

**UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CÂMPUS DE JABOTICABAL**

**DADOS AEROMAGNÉTICOS COMO COVARIÁVEL AMBIENTAL NO
PLANEJAMENTO AGRÍCOLA**

Thiago de Andrade Águas
Geógrafo

2020

**UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CÂMPUS DE JABOTICABAL**

**DADOS AEROMAGNÉTICOS COMO COVARIÁVEL AMBIENTAL NO
PLANEJAMENTO AGRÍCOLA**

Discente: Thiago de Andrade Águas

Orientador: Prof. Dr. Gener Tadeu Pereira

Coorientador: Dr. Diego Silva Siqueira

Coorientador: Prof. Dr. Tom Vanwallegem

Tese apresentada à Faculdade de Ciências Agrárias e Veterinárias – Unesp, Campus de Jaboticabal, como parte das exigências para a obtenção do título de Doutor em Agronomia (Ciência do Solo).

A282d Águas, Thiago de Andrade
 Dados aeromagnéticos como covariável ambiental no planejamento agrícola / Thiago de Andrade Águas. -- Jaboticabal, 2020
 101 f. : tabs., mapas

 Tese (doutorado) - Universidade Estadual Paulista (Unesp),
Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal
 Orientador: Gener Tadeu Pereira
 Coorientador: Diego Silva Siqueira

 1. Magnetismo. 2. Geologia. 3. Análise espacial (Estatística). 4.
Aprendizado do computador. 5. Agricultura de precisão. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.



UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal



CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: DADOS AEROMAGNÉTICOS COMO COVARIÁVEL AMBIENTAL NO PLANEJAMENTO AGRÍCOLA

AUTOR: THIAGO DE ANDRADE ÁGUAS
ORIENTADOR: GENER TADEU PEREIRA
COORDENADOR: TOM VANWALLEGHEM
COORDENADOR: DIEGO SILVA SIQUEIRA

Aprovado como parte das exigências para obtenção do Título de Doutor em AGRONOMIA (CIÊNCIA DO SOLO), pela Comissão Examinadora:

Prof. Dr. GENER TADEU PEREIRA
 Departamento de Ciências Exatas / FCAV / UNESP - Jaboticabal

Prof. Dr. RAFAEL MONTANARI (VIDEOCONFERÊNCIA)
 Departamento de Fitossanidade, Engenharia Rural e Solos-FEIS / Ilha Solteira/SP

Prof. Dr. MARCÍLIO VIEIRA MARTINS FILHO
 Departamento de Solos e Adubos / FCAV / UNESP - Jaboticabal

Dr. DIOGO MAZZA BARBIERI
 Athenas Consultoria Agrícola e Laboratório / Jaboticabal/SP

Profa. Dra. TERESA CRISTINA TARLE PISSARRA
 Engenharia Rural / FCAV / UNESP - Jaboticabal

Jaboticabal, 15 de janeiro de 2020

DADOS CURICULARES DO AUTOR

THIAGO DE ANDRADE ÁGUAS – Nasceu no dia 07 de novembro de 1988, no município de Marília, São Paulo. Graduado em Geografia pela Universidade Estadual de Maringá – UEM (2013). Mestre em Geografia (Dinâmica Ambiental e Planejamento) pela Universidade Federal de Mato Grosso do Sul – UFMS/CPTL (2015) (Título da dissertação: Análise ambiental nas áreas de influência direta e indireta da PCH Costa Rica, no alto curso da Bacia Hidrográfica do Rio Sucuriú – MS). Iniciou o doutorado em 2016 no Programa de Pós Graduação em Agronomia (Ciência do Solo/Bolsa CAPES) da Universidade Estadual Paulista – UNESP/FCAV (Título da tese: Dados aeromagnéticos como covariável ambiental no planejamento agrícola). O primeiro resumo da tese publicado em 2018 no International Meeting of Agrarian Science and Technology - IMAST recebeu menção honrosa como segundo colocado (Título: Estimation of soil attribute from Aeromagnetic covariable). Realizou o período de doutorado sanduíche (Bolsa CAPES/PDSE - Processo nº 88881.187336/2018-01) de novembro/18 à abril/19 na Universidad de Córdoba, Espanha, sob supervisão do Prof. Dr. Tom Vanwalleguem. Atua na área de agronomia com ênfase em mineração de dados, sensoriamento remoto, mapeamento digital de solos, pedometria e geoestatística.

“Alguns homens vêem as coisas como são, e dizem ‘Por quê?’. Eu sonho com as coisas que nunca foram e digo ‘Por que não?’ ”

George Bernard Shaw

A meu avô Andrade (*in memoriam*), por enxergar sempre o melhor das pessoas, em todos os sentidos.

DEDICO

AGRADECIMENTOS

A minha família, pelo apoio incondicional ao longo desses anos.

Ao meu orientador Gener Tadeu Pereira, pelo acompanhamento e orientação.
Ao Diego Siqueira (coorientador), pela amizade, ideias e incentivos.

Ao Tom Vanwallegem (coorientador), Juan Vicente Giráldez Cervera, Andrea Román Sánchez e equipe do grupo de pesquisa *Hidrología e Hidráulica Agrícola/UCO* pela preciosa contribuição durante a elaboração da tese e companheirismo durante a estadia em Córdoba.

Ao Amaro Luiz Ferreira (CPRM), Eder Cassola Molina (IAG/USP) e Roberto Paulo Zanon dos Santos (IAG/USP) por valiosas contribuições como: discussões, identificação dos dados, métodos de manipulação, dentre tantos outros temas ligados aos levantamentos aeromagnéticos. Quando não sabia por onde começar, os três estavam sempre prontos para ajudar no que fosse preciso.

Aos professores Marcílio Vieira Martins Filho e Alan Panosso pelas valiosas contribuições na banca de qualificação.

Aos professores e técnicos do departamento de Ciências Exatas: Glauco Rolim, Newton La Scala, Nelson Peruzzi, Zezé, Adriana e Shirlei.

Ao Prof. José Marques Júnior e equipe do CSME pela amizade e suporte nos momentos da pesquisa.

A segunda família que a UNESP/FCAV me deu: Andrea, Yane, Branca, Robson, Jonathan (porce), Jardel, Paulo (PH), Gabriela, Nayane, Lidia (Papa), Felipe (Comitiva), Gustavo, entre tantos outros que não cabem em um único parágrafo. Obrigado por deixarem esses anos mais alegre.

A *Unidad de Edafología/UCO*, em especial, Maria del Carmen del Campillo García, Maribel, José María Méndez, Antonio Rafael Sánchez Rodríguez, Adrian Gonzalez Guzman, Mónica Sánchez e Jéssica Oliveira (única brasileira!) pelos momentos agradáveis juntos.

Ao Programa de Pós Graduação em Agronomia (Ciência do Solo). O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

A *Universidad d Córdoba/UCO* pela recepção durante o período de doutorado sanduíche e a CAPES pela concessão da bolsa (Programa de Doutorado Sanduíche no Exterior – PDSE / Processo nº88881.187336/2018-01).

A todos aqueles que contribuíram de maneira direta ou indiretamente: meu muito obrigado!

SUMÁRIO

| | |
|---|-----|
| RESUMO | X |
| ABSTRACT | xii |
| LISTA DE ABREVIACIONES | xiv |
| LISTA DE TABELAS | xv |
| LISTA DE FIGURAS | xvi |
| CAPÍTULO 1 – Considerações gerais | 1 |
| 1.1 Introdução e justificativas | 1 |
| 1.2 Revisão de Literatura | 3 |
| 1.2.1 Relação solo-rocha para mapas de aplicação | 3 |
| 1.2.2 Usos dos dados aeromagnéticos na ciência do solo | 5 |
| 1.2.3 Dados legados e mapeamento digital dos solos..... | 8 |
| 1.2.4 Modelos baseados em árvores: Random Forest | 11 |
| Referências | 13 |
| CAPÍTULO 2 - Evaluating the potential of aeromagnetic data for mapping clay content at regional and local scales in Brazil | 20 |
| 2.1 Introduction..... | 21 |
| 2.2 Materials and methods | 23 |
| 2.2.1 Study area..... | 23 |
| 2.2.2 Sampling and measurement of clay | 25 |
| 2.2.3 Input datasets | 25 |
| 2.2.4 Data analysis | 26 |
| 2.3 Results and discussion..... | 27 |
| 2.4 Conclusions..... | 34 |
| Acknowledgements | 35 |
| References | 35 |
| CAPÍTULO 3 - Avaliação dos dados aeromagnéticos aplicados a agricultura em escala regional | 41 |
| 3.1. Introdução..... | 42 |
| 3.2 Material e Métodos | 43 |
| 3.2.1 Área de estudo e amostragem | 43 |
| 3.2.2 Banco de dados legados..... | 45 |
| 3.2.3 Conjuntos de dados de entrada | 46 |
| 3.2.4 Análise de dados | 47 |
| 3.3 Resultados e Discussão | 48 |
| 3.4. Conclusão..... | 59 |
| Referências | 59 |
| CAPÍTULO 4 - Considerações finais | 65 |
| APÊNDICES | 66 |

DADOS AEROMAGNÉTICOS COMO COVARIÁVEL AMBIENTAL NO PLANEJAMENTO AGRÍCOLA

RESUMO – O conhecimento das classes e atributos do solo é fundamental para um planejamento agrícola ideal. São necessárias técnicas adequadas para obter informações rápidas e confiáveis do solo em diferentes níveis de escala e complexidade. Os dados aeromagnéticos já são disponibilizados em escala nacional e tem relação com os minerais magnéticos da rocha, facilmente encontrados em solos tropicais. O estudo foi dividido em dois estágios: o primeiro estágio teve como objetivo utilizar os dados aeromagnéticos para estimar o conteúdo de argila como atributo do solo. No segundo estágio, foi utilizado os dados aeromagnéticos no planejamento agrícola, estimando a erodibilidade e o coeficiente de Freundlich para sorção de herbicida. O estudo foi realizado em duas diferentes escalas: em escala regional (Planalto Ocidental Paulista), com 130000 km², e escala local (município de Guatapar), com 8,7km². Foram coletadas 539 e 334 amostras, respectivamente, em profundidade 0.0-0.20m para a obteno do contudo de argila (g kg⁻¹). Ambos os locais de estudo esto localizadas em uma rea de contato de rocha arenito-basalto. As variveis de entrada utilizadas para estimar o contudo de argila foram os dados aeromagnticos em conjunto com a elevao, precipitao, TWI, solo e geologia. Foi avaliado o coeficiente de determinao (R²) do algoritmo *Random Forest* nas etapa de treinamento e teste e analisado as principais variveis preditoras. O estudo demonstrou que os dados aeromagnticos foram importantes em ambas as escalas, mas principalmente em escala local. Na escala regional, as variveis precipitao e elevao e dados aeromagnticos foram as mais importantes, com R² de treinamento e teste de 0.91 e 0.40, respectivamente. Na escala local, os dados aeromagnticos, elevao e solo, respectivamente, foram as variveis mais importantes, com R² de 0.90 de treinamento e 0.74 de teste. Ainda em escala regional, foram utilizados os dados aeromagnticos para estimar a erodibilidade e o coeficiente de Freundlich para soro de herbicida. As variveis de entrada utilizadas foram dados aeromagnticos em conjunto com solo, geologia, elevao, declividade, precipitao. Os mesmos procedimentos foram realizados na avaliao do modelo *Random Forest* e as variveis preditoras. Para estimar a erodibilidade, foram obtidos o R² de 0.93 para

treinamento e 0.48 para teste, com a precipitação, elevação e solo como as principais variáveis, respectivamente. Para o coeficiente de sorção de Freundlich, foram o R^2 de treinamento e teste foram 0.90 e 0.36, e as variáveis precipitação, dados aeromagnéticos e elevação, respectivamente, apareceram como as mais importantes. As análises realizadas em conjunto com os dados aeromagnéticos apresentaram estimativas satisfatórias, quando comparados com outros mapeamentos no mundo. São necessários novos modelos e novas variáveis para validar dos dados aeromagnéticos para mapeamento de solos em condições geológicas diferentes.

Palavras-Chave: dados aeromagnéticos, atributos do solo, modelagem do solo, random forest, mapeamento digital de solos.

AEROMAGNETIC DATA AS ENVIRONMENTAL COVARIABLE IN AGRICULTURAL PLANNING

ABSTRACT – Knowledge of soil classes and attributes is essential for optimal agricultural planning. Adequate techniques are needed to obtain fast and reliable soil information at different levels of scale and complexity. The aeromagnetic data are already available on a national scale and are related to the magnetic minerals of the rock, easily found in tropical soils. The study was divided into two stages: the first stage aimed to use aeromagnetic data to estimate the clay content as an attribute of the soil. In the second stage, aeromagnetic data was used in agricultural planning, estimating the erodibility and the Freundlich coefficient for herbicide sorption. The study was carried out on two different scales: on a regional scale (Western Paulista Plateau), with 130000 km², and on a local scale (municipality of Guatapar), with 8.7 km². 539 and 334 samples were collected, respectively, at 0.0-0.20m depth to obtain the clay content (g kg⁻¹). Both study sites are located in a sandstone-basalt rock contact area. The input variables used to estimate the clay content were aeromagnetic data together with elevation, precipitation, TWI, soil and geology. The determination coefficient (R²) of the Random Forest algorithm was evaluated in the training and test stages and the main predictor variables were analyzed. The study showed that aeromagnetic data were important at both scales, but mainly at the local scale. On the regional scale, the variables precipitation and elevation and aeromagnetic data were the most important, with R² of training and testing of 0.91 and 0.40, respectively. Still on a regional scale, aeromagnetic data were used to estimate erodibility and Freundlich coefficient for herbicide sorption. TWI, NDVI e EVI. The input variables used were aeromagnetic data together with soil, geology, elevation, slope, precipitation, TWI, NDVI and EVI. The same procedures were performed in the evaluation of the Random Forest model and the predictor variables. To estimate erodibility, R² was obtained for 0.93 for training and 0.48 for testing, with precipitation, elevation and soil as the main variables, respectively. For the Freundlich sorption coefficient, the training and test R² were 0.90 and 0.36, and the variables precipitation, aeromagnetic data and elevation, respectively, appeared as the most important. The analyzes carried out in conjunction with the aeromagnetic data showed satisfactory estimates when compared with other

mappings in the world. New models and new variables are needed to validate aeromagnetic data for mapping soils in different geological conditions.

Keywords: Aeromagnetic data, soil attributes, soil modeling, random forest, digital soil mapping.

LISTA DE ABREVIACOES

| | |
|-------------------|---|
| A | Alcance |
| $C_0 / [C_0+C_1]$ | Grau de dependncia espacial |
| C_0+C_1 | Patamar |
| C_0 | Efeito Pepita |
| CNEN | Comisso Nacional de Energia Nuclear |
| CPRM | Companhia de Pesquisa de Recursos Minerais |
| DNPM | Departamento Nacional de Produo Mineral |
| DSM | Digital Soil Mapping |
| EMBRAPA | Empresa Brasileira de Pesquisa Agropecuria |
| EVI | Enhanced Vegetation Index |
| GUA | Guatapar |
| K | Erodibilidade global |
| K_f | Coeficiente de soro de Freundlich |
| LS | Local Scale |
| MDE/DEM | Modelo Digital de Elevao/Digital Elevation Model |
| NA | Not Available |
| NDVI | Normalized Difference Vegetation Index |
| NUCLEBRS | Empresas Nucleares Brasileiras |
| PETROBRS | Petrleo Brasileiro S.A. |
| POP/WPP | Planalto Ocidental Paulista/Western Paulista Plateau |
| RF | Random Forest |
| RS | Regional Scale |
| SQR/RSS | Soma dos Quadrados dos Resduos/Residual Sum of Squares |
| TWI | Topographic Wetness Index |
| USLE | Universal Soil Loss Equation |

LISTA DE TABELAS**CAPÍTULO 2**

| | |
|---|----|
| Table 1 - Descriptive statistics for the predicted variables..... | 28 |
| Table 2 - Estimated models for semivariogram adjustment..... | 32 |

CAPÍTULO 3

| | |
|---|----|
| Tabela 1. Estatística descritiva para as variáveis..... | 50 |
| Tabela 2 - Semivariograma das variáveis K e Kf..... | 56 |

LISTA DE FIGURAS

CAPÍTULO 1

Figura 1 - Número de documentos publicados por ano, de 1981 até 2020, de acordo com a base de dados Scopus – Tópico: Aeromagnetic Data; Refinado por área de pesquisa: *Agricultural and Biological Sciences*. Data de acesso: 11 de novembro de 2019.6

Figura 2 - Número de documentos publicados por país, de acordo com a base de dados Scopus – Tópico: *Aeromagnetic Data*; Refinado por área de pesquisa: *Agricultural and Biological Sciences*. Data de acesso: 11 de novembro de 2019.7

CAPÍTULO 2

Figure 1 - Localization of Western Paulista Plateau (a), as RS and Guatapar (b), as LS, both in So Paulo, Brazil.....24

Figure 2 - Relative importance of variables for prediction of clay in Regional Scale (a) and Local Scale (b).29

Figure 3 - Scatter plot of training versus test values of clay content by random forest in Regional Scale (a) and Local Scale (b).31

Figure 4 - Spatial pattern of observed clay content (g kg^{-1}) (a and b), estimated by Random Forest (g kg^{-1}) (c and d), standard deviation (g kg^{-1}) (e and f) and relative error (%) (g and h) in Western Paulista Plateau and Guatapar.33

CAPÍTULO 3

Figura 1 - Mapa de localizao da rea experimental (esq.) e grade amostral (dir.). 44

Figura 2 - Importncia relativa das variveis preditoras do fator K (a) e do Coeficiente de Soro de Freundlich (b).53

Figura 3 - Grfico de disperso dos dados de treinamento vs teste do fator K (a) e do Coeficiente de Soro de Freundlich (b).55

Figura 4 - Padro espacial do fator K observado (a), estimado por Random Forest (c) e o erro relativo (%) (e); Coeficiente de soro de Freundlich observado (b), estimado por Random Forest (d) e o erro relativo (%) (f).58

CAPÍTULO 1 – Considerações gerais

1.1 Introdução e justificativas

É improvável a possibilidade de coleta e análise de todos os atributos do solo em todo o globo terrestre. Para a produção de aquisição de novos dados de solo é fundamental levar em consideração o gasto e o tempo utilizado, sem contar a utilização de equipamentos, a capacitação da equipe, o laboratório, os reagentes, entre tantos outros. A percepção da realidade é ainda mais clara em países tropicais em que a carência de informações do solo é encontrada em escala variada.

A escassez de informação dos atributos do solo pode facilitar o surgimento de problemas como o uso inadequado de herbicidas, erosão, compactação, entre outros (Pedron et al., 2004). Ainda mais importante que conhecer os atributos do solo é saber como utilizá-las de maneira estratégica, com fins de manejo específico para a agricultura.

O mapeamento digital de solos apresenta uma alternativa promissora para a predição de atributos do solo (Akpa et al., 2014). A proposta de McBratney et al. (2003) com o modelo scorpan estabelece relações do solo com fatores pedogenéticos. Esses fatores podem ser simulados a partir de covariáveis ambientais em conjunto com técnicas estatísticas e modelos geoestatísticos.

Dados legados obtidos a partir de levantamentos tradicionais de solo também contribuem para a calibração de modelos para a predição dos atributos do solo (Carré et al., 2007). No caso do Planalto Ocidental Paulista foi utilizado dados legados do grupo de pesquisa CSME (Caracterização do Solo para Fins de Manejo Específico), localizado na Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal – UNESP. A utilização de dados legados no Planalto Ocidental Paulista segue a discussão de Carré et al. (2007), em que: i) Espera-se utilizar dados legados em grandes áreas para evitar novos gastos com pesquisas de solo desnecessárias, e; ii) Os dados aeromagnéticos podem ser utilizados para estimar atributos do solo.

As imagens obtidas a partir de levantamentos aeromagnéticos podem ser utilizadas como covariável no planejamento agrícola, pois são disponibilizados gratuitamente pelo Serviço Geológico do Brasil, são amplamente explorados em todo o território brasileiro desde a década de 50 e identificam o material de origem oriundo

da formação do solo. Para Gray et al. (2016), covariáveis que simulam particularidades do material de origem são essenciais para os modelos e os mapas dos atributos do solo. Em países tropicais como o Brasil, a fração argila é responsável pelo armazenamento de minerais como os óxidos de ferro, que é propriedade encontrada na rocha matriz (Cornell e Schwertmann, 2003; Akpa et al., 2014).

O desafio científico é encontrar o modelo ideal capaz de explicar a relação da rocha (km/profundidade) com os atributos do solo (cm/profundidade) a partir dos dados aeromagnéticos. Além disso, foram utilizadas covariáveis ambientais que colaboram com os processos de formação do solo (relevo, vegetação, clima, etc).

O trabalho teve como objetivos: i) Identificar o potencial dos dados aeromagnéticos para estimar o conteúdo de argila em escala local e regional; ii) Avaliar o uso dos dados aeromagnéticos no planejamento agrícola para estimar a erodibilidade e o coeficiente de Freundlich para sorção do herbicida Imazaquim, ambos em escala regional.

1.2 Revisão de Literatura

1.2.1 Relação solo-rocha para mapas de aplicação

É de conhecimento que a formação do solo é basicamente consequência da pedogênese. O pensamento que os atributos do solo são formados por interações com os fatores do ambiente surge no século XIX, com Dokuchaev (1883) e Jenny (1941). Esses fatores de formação são: clima, organismos, topografia, material de origem e tempo. O entendimento desses fatores de formação auxilia na compreensão e no comportamento do solo no Brasil e no mundo.

O Brasil apresenta aproximadamente 58% do território ocupado por Latossolos e Argissolos com diversas finalidades de uso (Santos et al., 2018). Os Latossolos são solos profundos e intemperizados, com a presença de argila entre 15 a 85% e pouca disponibilidade de nutrientes, enquanto os Argissolos apresentam aumento da argila em profundidade e costumam ser susceptíveis à erosão (Eberhardt et al., 2008; Sanchez et al., 2009).

A fração argila do solo apresenta características estratégicas para o planejamento agrícola e ambiental. Devido a sua relevância, é utilizado como atributo diagnóstico no Sistema Brasileiro de Classificação de Solos (Santos et al., 2018). O conteúdo de argila é responsável pela retenção de água, nutrientes e outras atividades capazes de auxiliar nas funções do solo e no armazenamento dos principais minerais (Akpa et al., 2014).

Em regiões tropicais e subtropicais, os principais minerais encontrados na fração argila são óxidos de ferro e alumínio, variando seu conteúdo de acordo com a paisagem (Camargo et al., 2008). Essa variedade de minerais apresenta características da rocha matriz e da gênese do solo (Cornell e schwertmann, 2003; SILVA, 2016). Os principais minerais encontrados são caulinita, gibbsita, goethita e hematita que, na fração argila, participam da estrutura e das propriedades químicas do solo (Coelho e Vidal-Torrado, 2003; Inda Junior e Kämpf, 2005).

A maioria das rochas encontradas na natureza apresenta óxidos de ferro, variando em natureza e abundância (Cornell e Schwertmann, 2003). As rochas são as principais fontes de recursos naturais presentes na superfície terrestre e a maioria dos seus elementos circula ainda hoje. A rocha (ou material de origem) fornece a

matéria prima necessária para a formação do solo, podendo ser terra firme ou material inconsolidado (Gray et al., 2016). O material rochoso suportará fatores de formação do solo descritos por Jenny (1941) até a formação de um solo específico.

Com 130000km² e aproximadamente 45% do estado de São Paulo, a diversidade geológica do Planalto Ocidental Paulista condicionou a formação de solos em diferentes estágios de evolução. O Planalto Ocidental Paulista está localizado em uma zona de transição formada por rochas de características distintas. Estima-se que o Planalto Ocidental Paulista tenha se formado no período Cretáceo Superior, com idade entre 88 a 65 milhões de anos. Os derrames basálticos encontrados da área fazem parte da Formação Serra Geral enquanto as rochas areníticas fazem parte do Grupo Bauru (Fernandes e Coimbra, 2000). Da área total, 20000km² são de rochas basálticas e o restante de rochas sedimentares, sendo 74000km² da formação Vale do Rio do Peixe e 36000km² por outras formações geológicas (Fernandes, 2004; Fernandes et al., 2007). Com a diversidade geológica no Planalto Ocidental Paulista, surgem dezesseis classes de solo, sendo a maior ocorrência classificadas como Argissolo Vermelho-Amarelo, Latossolo Férrico, Neossolo Litólico, Nitossolo Vermelho e Gleissolo Háplico (Santos et al., 2018).

É importante conhecer a distribuição e funções do solo para a conservação e preservação da produtividade agrícola ideal (Terribile et al., 2011; Santra et al., 2017). Contudo, grandes áreas com heterogeneidade de solo podem ter a qualidade do trabalho comprometida de acordo com as amostras coletadas erroneamente em campo. Nesses casos seria necessário intensificar a grade amostral para verificar a variabilidade das propriedades requeridas (Akpa et al., 2016). Além disso, uma agricultura mais sustentável depende de informações rápidas e confiáveis dos atributos do solo. Análise de laboratório nem sempre são técnicas rápidas e de baixo custo (Santra et al., 2017).

