto, elevar os custos com o aumento da intensidade amostral na obtenção dos dados em campo.

REFERÊNCIAS

- ASSMANN, E. **The principles of forest yield study**. New York: Pergamon, 1970. 506p.
- BARROS, B.S.X.; GUERRA, S.P.S.; BARROS, Z.X.; CATITA, C.M.S.; FERNANDES, J. C. C. C. Uso de imagens de satélite para cálculo de volume em floresta de eucalipto no município de Botucatu/SP. **Revista Energia na Agricultura**, Botucatu, v. 30, n.1, p. 60-67, 2015.
- BARTOSZECK, A.C.P.; MACHADO, S.A.; FILHO, A.F.; OLIVEIRA, E.B. Modelagem da relação hipsométrica para bracatingais da região metropolitana de Curitiba-PR. **Revista Floresta**, Curitiba, v. 32, n. 2, p. 189-204, 2002.
- BATISTA, J.L.; COUTO, H.T.Z.; FILHO, D.F.S. Fundamentos de Silvimetria. In: -----
----- **Quantificação de recursos florestais: árvores, arvoredos e florestas**. São Paulo: Editora Oficina de Textos, 2014. cap. 13, p. 253-267.
- BERRA, E.F.; BRANDELERO, C.; PEREIRA, R. S.; SEBEM, E.; GOERGEN, L.C. G.; BENEDETTI, A.C.P.; LIPPERT, D.B. Estimativa do volume total de madeira em espécies de eucalipto a partir de imagens de satélite Landsat. **Ciência Florestal**, Santa Maria, v. 22, n. 4, p. 853-864, 2012.
- BIVAND, R.S.; PEBESMA, E.J.; GÓMEZ-RUBIO, V. Interpolation and Geostatistics. In: -----
----- **Applied Spatial Data Analysis with R**, 2nd ed. New York: Springer, 2011. chap.8, p.191-236.
- BLACKBRIDGE. High-Resolution Satellite Imagery. Disponível em:
<<http://www.blackbridge.com/rapideye/all-products/images.htm>>. Acesso em: 07 mar. 2015.
- CAMPOS, J.C.C.; LEITE, H.G. **Mensuração Florestal: Perguntas e Respostas**. 4th. Viçosa: UFV, 2013. 605p.
- CANAVESI, V.; PONZONI, F.J.; VALERIANO, M.M. Estimativa de volume de madeira em plantios de *Eucalyptus* spp. utilizando dados hiperespectrais e dados topográficos. **Revista árvore**, Viçosa, v. 34, n. 3, p. 539-549, 2010.
- CEPAGRI Clima dos Municípios Paulistas. Disponível em:
<http://www.cpa.unicamp.br/outras-informacoes/clima_muni_173.html>. Acesso em 05 de fev. 2015.
- CHILE'S, J.P.; DELFINER, P. **Geostatistics: modeling spatial uncertainty**. 2nd ed. New York: Wiley 2012. 158p.
- COUTO, H.T.Z. Sistema de Informações Geográficas: Aplicações florestais. **IPEF Série Técnica**, Piracicaba, v. 9, n. 28, p. 1-18, 1993.
- CRESSIE, N.A.C. **Statistics for Spatial Data: Revised Edition**. New York: Wiley Classics Library, 1993. 900p.

DIGGLE, P.J.; RIBEIRO JR, P.J. **Model-based Geostatistics** New York: . Springer Science, Business Media, LLC, 2007. 228p.

DIGITAL GLOBE. **Digitalglobe Constellation**: Geoeye-1. Disponível em: <www.digitalglobe.com>. Acesso em: 20 abr. 2015.

DORMANN C.F.; ELITH J.; BACHER S.; BURCHMANN C.; CARL G.; CARRÉ G.; MARQUÉZ J.R.G; GRUBER B.; LAFOURCADE B.; LEITÃO P.J.; MUNKEMULLER T.; MCCLEAN C.; OSBORNE P.E.; REINEKING B.; SCHRODER B.; SKIDMORE A.K.; ZURELL D.; LAUTENBACH S. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. **Ecography**, Copenhagen, v.36, p. 027-046, 2013.

EITEL, J.U.H.; LONG , D.S.; GESSLER , P.E.; SMITH, A.M.S. Using in-situ measurements to evaluate the new RapidEye satellite series for prediction of wheat nitrogen status. **International Journal of Remote Sensing**, Basingstoke, v. 28, p.1-8, 2007.

EITEL, J.U.H.; VIERTLING , L.A.; LITVAK , M.E.; LONG , D.S.; SCHULTHESS , U.; AGER, A.A.; KROFCHECK , D.J.; STOSCHECK , L. Broadband, red-edge information from satellites improves early stress detection in a New Mexico conifer woodland. **Remote Sensing of Environment**, New York, v.115, p.3640-3646, 2011.

ELACHI, C. **Introduction to the Physics and Techniques of Remote Sensing**, New York: Wiley, 1987. 413p.

EMBRAPA. Satélites de Monitoramento: Geoeye. Disponível em: <http://www.sat.cnpm.embrapa.br/conteudo/missao_geoeye.php>. Acesso em: 20 abr. 2015.

EARTH OBSERVATION TECHNOLOGIES – USD Price List for RapidEye Imagery Products - 2012. Disponível em: <http://www.eotec.com/images/Pricing_RapidEye_products_ver1.pdf>. Acesso em 08 de ago. 2015.

FIORELLA, M.; RIPPLE, W.J. Determining the successional stage of temperate coniferous forest with Landsat satellite data. **Photogrammetric Engineering and Remote Sensing**, Bethesda, v. 59, p. 239-246, 1993.

GARLIPP, R.; FOELKEL, C. O papel das florestas plantadas para atendimento das demandas futuras da sociedade. In: CONGRESSO FLORESTAL MUNDIAL/FAO, 13., 2009, Buenos Aires. **Anais...** Buenos Aires: Sociedade Brasileira de Sicultura, 2009. p.1-18.

GEOEYE - The World's Highest Resolution Commercial Earth-Imaging Satellite, 2008. Disponível em: <http://www.geoimage.com.au/media/satellite_pdfs/GeoEye_FactSheet_comb.pdf>. Acesso em: 29 abr. 2015.

GAUSS CENTRE FOR SUPERCOMPUTING. Environment and Energy: High Resolution Climate Modeling. Disponível em: <http://www.gausscentre.eu/gausscentre/EN/Projects/EnvironmentEnergy/2014/panitz_HRCM.html?nn=1345670>. Acesso em: 18 abr. 2015.

GUEDES, I.C.L.; OLIVEIRA, A.D.; MELLO, J.M.; SILVA, S.T.; MELLO, C.R.; SCOLFORO, J.R.S. Técnicas geoestatísticas e interpoladores espaciais na estratificação de povoamentos de *Eucalyptus sp.* **Ciência Florestal**, Santa Maria, v. 22, p. 541-550, 2012.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of Statistical Learning: Data Mining, Inference, and Prediction**. 2nd ed. Stanford, California: Springer, 2008. 146p.

HENGL, T. **A Practical Guide to Geostatistical Mapping**. 2nd ed. Amsterdam: University of Amsterdam, 2009. 290p.

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

HIJMANS, R.J.; ETTEN, J.V. **Raster: Geographic data analysis and modeling**. R package, 31p, 2012. Disponível em: <<https://cran.r-project.org/web/packages/raster/raster.pdf>>. Acesso em: 10 nov. 2014.

HOLMGREN, P.; THURESSON, T. Satellite remote sensing for forestry planning – A review. **Scandinavian Journal of Forest Research**, Alnarp, v. 13, n.1/4, p. 90-110, 1998.

IBÁ – **Industria Brasileira de Árvores**: Anuário estatístico 2015. Brasília, 2015. 64p.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An Introduction to Statistical Learning: with Applications in R**. New York: Springer, 2013. 426p.

JENSEN, J.R. **Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres**, Ed. Parêntese, 2009. 598p.

JI, L.; WYLIE, B. K.; NOSSOV, D. R.; PETERSON, B.; WALDROP, M. P.; MCFARLAND, J. W.; ROVER, J.; HOLLINGSWORTH, T. N. Estimating aboveground biomass in interior Alaska with Landsat data and field measurements. **International Journal of Applied Earth Observation and Geoinformation**, Enschede, v. 18, p. 451-461, 2012.

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

KANEGAE JUNIOR, H.; MELLO, J.M.; SCOLFORO, J.R.S.; OLIVEIRA, A.D.O. Avaliação da continuidade espacial de características dendrométricas em diferentes idades de povoamentos clonais de *Eucalyptus sp.* **Revista Árvore**, Viçosa-MG, v.31, n.5, p.859-866, 2007.

KANGAS, A. Design-based sampling and inference. In: KANGAS, A.; MALTAMO, M. (Ed.). **Forest Inventory: Methodology and Applications**. Dordrecht: Springer, 2006. chap. 2, p. 13-38.

KANGAS, A.; GOVE, J.H.; SCOTT, C.T. Introduction. In: KANGAS, A.; MALTAMO, M. (Ed.). **Forest Inventory: Methodology and Applications**. Dordrecht: Springer, 2006. chap. 1, p. 3-11.

KANGAS, A.; MALTAMO, M. (Ed.). **Forest Inventory: Methodology and Applications**. Dordrecht: Springer, 2006. 368p.

KÖHL, M. ; MAGNUSSEN, M. ; MARCHETTI, M. **Tropical Forestry: Sampling Methods, Remote Sensing and GIS Multiresource Forest Inventory**, Ed. Springer-Verlag Berlin Heidelberg, 2006, 373p.

LEAL, F.A.; MIGUEL, E.P.; MATRICARDI, E.A.T. Mapeamento de unidades produtivas utilizando a interpolação geoespacial krigagem a partir do inventário florestal em um povoamento de *Eucalyptus urophylla* S. T. Blake. **Enciclopédia Biosfera**, Centro Científico Conhecer – Goiania, v. 17, n.12, p. 727-745, 2011.

LEWIN-KOH, N.J.; BIVAND; Maptools package - Tools for reading and handling spatial objects. R packages. Disponível em: <<https://cran.r-project.org/web/packages/maptools/index.html>>. Acesso em: 14 out. 2012.

LOPES, D.M. **Estimating Net Primary Production in *Eucalyptus globules* and *Pinus Pinaster* Ecosystems in Portugal**. 2005. 291p. Tese (Doutorado na área de recursos ambientais) – Kingston University, Londres, 2005.

LUMLEY, T. **LEAPS**: R package version 2.9 for regression subset selection including exhaustive search. Washington: University of Washington, 2009. 8p. Disponível em: <<http://cran.r-project.org/web/packages/leaps/leaps.pdf>>. Acesso em: 08 nov. 2014.

MAGNUSSON, M. **Evaluation of remote sensing techniques for estimation of forest variables at stand level**. 2006. 38p. Tese (Doutorado na área de ciências florestais) - Swedish University of Agricultural Sciences, Umea, 2006.

MELLO, J.M. **Geoestatística aplicada ao inventário florestal**. 2004. 110p. Tese (Doutorado em Recursos Florestais) - Escola Superior de Agricultura “Luiz Queiroz”, Universidade de São Paulo, Piracicaba, 2004.

MELLO, J.M.; DINIZ, F.S.; OLIVEIRA, A.D.; SCOLFORO, J.R.S.; ACERBI JÚNIOR, F.W.; THIERSCH, C.R. Métodos de amostragem e geoestatística para estimativa do número de fustes e volume em plantios de *Eucalyptus grandis*. **FLORESTA**, Curitiba, v. 39, n. 1, p. 157-166, jan./mar. 2009.