Partindo dessa premissa, seria ideal uma fonte de dados já disponível e que quantificasse problemas encontrados na agricultura. Os dados aeromagnéticos são capazes de captar corpos magnéticos de grandes áreas em profundidade significativa (Kearey et al., 2009; Khalil et al. 2016). Métodos de varredura podem apresentar uma solução confiável para análise em grandes áreas (Santra et al., 2017).

O interesse por informações dos recursos minerais gerou ação conjunta de empresas como o Serviço Geológico do Brasil (CPRM), Departamento Nacional de Produção Mineral (CPRM/DNPM), Empresas Nucleares Brasileiras/Comissão Nacional de Energia Nuclear (NUCLEBRÁS/CNEN) e Petróleo Brasileiro (PETROBRAS). O resultado da ação foram as informações obtidas por aerolevantamentos em todo o território brasileiro. Os projetos geofísicos foram realizados de maneira fragmentada, sendo reunidos posteriormente em um único produto. As principais etapas dos aerolevantamentos geofísicos executados no Brasil ocorrem desde 1952 (Bizzi et al., 2003). A base de dados de projetos aerogeofísicos do Brasil foi disponibilizada para o público a partir de 1995 com o nome AERO. Os dados aeromagnéticos são amplamente utilizados na geociência para estudos em nível de rocha desde então (Stanton et al., 2010; Oliveira e Medeiros, 2018; Guimarães et al., 2013; Curto et al., 2014; Camacho e Sousa, 2017).

Dada à relação que o material de origem tem com os atributos do solo (Gray et al., 2016), os dados aeromagnéticos apresentam aplicabilidade para a agricultura em geral. Os óxidos de ferro encontrados na fração argila apresentam a mesma característica magnética da rocha (Cornell e Schwertmann, 2003). Até o presente momento não existem estudos ou pesquisas aplicadas na agricultura que utilizam os dados aeromagnéticos para a obtenção dos atributos do solo.

1.2.2 Usos dos dados aeromagnéticos na ciência do solo

Os primeiros projetos geofísicos no Brasil utilizando dados aeromagnéticos começaram a ser desenvolvidos no início da década de 50 (Bizzi et al., 2003) por empresas interessadas em extrair informações da rocha. Os resultados das ações foram *grids* magnetométricos que, posteriormente, foram unidos em uma única imagem abrangendo boa parte do território nacional. Esse material se tornou público a partir de 1995 com a criação da Base de Dados de Projetos Aerogeofísicos do Brasil, chamada AERO pela CPRM.

Não se sabe ao certo quando outros países começaram a desenvolver tecnologia semelhante. A primeira publicação (Scopus, 2019) que se tem conhecimento corrobora com o início dos levantamentos aeromagnéticos no Brasil.

Na década de 50, Smellie (1956) publica o primeiro artigo científico utilizando dados aeromagnéticos, focado principalmente no entendimento do recurso e formas de interpretar as anomalias magnéticas. Não muito distante, Domzalski (1957) destaca as principais limitações e cuidados com o processamento dos dados aeromagnéticos. As primeiras citações relacionando a rocha com os minerais vieram com Bichan (1959), que demonstrou a relação dos dados aeromagnéticos com os minerais presentes no solo. Leite e Leão (1985) foram os pioneiros no Brasil que automatizaram as interpretações quantitativas utilizando o algoritmo de Marquardt e a regressão Ridge. Observa-se que número de aplicações dos dados aeromagnéticos no Brasil e no mundo apresentam constante crescimento na área de geociência.

O conteúdo de argila em solos tropicais comumente apresenta óxidos de ferro em sua composição, características pedogenéticas detectadas pelos dados aeromagnéticos (Bichan, 1959; Cornell e Schwertmann, 2003; Akpa et al., 2014). Apesar da relação que os dados aeromagnéticos tem com a formação do solo, pouco ou nada é aproveitado em nível mundial (Scopus, 2019). O primeiro trabalho dentro das Ciências Agrárias é de 1981 (Figura 1), tal que após este ano houve uma inconstância na quantidade de publicações.

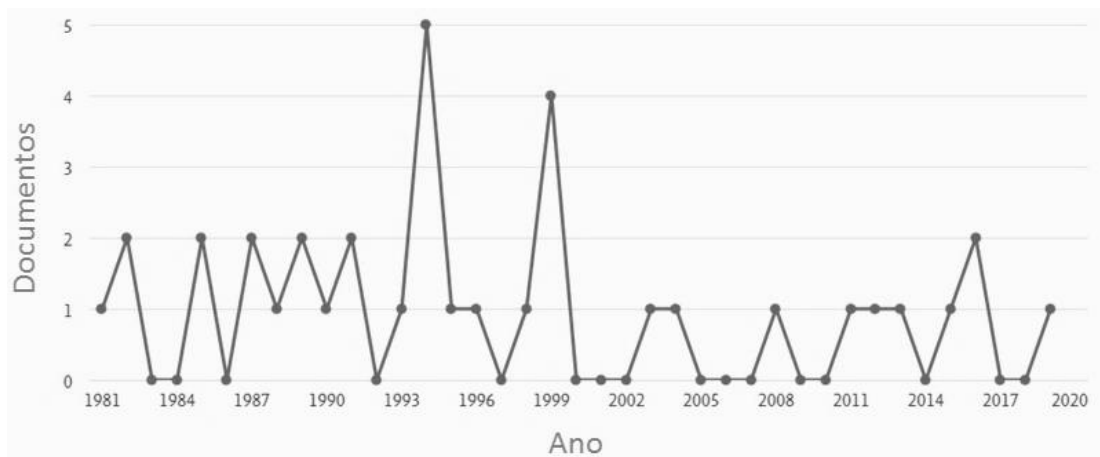


Figura 1 - Número de documentos publicados por ano, de 1981 até 2020, de acordo com a base de dados Scopus – Tópico: Aeromagnetic Data; Refinado por área de pesquisa: *Agricultural and Biological Sciences*. Data de acesso: 11 de novembro de 2019.

A quantidade de publicação nem sempre direciona à informação desejada. Em uma investigação mais profunda, dos poucos artigos que foram identificados utilizando

dados aeromagnéticos, a maioria não tem qualquer relação com a ciência do solo. Há exemplo como Warner et al. (1994) que utilizaram sensoriamento remoto e dado aeromagnéticos para detectar tipos de rocha e vegetação local.

Pela relação de países que se beneficiam dos dados aeromagnéticos, os Estados Unidos está três vezes à frente do segundo colocado, a Alemanha (Figura 2). Essa característica demonstra o interesse em que países desenvolvidos têm em aplicar tecnologias capazes gerar informação especializada. O Brasil não apresentou nenhuma publicação utilizando dados aeromagnéticos nas ciências agrárias (Scopus, 2019).

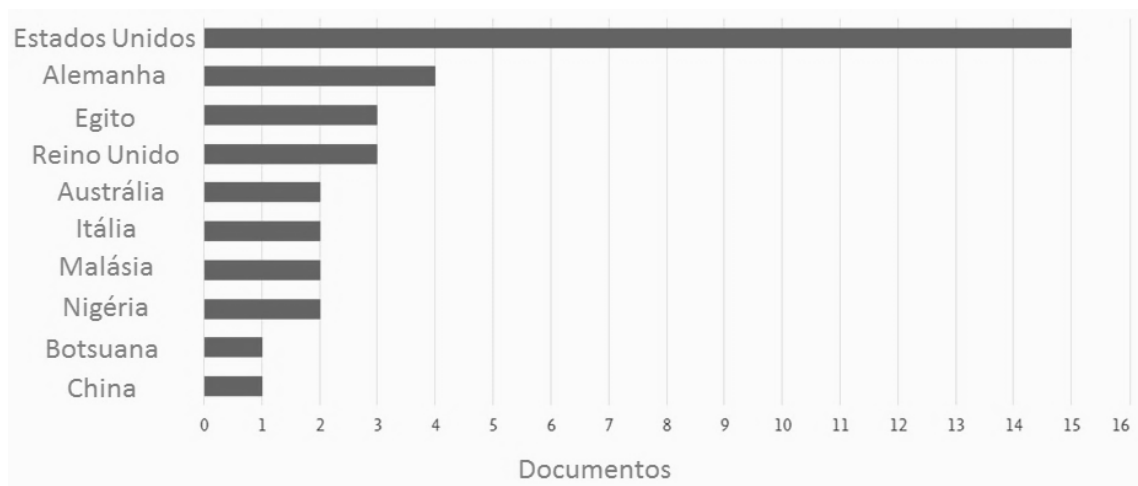


Figura 2 - Número de documentos publicados por país, de acordo com a base de dados Scopus – Tópico: *Aeromagnetic Data*; Refinado por área de pesquisa: *Agricultural and Biological Sciences*. Data de acesso: 11 de novembro de 2019.

O número de pesquisas que utilizam os dados aeromagnéticos nas ciências agrárias ainda é pequeno. Contudo, com o advento do Mapeamento Digital de Solos (MDS), existe uma lacuna muito grande de variáveis que possam auxiliar na estimativa dos atributos do solo. Os dados aeromagnéticos são ideais, pois identificam relações pedogenéticas com a granulometria do solo, ou seja, indicadores dos processos evolutivos. A ausência de publicações com dados aeromagnéticos na ciência do solo sustenta a originalidade do presente trabalho.

1.2.3 Dados legados e mapeamento digital dos solos

A capacidade produtiva dos solos é o recurso mais importante para a geração de alimentos em escala mundial. Estima-se que pouco mais de 6% da superfície terrestre possam ser utilizadas para a agricultura (Fao, 1980). Contudo, as terras cultiváveis apresentam perda superior a tolerável para seu equilíbrio, diminuindo assim a produtividade das culturas agrícolas (Fao, 2015; Teng et al., 2016). No Brasil, Veiga Filho et al. (1992) estimaram que aproximadamente 600 milhões de toneladas de solos apropriados para cultivo foram perdidos por erosão, sendo 194 milhões só no estado de São Paulo. Esses impactos não só aos solos nacionais, mas os tipos de ecossistemas presentes.

As técnicas de análise de dados em conjunto com a pedologia evoluíram muito nos últimos anos. A compatibilidade de diferentes procedimentos de estudo e descrição dos solos avançaram com o advento da informática, facilitando o acesso e manipulação da informação (Dai et al., 2019). Estima-se que 30% da superfície terrestre esta mapeada em escala 1: 1000000 por técnicas tradicionais (Hendriks et al., 2019) enquanto toda a superfície já esta mapeada em escala 1: 250000 por técnicas modernas, tais como o mapeamento digital de solos (Hengl et al., 2017).

Contudo, mesmo com os avanços na ciência do solo, ainda existe uma dependência relativa às técnicas de classificação. Os levantamentos tradicionais obtêm dados a partir de observações e análises realizadas em campo e/ou em laboratório para compreender atributos físicas e químicas do solo (Omuto et al., 2013; Bui, 2004). Quanto maior o nível de detalhamento, menor será as imperfeições e mais homogêneo será o mapa final (Sarmiento et al., 2014). Hengl; MacMillan (2019) caracterizaram como dados legados a soma de todas as informações obtidas desde que o solo foi estudado como objeto natural e independente. A composição de um banco de dados legado é realizada principalmente por levantamento tradicional de solos (Bui e Moran, 2001).

Nesse sentido, discussões sobre erodibilidade tem despertado interesse na agricultura, pois tem influência direta na perda de solo e nutriente em áreas agrícolas, devido a erosão (Freitas, 2014). Conhecer as causas e fatores que afetam a erodibilidade é essencial para a recomendação de boas práticas agrícolas. Além

disso, o conteúdo de argila auxilia na melhoria dos agregados devido seu papel cimentante nos agregados, ocasionando menor erodibilidade (Meyer; Harmon, 1984; Camargo et al., 2008). Para o cálculo da erodibilidade global (K, em $t h^{-1} MJ^{-1} mm^{-1}$) do modelo USLE foi utilizado à equação proposta por Denardin (1990) conforme a seguinte equação (Eq. 1):

$$K = 0,00000748 M + 0,00448059 P - 0,0631175 DMP + 0,01039567 REL \quad (1)$$

em que:

$M = (\% \text{ silte} + \% \text{ areia muito fina}) \times (100 - \% \text{ argila})$

$P =$ código de permeabilidade codificado conforme Wischmeier et al. (1971) (adimensional);

$DMP =$ diâmetro médio ponderado das partículas $< 2,0$ mm,

$REL = (\% \text{ matéria orgânica} \times \% \text{ nova areia})/100$

$\% \text{ nova areia} = \% \text{ areia total} - \% \text{ areia muito fina}$

O conteúdo de argila é a fração do solo com características essenciais para diversas aplicações na agricultura. A extração do conteúdo de argila no solo é feito pelo método da pipeta, com o emprego de NaOH 0.1 mol L^{-1} como dispersante químico e agitação mecânica por 16 horas em baixa velocidade (Teixeira et al., 2017). Além da praxe de análise da textura do solo em laboratório, é possível associar a argila com processos e propriedades físicas e químicas do solo (Amirian-Chakan et al., 2019), erodibilidade (Barbosa, 2014; Freitas, 2016; Teixeira, 2017), sorção de nutrientes e herbicidas (Rocha et al., 2003; Peluco, 2016), mineralogia (Silva, 2016), relação solo-rocha (Gray et al., 2016), entre tantos outros. A variabilidade da argila também está relacionada com clima e material de origem (Viscarra, 2011). Nesse sentido, além de validar dados aeromagnéticos com o conteúdo de argila, o presente trabalho tem como proposta utilizar a covariável para manejo específico na agricultura.

Outro ponto que merece atenção é a utilização de herbicida, pois é o insumo mais aplicado em volume e em área na agricultura (Peluco, 2016). O uso inadequado de herbicida é potencialmente perigoso para o meio ambiente devido sua ação contaminante (Lapworth e Goody, 2006). A adsorção de herbicida está associada

com diversos atributos do solo, tais como conteúdo de argila, pH, matéria orgânica, entre outros (Rocha et al., 2003). Para o cálculo do potencial de sorção de herbicida, foi utilizada a metodologia de meta-análise desenvolvida por Peluco (2016). O autor ajustou pesquisas semelhantes com diferentes classes de solos, teores de argila e concentração de herbicida a modelos matemáticos, obtendo assim o coeficiente de Freundlich (K_f) para sorção do herbicida Imazaquim.

Carré et al. (2007) salientaram que a só com a utilização de dados legados que pesquisas em grandes escalas podem ser realizadas, uma vez que novas despesas com pesquisas de solo são evitadas. O Brasil já conta com iniciativas como o *febr* de Samuel-Rosa (2019) e colaboradores espalhados pelo Brasil. O *febr* é um repositório livre direcionado a pesquisadores que disponibilizam seus dados estruturados de solo já publicados para outras pessoas utilizar, renovar (Rossiter, 2008) e atualizar (Kempen et al., 2009). Com a disponibilidade de dados legados sendo utilizados para outras aplicações, está claro que o MDS possa se beneficiar, pois dados primários fornecem o suporte necessário para calibração e validação de modelos que possam estimar atributos do solo (Carré et al., 2007; Rossiter, 2008).

Mapeamento digital de solos é definido como a criação e compartilhamento de sistema de informação espacial do solo, obtido por observações de campo e laboratório juntamente com sensores espaciais e não espaciais (Lagacherie e McBratney, 2006). O modelo empírico *SCORPAN*, proposto por McBratney et al. (2003), descreve as relações do solo com outros elementos responsáveis por sua formação. A principal função proposta por McBratney et al. (2003) foi contribuir para a produção de mapas digitais, tal que as classes e propriedades dos solos anteriormente mapeadas possam ser associadas a novos produtos e, ainda que novas áreas e suas incertezas possam ser realizadas. Essa relação com fatores do ambiente formam funções de predição espacial do solo. O modelo proposto por McBratney (2003) pode ser definido como (Eq. 2):

$$S = f(s, c, o, r, p, a, n) \quad (2)$$

em que:

S = Classe ou dos atributos do solo a ser estimado.

s = (*soil*) classes ou atributos do solo medido em um ponto;

c = (*climate*) clima ou propriedades climáticas em determinado ambiente;

o = (*organisms*) organismos, incluindo vegetação natural;

r = (*relief*) topografia e atributos e classes do terreno;

p = (*parent materials*) material de origem, litologia;

a = (*age / time*) idade / fator tempo;

n = (*spatial position*) posição espacial ou geográfica.

Regiões nas quais existam restrições para obter dados primários, estruturas genéricas podem ser utilizadas para substituir condições do ambiente, tais como modelos digitais de elevação e respostas espectrais de satélites (McBratney et al., 2003). O MDS se baseia nas relações entre os atributos do solo e variáveis auxiliares que são obtidas de maneira mais fáceis. Essas variáveis auxiliares explicitam quantitativamente o solo e podem ser amostradas exaustivamente (Teng et al., 2016). Confirmado a relação que os dados aeromagnéticos apresentam com fatores de formação do solo, resta propor a metodologia ideal para associar todas as características pedogenéticas.

1.2.4 Modelos baseados em árvores: Random Forest

Machine Learning (Aprendizado de Máquinas) é o termo utilizado quando o algoritmo é capaz de reconhecer padrões de aprendizado por ações bem sucedidas. Entre os algoritmos, existem aqueles que reconhecem padrões de dados a partir de classificações supervisionadas ou não supervisionadas. A aprendizagem supervisionada fornece ao algoritmo conjunto de dados de entrada e a saída desejada (Oshiro, 2013), enquanto em algoritmos não supervisionados são utilizados somente o conjunto de dados de entrada.

Entre os métodos de aprendizagem supervisionada, os *ensembles* utilizam múltiplos modelos mais simples de maneira agrupada em busca da melhor estimativa (Hastie et al., 2008). O Random Forest é um *ensemble* muito utilizado por apresentar menor ruído, melhor desempenho e ser capaz de amenizar a correlação entre as árvores, aumentando a precisão preditiva (Breiman, 2001).

Basicamente, a árvore de decisão utiliza a decomposição de um problema maior em problemas mais simples de serem resolvidos. O Random Forest emprega múltiplas árvores de decisão a partir de amostras aleatórias e independentes e a média de todos os valores obtidos por cada árvore gera o resultado final. A cada novo nó da árvore, um novo subconjunto de atributos é escolhido aleatoriamente e avaliado, no qual o melhor permanece na divisão (Oshiro, 2013).

O Random Forest divide o conjunto de dados de treinamento em várias amostras que irão formar as árvores. De cada uma dessas amostras o algoritmo retira pequenas amostras de maneira específica, chamada *bootstrapping* (Gentile, 2017). Dessa forma, o conjunto de árvores é treinado com diferentes atributos escolhidos inicialmente e de maneira aleatória, tornando cada árvore diferente das demais, utilizando amostras *bootstrapping* para quantificar incertezas encontradas no modelo (Efron, 1979; Breiman, 1996; Breiman, 2001). Esse processo permite que o algoritmo melhore seu desempenho sem a necessidade de um novo conjunto de dados durante o treinamento. O *bootstrapping* é considerado uma ferramenta estatística de avaliação da precisão do modelo (Hastie; Tibshirani; Friedman, 2008).

Contudo, um dos problemas das árvores de decisão é que seus estimadores possuem alta variância. Uma das soluções apresentadas é o *bagging* (ou *bootstrap aggregating*) para calcular a média das previsões de todas as árvores (James et al., 2013). Um dos benefícios do *bagging* é a impossibilidade de aumentar o conjunto de dados e a variância simultaneamente. O Random Forest ainda conta com o *feature bagging* para diminuir a correlação entre as árvores e aumenta o processo preditivo (Breiman, 2001).

Para validar o Random Forest, comumente é utilizado o *out-of-bag*, que é o resultado não utilizado pelo *bootstrap*, comparado e calculado a taxa de acerto (James et al., 2013). O estimador *out-of-bag* fornece o erro de generalização obtido pelo conjunto das árvores (Breiman, 2001). Contudo, existem outras formas de avaliar o

desempenho do modelo em MDS, como o coeficiente de determinação (Akpa et al., 2014) e validação cruzada (Pahlavan-Rad e Akbarimoghaddam, 2018).

O Random Forest é ideal para grandes bases de dados, pois apresenta bom desempenho, é capaz de utilizar variáveis numéricas e categóricas e tem pouca necessidade de ajuste para boa acurácia preditiva (Díaz-Uriarte e Alvarez, 2006). Diversas pesquisas em MDS utilizam o modelo Random Forest aplicado no Brasil (Bhering et al., 2016; Junior et al., 2016; Pinheiro et al., 2017) e no mundo (Akpa et al., 2014; Hengl et al., 2015; Gray et al., 2016; Pahlavan-Rad e Akbarimoghaddam, 2018; Zeraatpisheh et al., 2019).

Os resultados das variáveis mais importantes no modelo fornecido pelo pacote *randomForest* (Liaw e Wiener, 2002) aparecem de duas maneiras: i) *%IncMSE*, que é a porcentagem do aumento do *Mean Squared Accuracy* estimado a partir do *out-of-bag*, e; ii) *IncNodePurity*, cálculo da diminuição das impurezas entre os nós pelo *Gini Index*, (Akpa et al., 2016). A vantagem desses resultados é a possibilidade de conhecer quais as variáveis mais importantes utilizadas pelo Random Forest para chegar a determinado resultado, dando assim peso a cada uma delas.

Referências

Amirian-Chakan A, Minasny B, Taghizadeh-Mehrjardi R, Akbarifazl R, Darvishpasand Z, Khordehbin S.(2019) Some practical aspects of predicting texture data in digital soil mapping. **Soil and Tillage Research** 194:104289.

Akpa SIC, Odeh IOA, Bishop TFA, Hartemink AE (2014) Digital Mapping of Soil Particle-Size Fractions for Nigeria. **Soil Science Society of America Journal** 78:1953-1966.

Akpa SIC, Ugbaje SU, Bishop TFA, Odeh IOA (2016) Enhancing pedotransfer functions with environmental data for estimating bulk density and effective cation exchange capacity in a data-sparse situation. **Soil Use and Management** 32:644–658.

Barbosa R (2014) **Erodibilidade de latossolos predita pela suscetibilidade magnética e espectroscopia de reflectância difusa**. 78f.Tese (Doutorado em Agronomia) - Unesp, Jaboticabal.

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

Breiman, L (1996) Bagging predictors, **Machine Learning** 24:123-140.

Bichan, WJ (1959) Correlation of aeromagnetic data with source mineralogy; discussion. **Economic Geology** 54:512–5.

Bhering SB, Chagas CS, Junior WC, Pereira, NR, Filho BC, Pinheiro HS (2016) Mapeamento digital de areia, argila e carbono orgânico por modelos Random Forest sob diferentes resoluções espaciais. **Pesquisa Agropecuária Brasileira** 51:1359–1370.

Bizzi LA, Schobbenhaus C, Gonçalves JH, Vidoti RM, Gonçalves JH (Eds.) (2003) **Geologia, tectônica e recursos minerais do Brasil: texto, mapas & SIG**. Brasília: CPRM, .643p.

Bui EN, Moran CJ (2001) Disaggregation of polygons of surficial geology and soil maps using spatial modeling and legacy data. **Geoderma** 103:79–94.

Bui EN (2004) Soil survey as a knowledge system. **Geoderma** 120: 17-26.

Camacho, CR, Sousa FRF (2017) O arcabouço estrutural da Bacia Sedimentar do Araripe, Província Borborema, baseado em dados aeromagnetométricos. **Geologia USP. Série Científica** 17:149-161.

Camargo LA, Marques Jr J, Pereira GT, Horvat RA (2008) Variabilidade espacial de atributos mineralógicos de um Latossolo sob diferentes formas de relevo. **Revista Brasileira de Ciência do Solo** 32:2269-2277..

Carré F, McBratney AB, Minasny B (2007) Estimation and potential improvement of the quality of legacy soil samples for digital soil mapping. **Geoderma** 141:1–14.

Coelho MR, Vidal-Torrado P (2003) Caracterização e gênese de perfis plínticos desenvolvidos de arenito do Grupo Bauru. II – Mineralogia. **Revista Brasileira de Ciências do Solo** 27:495–507.

Cornell RM, Schwertmann U (2003) **The iron oxides: structure, properties, reactions, occurrence, and uses**. Wiley VCH: Weinheim, 664 p.

Curto JB, Vidotti RM, Fuck RA, Blakely RJ, Alvarenga CJS, Dantas EL (2014) The tectonic evolution of the Transbrasiliano Lineament in northern Paraná Basin, Brazil, as inferred from aeromagnetic data. **Journal of Geophysical Research: Solid Earth** 119:1544–1562.

Dai Y, Shanguan W, Wei N, Xin Q, Yuan H, Zhang S, Liu S, Lu X, Wang D, Yan, F (2019) A review of the global soil property maps for Earth system models. **SOIL** 5:137–158.

Denardin JE (1990) **Erodibilidade do solo estimado por meio de parâmetros físicos e químicos**. 114f. Tese (Doutorado em Agronomia) - USP, Piracicaba.

Díaz-Uriarte R, Alvarez AS (2006) Gene selection and classification of microarray data using random forest. **BMC Bioinformatics** 7:1-13.

Domzalski W (1957) Some problems of the aeromagnetic surveys. **Geophys Prospect** 5:469–79.

Dokuchaev VV (1883) **Russian Chernozems** (Russkii Chernozem), Israel Program for Scientific Translations: Jerusalem, 439p.

Efron B (1979) Bootstrap methods: Another look at the jackknife. **The Annals of Statistics** 7:1-26.

Eberhardt DN, Vendrame PRS, Becquer T, Fátima Guimarães M. Influência da granulometria e da mineralogia sobre a retenção do fósforo em latossolos sob pastagens no cerrado. **Revista Brasileira de Ciência do Solo** 32:1009–1016.

Fernandes LA, Coimbra AM (2000) The Late Cretaceous Caiuá Desert (Bauru Basin, Brazil). In: INTERNATIONAL GEOLOGY CONGRESS. Rio de Janeiro (Brazil). **Anais...** Rio de Janeiro: IUGS, 2000.

Fernandes LA (2004) Mapa litoestratigráfico da parte oriental da Bacia Bauru (PR, SP, MG), escala 1: 1.000.000. **Boletim Paranaense de Geociências** 55: 53–66.

Fernandes LA, Castro AB, Basilici G (2007) Seismites in continental sand sea deposits of the Late Cretaceous Caiuá Desert, Bauru Basin, Brazil. **Sedimentary Geology** 199: 51–64.

FAO 1980: Food and Agriculture Organization of the United Nations. The state of food and agriculture. Rome, 199p.

FAO 2015: Food and Agriculture Organization of the United Nations. Status of the world's soil resources: Main report. Rome, 607p.

Freitas L (2014) **Qualidade e erodibilidade de Latossolos sob mata e cultivo de cana-de-açúcar**. 110f. Tese (Doutorado em Agronomia) - Unesp, Jaboticabal.

Gentile N (2017) **Aprendizado de máquina e caracterização de aterosclerose subclínica: um estudo de caso**. 213f. Dissertação (Mestrado em Engenharia da Computação) – IPT, São Paulo.

Gray JM, Bishop TFA, Wilford JR (2016) Lithology and soil relationships for soil modelling and mapping. **Catena** 147: 429–440.

Guimarães SNP, Hamza VM, Silva JJ (2013) Airborne geophysical surveys in the north-central region of Goiás (Brazil): Implications for radiometric characterization of tropical soils. **Journal of Environmental Radioactivity** 116:10–18.

Hastie T, Tibshirani R, Friedman J (2008) **The Elements of Statistical Learning**. Springer, 745p.

Hendriks CMJ, Stoorvogel JJ, Lutz F, Claessens L (2019) When can legacy soil data be used, and when should new data be collected instead? **Geoderma** 348:181–188.

Hengl, T.; Heuvelink, G.; et al. (2015) Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. **PLoS ONE** 10:1–26.

Hengl T, Mendes de Jesus J, et al. (2017) SoilGrids250m: Global gridded soil information based on machine learning. **PLOS ONE** 12, n. 2:e0169748.

Hengl T, Macmillan, R (2019) **Predictive Soil Mapping with R**. OpenGeoHub foundation: Wageningen, 370 p.

Inda Junior AV, Kämpf N (2005) Variabilidade de goethita e hematita via dissolução redutiva em solos de região tropical e subtropical. **Revista Brasileira de Ciência do Solo** 29:851-866.

James G, Witten D, Hastie T, Tibshirani R (2013) **An Introduction to Statistical Learning with Applications in R**. Springer, 426p.

Jenny H (1941) **Factors of soil formation**. McGraw-Hill: New York, USA, 281p.

Junior WC, Filho BC, Chagas CS, Bhering SB, Pereira NR, Pinheiro HSK (2016) Regressão linear múltipla e modelo Random Forest para estimar a densidade do solo em áreas montanhosas. **Pesquisa Agropecuária Brasileira** 51:1428–1437.

Kearey P, Brooks, M HILL, I (2009) **An Introduction to Geophysical Exploration**. Blackwell Science: Malden, 272p.

Kempen B, Brus D, Heuvelink G, Stoorvogel J (2009) Updating the 1:50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. **Geoderma** 151:311–326.

Khalil A, Hafeez THA, Saleh HS, Mohamed WH (2016) Inferring the subsurface basement depth and the structural trends as deduced from aeromagnetic data at West Beni Suef area, Western Desert, Egypt. **NRIAG Journal of Astronomy and Geophysics** 5:380–392.

Lagacherie P, McBratney AB (2006) Spatial soil information systems and spatial soil inference systems. In: Perspectives for digital soil mapping. **Developments in Soil Science** 31:3-22.

Lapworth DJ, Goody DC (2006) Source and persistence of pesticides in a semi-confined chalk aquifer of southeast England. **Environmental Pollution** 144:1031-1044.

Leite LWB, Leao JWD (1985) Ridge regression applied to the inversion of two-dimensional aeromagnetic anomalies (Brazil). **Geophysics** 50:1294–1306.

Liaw A, Wiener M (2002) Classification and regression by random- Forest. **Rnews** 2:18-22.

McBratney AB, Mendonça Santos ML, Minasny B (2003) On digital soil mapping. **Geoderma** 117:3–52.

Meyer L, Harmon W (1984) Susceptibility of agricultural soils to interrill erosion. **Soil Science Society of America Journal** 48:1152- 1157.

Oliveira R G, Medeiros W (2018) E Deep crustal framework of the Borborema Province, NE Brazil, derived from gravity and magnetic data. **Precambrian Research** 315:45–65.

Oshiro, T (2013) **Uma abordagem para a construção de uma única árvore a partir de uma Random Forest para classificação de bases de expressão gênica**. 101f. Dissertação (Mestrado em Bioinformática) – USP, Ribeirão Preto.

Omuto C, Nachtergaele F, Rojas RV (2013) **State of the art report on global and regional soil information: Where are we? Where to go?** Global Soil Partnership Technical Report: Rome, FAO, 81p.

Pahlavan-Rad MR, Akbarimoghaddam A (2018) Spatial variability of soil texture fractions and pH in a flood plain (case study from eastern Iran). **Catena** 160:275–281.

Pedron F, Dalmolin R, Azevedo A, Kaminski J (2004) Solos urbanos. **Ciência Rural** 34:1647-1653.

Peluco RG (2016) **Magnetismo e cor na identificação de solos com diferentes potenciais de sorção de imazaquim**. 105f. Tese (Doutorado em Agronomia) - Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Jaboticabal, 2016.

Pinheiro HSK, Owens PR, Anjos LHC, Carvalho Júnior W, Chagas CS (2017) Tree-based techniques to predict soil units. **Soil Research** 55: 788–798.

Rocha W, Alleoni L, Regitano, J (2003) Energia livre da sorção de imazaquim em solos ácidos. **Revista Brasileira de Ciência do Solo** 27:239-246.

Rossiter D (2008) Digital Soil Mapping as a Component of Data Renewal for Areas with Sparse Soil Data Infrastructures. In: Hartemink AE, McBratney AB, Ahrens RJ (Eds.) **Digital Soil Mapping with Limited Data**, Springer, 69–80.

Samuel-Rosa A (2019) **febr: Free Brazilian Repository for Open Soil Data**. R package version 1.0.2.

Sanchez R B, Marques Júnior J, Pereira GT, Souza Z M, Martins Filho MV (2009) Variabilidade espacial de atributos do solo e de fatores de erosão em diferentes pedoformas. **Bragantia** 68:873–884.

Santos H, Jacomine P, Anjos L, Oliveira V, Lumbreras J, Coelho M, Almeida J, Araujo Filho J, Oliveira J, Cunha T (Eds.) (2018) **Sistema brasileiro de classificação de solos**. Brasília: Embrapa, 356p.

Santra P, Kumar M, Panwar NR, DAS BS (2017). Digital Soil Mapping and Best Management of Soil Resources: A Brief Discussion with Few Case Studies. In: Rakshit A, Abhilash PC, Singh HB, Ghosh S (Eds.) **Adaptive Soil Management: From Theory to Practices**. Springer, 3-38.

Sarmiento EC, Giasson E, Weber EJ, Flores CA, Rossiter DG, Hasenack H (2014) Caracterização de mapas legados de solos: uso de indicadores em mapas com diferentes escalas no Rio Grande do Sul. **Revista Brasileira de Ciência Do Solo** 38:1672–1680.

Scopus (2019). Disponível em: <https://www.scopus.com/results/results.uri?sort=plf-f&src=s&st1=%22aeromagnetic&st2=data%22&sid=4d3611d05beb55d667f3b6daa46a7737&sot=q&sdt=b&sl=39&s=TITLE-ABS-KEY-AUTH%28%22aeromagnetic+data%22%29&origin=browserextensionddm&editSaveSearch=&txGid=28c98bff183ac52415bb16c989016d75>. Data de acesso: 11 de novembro de 2019.

Silva LS (2016) **Mineralogia da fração argila dos solos do Planalto Ocidental Paulista**. 83f. Dissertação (Mestrado em Agronomia) - Unesp, Jaboticabal.

Smellie DW (1956) Elementary approximations in aeromagnetic interpretation. **Geophysics** 21:1021–40.

Stanton N, Schmitt R, Galdeano A, Maia M, Mane M (2010) Crustal structure of the southeastern Brazilian margin, Campos Basin, from aeromagnetic data: New kinematic constraints. **Tectonophysics** 490, 15–27.

Teixeira,P, Donagemma G, Fontana A, Teixeira W (Eds.) (2017) **Manual de métodos de análise de solo**. Brasília, Embrapa, 573p.

Terribile F, Coppola A, Langella G, Martina M (2011) Basile, A. Potential and limitations of using soil mapping information to understand landscape hydrology. **Hydrology and Earth System Sciences** 15:3895–3933.

Teng H, Viscarra Rossel RA, Shi Z, Behrens T, Chappell A, Bui E (2016) Assimilating satellite imagery and visible–near infrared spectroscopy to model and map soil loss by water erosion in Australia. **Environmental Modelling & Software** 77:156–167.

Teixeira, L (2017) **Mapeamento da erodibilidade de latossolos utilizando a cor**. 87f. Tese (Doutorado em Agronomia) - Unesp, Jaboticabal, 2017.

Rossel, RAV (2011) Fine-resolution multiscale mapping of clay minerals in Australian soils measured with near infrared spectra. **Journal of Geophysical Research** 116:F04023 .

Veiga Filho AA, Souza MCM, Martins NB, Yanaguishi CT, Matsunaga M (1992) Análise de investimentos em adoção de tecnologias e conservação do solo no estado de São Paulo. **Agricultura em São Paulo** 39:133- 154.

Warner, TA, Levandowski DW, Bell R, Cetin H (1994) Rule-based geobotanical classification of topographic, aeromagnetic, and remotely sensed vegetation community data. **Remote Sensing of Environment** 50:41–51.

Wischmeier WH, Johnson CB, Cross BV (1971) Soil erodibility nomograph for farmland and construction sites. **J Soil Water Conserv** 26:189–193.

Zeraatpisheh M, Ayoubi S, Jafari A, Tajik S, Finke P (2019) Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. **Geoderma** 338:445–452.

CAPÍTULO 2 - Evaluating the potential of aeromagnetic data for mapping clay content at regional and local scales in Brazil

ABSTRACT - The São Paulo State is one of the most important agricultural areas of Brazil, with highly heterogeneous soils. Geology exerts an important control on this variability, but existing geological maps are often coarse. While aeromagnetic surveys are available nationwide, and have been widely used for geological and petroleum exploration, at present they have not been used yet for digital soil mapping studies. This motivated this study to identify the potential of aeromagnetic data for the prediction of clay content. A study was made at two different scales: at regional (Western Paulista Plateau, $n = 539$) and at local scale (Guatapara, $n = 334$), using the Random Forest algorithm. The variables used for the prediction were aeromagnetic data, altitude, precipitation, topographic wetness index, soil and geology. Results show that aeromagnetic data can contribute to the spatial prediction of clay content at regional and local scales, although they are more important at the latter scale. At the regional scale, in the Western Paulista Plateau, the main variables were precipitation and altitude, respectively, with an R^2 of 0.4 for test data. These variables can be associated with soil formation processes, and therefore show how these are more important due to the extension and the complexity of the Western Paulista Plateau. At the local scale, for the Guatapara area, the main variables used in the prediction were aeromagnetic data and altitude, respectively, with an R^2 of 0.74 for the test data. This can be explained by the area being less complex. It is characterized by a contact area between two large geological formations (Sandstone/Basalt) and aeromagnetic data were able to pick this up. The results show clearly that the aeromagnetic data were useful to predict the clay content at both scales, so it has an important potential for digital soil mapping in data-poor regions.

Keywords: Aeromagnetic data, Soil attributes, Soil modeling, Spatial prediction, Digital soil mapping, Random Forest, clay content.

2.1 Introduction

Humanity depends on various soil services, and its capacity for retribution depends on the promotion of its ecosystem (Adhikari e Hartemink, 2016). Knowing the soil location, as well as its classes and attributes, is important for an adequate agricultural policy, management and monitoring (Mulder et al., 2011; Junior et al., 2016). However, some soil properties are difficult to measure and change over time. Important advances have occurred in the last three decades regarding the knowledge of soils, after the development of geostatistics, with more assertive predictions and error calculation. A rapid and assertive forecast is necessary to fill the lack of information about it (Minasny e Hartemink, 2011).

Brazil, despite the productive potential, presents coarse soil maps that are insufficient for strategic planning in agriculture. It is estimated that only 0.25% of Brazilian territory presents maps of first and second order, with a scale less than 1:30,000 (Sanchez et al., 2009; Mendonça-Santos e Santos, 2007). The greatest difficulty for a satisfactory mapping is the inability to perform continuous measurements of soil properties, especially in tropical countries, where the lack of pedological information at different scales is frequent (Minasny e Hartemink, 2011; Junior et al., 2016).

Digital Soil Mapping (DSM) has been growing rapidly within soil science. The term DSM consists of the creation of a geo-referenced soil system, where quantitative techniques for spatial prediction are used in conjunction with field and laboratory observations (Minasny e Hartemink, 2011; Minasny e McBratney, 2016; Marques et al., 2018). This is possible because of the relationship that soil variables have with other environmental components, such as climatic information, landscape forms, geology, among others (Sanchez et al., 2009; Hengl et al., 2015; Gray et al., 2016).

Among these soil properties, clay is one of the fundamental elements for agricultural strategic planning, since its characteristics correspond directly to agronomic and environmental issues. Such is its relevance that it is used as a diagnostic attribute in the American soil taxonomy (USDA, 1999) and in the Brazilian Soil Classification System (Santos et al., 2018). Clay content also reflects directly on the retention of water and nutrients, organic carbon content, porosity, among others,

and that affects several mechanical functions of the soil and its properties (Akpa et al., 2014). Because they are more stable to changes over a short period of time, clay content is excellent predictors of fertility, infiltration, drainage, and aeration of the soil (Minasny e Hartemink, 2011; Bhering et al., 2016). Besides that, the clay content presents strategic characteristics of the soil and is therefore used as a variable for the prediction of soil attributes at different scales (Hengl et al., 2015, Viscarra et al., 2016, Wood et al., 2016; Hengl et al., 2017). Field sampling and sample analysis is the most accurate way to obtain clay content, however, this method requires time and money (McBratney et al., 2003), increasing according to the study area.

Knowing the relation that the clay presents with the material of origin and its minerals (Barshad, 1957), what would be the possibility of using magnetic rock information as a parameter to estimate soil attributes at different scales? Aeromagnetic surveys present a fast, inexpensive and applicable form in regions of difficult access. Aeromagnetic surveying is one of the classical methods in geophysical prospecting, which makes use of airborne geophysical surveying gauges installed in an aircraft at a certain flight height range to acquire magnetic field strength from the subsurface. It is a passive technique to measure the ambient magnetic field, and has applications in geological reconnaissance, thickness of sedimentary cover, basement faults and uplifts, basin modeling, structure size, and depth (Fairhead, 2012; Zou, 2013).

Aeromagnetic images are able to synthesize large volumes of data in an easily understood visual format (Kearey et al., 2009). For Khalil et al. (2016) aeromagnetic survey is the most powerful technique of recognition of the geological surface, since its target is the position, depth, shape, and altitude of the magnetic bodies. In spite of its potential to differentiate rock types and their contact area (Domzalski, 1966), aeromagnetic data have been little explored in soil science. As far as we know, no studies exist on its application in digital soil mapping. However, given the importance of parent material composition for the spatial distribution of soil properties (Gray et al., 2016), it has a high potential for assisting in soil mapping and agriculture in general.

In view of its importance, it is necessary to have robust techniques capable of estimating the clay content at different scales. The use of the Random Forest (RF) algorithm consists of a data mining technique capable of classifying attributes in order of importance and their variation according to the target attribute. With a relatively

recent approach, its algorithm is able to improve prediction accuracy, presenting as advance resistance to overfitting and insensitivity (Breiman, 2001; Zhang et al., 2017).

We trained and tested six input variables and one output variables with the RF model. The coefficient of determination (R^2) was used to analyze the accuracy of the results, described throughout this thesis. The Spatial variability maps of the observed and estimated clay content of Western Paulista Plateau and Guatapara were also prepared. The relative error and standard deviation maps demonstrate the accuracy of the formulated model. Therefore, the objective of this study was to identify the potential of aeromagnetic data for the prediction of clay content at two different scales using the RF model.

2.2 Materials and methods

2.2. 1 Study area

The study spanned two areas with similar geological formations: Western Paulista Plateau, as regional scale (RS), and Guatapara, as local scale (LS), shown in Fig. 1. Both regions are located in a contact region between sedimentary and basaltic rocks. This characteristic shows similarity in the process of soil and landscape formation, differentiating only in the size of the area.

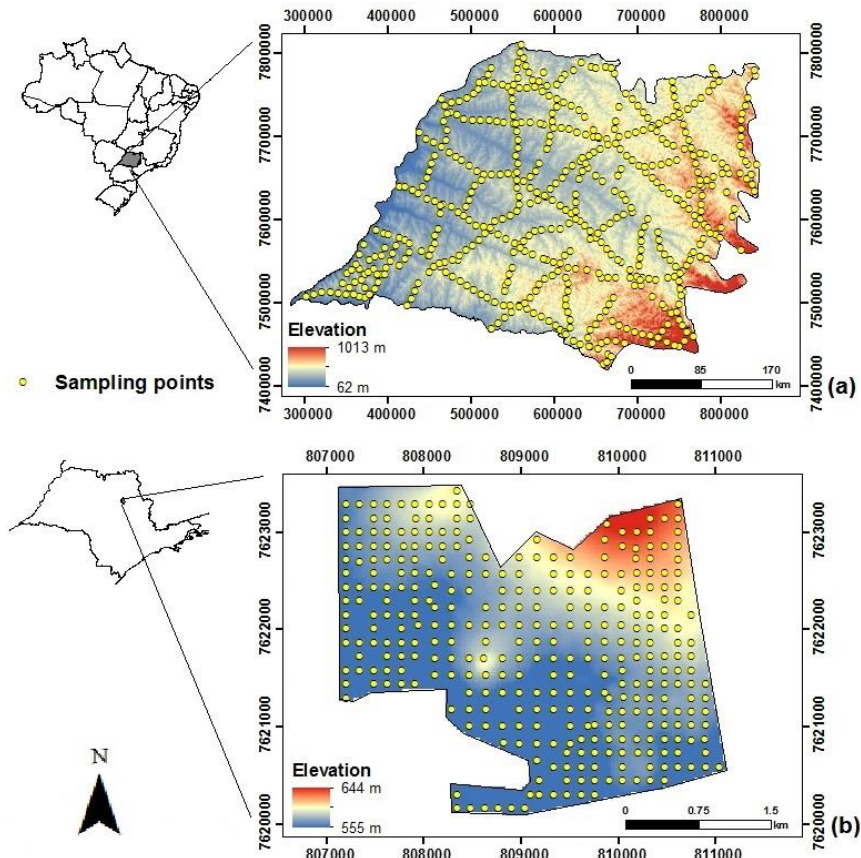


Figure 1 - Localization of Western Paulista Plateau (a), as RS and Guatapar (b), as LS, both in So Paulo, Brazil.

RS is an area of 130,000 km² (approximately 48% of the state of So Paulo), with a variation in altitude of 62 to 1013 m. The climate can be divided into three groups, according to Thornthwaite (1948). The north and northeast is characterized by tropical climate dry winters (C2rA'a '). In the southern, the temperate and humid climate predominates with the presence of hot summers (B4rB'4a). In the east and southeast of the RS the climate is temperate humid with dry winter and hot summer (B2rB'3a). The rocky substratum of the RS is basically composed of sedimentary rocks, especially the Rio do Peixe Formation. In the eastern region and the limits of the state of Paran are basaltic rocks of the Serra Geral Formation (Fernandes, 2004; Fernandes et al., 2007). The area presents sixteen soil classes, however, the ones with the highest occurrence are red yellow Argissol, ferric Oxisols, lithic Leptosol, red Nitisols and haplic Gleysols (Santos et al., 2018). The remaining natural vegetation is composed of Atlantic Forest to the west and Cerrado to the east. The main crops in the RS are sugar cane and pasture (LUPA, 2008).

The area of study in LS presents 8.7 km² with attitude variation of 550 to 644 m. According to Thornthwaite (1948), the climate is classified as humid mesothermic (B1rB'4a'), with little water deficiency and low annual evapotranspiration. The area is located on a rocky substrate composed by sedimentary rocks (Botucatu Formation) and basaltic rocks (Serra Geral Formation) (CPRM, 2014). According to the pedological map, the area presents mainly the units Ultisols, Oxisols and Entisols (Santos et al., 2018). The remaining natural vegetation in LS area is composed of semi-deciduous tropical forest, being used nowadays by sugarcane crop for more than 10 years (Teixeira et al., 2018).

2.2.2 Sampling and measurement of clay

In RS, 539 samples were collected between 0.0 - 0.20m, in environments with minimal anthropogenic interference. The sampling route plan followed the highways of the State of São Paulo, as can be seen in Fig.1a. The highway follows the main soil classes (southeast to northwest), guaranteeing representative samples. The minimum spacing between each sample was 10 km and the maximum spacing was 60 km. In LS, 334 samples were collected between 0.0 - 0.20m along a regular mesh, separated by minimum and maximum distances of 145 and 174 m, respectively, as can be seen in Fig.1b.

Laboratory analysis of clay content was done using the pipette method. The method consists of using NaOH 0.1 mol L⁻¹ as a chemical dispersant and mechanical stirring at low speed for 16 hours (Teixeira et al., 2017).

2.2.3 Input datasets

The elaboration of the RF model used as few input variables as possible, seeking to maintain the low correlation between the trees during the learning of the algorithm. The changes in the dataset may cause changes relevant to the model due to the sensitivity of the generated trees. The few input variables reduce the complexity of the model.

The images of the aeromagnetic data (Aeromag) were obtained from the Mineral Resources Research Company (Bizzi et al., 2003), with scale 2.500.000. For the soil (Soil) and geology (Geo) classes, the geo-referenced database of the Brazilian Institute of Geography and Statistics (IBGE, 2018) was used in the 1: 250,000 scale.

The Digital Elevation Model (DEM) TOPODATA (Valeriano, 2005), made available by the National Institute of Space Research (INPE), was used to calculate the altitude (Elev), with a scale of 1: 250,000. The DEM image was organized in mosaic with 13 images so that they could cover the whole RS and LS areas. The Topographic Wetness Index (TWI) (Beven e Kirkby, 1979) was calculated using TWI Tools in ESRI® ArcMap 10.3 based on the DEM.

For the calculation of precipitation (Precip), the data of the Normal Climatology of the Brazilian Institute of Meteorology (INMET, 2018) were used. The database presents the monthly average rainfall data in a period of 30 years (1981-2010). To include all study area, the stations present in São Paulo (16 weather stations) and states of Rio de Janeiro, Minas Gerais, Mato Grosso do Sul and Paraná (83 weather stations) were used. For the interpolation, ordinary kriging with 1: 250,000 scale was used. More details on this methodology can be found in the work of Lloyd (2005), who used a similar methodology.

After treatment of the individualized variables, all were imported into environment R (R Core Team, 2019) and standardized with the DEM in scale (1: 250,000) and pixel (30 m), using the *raster package* (Hijmans, 2019). The individual value of the predicted variables was assigned to each sample collected in the field, referenced in x and y, and transformed into a data frame.

2.2.4 Data analysis

Data of clay content and input variables analyzed by descriptive statistics (Table 1) were presented with the calculation of minimum, maximum, average, coefficient of variation and standard error of mean for RS and LS. Null values or outlier during data preprocessing were excluded from the analysis.

To evaluate the clay content, the samples were randomly divided into 70% for training and 30% for tests, using the *randomForest package* (Liaw e Wiener, 2002).

The algorithm creates decision trees (ntrees) and combines them by the average to obtain greater accuracy. Initially, 1000 ntrees were added to the model to avoid possible overfitting. This total of trees can reduce data variation, however, it does not correct bias. In the prediction process, six variables were classified, from the most influential (highest) to the least influential (lowest). This index illustrates the decrease of the impurities of the division node of each variable for the prediction, calculated by means of the mean and using all the attributes inserted in the model. The higher the value of the index, the greater the importance of the variable used in the regression. The impurity generated is measured by the mean square error, which may increase due to the random exchange of the variables used (Liaw e Wiener, 2002; Álvarez-Cabria et al., 2017). In addition, the importance graph of the predictor variables according to the RF model was generated. A model for each study scale (RS and LS) was generated, with different sample mesh and datasets. For the each model response, the *ithir package* (Malone, 2018) was used, aiming at the best result with R^2 (Akpa et al., 2014). The coefficient of determination (R^2) was used to explain the fit of the data within the obtained model, where the closer to zero, the lower the explanatory capacity of the model (James et al., 2013) for the prediction of clay content.