MENESES, P.R. ; ALMEIDA, T. **Introdução ao processamento de imagens de sensoriamento remoto**. Brasília: CNPq e UnB, 2012. 266p.

MORELI, T. RUFINO, C.; THIERSCH, C.R.; FERREIRA, S.O.; KANEGAE JUNIOR, H.; FAIS, D.; Uso da Geoestatística no Estudo da Relação entre Variáveis Dentrométricas de Povoamentos de *Eucalyptus* sp. e Atributos do Solo. **Ambiência**, Guarapuava, PR, v.2, p. 83-93, 2006.

NAIMI, B. **Package usdm**: Uncertainty Analysis for Species Distribution Models. R package version 1.1-15. Disponível em: <<https://cran.r-project.org/web/packages/usdm/index.html>>. Acesso em: 26 ago. 2015.

NOVO, E.M.L.M. **Sensoriamento remoto**: princípios e aplicações. 4 ed. São Paulo, Blucher, 2010. 387p.

PAYN, T.; CARNUS, J.; FREER-SMITH, P.; KIMBERLEY, M.; KOLLERT, W.; LIU, S.; ORAZIO, C.; RODRIGUEZ, L.; SILVA, L. N.; WINGFIELD, M. J. Changes in planted forests and future global implications. **Forest Ecology and Management**, Amsterdam, v. 352, p. 57-67, 2015.

PEBESMA, E.J. Gstat user's manual. Utrecht University, Utrecht, The Netherlands. Disponível em: <<http://www.gstat.org/gstat.pdf>>. Acesso em: 24 mar. 2014.

PEBESMA, E.J. Multivariable geostatistics in S: the gstat package. **Computers & Geosciences**, New York, v. 30, p. 683-691, 2004.

PEBESMA, E.J. BIVAND, R.S. **Classes and methods for spatial data in R. R News**, Vienna, v. 5, n.2, 2005. Disponível em: <<http://cran.r-project.org/doc/Rnews/>> Acesso em: 02 jan. 2013.

PELLISSARI, A. L.; CALDEIRA, S. F.; FILHO, A.F.; MACHADO, S.A. Propostas de mapeamentos da capacidade produtiva de sítios florestais por meio de análises geoestatísticas. **Scientia Forestalis**, Piracicaba, v. 43, n.107, p. 601-608, set. 2015.

PEÑA, E.A.; SLATE, E.H. Global validation of linear model assumptions. **Journal of the American Statistical Association**, New York, v. 101, n. 473, p. 341-354, 2006.

PEÑA, E.A.; SLATE, E.H. **gvlma**: Global Validation of Linear Models Assumptions. R package version 1.0.0.2. 2014. Disponível em: <<https://cran.r-project.org/web/packages/gvlma/index.html>>. Acesso em: 28 jan. 2015.

PICS ABOUT SPACE. Disponível em <<http://picsaboutspace.com/>>. Acesso em: 12 abr. 2014.

PONZONI, F.J.; SHIMABUKURO, Y.E.; KUPLICH, T.M. **Sensoriamento Remoto da Vegetação**, 2nd ed. São Paulo: Oficina de Textos, 2012. 160p.

PULITI, S.; ØRKA, H.O.; GOBAKKEN, T.; NÆSSET, E. Inventory of Small Forest Areas Using an Unmanned Aerial System. **Remote Sensing**, Basel, v. 7, p. 9632-9654, 2015.

QUEIROZ, W.T. **Amostragem em inventário florestal**. Belém: Edufra, 2012. 441p.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2015.

RAPIDEYE - **Delivering the World** - The RapidEye Red Edge Band, White Paper. Disponível em: <http://www.blackbridge.com/rapideye/upload/Red_Edge_White_Paper.pdf>. Acesso em: 21 abr. 2015.

RÉ, D.S.; THIERSCH, C.R.; ANDRADE, C.M. Avaliação do uso da krigagem na definição de estratos para o inventário florestal. In: SIMPÓSIO DE GEOESTATÍSTICA APLICADA EM CIÊNCIAS AGRÁRIAS, 2009, Botucatu-SP. **Anais...** Botucatu: Universidade Estadual de São Paulo, 2009. 4p.

RECIO, J.A., HELMHOLZ, P., MULLER, S. Potential evaluation of different types of images and their combination for the classification of GIS objects cropland and grassland. **Remote Sens. and Spatial Info. Sc.**, New York, v.38, 6p. 2011.

RIBEIRO, J.R.D.P.A **Detecção remota no inventário florestal: análise das potencialidades da utilização das imagens satélite.** 2007. 138p. Dissertação (Mestre em Ciências e Sistema de Informações Geográficas) – Universidade Nova de Lisboa, Lisboa, 2007.

RIBEIRO, M.S. **Geoestatística utilizada no planejamento estratégico de florestas de produção de *Pinus taeda* L.** 2014. 83p. Dissertação (Mestre em Engenharia Florestal) – Universidade Federal do Paraná, Curitiba, 2014.

ROSSITER, D.G. **A minimal introduction to geostatistics with R/gstat.** Version 3.1. Cornell University, School of Integrative Plant Science, Crop and Soil Sciences Section. , 2015. 135p.

SALES, M.H.; SOUZA JR, C.M.; KYRIAKIDIS, P.C.; ROBERTS, D.A.; VIDAL, E. Improving spatial distribution estimation of forest biomass with geostatistics: A case study for Rondonia, Brazil. **Ecological Modelling**, Amsterdam, v. 205, p. 221-230, 2007.

SANTANA, R.A. **Avaliação de técnicas geoestatísticas no inventário de povoamento de *Tectona grandis* L. f.** 2011. 43p. Dissertação (Mestrado em Ciências) – Universidade Federal de Viçosa, MG, 2011.

SCHREUDER, H. T.; ERNST, R.; RAMIREZ-MALDONADO, H. **Statistical Techniques for Sampling and Monitoring Natural Resources.** Gen. Tech. Rep. Fort Collins, United States Department of Agriculture, Rocky Mountain Station, 2004. 117p.

SCHUMACHER, F. X.; HALL, F.S. Logarithmic expression of timber-tree volume. **Journal of Agricultural Research**, Washington, D.C., v. 47, n. 9, p. 719-734, 1933.

SHINZATO, E.T. **Integração de dados laser scanner aerotransportado e imagens aéreas no inventário florestal para a estimativa volumétrica de madeira.** 2015. 85p., Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2015.

SILVA, A.G.P. **Estimativa da biomassa de lenho em povoamento de *Eucalyptus grandis* baseada em estatísticas do perfil de dossel geradas por escaneamento a laser aerotransportado.** 2014. 146p. Dissertação (Mestre em Recursos Florestais) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2014.

SOARES, C.P.B., PAULA NETO, F., SOUSA, A.L. **Dendrometria e inventário Florestal.** Viçosa: Universidade Federal de Viçosa, 2006. 276p.

SOUSA, A.M.O.; GOLÇALVES, A.C.; MESQUITA, P.; SILVA, J.R M. Biomass estimation with high resolution satellite images: A case study of *Quercus rotundifolia*. **ISPRS Journal of Photogrammetry and Remote Sensing**, Enschede, v. 101, p. 69-79, 2015.

SOUZA, A.L. **Estudo do efeito de suavização da krigagem ordinária em diferentes distribuições estatísticas**. 2007. 101p. Dissertação (Mestrado em Recursos Minerais e Meio Ambiente) - Universidade de São Paulo, São Paulo, 2007.

TROTTER, C.M., DYMOND, J.R.; GOULDING, C.J.; Estimation of timber volume in a coniferous plantation forest using Landsat TM. **International Journal of Remote Sensing**, Abingdon, v. 18, n.10, p. 2209-2223, 1997.

TSUI, O.W.; COOPS, N.C.; WULDER, M.A.; MARSHALL, P.L. Integrating airborne LiDAR and space-borne radar via multivariate kriging to estimate above-ground biomass. **Remote Sensing of Environment**, Abingdon, v. 139, p. 340-352, 2013.

TYC, G.; TULIP, J.; SCHULTEN, D.; KRISCHKE, M.; OXFORT, M. The RapidEye mission design. **Acta Astronautica**, Elmsford, v.56, p.213-219, 2005.

UNITED STATES GEOLOGICAL SURVEY, Landsat 8, Fact Sheet. 2013. v. 3060. Disponível em: < <http://pubs.usgs.gov/fs/2013/3060/pdf/fs2013-3060.pdf> >. Acesso em: 04 ago. 2015.

VALENTE, R.O. A. **Análise da estrutura da paisagem do rio Corumbataí**. 2004. 144p. Dissertação (Mestrado em Recursos Florestais) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2001.

WACKERNAGEL, H. **Multivariate Geostatistics: An Introduction with Applications**. Berlin: Springer, 2003. 387p.

WATANABE J.;YAMAMOTO J.K.; ROCHA M., FONSECA P.P. Estudo da influência da correlação inicial entre variáveis nos resultados de co-estimativas. **Geociências**, Rio Claro, v. 28, n.4, p. 467-484, 2009.

WU, C.; NIU, Z.; TANG, Q.; HUANG, W.; RIVARD, B.; FENG, J.; Remote estimation of gross primary production in wheat using chlorophyll-related vegetation indices. **Agricultural and Forest Meteorology**, Amsterdam, v. 149, p.1015-1021. 2009.

YAMAMOTO, J.K. Pós-processamento das Estimativas de Krigagem com Deriva Externa. **Geologia USP Série Científica**, São Paulo, v.9. n.3, p.89-101. 2009.

YAMAMOTO, J.K.; LANDIM, P.M.B. **Geoestatística: conceitos e aplicações**. São Paulo: Editora Oficina de Textos, 2013. 215p.

ZHANG, G.; GANGULY, S.; NEMANI, R.R.; WHITE, M.A.; MILESI, C.; HASHIMOTO, H.; WANG, W.; SAATCHI, S.; YU, Y.; MYNENI, R.B. Estimation of forest aboveground biomass in California using canopy height and leaf area index estimated from satellite data. **Remote Sensing of Environment**, Abingdon, v. 131, p. 44-55, 2014.