The interpolation of the observed and predicted clay content was done by ordinary kriging. The spatial variability of the clay was calculated from the experimental variogram modeling, based on the theory of regionalized variables (Isaaks e Srivastava, 1989). The definition of the best model to be used was based on RSS and R^2 .

2.3 Results and discussion

The clay content measured in the field in RS and LS ranged between 12.5 - 667 g kg⁻¹ and 49 - 693.82 g kg⁻¹, respectively, as presented in Table 1. The high coefficient of variation in clay content and aeromagnetic data demonstrates the high variability of the dataset. These characteristics may be related to the geological and pedological diversity found in the study area (Peluco et al., 2013). The Aeromagnetic data showed a similar behavior to clay, with an important variation and similar range

at the two spatial scales studied. Values for aeromagnetic data range from 33 - 164.67 and 40.33 - 140 grayscale in RS and LS, respectively.

Precipitation presented relevant variation only in RS due to its size, while in the much smaller area of LS precipitation does not show any important variation. The important variation of Elevation and TWI at both scales demonstrates the complexity of the terrain, although the variation at the RS is always highest, as expected. In both areas, RS and LS, categorical variables of soil classes ($n = 16$ and 3 , respectively) and geology ($n = 20$ and 2 , respectively) were also taken into account in the RF model.

Table 1 - Descriptive statistics for the predicted variables.

| | Study area | Min. | Max. | Mean | CV | SE |
|--|------------|-------|--------|--------|-------|------|
| Clay content (g kg^{-1}) | RS | 12.50 | 667.00 | 174.97 | 74.58 | 5.62 |
| | LS | 49.00 | 693.82 | 327.40 | 51.76 | 9.27 |
| Aeromagnetic data (Gray scale) | RS | 33.00 | 164.67 | 92.79 | 39.73 | 1.59 |
| | LS | 40.33 | 140.00 | 83.70 | 31.66 | 1.45 |
| Elevation (m) | RS | 280 | 902 | 495 | 21.69 | 4.63 |
| | LS | 520 | 646 | 563 | 4.38 | 1.35 |
| Precipitation (mm year^{-1}) | RS | 1376 | 1551 | 1468 | 2.68 | 1.69 |
| | LS | 1512 | 1518 | 1513 | 0.02 | 0.02 |
| Topographic Wetness Index (dimensionless) | RS | -1.19 | 10.31 | 6.53 | 54.03 | 0.15 |
| | LS | 2.19 | 10.67 | 4.59 | 35.08 | 0.09 |

CV = Coefficient of variation (%); SE = Standard Error.

At both scales, aeromagnetic data was identified as an important variable in the model, as shown in relevance of the predictor in RF model (Fig. 2). At the RS, precipitation was considered the most important variable to estimate the clay content in relation to the other variables. The reason for this can be in the impact of soil forming processes on clay content, as temperature and water dynamics directly affect the weathering processes and the production of clay in the soil (Huang et al., 2011). Rasmussen et al. (2010) state the significant formation of clays in tropical environments formed by basaltic rocks influenced by temperature and precipitation. The importance of the climate demonstrated by the RF model confirms its importance

as an element of soil formation (Budiman et al., 2003). The Elevation variable appears as the second most influential in the RF model. This further confirms the importance of soil weathering in this environment as elevation are directly related. This also corroborates with Basist et al. (1994), who state that elevation is capable of influencing wind patterns and condensation of rainwater. The Aeromagnetic data appeared as a relevant variable in the RF model in third place. Its importance is similar to that of Elevation and Soil, the next relative important in the model. Both variables are possibly related to the considerable amount of iron oxides present in tropical soils (Schwertmann e Kämpf, 1985). Obi et al. (2014) confirm the relationship between precipitation, topography elements and soil properties with soil formation in southeastern Nigeria in an area with 8412 km².

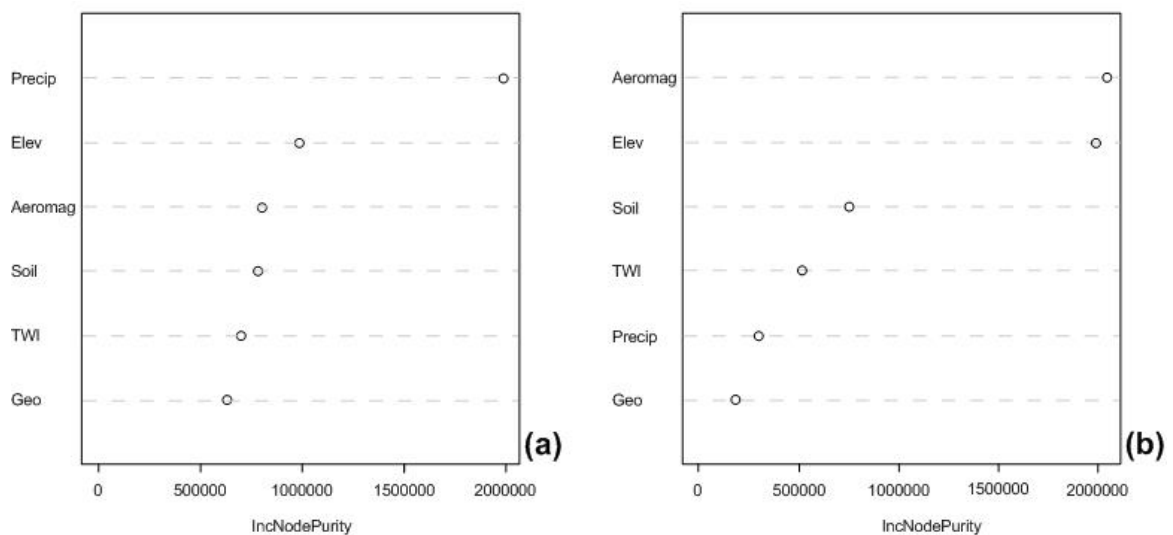


Figure 2 - Relative importance of variables for prediction of clay in Regional Scale (a) and Local Scale (b).

The TWI and Geology could be considered the least important variables in the environment for the prediction of clay content. However, the strong relationship between geology and soil forming processes is widely known, from similar studies in other environments. Gray et al. (2016), for example, demonstrated that the lithological variable was the most influential for the RF algorithm when applied in eastern Australia in an area of 81,000 hectares. It can be concluded that the proposed RF model

identified mostly variables related to different soil forming factors that control clay formation in the RS.

At the LS, as a less complex environment, aeromagnetic data was found to be the most important variable to estimate the clay content. LS are located between two important geological formations, formed by sedimentary and basaltic rocks. Both rocks in the area are of different developments. While the basaltic rock is characterized by the high concentration of iron oxides, consequence of the cooling of the magma, sedimentary rocks are formed by the deposition and consolidation of rock fragments from other areas. In this way, the minerals of the clay fraction are heavily influenced by the parent material composition, and differences in this composition are picked up by aeromagnetic data (Owliaie et al., 2006). These characteristics result in greater RS complexity in the process of soil formation and in the formation and origin of the clay content.

The Elevation variable presented a similar importance to that of the first variable, aeromagnetic data. This result corroborates with Pahlavan-Rad e Akbarimoghaddam (2018), where elevation was the second most important variable for the prediction of clay content in an area of 410 km² eastern Iran. It also corroborates the results obtained at the RS. However, the main difference here is that precipitation is found much less important. This is understood due to homogeneous rainfall variation in LS.

Soil and TWI appeared in sequence as variables of importance to the model. This indicates the importance of soil forming processes on the spatial distribution of clay content. A higher intensity of clay forming processes can be expected in those landscape positions where water collects, as indicated by TWI. The importance of the last two variables, Precipitation and Geology, in the model is low. As mentioned before, the reason for this is the low variation of these variables, due to the small size of the study area.

Fig. 3 shows the relation between the clay content observed in the field and the one predicted with the obtained RF models. The R² of the training dataset in both areas was similar, respectively 0.91 and 0.90 for RS and LS. However, it was in LS that the best results were obtained during the test. The model generated for the RS presented a R² of 0.40 and for LS of 0.74 for the test dataset. From the aeromagnetic data the RF model was able to generate more accurate predictions. The obtained values are in

the same range as those obtained by other authors in Brazil and world, using the RF model. Bhering et al. (2016), in the state of Mato Grosso do Sul, Brazil, obtained an R^2 of 0.40 in the prediction of clay content, in an area of 19,911 km². Akpa et al. (2014) in an area of 923, 76 km² located in Nigeria, obtained an R^2 of 0.57 in the prediction of clay content. The results demonstrate the potential to use such aeromagnetic data in areas such as DSM and precision agriculture.

The tests of the clay content from the aeromagnetic data in LS were efficient in the decision making of the model. A possible solution to improve estimation over large areas would be to use several small study areas within the same environment. This would enable the RF model to more easily identify more homogeneous areas. In addition, it provides knowledge of large territory on an ultra-detailed scale.

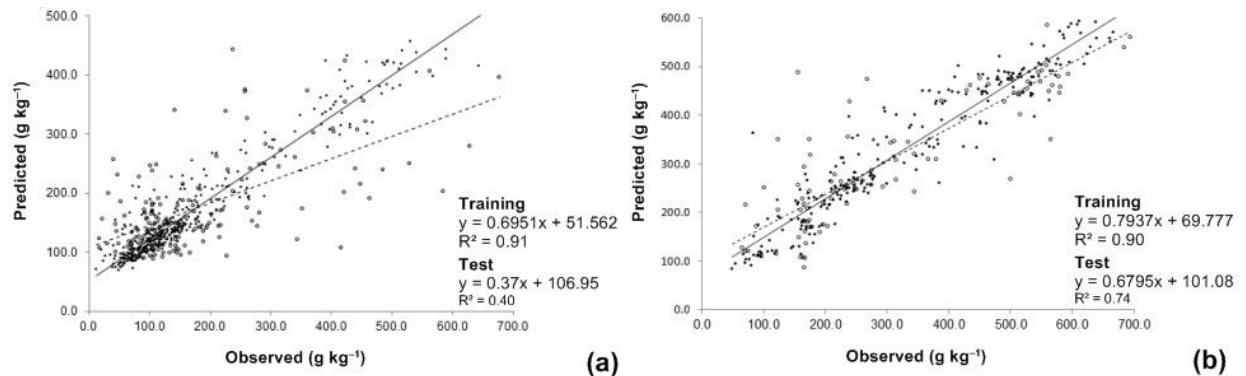


Figure 3 - Scatter plot of training versus test values of clay content by random forest in Regional Scale (a) and Local Scale (b).

The values of the clay content observed and predicted in both scales were submitted to geostatistical analysis. All samples were used to construct the semivariogram. The spherical model was the best fit in both cases, with R^2 equal to 0.65 and 0.61 for clay observed and predicted in RS and 0.71 and 0.81 for clay observed and predicted in LS. This type of model has characteristics of abrupt change in space and main for soil classification (Isaaks e Srivastava, 1989). The range indicates the area of influence of the variable in space. It can be useful for field sample planning, as their result shows the minimum distance between samples (Isaaks e Srivastava, 1989; Teixeira et al., 2018). The range values of the predicted clay were slightly higher when compared to the clay observed in the RS (145.8 and 123.41 km) and in LS (2.56 and 2 km). At both scales for clay content prediction the degree of

spatial dependence was strong. The prediction of the clay content observed presented moderate degree of spatial dependence in RS. All the others presented strong spatial dependence.

Table 2 - Estimated models for semivariogram adjustment.

| | Study area | Model | C₀^a | C₀+C₁^b | A^c | R^{2d} | C₀/[C₀+C₁]^e | RSS^f |
|--|-------------------|--------------|----------------------------------|--|----------------------|-----------------------|--|------------------------|
| Clay_{obs}^g | RS ⁷ | Spherical | 0.16 | 0.43 | 123.41 | 0.65 | 37 | 0.04 |
| | LS ⁸ | Spherical | 0.008 | 0.39 | 2 | 0.71 | 2 | 0.15 |
| Clay_{pred}^h | RS | Spherical | 0.03 | 0.16 | 145.8 | 0.61 | 18 | 0.03 |
| | LS | Spherical | 0.001 | 0.33 | 2.56 | 0.81 | 0.30 | 0.09 |

(a) Nugglet effect; (b) Sill; (c) Spatial Range (km); (d) Coefficient of determination; (e) Degree of spatial dependence (strong with $\leq 25\%$, moderate with 25–75%, weak with $> 75\%$); (f) Residual sum of squares; (g) Observed clay content, (h) Predicted clay content by Random Forest.

The spatial pattern of clay content observed in RS (Fig. 4a) and LS (Fig. 4b) were similar to clay predicted by the RF model (Fig. 4c and 4d, respectively). Hengl et al. (2015) point out that the variables that should be used in the DSM are those that best represent the soil formation process and the terrestrial characteristics. Despite close values in both areas of clay content (Table 1), the range of RS relative error (Fig. 4e) was higher relative to LS (Fig. 4f). This characteristic of the error range is related to the distance of the sampling mesh, thus increasing the prediction difficulty of the algorithm. The exposed standard deviation is used to calculate variability within the dataset. This behavior can be observed in RS (Fig 4g), ranging from 47 to 67% and LS (Fig. 4h), ranging from 31 to 39%. The uncertainty obtained by the RF model demonstrates the difficulty of prediction in large areas, requiring caution in choosing and processing datasets. The generated model obtained smaller uncertainties in LS due to the distance between the samples.

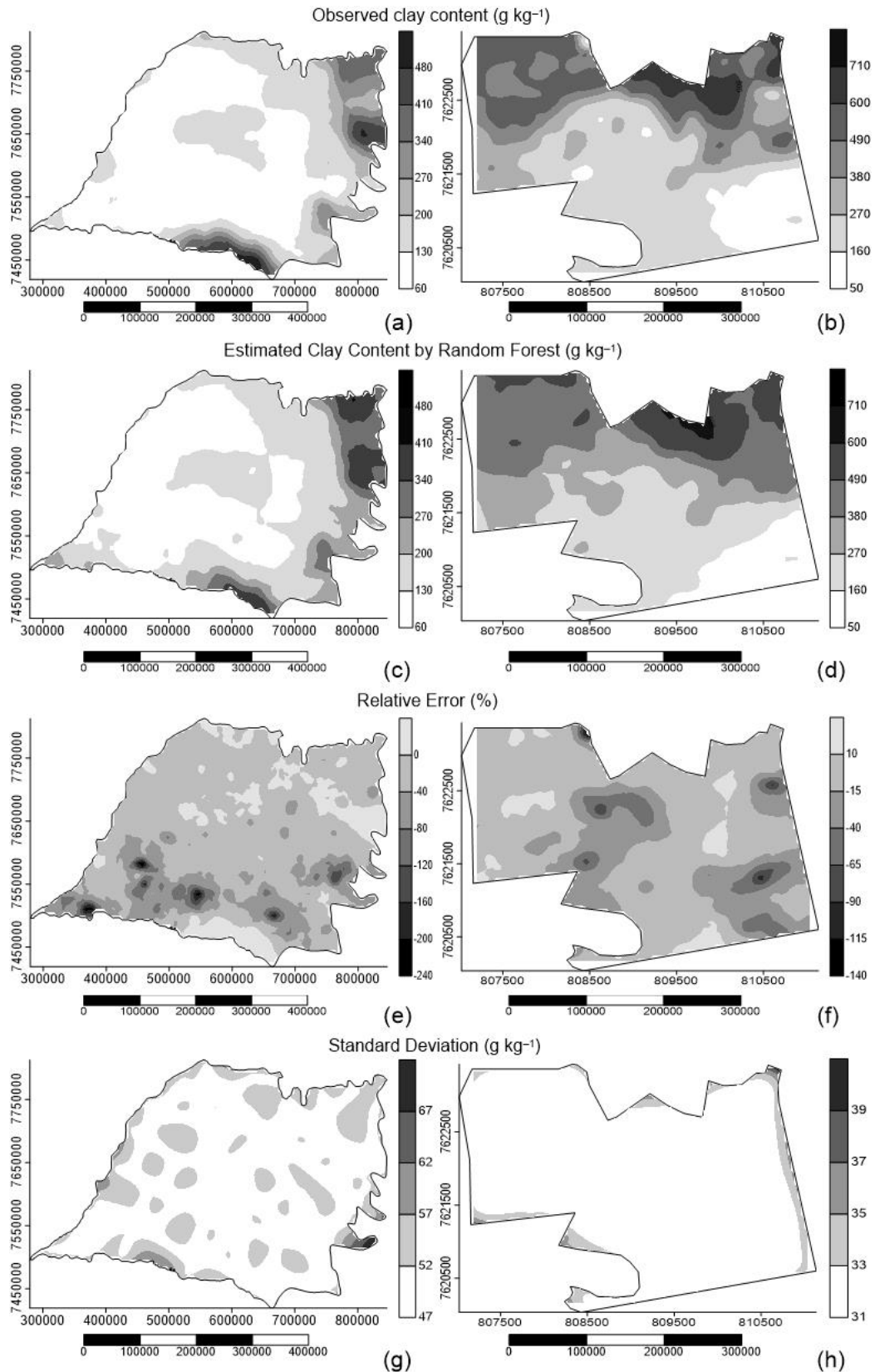


Figure 4 - Spatial pattern of observed clay content (g kg^{-1}) (a and b), estimated by Random Forest (g kg^{-1}) (c and d), standard deviation (g kg^{-1}) (e and f) and relative error (%) (g and h) in Western Paulista Plateau and Guatapara.

From these results it was clear that the aeromagnetic data, together with other input variables, can be used to predict clay content at different scales. The lack of soil information is a delay in the development of several sectors in agricultural production. The aeromagnetic data are able to compensate for the need that Brazil has of information on the characteristics of the soil. The studies using DSM commonly use landscape variables, but hardly incorporate soil forming processes (Ma et al., 2019). Only a few studies have explored this area. For example, in Brazil, Bonfatti et al. (2018) use a soil formation model to predict soil depth in a small area in Rio Grande do Sul. The relatively high accuracy presented by the RF model obtained in this study indicates that the use of aeromagnetic data, apart from predicting clay content, can possibly be extended to other soil properties and his functions as well. Having proved the relationship that aeromagnetic data has with the soil texture, this data can be tested in future studies for the knowledge of agricultural problems. With the availability and price of agricultural products and sensors, the aeromagnetic data can be a cost-effective and easily accessible alternative at different spatial scales. In addition, it can be used for other studies in tropical environments, as the RS and LS are representative for processes of soil formation.

2.4 Conclusions

The present study demonstrated the potential of aeromagnetic data to predict clay content at different scales. The Random Forest model was able to successfully predict clay content at different scales, consistent with studies performed in other parts of the world.

A better prediction with the Random Forest model was obtained at the local scale ($R^2 = 0.74$) followed by the regional scale ($R^2 = 0.40$) in the test procedure. At both scales, aeromagnetic data was a relevant variable, although its importance was much higher at the local scale. At the regional scale, precipitation, elevation, and aeromagnetic data indicate the influence of soil formation factors on the spatial variability of clay content. At the local scale, in LS, aeromagnetic data was more influential, indicating that the intrinsic characteristics of the parent material control the clay content variation in the area.

In both scales, the level of detail was decisive for the decision making of the model. Even with similar geological features at the local and regional scale, the regional scale proved to be much more complex. This complexity made it difficult to find patterns between aeromagnetic data and clay content. This first study confirms the potential of aeromagnetic data for mapping clay content at different scales. Further studies are needed to confirm the use of aeromagnetic data in other agricultural applications.

Acknowledgements

This study was conducted during a visiting scholar period at University of Córdoba, funded by CAPES/PDSE (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior/Programa de Doutorado Sanduíche no Exterior), nº 47/2017. This work was supported by CAPES (Proc. nº 149940); PROPE/UNESP (Pró-reitora de Pesquisa da Universidade Estadual Paulista, Edital Nº 15/2014); and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, Edital Universal 01/2016 – Faixa C, Proc.nº 402796/2016-0).

References

- Adhikari K, Hartemink AE (2016) Linking soils to ecosystem services — A global review. **Geoderma** 262:101–111.
- Álvarez-Cabria M, González-Ferreras AM, Peñas FJ, Barquín J (2017) Modelling macroinvertebrate and fish biotic indices: From reaches to entire river networks. **Science of the Total Environment** 577:308–318.
- Akpa SIC, Odeh IOA, Bishop TFA, Hartemink AE (2014) Digital Mapping of Soil Particle-Size Fractions for Nigeria. **Soil Science Society of America Journal** 78:1953-1966.
- Barshad I (1957) Factors Affecting Clay Formation. **Clays and Clay Minerals** 6:110–132.
- Basist A, Bell, GD, Meentemeyer, V (1994) Statistical Relationships between Topography and Precipitation Patterns. **Journal of Climate** 7:1305–1315.

Beven KJ, Kirkby MJ (1979) A physically based, variable contributing area model of basin. **Hydrological Sciences Bulletin** 24:43-69.

Bhering SB, Chagas CS, Junior WC, Pereira, NR, Filho BC, Pinheiro HS (2016) Mapeamento digital de areia, argila e carbono orgânico por modelos Random Forest sob diferentes resoluções espaciais. **Pesquisa Agropecuária Brasileira** 51:1359–1370.

Bizzi LA, Schobbenhaus C, Gonçalves JH, Vidoti RM, Gonçalves JH (Eds.) (2003) **Geologia, tectônica e recursos minerais do Brasil: texto, mapas & SIG**. Brasília: CPRM, .643p.

Bonfatti BR, Hartemink AE, Vanwalleggem T, Minasny B, Giasson E (2018) A mechanistic model to predict soil thickness in a valley area of Rio Grande do Sul, Brazil. **Geoderma** 309:17–31.

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

Budiman M, McBratney AB, Mendonça-Santos ML, Santos HG (2003) Revisão sobre funções de pedotransferência (PTFs) e novos métodos de predição de classes e atributos do Solo. Rio de Janeiro: Embrapa, 47p.

Domzalski W (1966) Importance of aeromagnetism in evaluation of structural control of mineralization. **Geophysical Prospecting** 14:273–291.

Fairhead JD (2012) Regional tectonics and basin formation: The role of potential field studies. In: Roberts DG, Bally AW (Eds.) *Regional Geology and Tectonics: Principles of Geologic Analysis*. Elsevier, 331-341.

Fernandes LA (2004) Mapa litoestratigráfico da parte oriental da Bacia Bauru (PR, SP, MG), escala 1: 1.000.000. **Boletim Paranaense de Geociências** 55: 53–66.

Fernandes LA, Castro AB, Basilici G (2007) Seismites in continental sand sea deposits of the Late Cretaceous Caiuá Desert, Bauru Basin, Brazil. **Sedimentary Geology** 199: 51–64.

Gray JM, Bishop TFA, Wilford JR (2016) Lithology and soil relationships for soil modelling and mapping. **Catena** 147: 429–440.

Hengl, T.; Heuvelink, G.; et al. (2015) Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. **PLoS ONE** 10:1–26.

Hengl T, Mendes de Jesus J, et al. (2017) SoilGrids250m: Global gridded soil information based on machine learning. **PLOS ONE** 12, n. 2:e0169748.

Hijmans RJ (2019) raster: Geographic Data Analysis and Modeling. R package version 2.8-19.

Huang C, Zhao W, Liu F, Tan W, Koopal LK (2011) Environmental significance of mineral weathering and pedogenesis of loess on the southernmost Loess Plateau, China. **Geoderma** 163:219–226.

IBGE 2019: Instituto Brasileiro de Geografia e Estatística. Mapa pedológico do Brasil. Disponível em: <https://www.ibge.gov.br/> . Data de acesso: : 22 de fevereiro de 2019

IBGE 2019: Instituto Brasileiro de Geografia e Estatística. Mapa geológico do Brasil. Disponível em: <https://www.ibge.gov.br/> . Data de acesso: : 22 de fevereiro de 2019.

INMET 2018: Instituto Nacional de Meteorologia. Normais Climatológicas do Brasil, 1981-2010. Disponível em: <http://www.inmet.gov.br/>. Data de acesso: : 22 de fevereiro de 2019.

Isaaks EH, Srivastava RM (1989) Applied Geostatistics. Oxford University Press: Nova York, 570p.

James G, Witten D, Hastie T, Tibshirani R (2013) **An Introduction to Statistical Learning with Applications in R**. Springer, 426p.

Junior WC, Filho BC, Chagas CS, Bhering SB, Pereira NR, Pinheiro HSK (2016) Regressão linear múltipla e modelo Random Forest para estimar a densidade do solo em áreas montanhosas. **Pesquisa Agropecuária Brasileira** 51:1428–1437.

Khalil A, Hafeez THA, Saleh HS, Mohamed WH (2016) Inferring the subsurface basement depth and the structural trends as deduced from aeromagnetic data at West Beni Suef area, Western Desert, Egypt. **NRIAG Journal of Astronomy and Geophysics** 5:380–392.

Kearey P, Brooks, M HILL, I (2009) **An Introduction to Geophysical Exploration**. Blackwell Science, Malden, 272p.

LUPA 2008: Levantamento Censitário das Unidades de Produção Agropecuária. Levantamento Censitário das unidades de produção agropecuária do estado de São Paulo. Disponível em: <http://www.cati.sp.gov.br/projetolupa>. Data de acesso: 22 de fevereiro de 2019.

Liaw A, Wiener M (2002) Classification and regression by random- Forest. **Rnews** 2:18-22.

Lloyd CD (2005) Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain. **Journal of Hydrology** 308:128–150.