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

ANEXO

ANEXO A -

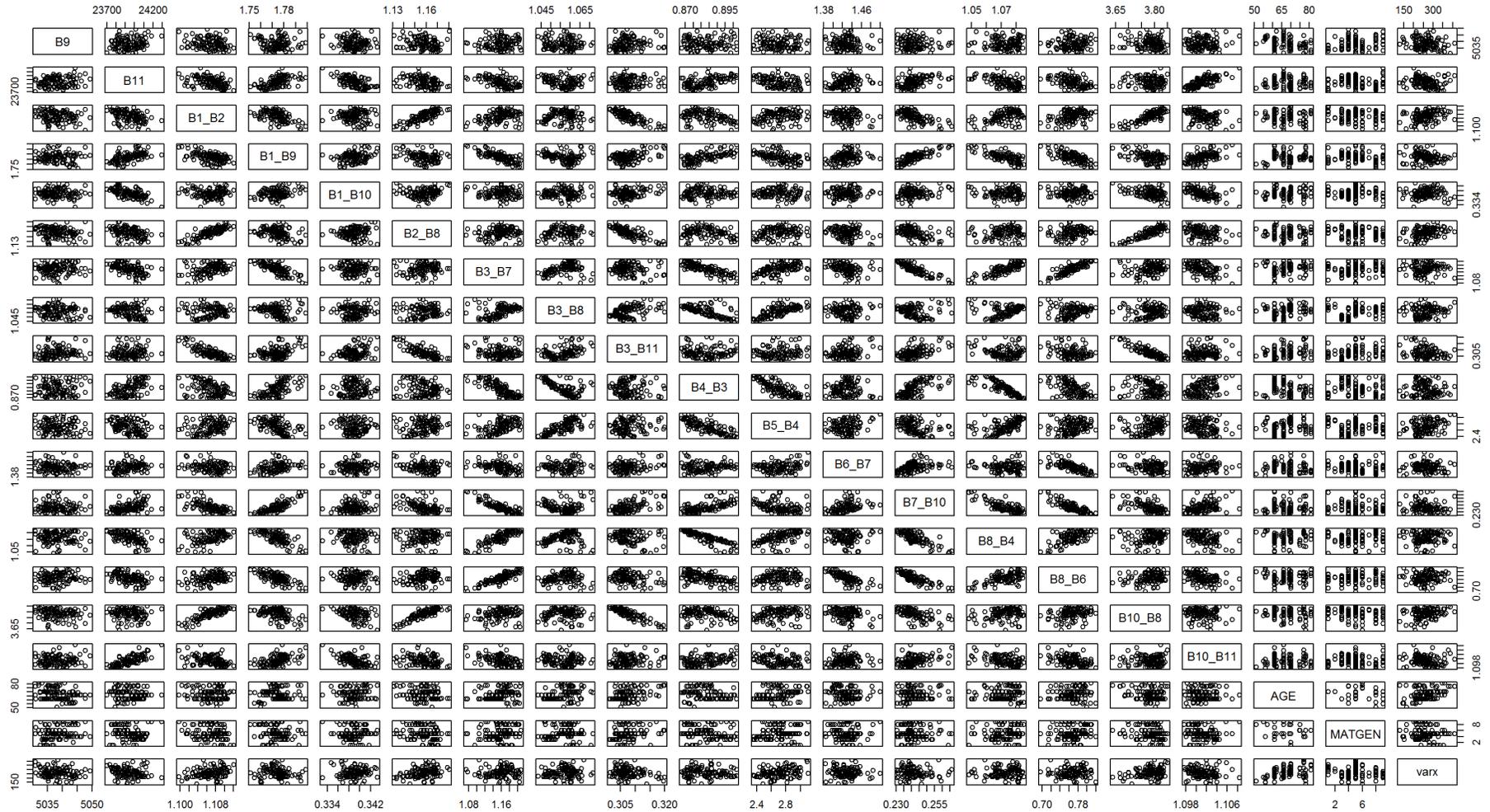


Figura 33 –Matriz de correlação das variáveis preditoras oriundas das imagens LANDSAT8 e submetidas ao processo de *Best Subset Selection*. Varx=VTCC

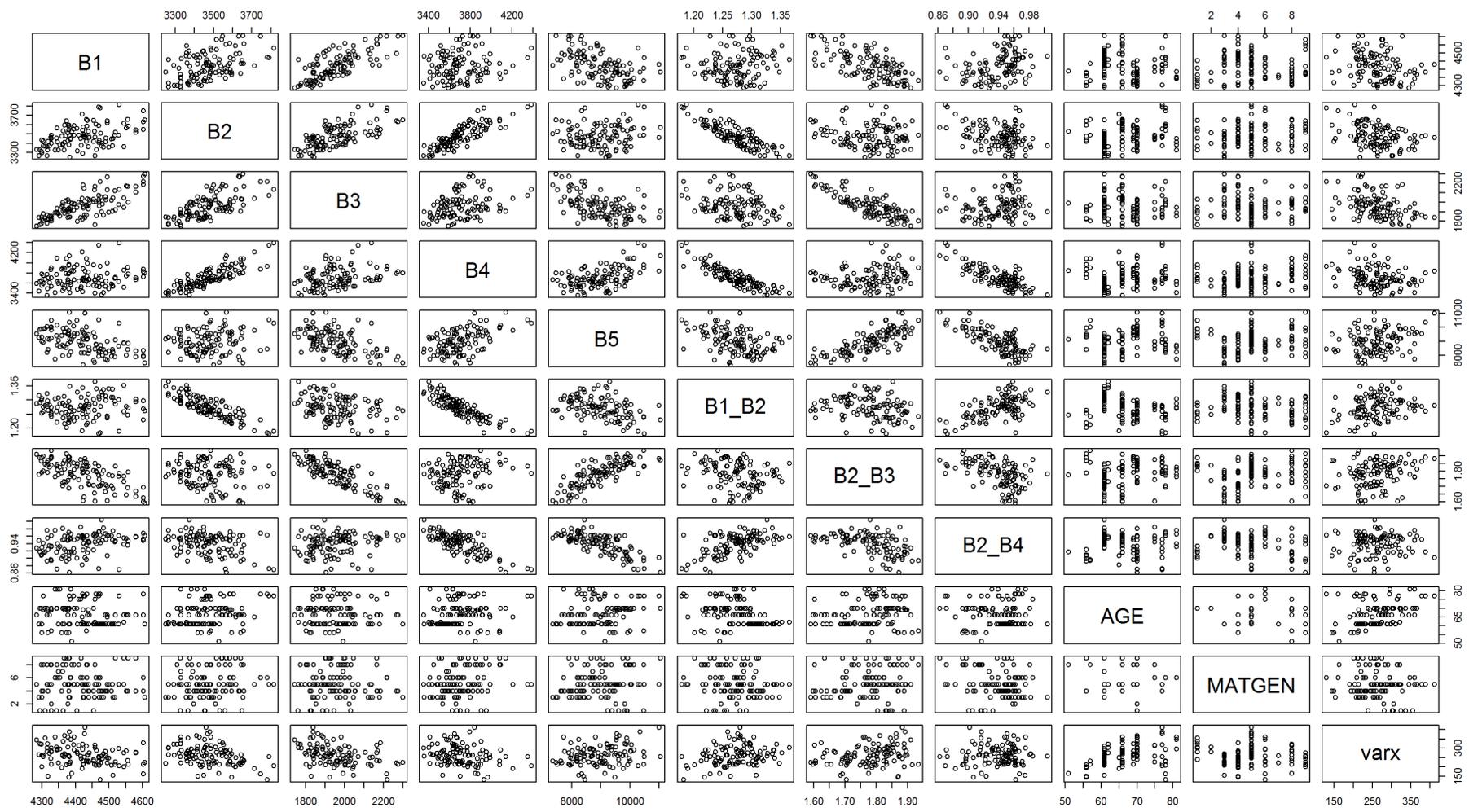


Figura 34 –Matriz de correlação das variáveis preditoras oriundas das imagens RAPIDEYE e submetidas ao processo de *Best Subset Selection*. Varx=VTCC

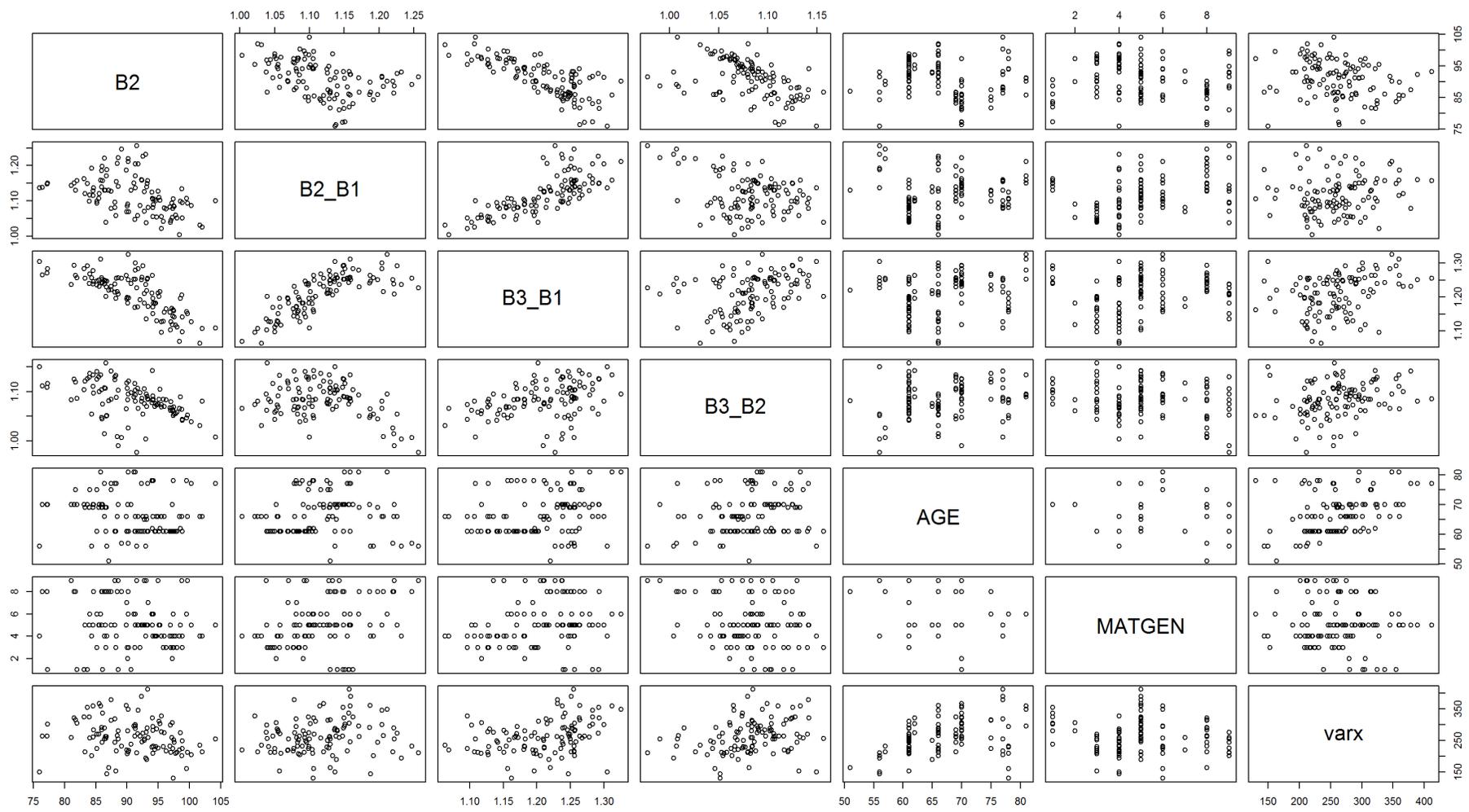


Figura 35 –Matriz de correlação das variáveis predictoras oriundas das imagens GEOEYE e submetidas ao processo de *Best Subset Selection*. Varx=VTCC