Ma Y, Minasny B, Welivitiya WDDP, Malone BP, Willgoose GR, McBratney AB (2019) The feasibility of predicting the spatial pattern of soil particle-size distribution using a pedogenesis model. **Geoderma** 341:195–205.

Malone B (2018) *ithir*: Soil data and some useful associated functions. R package version 1.0.

Marques KPP, Demattê JAM, Miller BA, Lepsch IF (2018) Geomorphometric segmentation of complex slope elements for detailed digital soil mapping in southeast Brazil. **Geoderma Regional** 14:1-9.

Mendonça-Santos ML, Santos HG (2007) The State of the Art of Brazilian Soil Mapping and Prospects for Digital Soil Mapping. **Developments in Soil Science** 31:39–55.

Minasny B, Hartemink AE (2011) Predicting soil properties in the tropics. **Earth-Science Reviews** 106:52–62.

Minasny B, McBratney, AB (2016) Digital soil mapping: A brief history and some lessons. **Geoderma** 264:301–311.

Mulder VL, Bruin S, Schaepman ME, Mayr, TR (2011) The use of remote sensing in soil and terrain mapping - A review. **Geoderma** 162:1–19.

McBratney AB, Mendonça Santos ML, Minasny B (2003) On digital soil mapping. **Geoderma** 117:3–52.

Obi JC, Ogban PI, Ituen UJ, Udoh BT (2014). Development of pedotransfer functions for coastal plain soils using terrain attributes. **Catena** 123:252–262.

Owliaie HR, Abtahi A, Heck RJ, (2006) Pedogenesis and clay mineralogical investigation of soils formed on gypsiferous and calcareous materials, on a transect, southwestern Iran. **Geoderma** 134:62–81.

Pahlavan-Rad MR, Akbarimoghaddam A (2018) Spatial variability of soil texture fractions and pH in a flood plain (case study from eastern Iran). **Catena** 160:275–281.

Peluco RG, Júnior JM, et al. (2013). Suscetibilidade magnética do solo e estimação da capacidade de suporte à aplicação de vinhaça. **Pesquisa Agropecuaria Brasileira** 48:661–672.

Rasmussen C, Dahlgren RA, Southard RJ (2010) Basalt weathering and pedogenesis across an environmental gradient in the southern Cascade Range, California, USA. **Geoderma** 154:473–485.

R Development Core Team (2019) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna Austria. <http://www.Rproject.org/>.

Santos H, Jacomine P, Anjos L, Oliveira V, Lumbrellas J, Coelho M, Almeida J, Araujo Filho J, Oliveira J, Cunha T (Eds.) (2018) **Sistema brasileiro de classificação de solos**. Brasília: Embrapa, 356p.

Sanchez PA, Ahamed S, et al. (2009). Digital Soil Map of the World. **Science** 325:680–681.

CPRM 2014: Serviço Geológico do Brasil. Mapa Geológico do Brasil. Disponível em: <http://geobank.sa.cprm.gov.br/>. Data de acesso: : 22 de fevereiro de 2019.

Soil Survey Staff (1999). Soil Taxonomy: a Basic System of Soil Classification of Making and Interpreting Soil Surveys. USDA - Natural Resources Conservation Service: Washington, 886p.

Schwertmann U, Kämpf N (1985). Properties of goethite and hematite in kaolinitic soils of southern and central Brazil. **Soil Science** 139:344-350.

Teixeira DB, Marques Júnior J, Siqueira DS, Vasconcelos V, Carvalho OA, Martins ES, Pereira GT (2018) Mapping units based on spatial uncertainty of magnetic susceptibility and clay content. **Catena** 164:79–87.

Teixeira,P, Donagemma G, Fontana A, Teixeira W (Eds.) (2017) **Manual de métodos de análise de solo**. Brasília, Embrapa, 573p.

Thornthwaite CW (1948) An Approach toward a Rational Classification of Climate. **Geographical Review** 38:55-94.

Valeriano MM (2005) Modelo digital de variáveis morfométricas com dados SRTM para o território nacional: o projeto TOPODATA. In: XII SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO. **Anais eletrônico...** Goiânia: INPE, p. 1-8.

Viscarra Rossel RA, Bui EN (2016) A new detailed map of total phosphorus stocks in Australian soil. **Science of the Total Environment** 542:1040–1049.

Wood EF, Nauman TW, Brungard CW, McBratney AB, Hempel JW, Chaney NW, Odgers NP (2016) POLARIS: A 30-meter probabilistic soil series map of the contiguous United States. **Geoderma** 274:54–67.

Zhang H, Wu P, Yin A, Yang X, Zhang M, Gao C, (2017) Prediction of soil organic carbon in an intensively managed reclamation zone of eastern China: A comparison

of multiple linear regressions and the random forest model. **Science of the Total Environment** 592:704–713.

Zou C, Zhang G, Zhu R, Yuan X, Zhao X, Hou L, Wen B, Wu X (2013) Volcanic reservoirs in petroleum exploration. Elsevier, 196p.

CAPÍTULO 3 - Avaliação dos dados aeromagnéticos aplicados a agricultura em escala regional

RESUMO - Boas práticas agrícolas são necessárias para diminuir a degradação do solo, bem como aproveitar o seu potencial. O objetivo do trabalho é avaliar o potencial dos dados aeromagnéticos como variável de entrada para estimar a erodibilidade e o coeficiente de Freundlich para sorção de herbicida nos solos do Planalto Ocidental Paulista. Foi utilizado um banco de dados legados com 549 amostras de solo, incluindo o atributo argila, erodibilidade e coeficiente de Freundlich para sorção de herbicida. Para a elaboração do modelo foram utilizados os dados aeromagnéticos e as variáveis solo, geologia, elevação, declividade, precipitação, TWI, NDVI e EVI. As variáveis passaram por tratamento de valores nulo ou banco. O modelo utilizado foi o Random Forest com hiperparâmetro de 1000 árvores (*ntrees*). Os resultados do coeficiente de determinação para treinamento e teste do modelo de erodibilidade estimada foram de 0.93 e 0.48, respectivamente. As variáveis precipitação, elevação, solo e os dados aeromagnéticos foram as mais importantes para o modelo Random Forest. Para o coeficiente de Freundlich para sorção de herbicida, o resultado estimado de treinamento e teste foi de 0.90 e 0.36, respectivamente. As variáveis precipitação, dados aeromagnéticos e a elevação se mostraram as mais importantes para o modelo Random Forest. Os resultados demonstram a relação que os dados aeromagnéticos apresentam em área de transição geológica de rochas sedimentares e basálticas. Sendo assim, os dados aeromagnéticos apresentaram potencial para estimar a erodibilidade e o coeficiente de Freundlich para sorção de herbicida, possibilitando a aplicação em outros ambientes de transição geológica.

Palavras-chave: Dados aeromagnéticos, Modelagem do solo, Random Forest, Erodibilidade, Herbicida.

3.1. Introdução

Informações detalhadas sobre o uso e cobertura do solo são essenciais para as ciências agrárias em geral. Estima-se que as áreas de agricultura e a pastagem juntas somam mais de 40% das zonas da superfície terrestre de área ocupada (Foley et al., 2005). Boas práticas agrícolas são necessárias para diminuir a degradação do solo, bem como prevenção de erosão e manutenção da fertilidade (Bogner et al., 2018). Contudo, a degradação do solo vem crescendo em grandes áreas (Nguyen et al., 2019). São necessárias técnicas com abordagens rápidas e adequadas para prevenção e controle do uso e ocupação do solo em escala e níveis de complexidade diversos.

O Brasil, apesar do potencial agrícola, apresenta mapeamentos de escala grosseira para aplicação em campo. Estima-se que apenas 0.25% do território nacional apresente algum tipo de mapa de solos de primeira ou segunda ordem com escala menor que 1: 30000 (Mendonça-Santos e Santos, 2007; Sanchez et al., 2009). A maior dificuldade é a falta de verificações contínuas dos atributos do solo. A falta de informações pedológicas em escalas detalhada é frequente para regiões de solos tropicais (Minasny e Hartemink, 2011; Junior et al., 2016).

É de suma importância que o aumento da produção agrícola venha em conjunto com o desenvolvimento sustentável, minimizando os impactos gerados pelas atividades agrícolas (Hobbs, 2007). Dos diversos problemas encontrados relacionados ao solo, dois merecem destaque na agricultura: a erosão e a sorção de herbicida do solo. A erodibilidade tem relação direta com a perda do solo e seus nutrientes (Freitas, 2014), que são os mecanismos mais prejudiciais de degradação do solo e o principal motivo de declínio na produção agrícola (Cogo et al., 2003; Meirelles et al., 2005). Os herbicidas são os insumos mais aplicados na agricultura mundial, tanto em consumo como de área tratada (Peluco, 2016). A retenção do herbicida pode apresentar problemas ao ambiente, caso seja aplicado de modo incorreto. Esses problemas podem ser desde a disponibilidade para as plantas até lixiviação para águas subterrâneas (Koskinen e Harper, 1990; Oliveira Junior et al., 1999).

Dados legados constituem um recurso fundamental para o mapeamento digital de solos e calibração dos modelos, devido sua origem ser dos mapeamentos tradicionais (Carré et al., 2007). Estimar atributos do solo em escala detalhada pode melhorar o tempo e custo do trabalho. (McBratney, 1998). Essa possibilidade se dá por meio da inserção de variáveis primárias que podem extrair informações a partir de dados detectados remotamente e dados auxiliares de campo (Cheng et al, 2018).

Os dados aeromagnéticos são capazes de sintetizar grandes áreas e volume de dados de maneira fácil de visualizar e interpretar (Kearey et al., 2009). Os aerolevantamentos são ideais para o reconhecimento de superfícies geológicas, uma vez que seu alvo, entre tantos, são rochas com minerais magnéticos (Khalil et al., 2016). Relacionados à importância do material de origem e a pedologia (Gray et al., 2016), os dados aeromagnéticos apresentam potencial para auxiliar no mapeamento dos solos e no manejo específico no campo. Apesar do seu potencial para identificação geológica em zonas de contato (Domzalski, 1966), os dados aeromagnéticos são pouco explorados na ciência do solo. Nesse contexto, o objetivo deste trabalho foi avaliar o potencial dos dados aeromagnéticos como variável de entrada para estimar a erodibilidade e o coeficiente de sorção de herbicida no Planalto Ocidental Paulista.

3.2 Material e Métodos

3.2.1 Área de estudo e amostragem

O Planalto Ocidental Paulista (Figura 1) corresponde a uma região de 130000 km² (aproximadamente 45% do estado de São Paulo), com variação de 62 a 1013 m de altitude. Está localizado em uma área de transição de diferentes formações rochosas. Dos 130000 km², 20000 km² são compostas por rochas basálticas, 74000 km² pela formação Vale do Rio do Peixe e 36000 km² por outras formações de rochas sedimentares. (Fernandes, 2004; Fernandes et al., 2007).

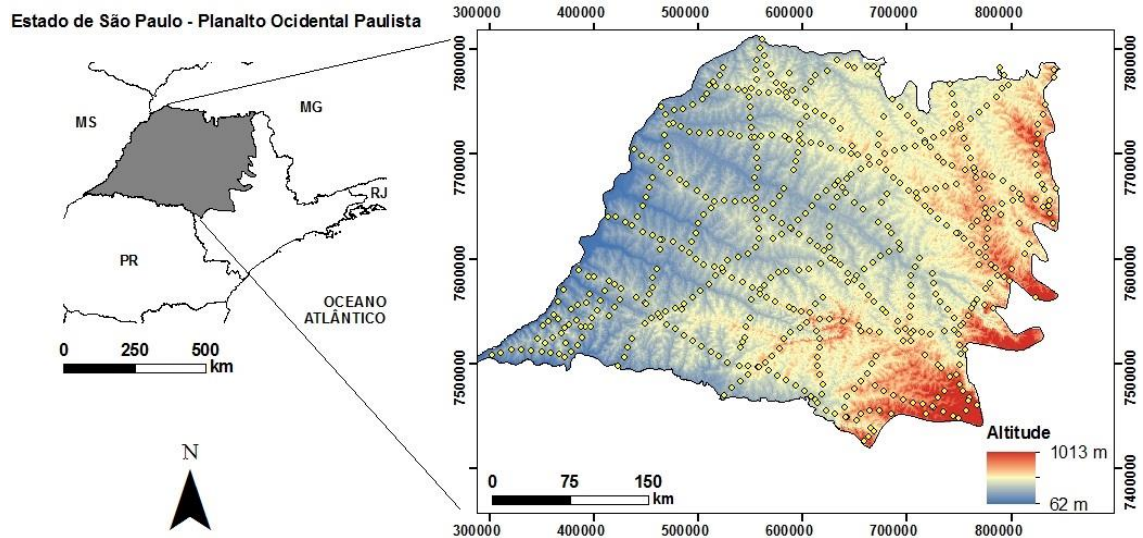


Figura 1 - Mapa de localização da área experimental (esq.) e grade amostral (dir.).

O Planalto Ocidental Paulista apresenta dezesseis classes de solos, contudo, a maior ocorrência na região é classificada como Argissolo Vermelho-Amarelo, Latossolo férrico, Neossolo Litólico, Nitossolo Vermelho e Gleissolo Háptico (Santos et al., 2018). O clima do Planalto Ocidental Paulista pode ser dividido em três grupos, de acordo com a classificação de Thornthwaite (1948). O norte e nordeste o clima é caracterizado como tropical com estação seca no inverno (C2rA'a'). Ao sul, o clima é temperado úmido, com a presença de verões quentes (B4rB'4a), e a leste e sudeste, o clima é temperado úmido com verões quentes e invernos secos (B2rB'3a). A vegetação natural é composta por Mata Atlântica a oeste e Cerrado a leste e as principais culturas encontradas são cana-de-açúcar e pastagem (LUPA, 2008).

Foram utilizadas 539 amostras coletadas em profundidade de 0.0 - 0.20m com o mínimo de interferência antrópica. A densidade amostral apresentada é de uma amostra para 240 km² (ou 24,000 hectares). O planejamento amostral percorreu as principais rodovias do estado de São Paulo, cruzando as principais classes de solo do Planalto Ocidental Paulista no sentido sudeste a noroeste. Para garantir representatividade das amostras, o espaçamento mínimo foi de 10 km e o máximo foi de 60 km, garantindo representatividade das amostras.

3.2.2 Banco de dados legados

Banco de dados legados costumam amparar o mapeamento digital com dados obtidos a partir de métodos tradicionais, auxiliando na produção de novas informações do solo (Bui e Moran, 2001; Carré et al., 2007). As amostras utilizadas para a obtenção da erodibilidade (fator K) e o coeficiente de Freundlich para sorção de herbicida (K_f) utilizado pelo modelo Random Forest (RF) faz parte do projeto de dados legados do grupo CSME (Caracterização do Solo para Fins de Manejo Específico), localizado na Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal – UNESP.

Para a realização do cálculo proposto por Denardin (1990) para fator de erodibilidade global (K) do modelo USLE, foi utilizada a seguinte equação (Eq. 1):

$$K = 0,00000748 M + 0,00448059 P - 0,0631175 DMP + 0,01039567 REL \quad (1)$$

em que:

$M = (\% \text{ silte} + \% \text{ areia muito fina}) \times (100 - \% \text{ argila})$

$P = \text{código de permeabilidade codificado conforme Wischmeier et al. (1971)}$
(adimensional);

$DMP = \text{diâmetro médio ponderado das partículas} < 2,0 \text{ mm,}$

$REL = (\% \text{ matéria orgânica} \times \% \text{ nova areia})/100$

$\% \text{ nova areia} = \% \text{ areia total} - \% \text{ areia muito fina}$

Para o cálculo do coeficiente de sorção de herbicida, foi utilizada a metodologia de Peluco (2016) desenvolvida a partir de meta-análise. O autor relacionou a concentração do herbicida com o conteúdo de argila encontrada nos solos, ajustando assim a equação de Freundlich (Green e Karickhoff, 1990). Como resultado, foi elaborado o modelo matemático para o coeficiente de Freundlich para sorção de herbicida que, posteriormente, foi aplicado no Planalto Ocidental Paulista.

3.2.3 Conjuntos de dados de entrada

Na elaboração do modelo RF, optou-se por utilizar o mínimo de variáveis de entrada possível para reduzir a complexidade do modelo. Essa estratégia é uma das causas da baixa correlação entre as árvores durante o aprendizado do algoritmo. Qualquer tipo de mudança nos dados de entrada poderia gerar alterações ao modelo devido à sensibilidade das árvores.

Os dados aeromagnéticas (Bizzi et al., 2003) foram obtidos a partir do banco de dados do Serviço Geológico do Brasil (CPRM), com escala de 1: 2500000. O Modelo Digital de Elevação (MDE) – TOPODATA (Valeriano, 2005) foi obtido pelo Instituto Nacional de Pesquisas Espaciais (INPE), com escala 1: 250000. Para que fosse possível cobrir toda a área do Planalto Ocidental Paulista, foi realizado um mosaico com 13 imagens MDE. A partir desse mosaico, obteve-se os valores de elevação para cada ponto amostrado. Também foi calculado o grau de declividade, em porcentagem (Santos et al., 2018) e o índice TWI (*Topographic Wetness Index*) (Beven e Kirkby, 1979). Todo o processamento das imagens MDE foi realizado no software ESRI® ArcGIS 10.3.

A Normal Climatológica disponibilizada pelo Instituto Brasileiro de Meteorologia (INMET, 2018) foi utilizada para calcular a Precipitação. O conjunto de dados apresenta a média mensal de chuvas em um período de 30 anos (1981-2010). Para a interpolação da precipitação de todo o Planalto Ocidental Paulista, foram utilizadas as estações meteorológicas do estado de São Paulo (n=16) e dos estados de Rio de Janeiro, Minas Gerais, Mato Grosso do Sul e Paraná (n=83). Para a interpolação, foi utilizada a técnica de krigagem ordinária com escala de 1:250000. Os detalhes teóricos são descritos em Lloyd (2005).

Como índices de vegetação, as médias do NDVI (*Normalized Difference Vegetation Index*) e EVI (*Enhanced Vegetation Index*) foram obtidas na plataforma Google Earth Engine – GEE (GEE, 2019) entre fevereiro/2018 a fevereiro/2019, com escala aproximada de 1: 2500000. Para as classes de solo (n=16) e geologia (n = 20) foi utilizado o banco de dados georreferenciado do Instituto Brasileiro de Geografia e Estatística (IBGE, 2019), com escala de 1: 250000.

Após o tratamento de cada variável de maneira individual, todas foram importadas para o ambiente R (R Core Team, 2019) e padronizadas com o MDE em escala (1: 250000) e tamanho do pixel (30m) com o auxílio do pacote *Raster* (Hijmans, 2019) em ambiente R (R Core Team, 2019). Foi possível atribuir novos valores para cada ponto na grade amostral do Planalto Ocidental Paulista a partir da Latitude e Longitude e realizar a conversão de imagem (ou Raster) para tabela (ou Data Frame).

3.2.4 Análise de dados

O conjunto de dados foi analisado pela estatística descritiva, sendo calculados os valores de mínimo, máximo, média, mediana, erro padrão da média, coeficiente de variação e intervalo de confiança. Todas as variáveis foram submetidas ao teste t de Student. Foram elaboradas três etapas para o pré-processamento dos dados. Inicialmente, todas as variáveis foram submetidas ao teste de multicolinearidade (Variance Inflation Factor, VIF). O VIF é utilizado para medir o grau de multicolinearidade da variável dependente em relação às demais variáveis independente no modelo de regressão. VIF maior que 10 e tolerância menor que 0.1 indica a multicolinearidade da variável independente (O'Brien, 2007; Kalantar et al., 2019), sendo excluída do modelo RF. Para o cálculo do VIF foi utilizado o pacote *olsrr* (Hebbali, 2018) em ambiente R (R Core Team, 2019). Posteriormente, foram analisadas todas as colunas em busca de valores nulos (NA) e a presença de outliers. Essas informações foram investigadas e, quando necessário, excluídas ou corrigidas no banco de dados. Todo o processo de modelagem das variáveis foi realizado em ambiente R (R Core Team, 2019).

Os dados da malha amostral foram separados aleatoriamente em 70% para treinamento e 30% para teste. Foi utilizado o pacote *randomForest* (Liaw e Wiener, 2002), em ambiente R (R Core Team, 2019). O algoritmo é responsável pela geração de árvores de decisão (*ntrees*) em que é obtida a média da combinação de resultados, gerando maior precisão. É necessário um grande número de árvores para um bom desempenho do modelo e para que o mesmo não ceda a sobreajustes. Foram utilizadas 1000 árvores como hiperparâmetro para cada modelo, conforme sugere Kuhn e Johnson (2013).

No processo de predição, as nove variáveis de entrada são classificadas de acordo com sua ordem de importância, da mais influente para a menos influente, durante a construção do modelo RF. Essa classificação apresenta um índice de impureza dos nós com a divisão de cada elemento durante da predição, obtida por meio do resultado médio após utilizar todas as variáveis. Essa impureza é calculada a partir do erro quadrático médio e pode aumentar com a aleatoriedade da mudança das variáveis (Liaw e Wiener, 2002; Álvarez-Cabria et al., 2017). Para a classificação das variáveis mais importantes, o modelo RF gerou um gráfico para cada variável a ser mapeada (fator K e K_f) a partir da mesma malha amostral. Os modelos gerados para treinamento e teste tiveram os resultados validados a partir do Coeficiente de Determinação (R^2), obtido pelo pacote *ithir* (Malone, 2018) em ambiente R (R Core Team, 2019). O R^2 foi utilizado para explicar a capacidade preditiva do modelo, em que quanto mais próximo de 1, melhor a capacidade de acerto (James et al., 2013).

A elaboração dos mapas de treinamento, teste e erro do fator K e K_f no Planalto Ocidental Paulista foi feito por krigagem ordinária. A dependência espacial foi estimada a partir do modelo ajustado ao semivariograma experimental, baseado na teoria das variáveis regionalizadas (Isaaks e Srivastava, 1989). A escolha do melhor modelo ajustado aos semivariograma experimentais foi com base na menor soma de Quadrados dos Resíduos (SQR) e maior R^2 . Os variogramas e os mapas finais foram elaborados no software Geostatistics for Environmental Sciences - GS+10.0 (Robertson, 1998).

3.3 Resultados e Discussão

Os resultados obtidos pela estatística descritiva podem ser observados na tabela 1. Os valores de mínimo e máximo encontrados para o fator K variam entre 0,020 e 0,048 t h MJ⁻¹ mm⁻¹ no Planalto Ocidental Paulista. Esses valores estão de acordo com resultados encontrados em latossolos e argissolos brasileiros, variando entre 0,002 a 0,034 t h MJ⁻¹ mm⁻¹ e 0,008 a 0,045 t h MJ⁻¹ mm⁻¹, respectivamente (Sá et al., 2004; Tomazoni e Guimarães, 2005). Apesar do tamanho da área experimental, o fator K apresentou coeficiente de variação de 16,37%. O K_f apresentou valores mínimo e máximo de 0,24 a 3,67 mg mL⁻¹, com coeficiente de variação de 8,31%.

Essa característica demonstra a heterogeneidade da área, sendo de extrema importância o conhecimento dos solos para aplicação de herbicida preservando o ambiente e diminuindo custo de utilização. Carré et al. (2007) salienta que com o uso de dados legados é possível realizar levantamentos a nível nacional e evitar despesas de novas pesquisas de solo, principalmente para aplicação.

Tabela 1. Estatística descritiva para as variáveis.

| Variáveis | Mínimo | Máximo | Média | Mediana | EPM ⁽¹⁾ | CV ⁽²⁾ | IC ⁽³⁾ | VIF ⁽⁴⁾ | Tolerância |
|--|--------|--------|-------|---------|--------------------|-------------------|----------------------|--------------------|------------|
| Fator K (t h MJ ⁻¹ mm ⁻¹) | 0,020 | 0,048 | 0,035 | 0,037 | 2.40 E-04 | 16,37 | 0,034 – 0,035 | - | - |
| Coef. Sorção Freundlich (mg mL ⁻¹) | 0,24 | 3,67 | 0,56 | 0,39 | 0,02 | 82,31 | 0,52 – 0,60 | - | - |
| Dados Aeromagnéticos (Escala de cinza) | 33,00 | 164,67 | 92,79 | 91,33 | 1,59 | 39,73 | 89,66 – 95,90 | 0,99 | 1,01 |
| Elevação (m) | 280 | 902 | 495 | 479 | 4,63 | 21,69 | 480,80 – 505,00 | 0,60 | 1,66 |
| Declividade (%) | 0,00 | 19,40 | 5,60 | 5,14 | 0,15 | 60,22 | 5,31 – 5,88 | 0,97 | 1,03 |
| Precipitação (mm ano ⁻¹) | 1375 | 1550 | 1467 | 1468 | 1,69 | 2,68 | 1464,21 – 1470,87 | 0,60 | 1,64 |
| TWI (Adimensional) | -1,19 | 10,31 | 6,53 | 9,40 | 015 | 54,03 | 6,23 – 6,82 | 0,98 | 1,02 |
| NDVI (Adimensional) | 0,00 | 0,90 | 0,66 | 0,69 | 5.70 E-03 | 20,26 | 0,65 – 0,67 | 0,41 | 2,43 |
| EVI (Adimensional) | 0,00 | 0,85 | 0,45 | 0,45 | 5.10 E-03 | 26,73 | 0,43 – 0,45 | 4,20 | 2,38 |

Onde: (1) Erro Padrão da Média; (2) Coeficiente de Variação, em %; (3) Intervalo de confiança em 95%; (4). Variance Inflation Factor.