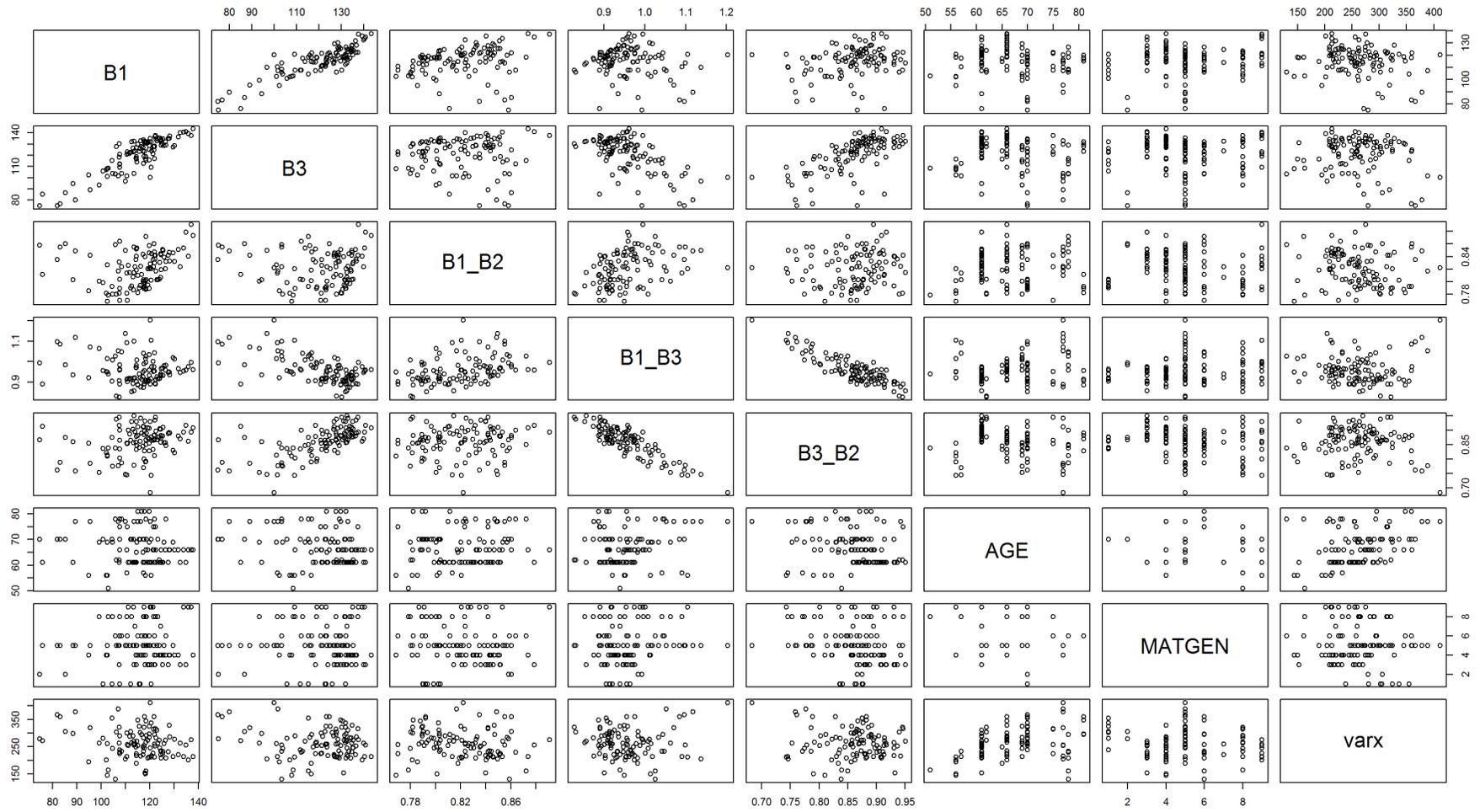


Figura 36 –Matriz de correlação das variáveis predictoras oriundas das imagens Aerotransportadas e submetidas ao processo de *Best Subset Selection*.

Varx=VTCC

Tabela 12 - Estatísticas do ajuste de RLM para o conjunto de dados LANDSAT8. *Valores de R² Ajustado e RMSE% obtidos a partir da validação cruzada LOOCV

Variável	Parâmetro (β_i)	Erro Padrão	R² Ajustado*	RMSE%*
<i>Intercepto</i>	-5101,42	733,92		
<i>B4_B3</i>	-3266,54	630,32		
<i>B5_B4</i>	99,13	36,93		
<i>B9_B3</i>	2883,68	354,92		
<i>AGE</i>	3,75	0,67		
<i>MATGEN2</i>	-26,31	25,22		
<i>MATGEN3</i>	-55,41	16,54	0,58	13,31
<i>MATGEN4</i>	-45,76	15,43		
<i>MATGEN5</i>	6,64	13,28		
<i>MATGEN6</i>	-50,21	18,35		
<i>MATGEN7</i>	-67,16	26,34		
<i>MATGEN8</i>	-22,35	14,57		
<i>MATGEN9</i>	-34,03	16,47		

Tabela 13 - Estatísticas do ajuste de RLM para o conjunto de dados RAPIDEYE. *Valores de R² Ajustado e RMSE% obtidos a partir da validação cruzada LOOCV

Variável	Parâmetro (β_i)	Erro Padrão	R² Ajustado*	RMSE%*
<i>Intercepto</i>	-383,44	237,23		
<i>B1</i>	-0,10	0,04		
<i>B1_B2</i>	621,92	92,18		
<i>AGE</i>	4,60	0,69		
<i>MATGEN2</i>	-17,38	26,98		
<i>MATGEN3</i>	-50,16	17,39		
<i>MATGEN4</i>	-43,81	16,14	0,50	14,55
<i>MATGEN5</i>	-6,27	14,64		
<i>MATGEN6</i>	-83,93	17,32		
<i>MATGEN7</i>	-47,79	27,79		
<i>MATGEN8</i>	-21,23	15,74		
<i>MATGEN9</i>	-37,40	17,92		

Tabela 14 - Estatísticas do ajuste de RLM para o conjunto de dados GEOEYE. *Valores de R² Ajustado e RMSE% obtidos a partir da validação cruzada LOOCV

Variável	Parâmetro (β_i)	Erro Padrão	R² Ajustado*	RMSE%*
<i>Intercepto</i>	6687,94	1783,04		
<i>B2</i>	-76,73	19,69		
<i>B3</i>	70,57	18,07		
<i>B3_B2</i>	-6115,60	1633,98		
<i>AGE</i>	3,72	0,78		
<i>MATGEN2</i>	8,53	31,52		
<i>MATGEN3</i>	-40,65	20,53	0,37	16,39
<i>MATGEN4</i>	-32,20	18,63		
<i>MATGEN5</i>	0,11	17,10		
<i>MATGEN6</i>	-79,24	20,20		
<i>MATGEN7</i>	-38,81	32,11		
<i>MATGEN8</i>	-24,40	18,42		
<i>MATGEN9</i>	-41,32	20,34		

Tabela 15 - Estatísticas do ajuste de RLM para o conjunto de dados Aerotransportados. *Valores de R² Ajustado e RMSE% obtidos a partir da validação cruzada LOOCV

Variável	Parâmetro (β_i)	Erro Padrão	R² Ajustado*	RMSE%*
<i>Intercepto</i>	-1073,81	470,31		
<i>B1</i>	-11,24	4,29		
<i>B3</i>	-11,04	4,24		
<i>B1_B3</i>	1166,91	498,02		
<i>AGE</i>	3,20	0,97		
<i>MATGEN2</i>	-4,15	36,17		
<i>MATGEN3</i>	-54,43	22,61	0,32	16,96
<i>MATGEN4</i>	-50,92	20,27		
<i>MATGEN5</i>	-4,00	18,79		
<i>MATGEN6</i>	-78,20	21,44		
<i>MATGEN7</i>	-40,68	35,34		
<i>MATGEN8</i>	-32,99	20,29		
<i>MATGEN9</i>	-47,37	23,05		

Tabela 16 – Seleção de modelos geoestísticos candidatos da KO. RMSE%: Raíz do erro quadrático médio em porcentagem; ME: Erro médio; COR1: Correlação de Pearson (Observado x Estimado) e COR2: Correlação de Pearson (Estimado x Resíduo)

Modelo	RMSE%	ME	COR1	COR2
Exp	13,89	0,42	0,74	0,11
Bes	14,63	0,51	0,70	0,10
Pow	14,89	0,42	0,74	0,11
Log	15,24	0,34	0,68	0,17
Pen	15,69	0,47	0,64	0,08
Sph	16,11	0,44	0,62	0,06
Cir	16,71	0,40	0,58	0,02
Gau	17,32	0,49	0,53	-0,04
Lin	18,40	0,27	0,45	-0,12
Per	20,63	0,00	-1,00	-1,00

Tabela 17 – Seleção de modelos geoestísticos candidatos da KDE (LANDSAT8). RMSE%: Raíz do erro quadrático médio em porcentagem; ME: Erro médio; COR1: Correlação de Pearson (Observado x Estimado) e COR2: Correlação de Pearson (Estimado x Resíduo)

Modelo	RMSE%	ME	COR1	COR2
Bes	12,30	0,19	0,80	-0,05
Pen	12,41	0,05	0,80	-0,05
Pow	12,43	0,10	0,79	-0,05
Exp	12,43	0,10	0,79	-0,05
Sph	12,57	-0,05	0,79	-0,06
Cir	12,67	-0,09	0,79	-0,06
Lin	12,84	-0,12	0,78	-0,07
Log	12,85	-0,04	0,78	-0,07
Gau	12,96	0,21	0,77	-0,07
Per	13,31	-0,12	0,76	-0,09

Tabela 18 – Seleção de modelos geoestísticos candidatos da KDE (RAPIDEYE). RMSE%: Raíz do erro quadrático médio em porcentagem; ME: Erro médio; COR1: Correlação de Pearson (Observado x Estimado) e COR2: Correlação de Pearson (Estimado x Resíduo)

Modelo	RMSE%	ME	COR1	COR2
Bes	13,41	0,07	0,76	-0,08
Pow	13,44	0,07	0,76	-0,08
Exp	13,44	0,07	0,76	-0,08
Gau	13,46	-0,09	0,75	-0,08
Pen	13,47	-0,06	0,75	-0,08
Sph	13,50	-0,08	0,75	-0,08
Cir	13,54	-0,09	0,75	-0,08
Log	13,76	-0,05	0,74	-0,09
Lin	14,45	-0,14	0,71	-0,13
Per	19,52	-0,69	0,55	-0,49

Tabela 19 – Seleção de modelos geostísticos candidatos da KDE (GEOEYE). RMSE%: Raíz do erro quadrático médio em porcentagem; ME: Erro médio; COR1: Correlação de Pearson (Observado x Estimado) e COR2: Correlação de Pearson (Estimado x Resíduo)

Modelo	RMSE%	ME	COR1	COR2
Pow	14,34	0,16	0,71	-0,06
Exp	14,34	0,16	0,71	-0,06
Bes	14,44	0,11	0,71	-0,07
Pen	14,82	0,11	0,69	-0,08
Log	14,94	0,10	0,68	-0,08
Sph	15,07	0,08	0,68	-0,10
Gau	15,14	0,07	0,68	-0,10
Cir	15,29	0,06	0,67	-0,11
Lin	15,73	0,01	0,65	-0,14
Per	16,39	-0,12	0,61	-0,15

Tabela 20 – Seleção de modelos geostísticos candidatos da KDE (Aerotransportada). RMSE%: Raíz do erro quadrático médio em porcentagem; ME: Erro médio; COR1: Correlação de Pearson (Observado x Estimado) e COR2: Correlação de Pearson (Estimado x Resíduo)

Modelo	RMSE%	ME	COR1	COR2
Gau	14,09	0,29	0,72	0,00
Bes	14,22	0,42	0,72	-0,01
Lin	14,24	0,40	0,72	-0,02
Sph	14,26	0,45	0,72	-0,01
Pen	14,27	0,45	0,72	-0,01
Exp	14,27	0,47	0,72	-0,01
Cir	14,28	0,44	0,72	-0,01
Pow	14,29	0,47	0,72	-0,01
Log	15,06	0,43	0,68	-0,03
Per	16,96	0,16	0,57	-0,13

APÊNDICE

APÊNDICE A -

Aplicação da KDE – Estudo dirigido

A seguir será demonstrado um estudo dirigido da aplicação da KDE, no ambiente computacional R, para a predição volumétrica em uma floresta jovem de *Eucalyptus* spp. na região centro-oeste do Brasil.

Obs¹: Algumas funções utilizadas ainda estão fase alfa/beta de desenvolvimento pelo autor, entre em contato para saber se existe alguma versão mais recente do material. São utilizados alguns arquivos hospedados online, no caso de indisponibilidade entre em contato com o autor – esthevan@florestal.eng.br.