Os dados aeromagnéticos apresentaram valores mínimo e máximo de 33 e 164,67, respectivamente, com coeficiente de variação de 39,73%. A variação dentro do Planalto Ocidental Paulista é consequência da área de transição entre rochas basálticas e areníticas. As rochas possuem diferentes concentrações de óxidos de ferro, ocasionando alta variação do comportamento magnético. Os valores encontrados na elevação, declividade e TWI demonstram a complexidade do Planalto Ocidental Paulista. A elevação apresentou valor mínimo e máximo de 280 e 902 de altitude, respectivamente. O coeficiente de variação de Declividade (60,22%) e TWI (54,03%) mostram ainda mais isso, uma vez que sofrem diretamente os processos de formação da paisagem. A precipitação foi a variável que apresentou o menor coeficiente de variação, com apenas 2,68%. Essa característica está relacionada aos fatores macros do clima, onde normalmente ocorrem em grandes áreas.

O NDVI e EVI são muito utilizados para o monitoramento das condições da vegetação. É possível observar que os valores de mínimo e máximo do NDVI (0,00 e 0,90) e EVI (0,00 e 0,85) são semelhantes, inclusive, no coeficiente de variação. Isso ocorre devido o cálculo do EVI considerar as bandas do vermelho e infravermelho, assim como o NDVI, e ainda utilizar a banda azul para amenizar as influências atmosféricas. Ambos demonstram a ausência de vegetação e presença de vegetação, com valores de zero até o valor máximo encontrado.

Os dados aeromagnéticos e a declividade apresentam a média ligeiramente maior que a mediana, indicando a curva de distribuição levemente assimétrica positiva. O K_f e a elevação também demonstraram a assimetria positiva, de forma mais acentuada, indicando a alta frequência dos valores mais baixos. O fator K, precipitação e NDVI apresentaram a média ligeiramente menor que a mediana, indicando a curva de distribuição ligeiramente negativa. O TWI também demonstrou a assimetria negativa de forma mais acentuada, indicando a baixa frequência dos valores mais altos. O EVI foi a única variável utilizada com a média e mediana iguais, demonstrando a simetria dos dados. Somente a elevação apresentou valores acima do esperado para o erro padrão da média. Valores baixos ou nulos são consequência das casas decimais utilizadas. Todas as amostras foram significativas a 95% conforme o teste t de student.

A verificação da multicolinearidade dos dados teve como objetivo remover fatores altamente correlatos do processo que poderiam vir a interferir no aprendizado do modelo (Garosi et al., 2019). Todas o conjunto de dados foram submetidas ao teste de multicolinearidade. Com a utilização do fator K e K_f como variável dependente, não houve a necessidade de exclusão de nenhuma variável independente no ajuste do modelo. O solo e geologia não entraram na análise estatística e no teste de multicolinearidade por serem variáveis categóricas. Contudo, devido à complexidade do Planalto Ocidental Paulista e a importância na agricultura, ambos foram mantidas no processo de desenvolvimento do modelo RF.

O conjunto de dados passou por tratamento e, quando necessário, exclusão da amostra durante a análise exploratória do conjunto de dados. Foram detectados valores faltantes (ou NA) em NDVI, EVI e Declividade ($n = 6, 6$ e 14 , respectivamente). Esse tipo de erro ocorre devido a falhas na captação da imagem ou no período de pré-processamento. Foram detectados valores outliers nas variáveis NDVI, EVI, Declividade e Elevação ($n = 31, 29, 17$ e 12 , respectivamente). Essa variação ocorre devido à complexidade do ambiente encontrado no Planalto Ocidental Paulista.

O aprendizado para o modelo RF no Planalto Ocidental Paulista foi satisfatório no caso do fator K considerando as amostras, enquanto que para o K_f foi necessário um ajuste, mas permaneceu com 449 pontos. É possível observar na figura 2a e 2b a importância das variáveis preditoras para o fator K e K_f , respectivamente.

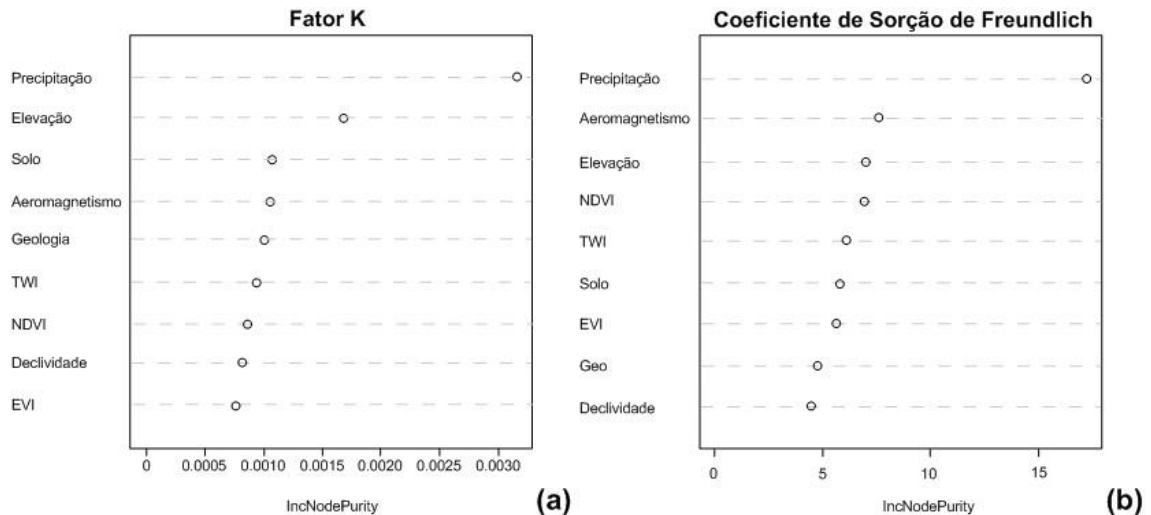


Figura 2 - Importância relativa das variáveis preditoras do fator K (a) e do Coeficiente de Sorção de Freundlich (b).

Para o fator K, a precipitação obteve o resultado superior aos demais, seguido da elevação. O clima costuma ser o elemento de destaque devido sua importância no processo de formação do solo (Budiman et al., 2003). Além disso, a elevação auxilia nos fatores climáticos, influenciando nos padrões do vento e condensação da chuva (Basist et al., 1994).

As variáveis solo, os dados aeromagnéticos e a geologia trazem informações importantes, mesmo demonstrando menos influência no modelo RF. Barbosa (2014) demonstrou em seu estudo que o fator K apresenta relação com o magnetismo do solo. Os solos tropicais costumam ser ricos em óxidos de ferro (Schwertmann e Kämpf, 1985) como herança da rocha matriz e os levantamentos aeromagnéticos são capazes de reconhecer os corpos magnéticos da superfície geológica (Khalil et al., 2016). Gray et al. (2016) demonstram que a variável geologia foi a mais importante para o modelo RF na Austrália, em uma área de 810 km², para mapear propriedades do solo.

O modelo RF obteve menor importância para as variáveis da paisagem (declividade e TWI) e da vegetação (NDVI e EVI). Apesar do resultado obtido, Estrada-Carmona et al. (2017) demonstraram com dados legados dos Estados Unidos, Costa Rica e Europa que as variáveis de topografia (fator LS) e cobertura do solo (fator C) se mostraram as mais importantes para aplicação do modelo RF. Cheng et al (2018)

salia que a cobertura e as condições do solo, da topografia e do clima são as principais variáveis em modelos de perda de solo.

A precipitação também apareceu como variável mais importante para a geração do modelo de predição do K_f . Os dados aeromagnéticos aparecem como segunda variável mais importante para a geração do modelo RF. Esse resultado corrobora Peluco (2016), o qual encontrou relação entre a susceptibilidade magnética do solo com a sorção de herbicida Imazaquim, em uma área experimental de 9,08 km² de transição geológica arenito-basalto.

A elevação aparece como terceira variável mais importante no modelo RF. A relação entre elevação e precipitação corrobora com Obi et al. (2014), os quais confirmaram essa relação na formação do solo no sudeste da Nigéria, em uma área de 8412km². A presença do NDVI como quarta variável mais importante pode ter relação com a vegetação local e a disponibilidade do herbicida Imazaquim retido no solo (Oliveira et al., 2004). A resposta do índice NDVI é capaz de relacionar as condições da planta como resultado da qualidade do solo. As variáveis TWI, solo, geologia, declividade e EVI apresentaram menor importância relativa para a obtenção do modelo RF.

Os valores da validação do modelo RF para a predição do fator K (Figura 3a) e K_f (Figura 3b) apresentaram valores semelhantes durante a etapa de treinamento, com R^2 de 0.93 e 0.90, respectivamente. Esses altos valores ocorrem devido ao aprendizado da máquina para ajustar ao modelo às variáveis de entrada com as variáveis dependentes do fator K e K_f . O algoritmo RF é um conjunto de árvores de decisão que, a cada nova amostra de treinamento, é criada uma árvore que se ajusta aos dados (Bogner et al., 2018). Contudo, o resultado dos testes é o que demonstra a eficácia do modelo. Nessa etapa o algoritmo RF aplica diferentes variáveis independentes como entrada no modelo gerado durante o treinamento e valida os resultados obtidos com os resultados esperados.

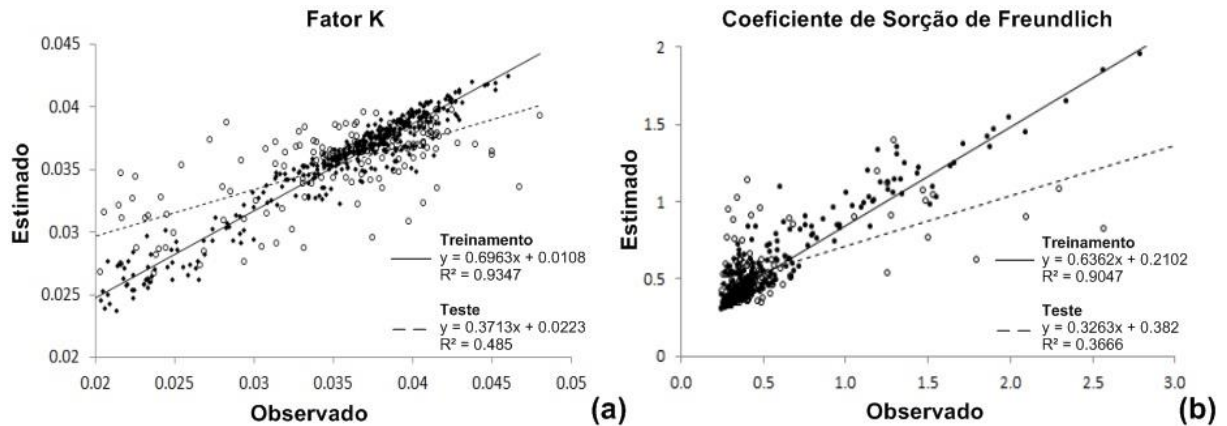


Figura 3 - Gráfico de dispersão dos dados de treinamento vs teste do fator K (a) e do Coeficiente de Sorção de Freundlich (b).

O fator K teve a variabilidade dos dados explicada de maneira moderada no processo de teste do modelo, com R^2 de 0.48 e do K_f um pouco inferior, apresentando R^2 de 0.36. Em ambos os casos houve concentração dos valores de treinamento e teste. Essa concentração demonstra a variabilidade da geologia e do solo, onde majoritariamente é coposta por rochas sedimentares.

Quando comparados, é possível observar que os dados de treinamento e teste do fator K são melhores distribuídos ao longo do eixo que os dados de K_f . Mesmo com o pré-processamento, o conjunto de dados preditos do K_f permaneceram em 61% de 0.0 a 0.5 mg mL⁻¹, 29% até 1.0mg mL⁻¹ e 10% para dados maiores que 1.0mg mL⁻¹. Essa característica demonstra a complexidade do ambiente e a magnitude em relação ao tamanho e, possivelmente, foi o que interferiu na melhora na fase de teste. Uma possível solução para o Planalto Ocidental Paulista seria recortar em pequenos ambientes e realizar estudos setorizados para, posteriormente, juntar todos em um único banco de dados. Esse processo facilitaria o aprendizado do algoritmo RF e criaria um banco de dados legado de escala ultra detalhada. Os resultados iniciais demonstraram o potencial da utilização dos dados aeromagnéticos para aplicação no setor agrícola. Mesmo com R^2 moderado o modelo RF foi eficiente para estimar o fator K e K_f no Planalto Ocidental Paulista.

Na tabela 2 são apresentados os parâmetros dos variogramas ajustados aos semivariograma experimentais das variáveis K e K_f . O modelo esférico foi o modelo ajustado às variáveis K_{Obs} , K_{Est} , $K_{f Obs}$ e $K_{f Est}$, indicando que estas variáveis tem um comportamento menos errático na pequena escala (Isaaks e Srivastava, 1989),

indicando transição suave entre os diferentes níveis, com R^2 de 0.76 e 0.79 para o fator K e 0.59 e 0.69 observado e estimado, respectivamente. Esse tipo de modelo é característico para mapeamento de classes de solos (Isaaks e Srivastava, 1989). O alcance indica a área de influência em que determinada amostra tem no espaço. A diferença do alcance das variáveis estimadas em relação a variável observada corrobora com o erro encontrado durante o período de teste. Com exceção do K_f estimado, todas as outras variáveis apresentaram grau de dependência espacial moderado (Cambardella et al., 1994). Essa característica, em especial do fator K estimado, demonstra moderada dependência espacial no Planalto Ocidental Paulista, demonstrando o potencial para planejamento estratégico na agricultura.

Tabela 2 - Semivariograma das variáveis K e K_f .

| | Modelo | Patamar | | Alcance (km) | GDE ⁽¹⁾ $C_0/[C_0+C_1]$ | R^2 | SQR ⁽²⁾ |
|-------------|----------|-----------|-----------|-----------------|---------------------------------------|-------|------------------------|
| | | C_0 | C_0+C | | | | |
| K_{Obs} | Esférico | 0.01632 | 0.04162 | 156.31 | 39% | 0.76 | 1.488×10^{-3} |
| K_{Est} | Esférico | 0.0000082 | 0.0000207 | 213.17 | 39% | 0.79 | 2.863×10^{-3} |
| K_f_{Obs} | Esférico | 0.0738 | 0.2406 | 127.57 | 30% | 0.59 | 0.0563 |
| K_f_{Est} | Esférico | 0.02513 | 0.1294 | 156.01 | 19% | 0.67 | 0.0192 |

Onde: (1) Grau de dependência espacial (forte com $\% \leq 25\%$, moderado com 25 a 75%, e fraco $>75\%$); (2) Soma dos quadrados dos resíduos.

Na Figura 4a e 4b são ilustrados os mapas espaciais da estimativa do fator K e K_f . O mapa de distribuição espacial Os mapas da distribuição espacial do fator K observado (Figura 4a) e estimado (Figura 4c) são muito semelhantes, existindo maior número de classes encontrado no fator K estimado, principalmente na parte central do Planalto Ocidental Paulista. Nessas regiões foi observado o maior erro relativo (Figura 4e), nas quais o fator K foi subestimado em 32% Essa região merece atenção, pois tem indicadores pedoambientais que indicam potencial de erosão do solo devido sua formação arenítica. O mapa do fator K estimado apresenta maior transição de classes a leste, zona geológica formada por basaltos.

O K_f Observado (Figura 4b) e estimado (Figura 4d) mantiveram os valores mais altos nas regiões leste e sudeste, em que são localizadas principalmente formações basálticas. Por ser área de transição basáltica, os minerais geológicos encontrados

apresentam maior concentração de óxidos de ferro que, por sua vez, apresenta relação com a sorção de herbicida (Peluco, 2016). O erro relativo (Figura 4e) encontrado no Planalto Ocidental Paulista subestimou o K_f em até 90%. A variabilidade dos processos químicos e físicos do solo pode ocasionar retenção diferencial do herbicida, variando assim no controle de plantas daninhas, principalmente em grandes áreas, em que se costuma aplicar dose única do herbicida (Gerstl, 2000; Oliveira et al., 2004).

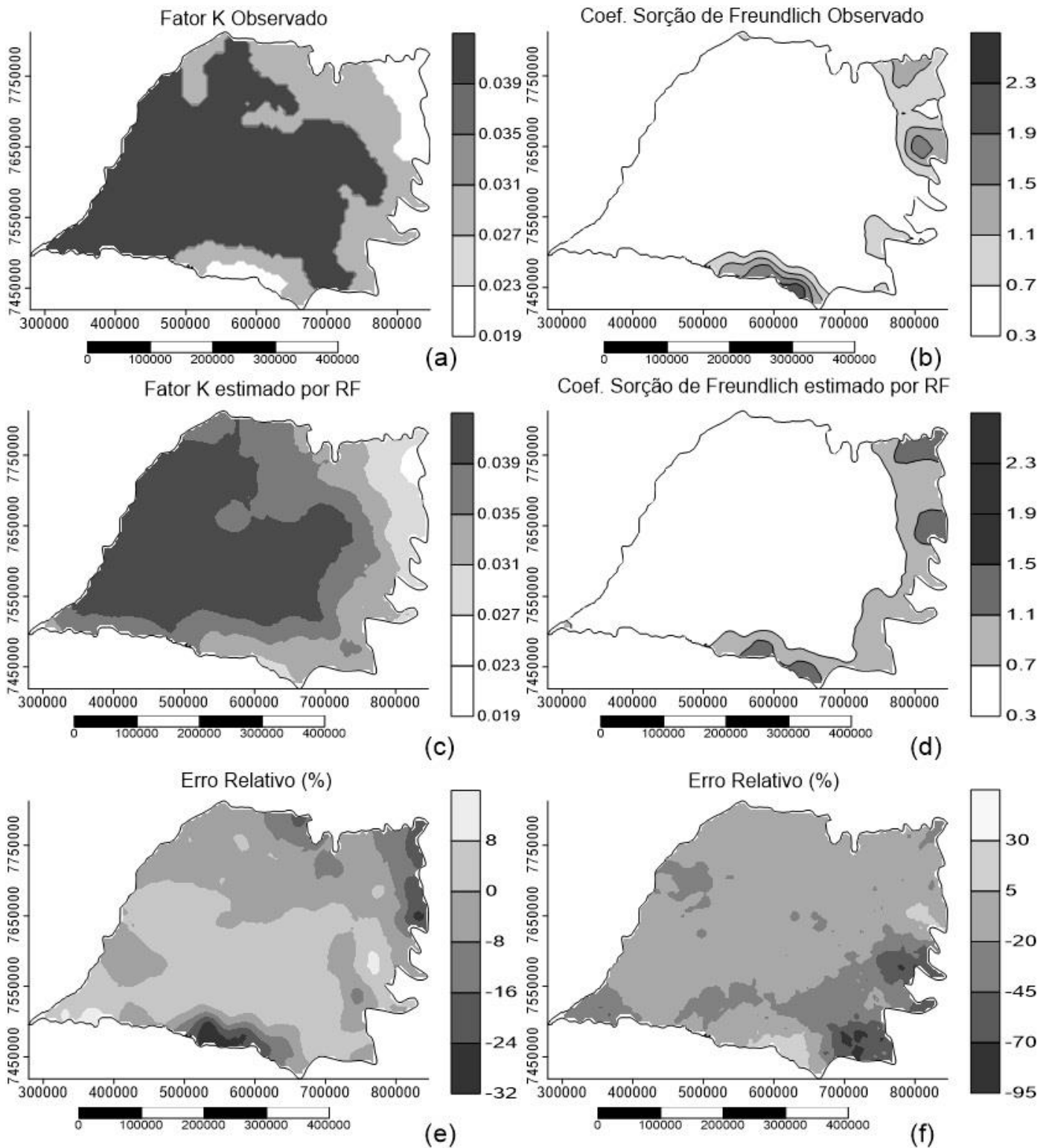


Figura 4 - Padrão espacial do fator K observado (a), estimado por Random Forest (c) e o erro relativo (%) (e); Coeficiente de sorção de freundlich observado (b), estimado por Random Forest (d) e o erro relativo (%) (f).

A proposta de estimar o fator K e K_f no Planalto Ocidental Paulista a partir de dados aeromagnéticos é favorável para ambientes com formações geológicas semelhantes. As estimativas de fatores como o K e K_f são capazes de oferecer novas informações para melhorar as condições econômicas e ambientais de agricultores locais em escala local e regional. As incertezas geradas durante a modelagem dos dados podem ser supridas utilizando novas técnicas ou variáveis de entrada.

Como o fator K e K_f estão relacionados às variáveis de formação do solo, conforme demonstrado pelo modelo RF, inúmeros outros problemas podem ser calculadas a partir destas fontes de dados. Resultados com eficiência moderada reforçam a necessidade de mais estudos que avaliem as variáveis preditoras para aplicação no Brasil e no mundo.

Esses valores podem ser melhorados com o auxílio de novas variáveis de entrada e novos ajustes no conjunto de dados. Novos estudos devem ser realizados no Planalto Ocidental Paulista para descobrir as melhores variáveis de entrada para estudos em zonas de transição geológica arenito-basalto.

3.4. Conclusão

Os resultados obtidos com os dados aeromagnéticos demonstraram sua aplicabilidade na agricultura, em especial para estimar a erodibilidade e sorção de herbicida. O coeficiente de determinação obtido nos resultados de validação são considerados aceitáveis, conforme outros estudos de mapeamento digital de solos no mundo. O modelo Random Forest foi capaz de elencar as variáveis mais importantes utilizadas ao considerar a situação da área em questão. Apesar do potencial dos dados aeromagnéticos para a agricultura, novas pesquisas devem ser realizadas com novas variáveis de teste para que possa melhorar o modelo.

Referências

Álvarez-Cabria M, González-Ferreras AM, Peñas FJ, Barquín J (2017) Modelling macroinvertebrate and fish biotic indices: From reaches to entire river networks. **Science of the Total Environment** 577:308–318.

Basist A, Bell, GD, Meentemeyer, V (1994) Statistical Relationships between Topography and Precipitation Patterns. **Journal of Climate** 7:1305–1315.

Barbosa R (2014) **Erodibilidade de latossolos predita pela suscetibilidade magnética e espectroscopia de reflectância difusa**. 78f.Tese (Doutorado em Agronomia) - Unesp, Jaboticabal.

Beven KJ, Kirkby MJ (1979) A physically based, variable contributing area model of basin. **Hydrological Sciences Bulletin** 24:43-69.

Bizzi LA, Schobbenhaus C, Gonçalves JH, Vidoti RM, Gonçalves JH (Eds.) (2003) **Geologia, tectônica e recursos minerais do Brasil: texto, mapas & SIG**. Brasília: CPRM, .643p.

Bogner C, Seo B, Rohner D, Reineking B (2018) Classification of rare land cover types: Distinguishing annual and perennial crops in an agricultural catchment in South Korea. **PLOS ONE** 13:e0190476.

Bui EN, Moran CJ (2001) Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. **Geoderma** 103:79–94.

Budiman M, McBratney AB, Mendonça-Santos ML, Santos HG (2003). **Revisão sobre funções de pedotransferência (PTFs) e novos métodos de predição de classes e atributos do Solo**. Rio de Janeiro: Embrapa, 47p.

Cambardella CA, Moorman TB, Novak JM, Parkin TB, Karlen DL, Turco RF, Konopka AE (1994). Field-scale variability of soil properties in Central Iowa Soils. **Soil Science Society of America Journal** 58:1501-1511.

Carré F, McBratney AB, Minasny B (2007) Estimation and potential improvement of the quality of legacy soil samples for digital soil mapping. **Geoderma** 141: 1–14.

Cheng Z, Lu D, Li G, Huang J, Sinha N, Zhi J, Li S (2018). A Random Forest-Based Approach to Map Soil Erosion Risk Distribution in Hickory Plantations in Western Zhejiang Province, China. **Remote Sensing** 10:1899.

Cogo N P, Levien R, Schwarz RA (2003). Perdas de solo e água por erosão hídrica influenciadas por métodos de preparo, classes de declive e níveis de fertilidade do solo. **Revista Brasileira de Ciência do Solo** 27:743-753.

Denardin JE (1990) **Erodibilidade do solo estimado por meio de parâmetros físicos e químicos**. 114f.Tese (Doutorado em Agronomia) - USP, Piracicaba.

Domzalski W (1966) Importance of aeromagnetism in evaluation of structural control of mineralization. **Geophysical Prospecting** 14:273–291.

EMBRAPA 1999: Empresa Brasileira De Pesquisa Agropecuária. **Sistema brasileiro de classificação de solos**. Brasília: Embrapa, 412p.

Estrada-Carmona N, Harper EB, Declerck F, Fremier AK (2017) Quantifying model uncertainty to improve watershed-level ecosystem service quantification: a global sensitivity analysis of the RUSLE. *International Journal of Biodiversity Science, Ecosystem Services & Management* 13:40–50.

Fernandes LA (2004) Mapa litoestratigráfico da parte oriental da Bacia Bauru (PR, SP, MG), escala 1: 1.000.000. **Boletim Paranaense de Geociências** 55: 53–66.

Fernandes LA, Castro AB, Basilici G (2007) Seismites in continental sand sea deposits of the Late Cretaceous Caiuá Desert, Bauru Basin, Brazil. **Sedimentary Geology** 199: 51–64.

Foley JA, Defries R, et al. Global consequences of land use. **Science** 309:570–574.