Obs²: Os comandos a seguir servem como uma orientação e foram testados para a base disponibilizada, sendo necessário ajustes em diversas funções ao se trabalhar com outra base de dados. Portanto, o estudo considera que o usuário já possui conhecimentos básicos da utilização do software R.

Obs³: Todo material a seguir está protegido sob a licença Creative Commons CC BY-NC-SA 3.0.

1) ETAPA 1: PREPARAÇÃO DA BASE DE DADOS

Após a instalação do programa e de uma interface IDE, caso o usuário desejar, e conectado na internet, execute as funções abaixo. Os comandos possuem o objetivo de carregar todos os pacotes e funções necessárias para memória no computador.

```
> getwd() #diretório de trabalho, se desejar mudar utilize setwd()
> download.file("http://goo.gl/OtD4Gv", destfile="pacotes_funcoes.R") #download de rotinas
> source("pacotes_funcoes.R") #carrega na memória
```

Em seguida será realizado o download dos dados e das imagens que utilizaremos neste tutorial. O conjunto de funções abaixo é capaz de realizar essa tarefa. O material pode também ser baixado manualmente neste link: <<https://www.dropbox.com/s/92sepp9musku8at/dataset.zip?dl=0>>

```
> download.file("http://goo.gl/XBKKjj", destfile="dataset.zip", mode="wb") #Download do conjunto de dados
> unzip("dataset.zip", exdir = getwd()) #descompacta no diretório
```

A próxima etapa será carregar as imagens, shapes e dados necessários para o trabalho, para dentro do programa.

```
#IMAGENS
> Imagem <- brick('LS8_composite_15m.tif') #Le arquivo multispectral
#SHAPES
> shpTalhoes <- readShapePoly('Talhoes.shp') #Shape talhoes
> shpFazenda <- readShapePoly('Fazenda.shp') #Shape talhoes
> datum <- projection(Imagem) #Chamando Datum
> proj4string(shpTalhoes) <- datum #Adicionando Projeção
#TABELAS
> dfParcelas <- read.csv2("inventario_parcelas_final.csv")
```

Conhecendo os dados: O conjunto de função a seguir permite explorarmos um pouco mais os dados.

```
> head(dfParcelas) #observando os dados do inventário
> (Imagem) #observando estatísticas da imagem
> plotRGB(Imagem, r=6,g=4,b=3)
> plot(shpFazenda, add=T)
```

Output:

```
class       : RasterBrick
dimensions  : 691, 620, 428420, 7 (nrow, ncol, ncell, nlayers)
resolution  : 15, 15 (x, y)
extent      : 209711.9, 219011.9, 7767298, 7777663 (xmin, xmax, ymin, ymax)
coord. ref. : +proj=utm +zone=22 +south +ellps=GRS80 +units=m +no_defs
data source : in memory
names       : LS8_composite_15m.1, LS8_composite_15m.2, LS8_composite_15m.3,
LS8_composite_15m.4, LS8_composite_15m.5, LS8_composite_15m.6,
LS8_composite_15m.7
min values  :          0,          475,          1082,
5295,      8164,          5184,          5063
max values  :      11952,      10093,          9840,
18981,      12578,      21013,      19269
```

A imagem disponibilizada refere-se a uma imagem LANDSAT-8, georeferenciada, fusionada com 15m de resolução espacial e 7 bandas espectrais. As operações de georreferenciamento, fusionamento e composição colorida podem ser realizadas em um sistema de informação geográfica (SIG) ou no ambiente R, por meio de algoritmos mais avançados não demonstrados aqui. As imagens LS8 podem ser obtidas no site da USGS Earth Explorer <http://earthexplorer.usgs.gov/>.



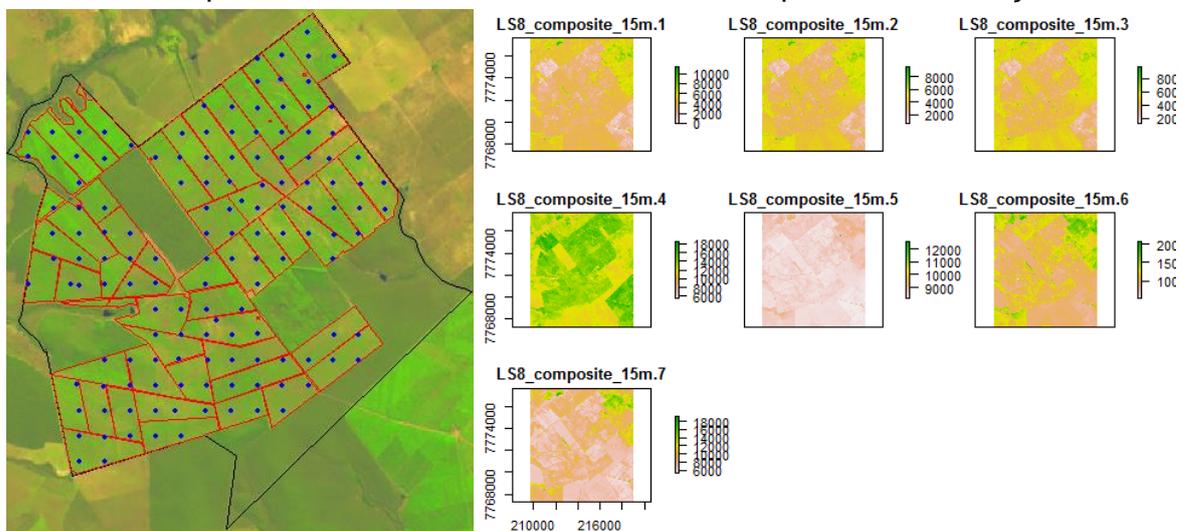
A visualização da imagem nos mostra que a nossa área de interesse (fazenda) representa apenas uma fração da imagem total. Dessa forma, o próximo passo será recortar a área de interesse, isso permite que o processamento se torne mais rápido.

```

#Recorte
> clip <- cborder(shpFazenda, buffer=100) #Cria um perimetro de 100 metros em torno da fazenda
> Imagem <- crop(Imagem, extent(clip)) #Recorta imagem com base no perimetro
#Visualização das imagens
> plotRGB(Imagem, r=6,g=4,b=3)
> plot(shpFazenda, add=T)
> plot(shpTalhoes, border="red", add=T)
> points(dfParcelas$x, dfParcelas$y, cex=0.5, pch=19, col="blue")
> plot(Imagem) #Observando cada banda espectral

```

Output: Como resultado podemos ver com mais detalhes a área de interesse, os talhões e a localização das parcelas do inventário florestal. Além disso, pode-se observar o comportamento de cada uma das bandas espectrais em relação ao alvo.



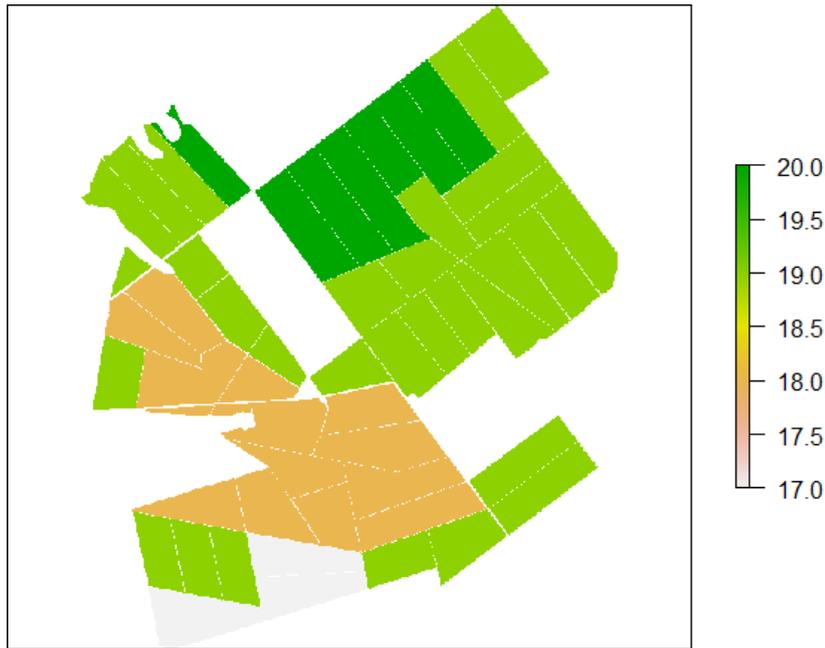
Em seguida recortaremos a imagem mais uma vez, porém agora utilizando o *shape* de talhões para o recorte. Como cada um dos talhões possui informações cadastrais de Idade, podemos utiliza-las para criar uma imagem das idades, isso é necessário ao se utilizar variáveis cadastrais na modelagem.

```

#Criando uma Lista de Imagens
> Img_list <- list(B1=Imagem[[1]], B2=Imagem[[2]], B3=Imagem[[3]], B4=Imagem[[4]], B5=Imagem[[5]],
B6=Imagem[[6]], B7=Imagem[[7]]) #criando base de dados
#Recorte
> Img_list <- cliprasteres(Img_list, shape=shpTalhoes, parallel=T) #recortando área de interesse
#Variável cadastral espacializada
> Img_idade <- shp2raster(shpTalhoes, Img_list[[1]], label="IDADE", invalue=shpTalhoes$IDADE)
> plot(Img_idade)

```

Output: Neste exemplo estamos utilizando apenas a idade como variável cadastral entretanto, o procedimento é o mesmo ao se utilizar material genético ou outra variável disponível no *shape* de talhões e que se relacione com a variável de interesse. A imagem abaixo demonstra a espacialização da idade (meses), seguindo o mesmo padrão de (resolução espacial e *datum*) fornecida pela imagem LS8.



Todo o procedimento de modelagem e interpolação acontece no formato tabular, ou seja na próxima etapa iremos converter as imagens no formato *raster* para tabelas.

```
#Criando nova lista de imagens
> Img_list <- list(B1=Img_list[[1]], B2=Img_list[[2]], B3=Img_list[[3]], B4=Img_list[[4]],
B5=Img_list[[5]], B6=Img_list[[6]], B7=Img_list[[7]], Idade=Img_idade)
#Conversão
> dfImagens <- raster2data(Img_list, pixel=c(15,15), parallel=T)
```

Output: Os resultados (abaixo) de saída das 6 primeiras linhas, deixam mais clara essa operação. Converteremos o conjunto com 8 imagens em uma tabela com as coordenadas x e y de cada *pixel*, bem como os valores de cada banda espectral e a variável cadastral idade na respectiva coordenada.

x	y	B1	B2	B3	B4	B5	B6	B7	Idade
211166.9	7768498	7868	7055	6881	15124	9549	15154	12953	17
211181.9	7768498	6608	6327	6518	13992	9055	14663	11857	17
211196.9	7768498	6055	5774	5965	13992	9055	14663	11857	17
211211.9	7768498	6736	6120	6118	14184	9290	15447	12704	17
211226.9	7768498	6942	6326	6324	14184	9290	15447	12704	17
211166.9	7768513	7868	7055	6881	15124	9549	15154	12953	17...

O formato tabular permite o cálculo de índices de vegetação de forma muito mais eficiente computacionalmente no ambiente R, ao se comparar com o formato raster. Em nosso próximo passo, iremos calcular a razão simples entre todas as combinações possíveis de banda, além do índice NDVI.

```
> dfImagens$B1[dfImagens$B1<1] <- 1 #Tratativa para alguns pixels com valor 0
> dfSR <- srcombine(dfImagens[,3:9]) #Simple Ratio combinações entre bandas
> dfFinal <- cbind(dfImagens[,],dfSR[,8:49]) #Juntando data.frames
> dfFinal$NDVI <- (dfFinal$B5-dfFinal$B4)/(dfFinal$B5+dfFinal$B4) #Índice de Vertação NDVI
> dfFinal <- na.omit(dfFinal) #Exclui as linhas com observações NA
> head(dfFinal)
```

Iremos agora extrair as informações de *pixel* que coincidam com as coordenadas das parcelas. Ou seja para cada uma das parcelas do inventário florestal teremos os valores da variável de interesse “varx” (vtcc – variável resposta) e as variáveis preditoras.

```
> img <- data2raster(dfFinal, datum, parallel=T)
> head(dfParcelas)
> dados <- extractpoints(img, dfParcelas$vtcc, dfParcelas$x,
                        dfParcelas$y, 20, parallel=T) #Extraindo Pixels
> head(dados)
```

Na parte final da ETAPA 1, iremos salvar os resultados em arquivos .csv, isso nos dará maior flexibilidade. De maneira que se quisermos reprocessar a ETAPA 2, não temos que processar novamente a ETAPA 1.

```
> write.csv2(dfFinal, 'database_15m_landsat.csv', row.names=F)
> write.csv2(dados, 'parcdatabase_20m_landsat.csv', row.names=F)
```

Concluimos essa fase com dois arquivos.

database_15m_landsat.csv: estão todas as informações das variáveis auxiliares em área total. Este arquivo será utilizado na próxima fase para a predição do volume em área total.

parcdatabase_20m_landsat.csv: estão as informações das parcelas, ou seja, coordenada das parcelas, variável de interesse e variáveis preditoras. Este arquivo é o *input* principal para a fase de modelagem estatística.

2) ETAPA 2: MODELAGEM E PREDIÇÃO

Iniciaremos limpando a memória do sistema e chamando novamente, apenas os arquivos de interesse. Entretanto essa etapa é optativa.

```
> rm(list=ls(all=TRUE)) #Remove qualquer coisa da memoria antes de começar;
> gc()
> download.file("http://goo.gl/OtD4Gv", destfile="pacotes_funcoes.R")
> source("pacotes_funcoes.R") #carrega na memória
```

Em seguida chamamos para a memória do sistema os arquivos de saída da ETAPA 1, bem como os *shapes* de talhões e fazenda.

```
> dfParc <- read.csv2('parcdatabase_20m_landsat.csv') #pontos das parcelas
> dfImag <- read.csv2('database_15m_landsat.csv', comment.char="", colClass="numeric") #df de imagens
> shpTalhoes <- readShapePoly('Talhoes.shp') #Shape talhoes
> shpFazenda <- readShapePoly('Fazenda.shp') #Shape talhoes
> datum <- "+proj=utm +zone=22 +south +ellps=GRS80 +units=m +no_defs"
> proj4string(shpTalhoes) <- datum
> proj4string(shpFazenda) <- datum
```

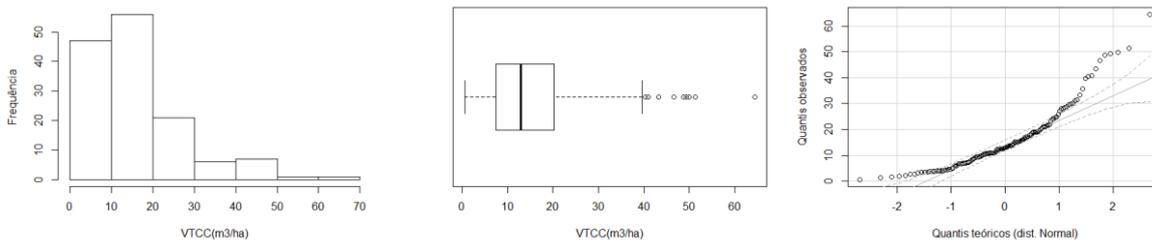
A seguir faremos uma breve análise exploratória da variável resposta vtcc, para ver se esta apresenta *outliers* e distribuição normal. A distribuição normal é uma das premissas da interpolação geoestística.

```

> hist((dfParc$varx), main="", xlab="VTCC(m3/ha)", ylab="Frequência")
> boxplot((dfParc$varx), horizontal=T, xlab="VTCC(m3/ha)")
> qqPlot((dfParc$varx), col.lines="darkgrey", lwd=1, distribution="norm", xlab="Quantis teóricos (dist. Normal)", ylab="Quantis observados")

```

Output: Os resultados da análise exploratória mostram a necessidade de normalização dos dados. Faremos a transformação em uma etapa um pouco mais a frente.



Iremos agora eliminar as variáveis resposta com elevada colinearidade, ou seja que apresentam coeficiente de correlação de *Pearson* superior ou inferior a 0,9 e -0,9, respectivamente. Será gerada também uma matriz de correlação entre as variáveis restantes.

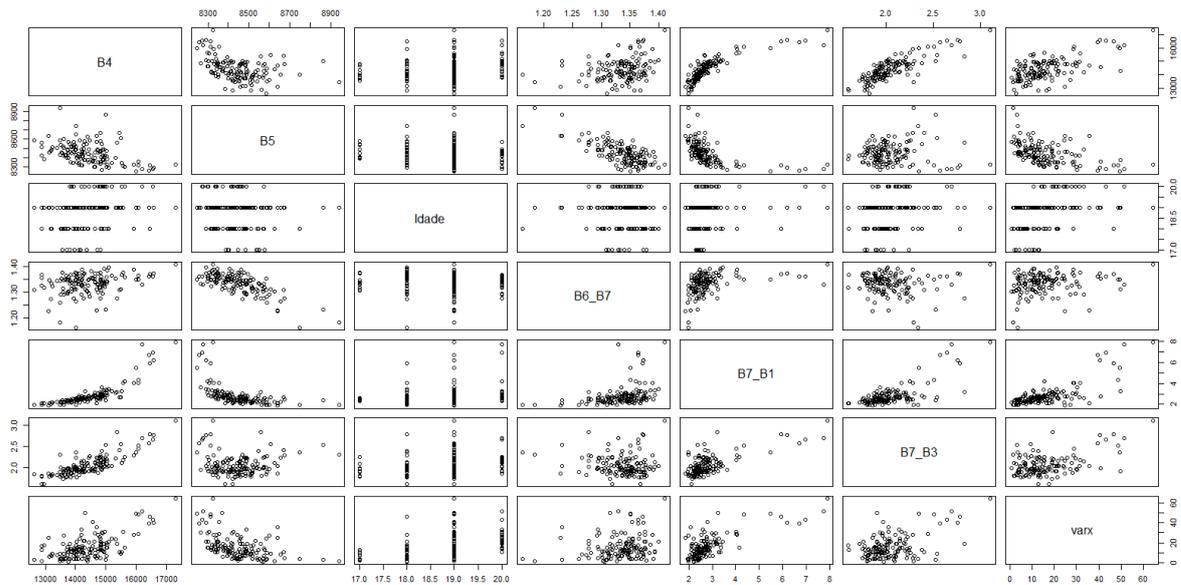
```

> dfRemov <- vifcor(dfParc[,c(-1:-2,-53)], th=0.90)
> dfParctp <- exclude(dfParc, dfRemov)
> dfParctp$x <- dfParc$x
> dfParctp$y <- dfParc$y
> dfParctp$varx <- dfParc$varx
> dfParc <- dfParctp
> head(dfParc)
> plot(dfParc[, -8:-9])

```

Output: Após a remoção inicial de variáveis com colinearidade restaram-se 6 variáveis preditoras, sendo elas: B4, B5, Idade, B6/B7, B7/B1 e B7/B3.

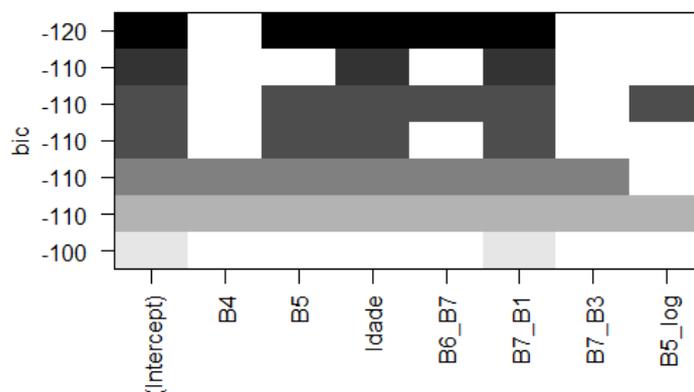
B4	B5	Idade	B6_B7	B7_B1	B7_B3	varx	x	y
12959.00	8381.800	19	1.319258	2.109283	1.602337	17.4	211728	7769731
14014.33	8654.500	19	1.265475	2.265744	2.225552	8.2	211719	7770212
13580.71	8517.286	19	1.316282	2.178491	1.866303	3.8	212227	7769231
13756.17	8396.667	19	1.352039	2.403964	1.844799	10.7	212224	7769731
13482.80	8433.200	19	1.344659	2.237567	1.783275	8.8	214218	7769731
14414.00	8440.000	19	1.348779	2.538949	2.029761	4.6	214722	7769729...



Com as variáveis restantes iniciaremos a etapa de seleção de modelos por meio de busca exaustiva.

```
> x.best <- regsubsets(varx~., data=dfParc[,c(-8:-9)], nbest=1, nvmax=20,
  method='exhaustive', really.big=F) # "exhaustive", "backward", "forward", "seqrep"
> plot(x.best)
```

Output: Na figura abaixo é apresentado os melhores modelos para cada tamanho de modelo possível, bem como o respectivo valor de BIC (*Bayesian Information Criterion*). O BIC é uma estatística de avaliação de modelos muito semelhante ao AIC (*Akaike Information Criterion*), onde valores menores indicam modelos melhores. Dessa forma, observa-se que o melhor modelo é aquele que utiliza os preditores B5, Idade, B6/B7 e B7/B1.



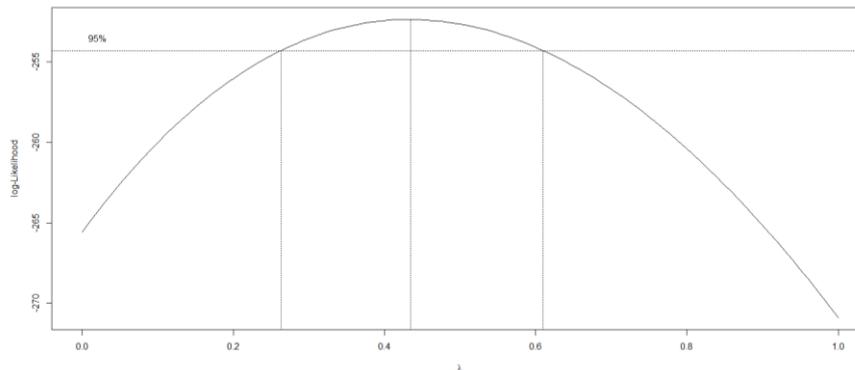
Definido as variáveis a serem utilizadas iremos ajustar um modelo linear múltiplo. O modelo servirá de base para a transformação de normalização Box-Cox.

```

> model <- as.formula('varx ~ B5 + Idade + B6_B7 + B7_B1')
> ajust <- lm(model, dfParc)
> summary(ajust)
> boxcox(ajust, plotit=T)
> boxcox(ajust, lam=seq(0, 1, 1/10))

```

Output: Por meio do gráfico abaixo, podemos observar que o valor ideal de lambda para transformação Box-Cox, gira em torno de 0,43.