Freitas L (2014) **Qualidade e erodibilidade de Latossolos sob mata e cultivo de cana-de-açúcar**. 110f. Tese (Doutorado em Agronomia) - Unesp, Jaboticabal.

Garosi Y, Sheklabadi M, Conoscenti C, Pourghasemi HR, Van Oost K (2019). Assessing the performance of GIS- based machine learning models with different accuracy measures for determining susceptibility to gully erosion. **Science of The Total Environment** 664:1117–1132.

Gerstl Z (2000) An update on the Koc concept in regard to regional scale management. **Crop Protection** 19: 643-648.

GEE 2019: Google Earth Engine. Disponível em: <https://earthengine.google.org/#intro>. Data de acesso: 22 de Fevereiro de 2019.

Gray JM, Bishop TFA, Wilford JR (2016) Lithology and soil relationships for soil modelling and mapping. **Catena** 147: 429–440.

Green RE, Karickhoff SW (1990) Sorption estimates for modeling. In: **Pesticides in the soil environment: processes, impacts and modeling**. Madison: Soil Science Society of America, p.79-101.

Hebbali A (2018) olsrr: **Tools for Building OLS Regression Models**. R package version 0.5.2.

Hijmans RJ (2019) raster: **Geographic Data Analysis and Modeling**. R package version 2.8-19.

Hobbs PR (2007) Conservation agriculture: what is it and why is it important for future sustainable food production? **Journal of Agricultural Science** 145:127-137.

IBGE 2019: Instituto Brasileiro de Geografia e Estatística. **Mapa pedológico do Brasil**. Disponível em: <https://www.ibge.gov.br/> . Data de acesso: : 22 de fevereiro de 2019

IBGE 2019: Instituto Brasileiro de Geografia e Estatística. **Mapa geológico do Brasil**. Disponível em: <https://www.ibge.gov.br/> . Data de acesso: : 22 de fevereiro de 2019.

INMET 2018: Instituto Nacional de Meteorologia. **Normais Climatológicas do Brasil, 1981-2010**. Disponível em: <http://www.inmet.gov.br/>. Data de acesso: : 22 de fevereiro de 2019.

Isaaks EH, Srivastava RM (1989) **Applied Geostatistics**. Oxford University Press: Nova York, 570p.

James G, Witten D, Hastie T, Tibshirani R (2013) **An Introduction to Statistical Learning with Applications in R**. Springer, 426p.

Junior WC, Filho BC, Chagas CS, Bhering SB, Pereira NR, Pinheiro HSK (2016) Regressão linear múltipla e modelo Random Forest para estimar a densidade do solo em áreas montanhosas. **Pesquisa Agropecuaria Brasileira** 51:1428–1437.

Kearey P, Brooks, M HILL, I (2009) **An Introduction to Geophysical Exploration**. Blackwell Science, Malden, 272p.

Khalil A, Hafeez THA, Saleh HS, Mohamed WH (2016) Inferring the subsurface basement depth and the structural trends as deduced from aeromagnetic data at West Beni Suef area, Western Desert, Egypt. **NRIAG Journal of Astronomy and Geophysics** 5:380–392.

Koskinen WC, Harper SS (1990) The retention process: mechanisms. In: **Pesticides in the soil environment: processes, impacts and modeling**. Madison: Soil Science Society of America, p.51-77.

Kuhn M, Johnson K (2013) **Applied predictive modeling**. Springer, 600p.

LUPA 2008: Levantamento Censitário das Unidades de Produção Agropecuária. Levantamento Censitário das unidades de produção agropecuária do estado de São Paulo. Disponível em: <http://www.cati.sp.gov.br/projetolupa>. Data de acesso: 22 de fevereiro de 2019.

Liaw A, Wiener M (2002) Classification and regression by random- Forest. **Rnews** 2:18-22.

Lloyd CD (2005) Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain. **Journal of Hydrology** 308:128–150.

Malone B (2018) **Soil data and some useful associated functions**. R package version 1.0.

McBratney AB (1998) Some considerations on methods for spatially aggregating and disaggregating soil information. **Nutrient Cycling in Agroecosystems** v50:51–62.

Mendonça-Santos ML, Santos HG (2007) The State of the Art of Brazilian Soil Mapping and Prospects for Digital Soil Mapping. **Developments in Soil Science** 31:39–55.

Meirelles MSP, Bueno MCD, Dias TCS, Coutinho HLC (2005) Sistema de suporte a decisão para avaliação do risco de impactos ambientais em bacias hidrográficas por redes de dependência e lógica fuzzy. In: XII SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO. **Anais eletrônico...** Goiânia: INPE, p.2259-2266.

Minasny B, Hartemink AE (2011) Predicting soil properties in the tropics. **Earth-Science Reviews** 106:52–62.

Nguyen KA, Chen W, Lin BS, Seeboonruang U, Thomas K. (2019) Predicting Sheet and Rill Erosion of Shihmen Reservoir Watershed in Taiwan Using Machine Learning. **Sustainability** 11:3615.

Ogban O'BRIEN, R. M. A Caution Regarding Rules of Thumb for Variance Inflation Factors. **Quality & Quantity**, v. 41, n. 5, p. 673–690, 2007.

Oliveira MF, Colonna I, Prates HT, Mantovani EC, Gomide RL, Oliveira Júnior RS (2004). Sorção do herbicida imazaquim em Latossolo sob plantio direto e convencional. **Pesquisa Agropecuária Brasileira, Brasília** 39:787-793.

Oliveira Junior R S, Koskinen WC, Ferreira FA, Khakural BR, Mulla DJ, Robert, PJ (1999) Spatial variability of imazethapyr sorption in soil. **Weed Science** 47:243-248.

Peluco RG (2016) **Magnetismo e cor na identificação de solos com diferentes potenciais de sorção de imazaquim**. 105f. Tese (Doutorado em Agronomia) - Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Jaboticabal, 2016.

R Development Core Team (2019) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna Austria. <http://www.Rproject.org/>.

Robertson GP (1998) **GS+: geostatistics for the environmental sciences – GS+ user's guide**. Plainwell: Gamma Design Software, 152p.

Sá M, Lima J, Curi N, Massaroto J, Marques J (2004) Estimativa da erodibilidade pela desagregação por ultra-som e atributos de solos com horizonte B textural. **Pesquisa Agropecuária Brasileira** 39:691-699.

Santos H, Jacomine P, Anjos L, Oliveira V, Lumbreras J, Coelho M, Almeida J, Araujo Filho J, Oliveira J, Cunha T (Eds.) (2018) **Sistema brasileiro de classificação de solos**. Brasília: Embrapa, 356p.

Sanchez PA, Ahamed S, et al. (2009). Digital Soil Map of the World. **Science** 325:680–681.

Schwertmann U, Kämpf N (1985). Properties of goethite and hematite in kaolinitic soils of southern and central Brazil. **Soil Science** 139:344-350.

Tomazoni JC, Guimarães E (2005). A sistematização dos fatores da EUPS em SIG para quantificação da erosão laminar na bacia do Rio Jirau. **Revista Brasileira de Cartografia** 57:235-244.

Thornthwaite CW (1948) An Approach toward a Rational Classification of Climate. **Geographical Review** 38:55-94.

Valeriano MM (2005) Modelo digital de variáveis morfométricas com dados SRTM para o território nacional: o projeto TOPODATA. In: XII SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO. **Anais eletrônico...** Goiânia: INPE, p. 1-8.

Wischmeier WH, Johnson CB, Cross BV (1971) Soil erodibility nomograph for farmland and construction sites. **J Soil Water Conserv** 26:189–193.

CAPÍTULO 4 - Considerações finais

O estudo demonstrou que os dados aeromagnéticos servem como covariável para estimar os atributos do solo, como o conteúdo de argila. Os mapas apresentados também demonstraram a capacidade de aplicação dos dados aeromagnéticos para estimar a erodibilidade ou o coeficiente de sorção de herbicida.

O modelo Random Forest foi eficaz na indicação das principais variáveis preditoras, determinando se a variável aeromagnética era importante ou não para o modelo em relação às demais. O coeficiente de determinação encontrado para estimar argila, erodibilidade e coeficiente de sorção de Freundlich foi moderado durante os períodos de teste, corroborando com resultados encontrados em mapeamentos de solos em outras partes do Brasil e do mundo.

São necessários novos testes de aplicação dos dados aeromagnéticos em outros ambientes, com escala e variáveis pedogenéticas diferentes, para avaliar novos problemas recorrentes na agricultura.

APÊNDICES

Apêndice I - Rotina dos resultados obtidos em ambiente R.

```
#### DATA #####  
#Do not forget to change the search path#  
  
setwd("C:\\[A]Aerogeofisico\\...\\THESIS")  
  
pop<-read.table("pop7.txt", head=T, dec='.')  
head(pop, 10)  
str(pop)  
attach(pop)  
  
#### EXPLORING THE DATA ####  
  
#VIF#  
#install.packages("olsrr")  
library(olsrr)  
model <- lm(Clay ~ Soil+Geo+Aeromag+Precip+TWI,  
            data = pop)  
ols_vif_tol(model)  
  
#DESCRIPTIVE STATISTICS#  
  
ed<-function (x){  
  media<-mean(x)  
  dp<-sd(x)  
  minimo<-min(x)  
  maximo<-max(x)  
  q1<-quantile(x,0.25)  
  mediana<-median(x)  
  q3<-quantile(x,0.75)  
  cv<-sd(x)/mean(x)*100
```

```

n<-length(x)
num.ass<-(1/n)*sum((x-mean(x))^3)
den.ass<-((1/n)*sum((x-mean(x))^2))^(3/2)
ass<- num.ass/den.ass
num.ct<-(1/n)*sum((x-mean(x))^4)
den.ct<-((1/n)*sum((x-mean(x))^2))^2
ct<-num.ct/den.ct-3
epm<-sd(x)/sqrt(n)

c(mínimo=minimo,Q1=q1,média=media,mediana=mediana,desv_pad=dp,
  EPM=epm,Q3=q3,máximo=maximo,CV=cv,Assimetria=ass,Curtosis=ct)
}

round(ed(Clay),3)
var(Clay,na.rm=TRUE)

#DO THE SAME WITH THE OTHER ATTRIBUTES#

#NORMALIZING THE DATA#
#PAY ATTENTION! - ONLY IF NECESSARY FOR MODEL#

#normalize <- function(x) {
# return ((x - min(x)) / (max(x) - min(x)))
#}

#hist(pop$Clay)
#pop$Clay_norm<-normalize(pop$Clay)

#x11()
#par(mfrow=c(1,2))
#hist(pop$Clay, xlab="Freyquency" , ylab="Clay" , main="Clay (g kg-1)")

```

```
#hist(pop$Clay_norm, xlab="Frequency" , ylab="Clay" , main="Normalized Clay (g
kg-1)")
```

```
#DO THE SAME WITH THE OTHER ATTRIBUTES#
```

```
#REMOVING NA's#
```

```
sapply(pop, function(x) sum(is.na(x)))
sum(is.na(pop))
```

```
#pop[pop==0]<-NA
#pop[pop==99]<-NA
```

```
sapply(pop, function(x) sum(is.na(x)))
pop<-na.omit(pop)
sum(is.na(pop))
```

```
#REMOVING OUTLIERS#
```

```
boxplot(Clay)
boxplot(Clay, plot=FALSE)$out
```

```
outliers1 <- boxplot(Clay, plot=FALSE)$out
print(outliers1)
```

```
pop[which(Elev %in% outliers1),]
```

```
pop<-pop[-which(Elev %in% outliers1),]
boxplot(pop$Clay)
```

```
#DO THE SAME WITH THE OTHER ATTRIBUTES#
```

```
##### CLAY PREDICTION - Random Forest #####

#To install the ithir package, it is recommended to use the following function:#
#install.packages("devtools")
#library(devtools)
#install_bitbucket("brendo1001/ithir/pkg")

library(randomForest)
library(ithir)
library(raster)

set.seed(123)
training <- sample(nrow(pop), 0.7 * nrow(pop))

# fit the model

pop.RF.Exp <- randomForest(Clay ~ Soil+Geo+Aeromag+Precip+
                          TWI+Elev,
                          data = pop[training, ],
                          importance = TRUE, ntree = 1000)
#str(pop.RF.Exp)

print(pop.RF.Exp)
importance(pop.RF.Exp)

varImpPlot(pop.RF.Exp)
varImpPlot(pop.RF.Exp, type=2)

# Training

RF.pred.C <- predict(pop.RF.Exp, newdata = pop[training, ])
goof(observed = pop$Clay[training], predicted = RF.pred.C)
```

```
#write.table(RF.pred.C,"RF.IV.txt",quote=F,row.names = T)
```

```
# Test
```

```
RF.pred.V <- predict(pop.RF.Exp, newdata = pop[-training, ])  
goof(observed = pop$Clay[-training], predicted = RF.pred.V)  
#write.table(RF.pred.V,"RF.EV.txt",quote=F,row.names = T)
```

```
##### CREATING A MAP #####
```

```
map.RF <- predict(pop.RF.Exp, newdata = pop)  
mat<-cbind(data.frame(pop$x,pop$y, map.RF))  
#write.table(mat,"mat.txt",quote=F,row.names = F)
```

```
#VARIOGRAPHIC ANALYSIS#
```

```
library(sp)
```

```
library(maptools)
```

```
library(raster)
```

```
library(gstat)
```

```
library(graphics)
```

```
library(lattice)
```

```
mat
```

```
names(mat)
```

```
coordinates(mat)=~ pop.x+pop.y
```

```
form<-map.RF~1
```

```
v.arg.RF<-variogram(form,data=mat)
```

```
#x11()
```

```
plot(v.arg.RF,pl=F,pch=16,col=1)
```

```
#MODELING#
```

```
m.arg.RF <- fit.variogram(v.arg.RF,vgm(6000,"Sph",150000,2000))
m.arg.RF
sqr<-attr(m.arg.RF, "SSErr")
sqr
#x11()
plot(v.arg.RF,model=m.arg.RF, col=1,pl=F,pch=16,main ="Clay variogram by
Random Forest
  \n esf(1869.688; 5052.787; 183279.3; 2.219)")
```

```
#CREATING GRID#
```

```
x<-mat$pop.x
y<-mat$pop.y
dis <- (max(x)-min(x))/109 #Distância entre pontos
grid <- expand.grid(X=seq(min(x),max(x),dis), Y=seq(min(y),max(y),dis))
gridded(grid) = ~ X + Y
#plot(grid)
```

```
#CREATING CONTOUR
```

```
contorno<-read.table("contorno.pop.txt", head=T)
```

```
p = Polygon(contorno)
ps = Polygons(list(p),1)
contorno = SpatialPolygons(list(ps))
#plot(contorno)
```

```
#ORDINARY KRIGING#
```

```
ko.arg.RF<-krige(formula=form, mat, grid, model=m.arg.RF,
```

```
nmin=5,
nmax=20,
block=c(0,0),
nsim=0,
na.action=na.pass,
debug.level=-1)

# Making kriging file conversions

ko.arg.RF <- as.data.frame(ko.arg.RF)

#ATTENTION FOR THIS LINE!!
#write.table(ko.arg.RF,"ko.arg.RF.txt",quote=F,row.names = F)
#This is important for do de error map

head(ko.arg.RF)
coordinates(ko.arg.RF)=~X+Y #conversion
gridded(ko.arg.RF)=TRUE #conversion
ko.arg.RF <- raster(ko.arg.RF) #conversion
head(ko.arg.RF)

#PRINTING THE MAP OF VALUES PREDICTED BY RANDOM FOREST#

ko.arg.RF <- mask(ko.arg.RF, contorno, inverse=FALSE)
#spplot(ko.arg.RF)
#x11()
plot(ko.arg.RF,main="Predicted Clay by Random Forest",
      xlab="Latitude",ylab="Longitude")
plot(contorno, add=T)
contour(ko.arg.RF, add=T, nlevels = 6)

####CREATING AN ERROR MAP####
```

```
Error<-read.table("ko.arg.RF.txt", head=T)
Error<-Error[,c(1:2,4)]
head(Error)

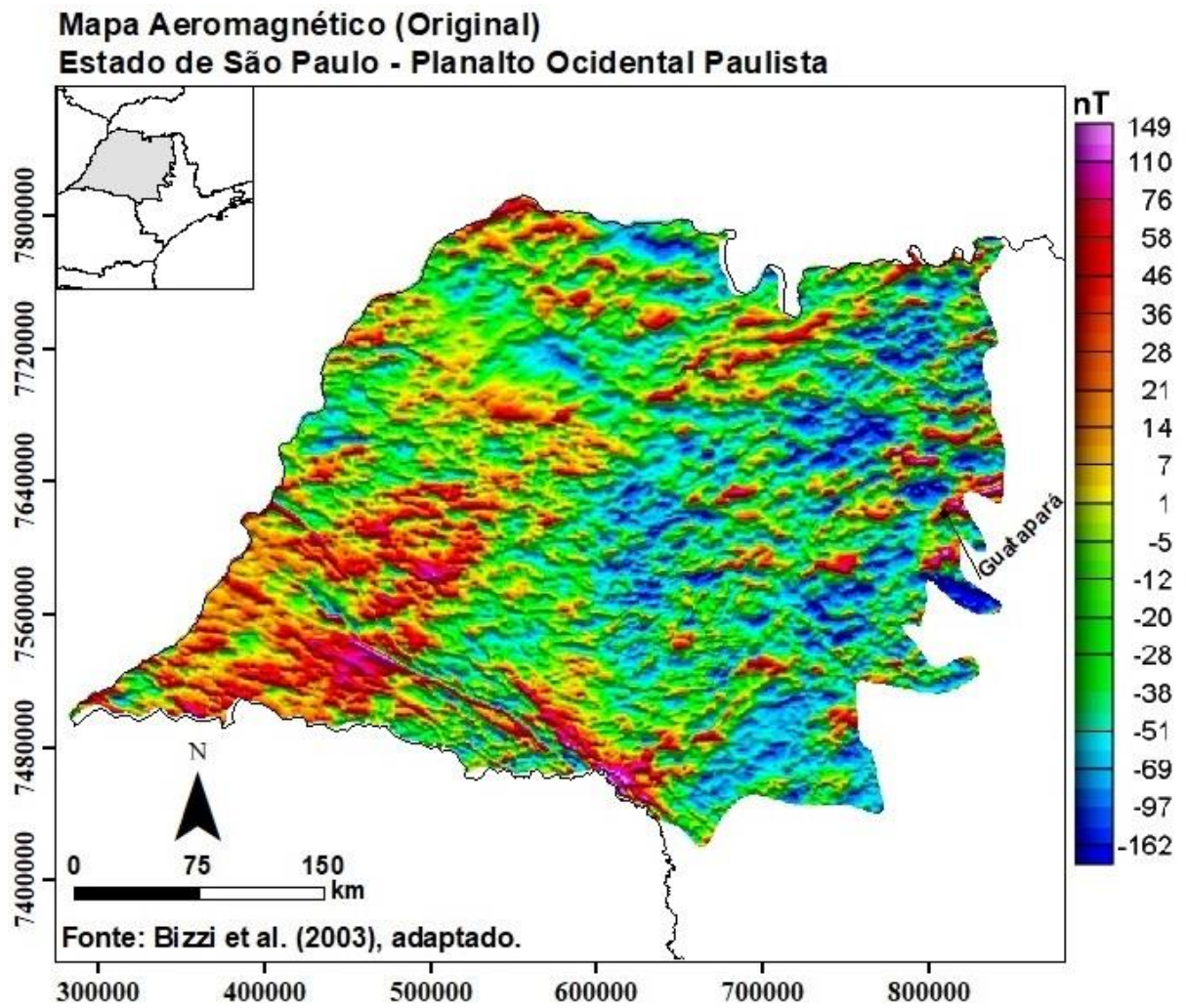
ko.arg.RF.Error <- as.data.frame(Error)
head(ko.arg.RF.Error)
coordinates(ko.arg.RF.Error)=~X+Y #conversion
gridded(ko.arg.RF.Error)=TRUE #conversion
ko.arg.RF.Error <- raster(ko.arg.RF.Error) #conversion
head(ko.arg.RF.Error)

# Printing the Error map
ko.arg.RF.Error <- mask(ko.arg.RF.Error, contorno, inverse=FALSE)
#spplot(ko.arg.RF.Error)
#x11()
plot(ko.arg.RF.Error,main="Variance Map",
     xlab="Latitude",ylab="Longitude")
plot(contorno, add=T)
contour(ko.arg.RF.Error, add=T, nlevels = 6)

# END #

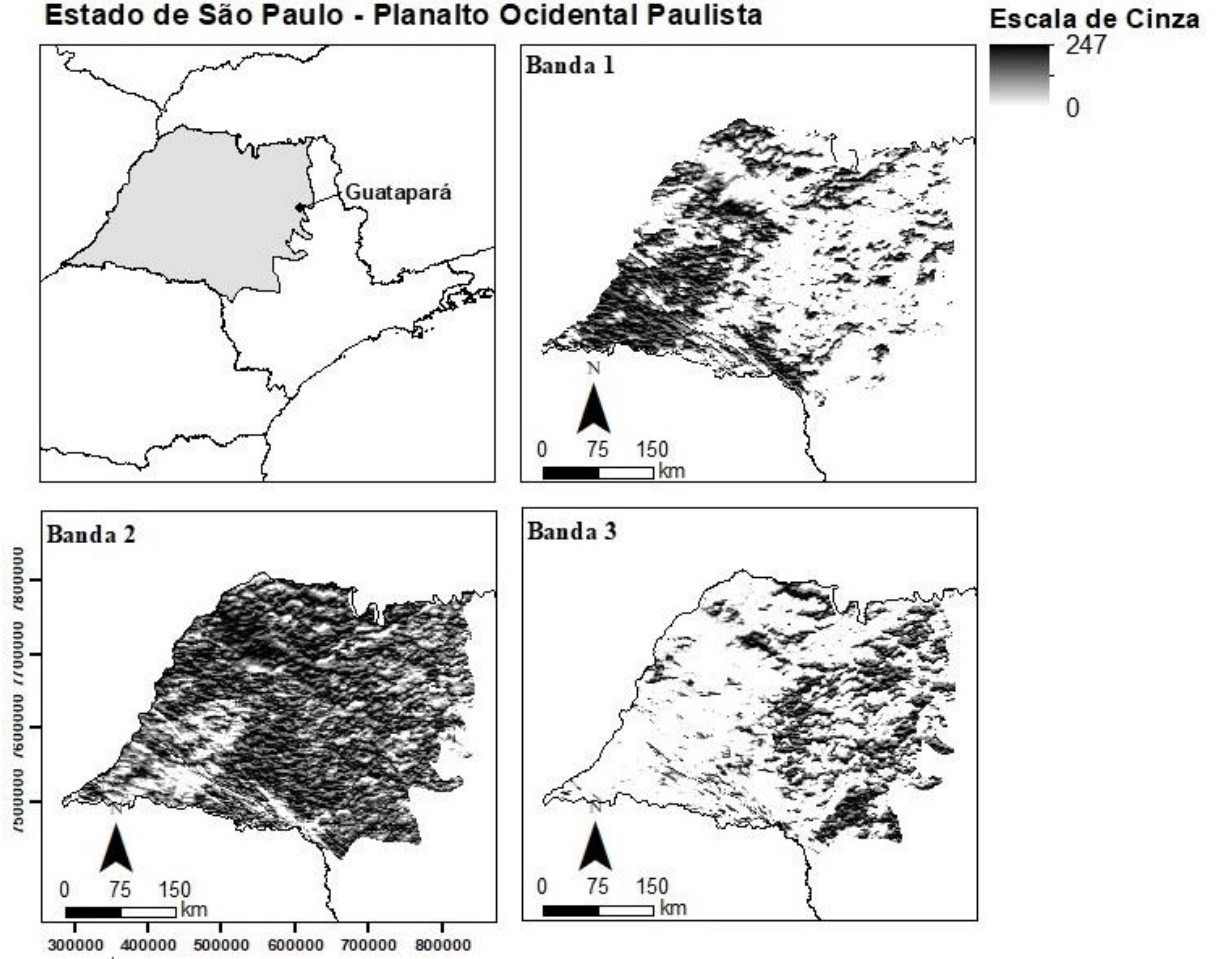
#Any new changes to the script will be made available at:
#https://github.com/thdeandrade/thesis
```

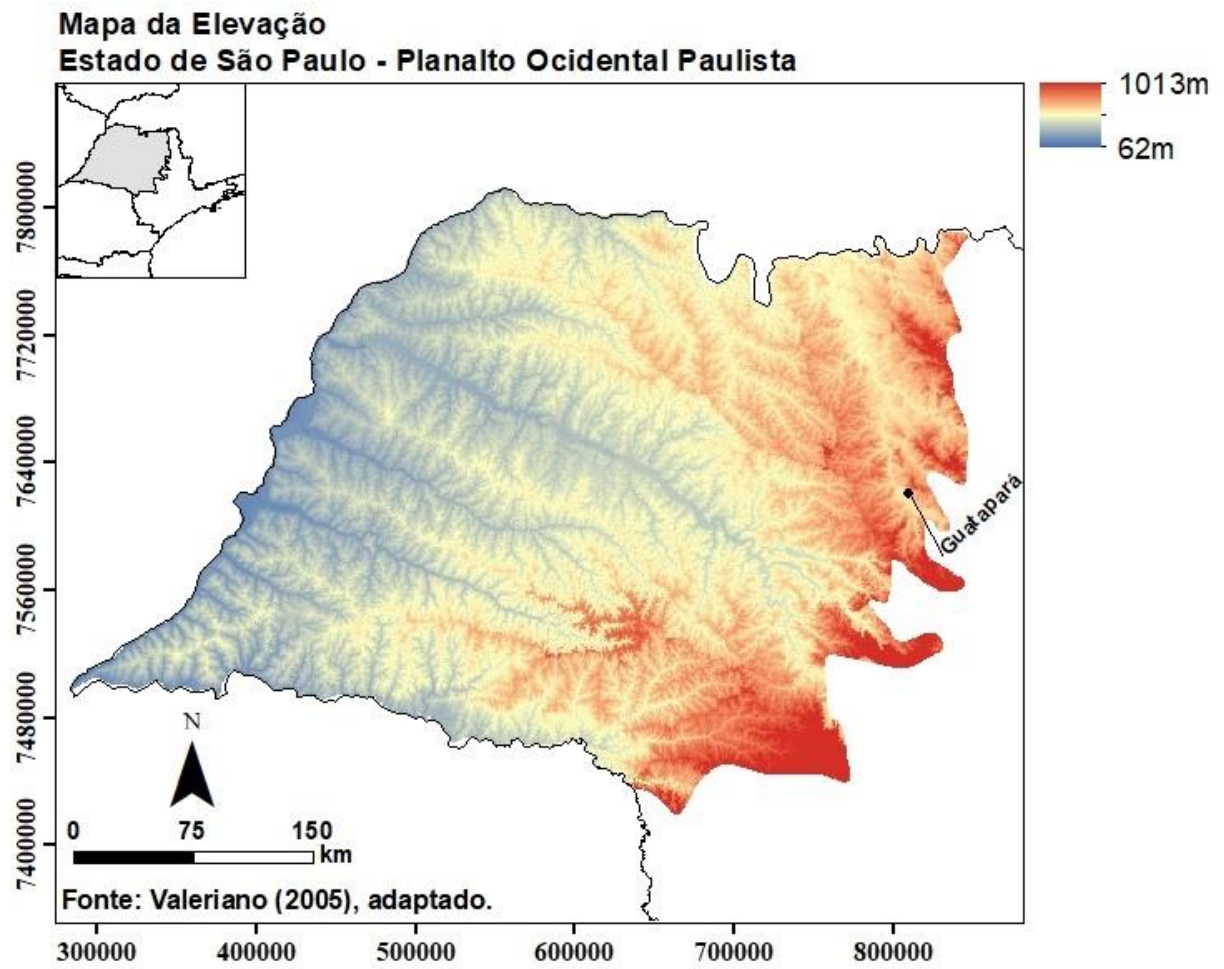
Apêndice II – Variáveis preditoras utilizadas. Continua...