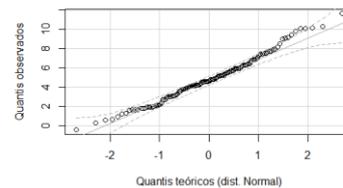
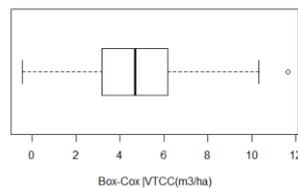
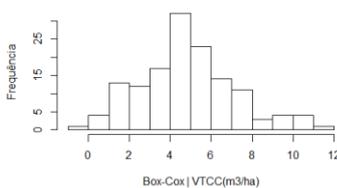


```

> lambda <- 0.43
> dfParc$varx <- (dfParc$varx^(lambda - 1))/lambda
> hist(dfParc$varx, main="", xlab="Box-Cox | VTCC(m3/ha)", ylab="Frequência")
> boxplot(dfParc$varx, horizontal=T, xlab="Box-Cox | VTCC(m3/ha)")
> qqPlot(dfParc$varx, col.lines="darkgrey", lwd=1, distribution="norm", xlab="Quantis teóricos (dist. Normal)", ylab="Quantis observados")

```

Output: Após a transformação podemos observar a presença de uma única observação discrepante. No exercício iremos removê-la entretanto tal procedimento deve ser analisado caso a caso.



Remoção de outlier e novo ajuste.

```

> grafico_bp <- boxplot(dfParc$varx)
> dfParc <- subset(dfParc, dfParc$varx < grafico_bp$out)
> ajust <- lm(model, dfParc)
> summary(ajust)

```

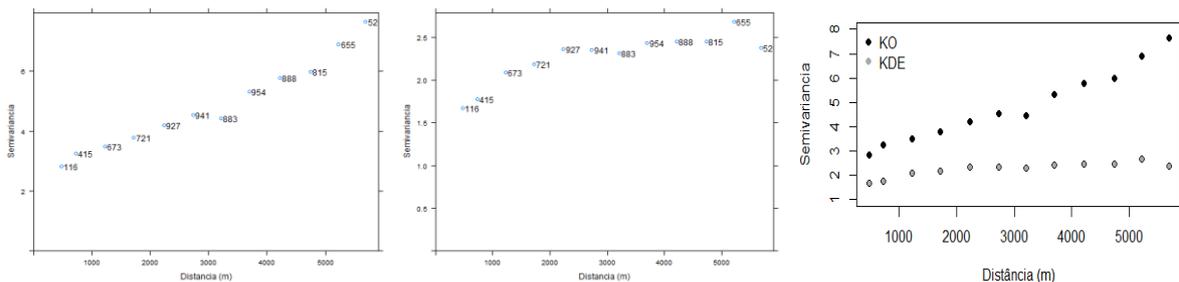
Iniciaremos agora a fase de modelagem geoestística. Para tanto precisamos converter os arquivos *data.frame* em arquivos espaciais.

```
#Conversão
> coordinates(dfParc) <- ~ x + y
> coordinates(dfImg) <- ~ x + y
> gridded(dfImg) <- T
> proj4string(dfParc) <- datum #Adicionando Projeção
> proj4string(dfImg) <- datum #Adicionando Projeção
#Recorte
> t <- shpTalhoes
> grd <- overlay(dfImg, t)
> dfImg <- dfImg[!is.na(grd),]
```

Faremos uma comparação do semivariograma da KO e da KDE utilizando as variáveis auxiliares.

```
#Ordinary Kriging
> svpointko <- variogram(varx ~ 1, data=dfParc, cutoff=6000, width=500,
  cressie=F); plot(svpointko, pl=T,xlab='Distancia (m)',ylab='Semivariância')
#Kriging with external drift
> svpoint <- variogram(model, data=dfParc, cutoff=6000, width=500,
  cressie=F); plot(svpoint, pl=T,xlab='Distancia (m)',ylab='Semivariância')
#Comparando
> plot(svpointko$dist, svpointko$gamma, bg="black", pch=21,
  ylim=c(1,8), xlab="Distância (m)", ylab="Semivariância")
> points(svpoint$dist, svpoint$gamma, bg="darkgrey", pch=21)
> legend("topleft", legend=c("KO", "KDE"), col=c('black','darkgrey'),
  bty="n", pch=c(19,19))
> var_ko <- mean(svpointko$np*svpointko$gamma);var_ko
> var_kde <- mean(svpoint$np*svpoint$gamma);var_kde
> 100-(var_kde/var_ko*100) #redução na variância
```

Output: Observa-se uma redução na variância na ordem de 53% com a adoção das variáveis auxiliares.



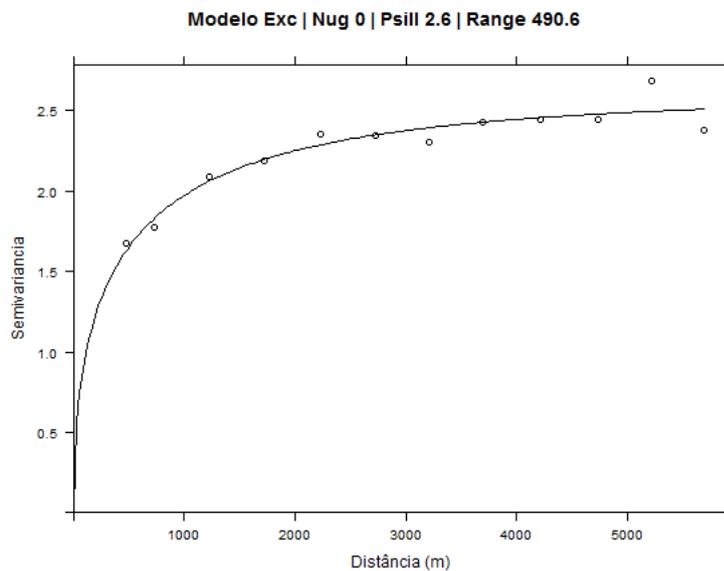
Na sequência ajustaremos vários modelos geoestísticos para o variograma da KDE e escolheremos o melhor com base na LOOCV. A função *bestfit* testa todos modelos implementados no pacote *gstat*, com diferentes métodos de ajuste, e realiza uma LOOCV para sugerir qual o melhor modelo.

```
#Ajuste visual de parametros
> Nug <- min(svpoint$gamma)*0.6 #Efeito Pepita
> Range <- mean(svpoint$dist)*1.4 #Alcance da Dependencia Espacial
> Psill <- mean(svpoint$gamma)*1.4 #SemiVariância Maxima
> models <- bestfit(model=model, variogram=svpoint, data=dfParc, autostart=T, parallel=T)
```

No objeto *models* podem ser acessadas as estatísticas de ajuste de cada modelo testado. Conforme informação de *output* ficaremos com o modelo Exc com o método de ajuste 2.

```
> x<-1 #primeiro modelo da tabela models
> Mods <- as.character(models$model[x]) # Seleção do melhor modelo
> parmol <- vgm(Psill,Mods,Range,Nug) #Modificar Modelo Aqui
> plot(svpoint, pl=F, model=parmol, xlab='Distancia (m)',ylab='Semivariância')
> parfit <- fit.variogram(svpoint, parmol, fit.method=paste(models$metodo[x],sep=""))
> plot(svpoint, pl=F, model=parfit, main=paste('Modelo ',parfit[2,1],' | Nug ',round(parfit[1,2],1),' | Psill ',
round(parfit[2,2],1),' | Range ', round(parfit[2,3],1), sep=""),
xlab='Distância (m)',ylab='Semivariância', col="black")
```

Output: Podemos observar o resultado do semivariograma ajustado.



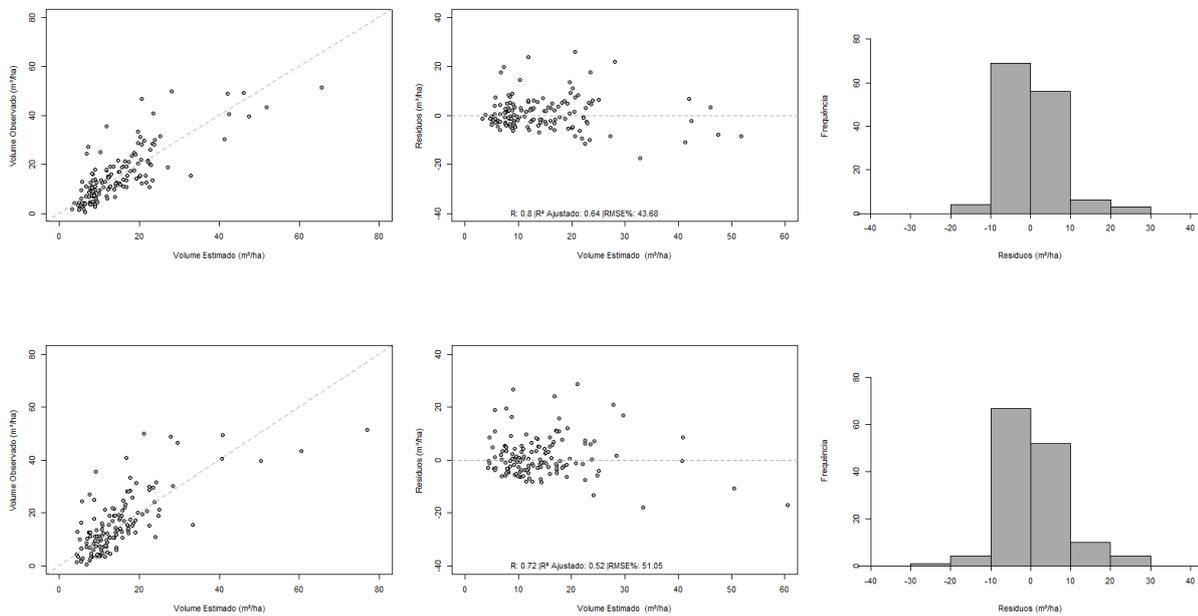
Iremos agora executar uma série de funções com o intuito de fazer uma análise comparativa entre a LOOCV utilizando RLM e KDE.

```

#ESTADÍSTICAS KDE
> cros_val <- krige.cv(model, dfParc, nfold=nrow(dfParc),
  model=parfit) #Aplica a krigagem
> cros_val <- as.data.frame(cros_val)
> cros_val$var1.pred = (cros_val$var1.pred*lambda+1)^(1/lambda)
> cros_val$observed = (cros_val$observed*lambda+1)^(1/lambda)
> cros_val$residual = cros_val$observed-cros_val$var1.pre
> err_kde <- sqrt(mean(cros_val$residual^2)); err_kde #RMSE (precisão)
> err_perc_kde <- err_kde/mean(cros_val$observed)*100; err_perc_kde# Erro em porcentagem
> ajust_kde <- lm(observed~var1.pred, cros_val)
> r_kde <- cor(cros_val$var1.pred,cros_val$observed); r_kde
> r2_kde <- summary(ajust_kde)[[9]]; r2_kde
> write.csv2(cros_val, 'crossval_landsat.csv', row.names=F)
#ESTADÍSTICAS RLM
> temp <- loocv.lm(model, dfParc)
> head(temp)
> temp$varx = (temp$varx*lambda+1)^(1/lambda)
> temp$loocv = (temp$loocv*lambda+1)^(1/lambda)
> err_rlm <- sqrt(mean((temp$varx-temp$loocv)^2)); err_rlm #RMSE (precisão)
> err_perc_rlm <- err_rlm/mean(temp$varx)*100; err_perc_rlm# Erro em porcentagem
> ajust_rlm <- lm(varx~loocv, temp)
> r_rlm <- cor(temp$loocv,temp$varx); r_rlm
> r2_rlm <- summary(ajust_rlm)[[9]]; r2_rlm
#Graficos KDE
#Plot 1
> plot(cros_val$var1.pred,cros_val$observed, xlab="Volume Estimado (m³/ha)", ylab="Volume Observado
(m³/ha)", xlim=c(0,80), ylim=c(0,80), pch=21, bg="darkgray")
> abline(0,1, col="darkgrey", lty=2)
#Plot 2
> plot(cros_val$residual~cros_val$var1.pred, ylim=c(-40,40), xlim=c(0,60),xlab="Volume Estimado
(m³/ha)", ylab="Resíduos (m³/ha)", pch=21, bg="darkgray")
> abline(h=0, col="darkgrey", lty=2)
> legend("bottomleft", legend=c(paste("R:", round(r_kde,2), "| R² Ajustado:", round(r2_kde,2), "| RMSE%:",
round(err_perc_kde ,2))),bty="n")
#Plot 3
> xhist <- c(-40, -30, -20, -10, 0, 10, 20, 30, 40)
> hist(cros_val$residual, col="darkgrey", xlab="Resíduos (m³/ha)", ylab="Frequência",
  ylim=c(0,90), main="", xlim=c(-40,40), axes=F, breaks=xhist)
> axis(1,at = seq(-150,150,10),labels = TRUE,pos = 0); axis(2,pos = -40)
#Graficos RLM
#Plot 1
> plot(temp$loocv,temp$varx, xlab="Volume Estimado (m³/ha)", ylab="Volume Observado (m³/ha)",
xlim=c(0,80), ylim=c(0,80), pch=21, bg="darkgray")
> abline(0,1, col="darkgrey", lty=2)
#Plot 2
> plot((temp$varx-temp$loocv)~temp$loocv, ylim=c(-40,40), xlim=c(0,60),xlab="Volume Estimado
(m³/ha)", ylab="Resíduos (m³/ha)", pch=21, bg="darkgray")
> abline(h=0, col="darkgrey", lty=2)
> legend("bottomleft", legend=c(paste("R:", round(r_rlm,2), "| R² Ajustado:", round(r2_rlm,2), "| RMSE%:",
round(err_perc_rlm ,2))),bty="n")
#Plot 3
> xhist <- c(-40, -30, -20, -10, 0, 10, 20, 30, 40)
> hist((temp$varx-temp$loocv), col="darkgrey", xlab="Resíduos (m³/ha)", ylab="Frequência",
  ylim=c(0,90), main="", xlim=c(-40,40), axes=F, breaks=xhist)
> axis(1,at = seq(-150,150,10),labels = TRUE,pos = 0); axis(2,pos = -40)