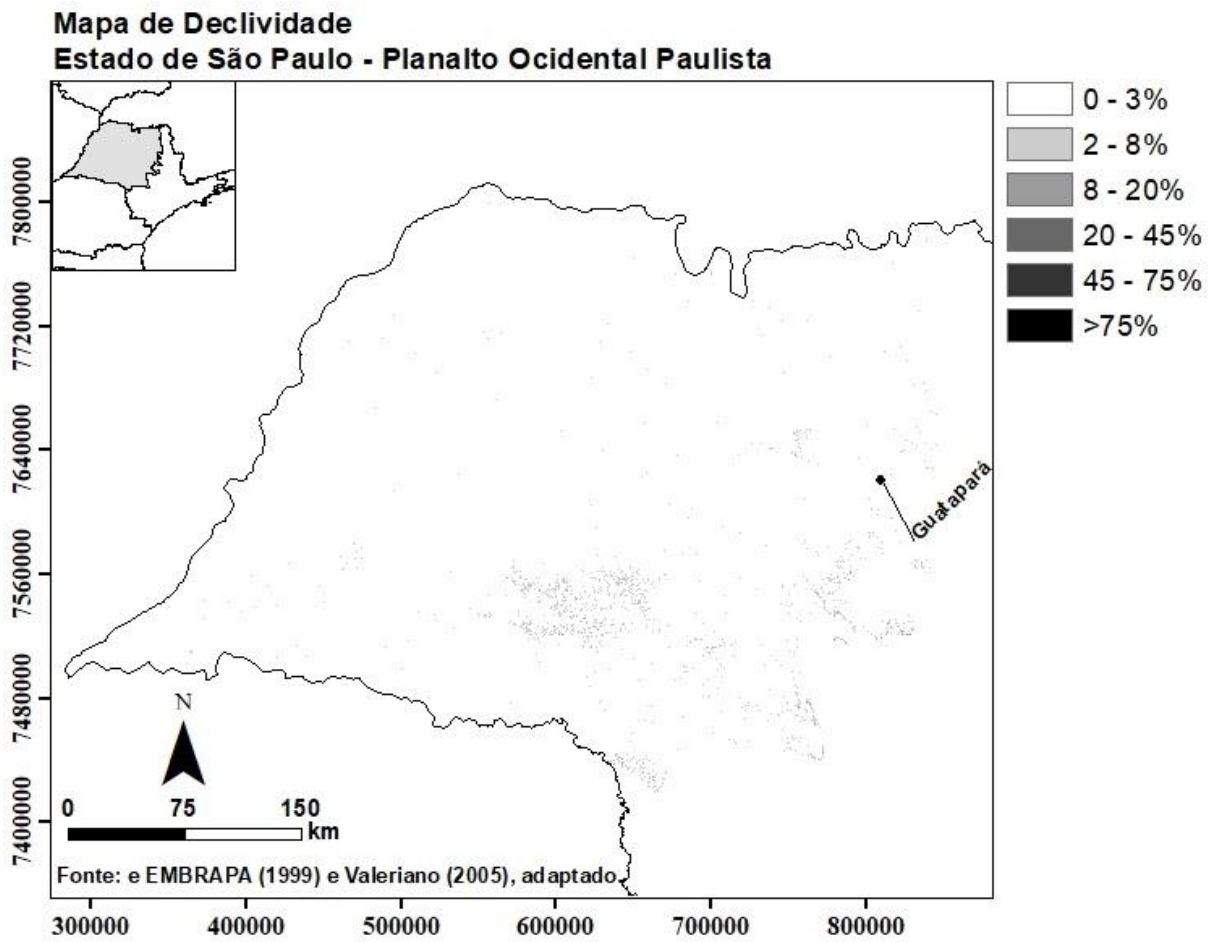
Apêndice II – Continuação.

Mapas Aeromagnéticos (Adaptado) Estado de São Paulo - Planalto Ocidental Paulista

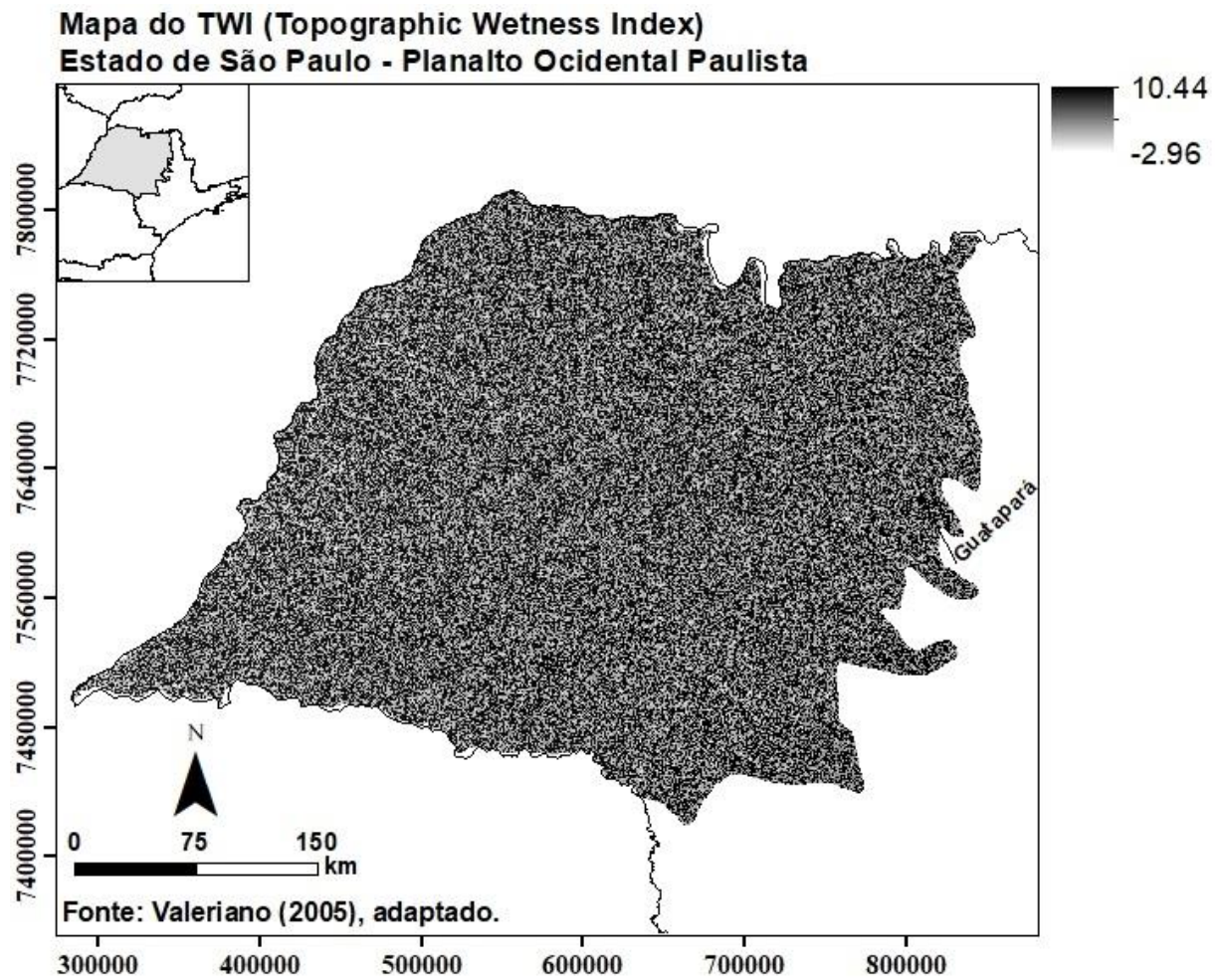


Apêndice II – Continuação.

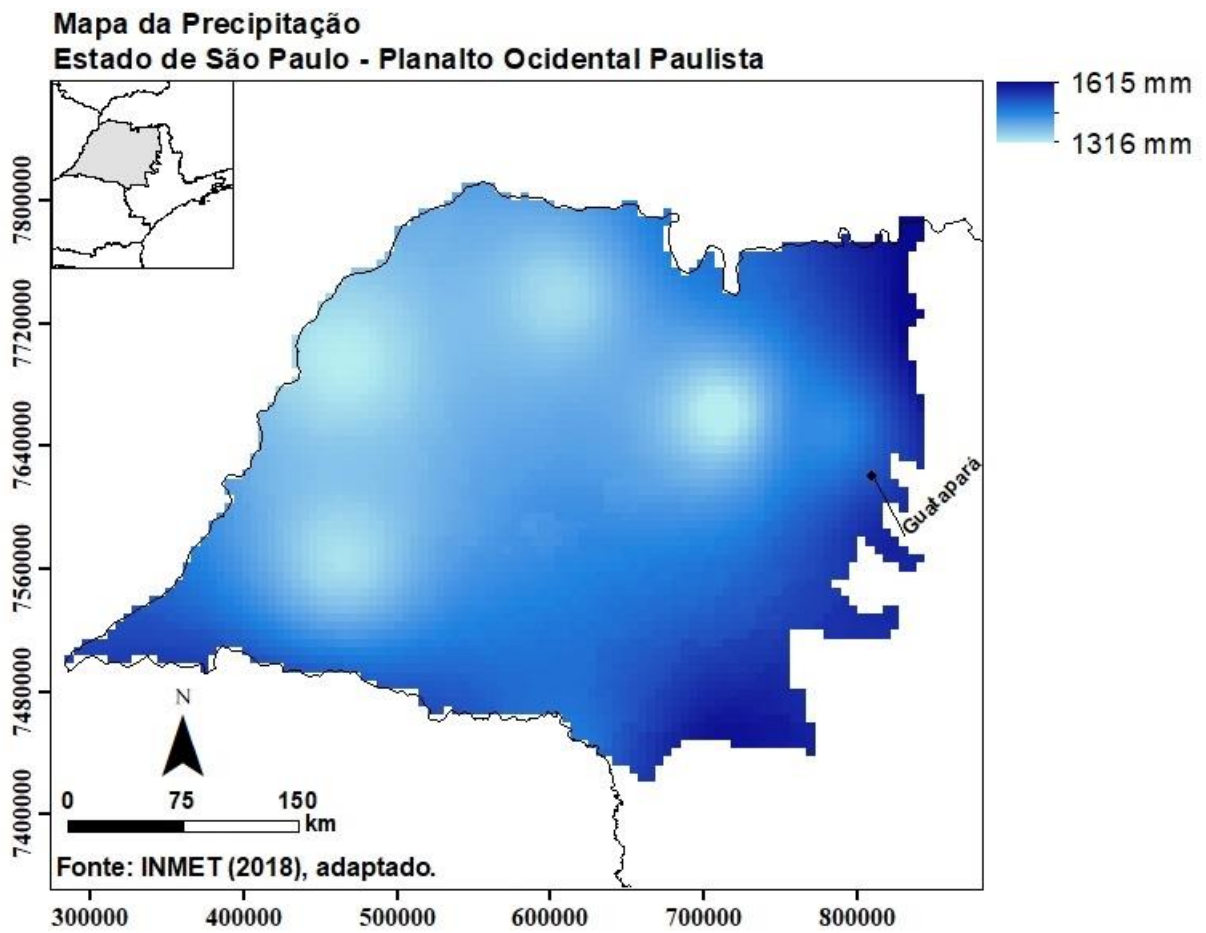
Apêndice II – Continuação.



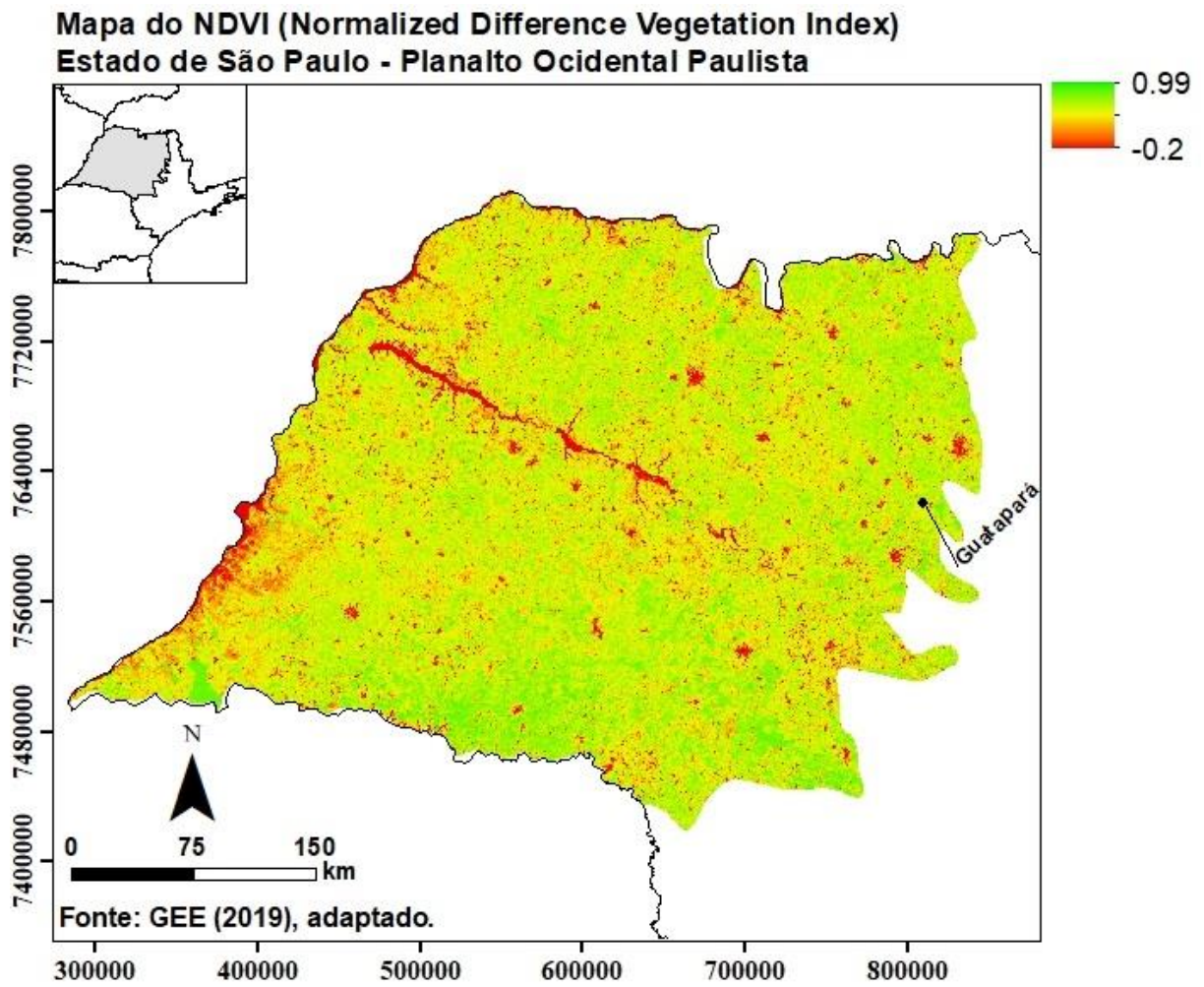
Apêndice II – Continuação.



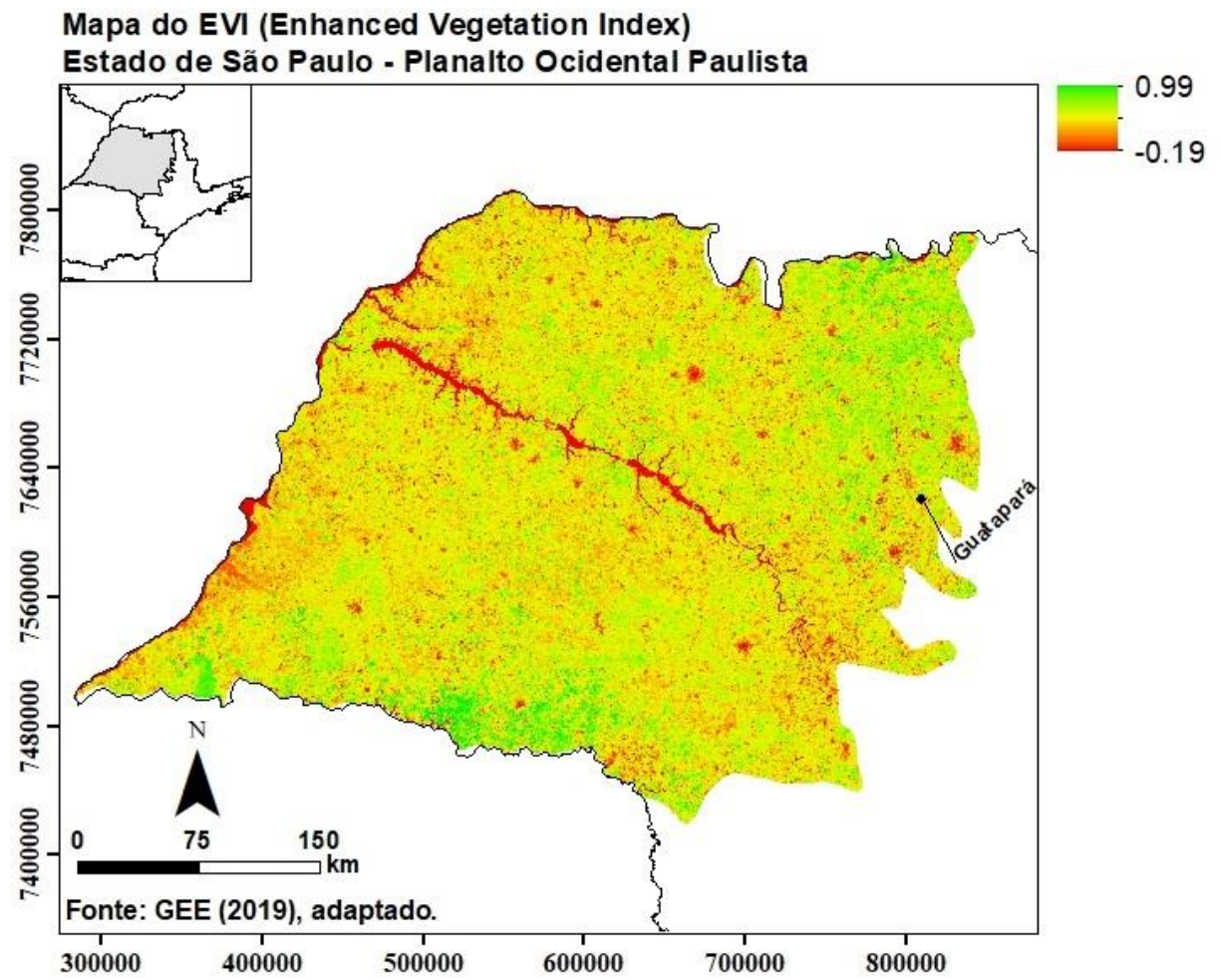
Apêndice II – Continuação.



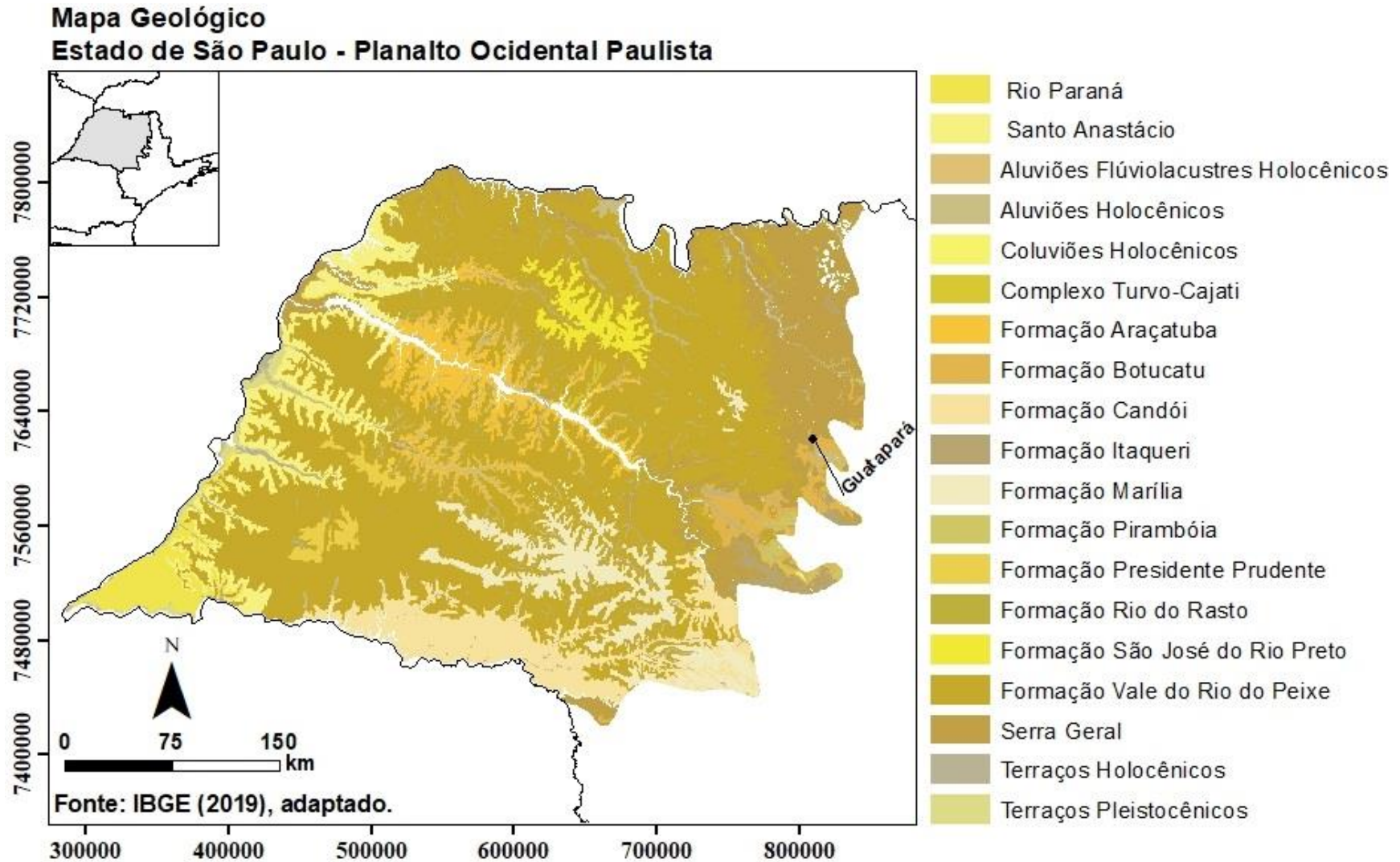
Apêndice II – Continuação.



Apêndice II – Continuação.



Apêndice II – Continuação.



Apêndice II – Continuação.