```

Output: Observa-se que a KDE (em cima) apresentou resultados superiores em relação a RLM (em baixo).



A sequência da análise se ocorre pela aplicação do modelo geostatístico em área total.

```
> krig_kde <- krige(model,locations=dfParc, newdata=dfImg, model=parfit) #Aplicando a Krigagem KDE
> Predict <- as.data.frame(krig_kde)
> dfImg$VOL_KDE <- (Predict$var1.pred*lambda+1)^(1/lambda)
> test <- over(dfImg, shpTalhoes)
> dfImg$CD_TAL <- test$CODTALHAO
> dfImg$AREA <- test$Hectares
> gridded(dfImg)=TRUE #converções
> proj4string(dfImg) <- datum #Adicionando Projeção
#Preparando mapa
> KDE_MAP <- raster(dfImg[, 'VOL_KDE']) #converções
> KDE_MAP[KDE_MAP < 0] <- 0
> KDE_MAP[KDE_MAP > 80] <- 80
> ATOTAL <- round(sum(shpTalhoes$AREA),2)
> VTKDE <- round(mean(as.vector(KDE_MAP), na.rm = T)*ATOTAL,2)
> VMKDE <- round(mean(as.vector(KDE_MAP), na.rm = T),2)
> RMSE_KDE <- round(err_perc_kde, 2)
> TIPO_KDE <- "KDE-Landsat"
> MODELO_KDE <- "Modelo: VOL~f(B5,Idade,B6/B7,B7/B1) | KDE: Exc"
```

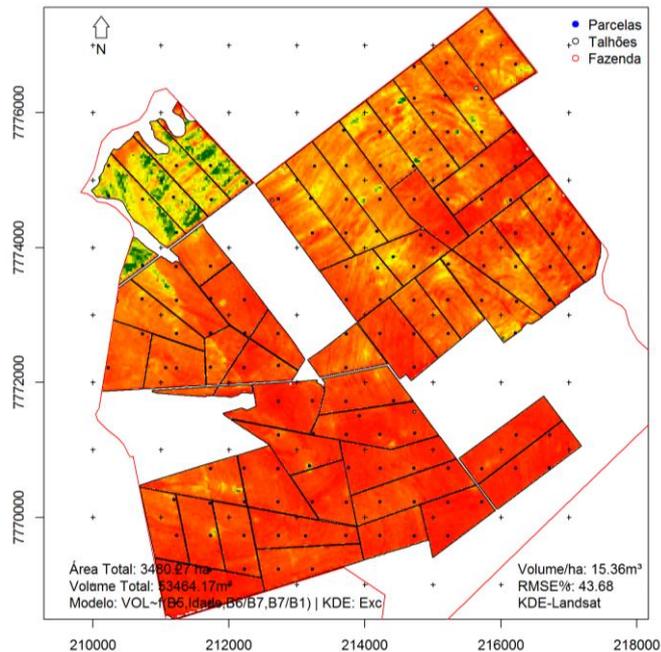
Após a aplicação em área total pode-se gerar um mapa da distribuição do volume em área total.

```

> writeRaster(KDE_MAP,"kde_landsat.tif","GTiff", overwrite=TRUE) #salva .tif (Abrir em SIG)
> MAPA <- KDE_MAP
> my_palette <- colorRampPalette(c("red", "yellow", "darkgreen"))(n = 50) #KDE
> plot(MAPA, legend=TRUE, col=my_palette, main=paste(" Predição Volume/ha | Landsat 8"),asp=1,
zlim=c(0,80))
> par(bg="transparent")
> yb <- (extent(MAPA)[3]+(extent(MAPA)[4]-extent(MAPA)[3])*0.95)
> xb <- (extent(MAPA)[1]+(extent(MAPA)[2]-extent(MAPA)[1])*0.02)
> north.arrow(xb=xb, yb=yb, len=80, lab="N")
> plot(shpTalhoes,border='black',col='transparent',add=TRUE, asp=1)
> plot(shpFazenda,border='red',col='transparent',add=TRUE, asp=1)
> plot(dfParc, bg='blue', cex=0.5, pch=21, asp=1, add=T)
> x_m <- seq(signif(xmin(MAPA),3),signif(xmax(MAPA),3),1000) #Montando GRID
> y_m <- seq(signif(ymin(MAPA),4),signif(ymax(MAPA),4),1000) #Montando GRID
> grid_plot <- expand.grid(x=x_m,y=y_m) #Montando GRID
> points(grid_plot$x,grid_plot$y,col='black', pch=3, asp=1, cex=0.5)
> legend("topright", legend=c("Parcelas","Talhões","Fazenda"), col=c('blue','black',"red"), bty="n",
pch=c(19,21,21))
> legend("bottomright", legend=c(paste("Volume/ha: ",VMKDE,"m³",sep=""),paste("RMSE%:
',RMSE_KDE,sep=""), TIPO_KDE), bty="n")
> legend("bottomleft", legend=c(paste("Área Total: ',ATOTAL, " ha",sep=""), paste("Volume Total:
',VTKDE,"m³",sep=""), MODELO_KDE), bty="n")

```

Output: A figura abaixo demonstra a aplicação da KDE na área total.



Para finalizar iremos calcular os valores de erro amostral utilizando a AAS e o estimador de regressão com base na variável auxiliar gerada pela KDE.

```

> DADOS <- as.data.frame(dfImg); DADOS <- DADOS[complete.cases(DADOS), ]
> dftotal <- DADOS; dfParc <- cros_val
#AAS
> med_acs <- mean(dfParc$observed)
> Sy2=var(dfParc$observed); Sy2; #Variância de volume/ha (m3/ha)^2
> Sy=sqrt(Sy2); Sy; #Desvio-padrão do volume/ha (m3/ha)
> med=mean(dfParc$observed); med; #Média de volume/ha (m3/ha)
> cv_perc=Sy/med*100; cv_perc; #Coeficiente de variação em % (CV%)
> n=length(dfParc$observed); n; #Conta o numero de parcelas utilizadas (U.N.)
> N=ATOTAL/(500/10000); N #Numero de parcela possiveis
> erro_pad_med=(Sy/sqrt(n))*sqrt(1-(n/N));erro_pad_med #Erro-padrão da média com correção para
populações finitas (m3)
> ns=0.05; #Nível de Significância;
> erro_unid=abs(qt(ns/2,n-1)*erro_pad_med); erro_unid; #Erro na unidade (m3)/ha
> erro_perc=erro_unid/med*100; erro_perc; #Erro do Inventario em porcentagem (%)
> volume_total=med*(ATOTAL);volume_total #Volume total do Fragmento = (média volume/ha)*(16.06 ha)
(m3)
> erro_frag=erro_unid*(ATOTAL); erro_frag #Erro total do Inventário para toda a área
> ICI=volume_total-erro_frag; ICI #Intervalo de confiança de todo o inventário(Limite Inferior)(m3)
> ICs=volume_total+erro_frag; ICs #Intervalo de confiança de todo o inventário(Limite Superior)(m3)
#ESTIMADOR DE REGRESSAO (ER)
> ux <- VMKDE; ux #média populacional
> N <- ATOTAL/(500/10000); N #tamanho populacional
> n <- length(dfParc$observed); n #tamanho da amostra
> u_est_y <- mean(dfParc$observed); u_est_y # média amostral da variavel de interesse
> u_est_x <- mean(dfParc$var1.pred); u_est_x # média amostral da variavel da medida auxiliar
> ajust <- lm(dfParc$observed~dfParc$var1.pred) #ajuste do Beta
> summary(ajust)
> ul <- u_est_y+coef(ajust)[2]*(ux-u_est_x); ul #estimador de regressao para o volume ## (MEDIA/HA)
> var_pop <- (sum((dfParc$observed-ul)^2)-coef(ajust)[2]^2*sum((dfParc$var1.pred-u_est_x)^2))/(n-2);
> var_pop #estimativa da variancia populacional
> var_red <- (var_pop/n)*(1-(n/N)); var_red #variancia da estimativa de regressao
> t <- abs(qt(ns/2,n-1))
> ic_m3 <- t*sqrt(var_red); ic_m3 #intervalo de confiança da média (+ou-)
> ic_perc <- ic_m3/ul*100; ic_perc
> v_tot <- ul*ATOTAL
> er_tot <- ic_m3*ATOTAL
#SUFICIENCIA AMOSTRAL
> nnew=((qt(ns,n-1)^2)*(cv_perc^2))/((ic_perc^2)+(qt(ns,n-1)^2*cv_perc^2)/N)
> nnew - n #número de parcelas a mais necessárias
> mtd <- c("AAS", "ER-KDE")
> vol_ha <- c(med, ul)
> vol_tot <- c(volume_total, v_tot)
> erro_p <- c(erro_perc, ic_perc)
> resumo <- data.frame(metodo=mtd, vol_ha=vol_ha, vol_total=vol_tot, erro_amostral=erro_p)
> rownames(resumo) <- c()
> resumo

```

Output: Os resultados finais demonstram que houve uma redução de 12% para 7% no erro amostral utilizando o estimador de regressão + KDE em comparação com a AAS, permitindo uma economia de 142 parcelas.

	metodo	vol_ha	vol_total	erro_amostral
1	AAS	15.57971	54221.60	12.070238
2	ER-KDE	16.00601	55705.24	7.092861