

UFRRJ
INSTITUTO DE GEOCIÊNCIAS - IGEO
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM E EVOLUÇÃO
GEOLÓGICA – PPGMEG

DISSERTAÇÃO

USO DE DADOS AEROGEOFÍSICOS NA MODELAGEM PREDITIVA DE
ATRIBUTOS DO SOLO NO MUNICÍPIO DE BOM JARDIM - RJ

Blenda Pereira Bastos

2023



**UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
INSTITUTO DE GEOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM E EVOLUÇÃO
GEOLÓGICA**

**USO DE DADOS AEROGEOFÍSICOS NA MODELAGEM PREDITIVA DE
ATRIBUTOS DO SOLO NO MUNICÍPIO DE BOM JARDIM – RJ**

BLEND A PEREIRA BASTOS

Sob a orientação da Professora

Dra. Helena Saraiva Koenow Pinheiro

e Coorientação do Professor

Dr. Francisco José Fonseca Ferreira

Dissertação submetida como requisito para obtenção do grau de **Mestre em Geociências**, no Curso de Pós-Graduação em Modelagem e Evolução Geológica, Área de Concentração Modelagem de Processos Geológicos e Geodinâmicos, na linha de Pesquisa de Processos Exógenos e Geoambientais.

Seropédica, RJ

Outubro de 2023

Universidade Federal Rural do Rio de Janeiro
Biblioteca Central / Seção de Processamento Técnico

Ficha catalográfica elaborada
com os dados fornecidos pelo(a) autor(a)

B327u Bastos, Blenda Pereira, 1994-
 Uso de dados aerogeofísicos na modelagem preditiva
de atributos do solo no Município de Bom Jardim - RJ /
Blenda Pereira Bastos. - Seropédica, 2023.
 87 f.: il.

 Orientadora: Helena Saraiva Koenow Pinheiro.
 Coorientador: Francisco José Fonseca Ferreira.
 Dissertação (Mestrado). -- Universidade Federal
Rural do Rio de Janeiro, Programa de Pós-Graduação em
Modelagem e Evolução Geológica, 2023.

 1. mapeamento digital de solos. 2. dados
gamaespectrométricos. 3. dados magnetométricos. 4.
material de origem. I. Saraiva Koenow Pinheiro,
Helena, 1985-, orient. II. Fonseca Ferreira,
Francisco José, -, coorient. III Universidade Federal
Rural do Rio de Janeiro. Programa de Pós-Graduação em
Modelagem e Evolução Geológica. IV. Título.



TERMO Nº1215/2023-PPGMEG (11.39.00.12)

(NºdoProtocolo:NÃO PROTOCOLADO)

Seropédica-RJ, 04 de outubro de 2023.

**UNIVERSIDADE FEDERAL RURAL DO RIO DE
JANEIRO INSTITUTO DE AGRONOMIA
CURSO DE PÓS-GRADUAÇÃO EM MODELAGEM E EVOLUÇÃO GEOLÓGICA**

NOME DO(A) PÓS-GRADUANDO(A): BLENDA PEREIRA BASTOS

Dissertação de mestrado submetida como requisito parcial para obtenção do grau de Mestre em Geociências no Curso de Pós-Graduação em Modelagem e Evolução Geológica, área de Concentração em Modelagem de Processos Geológicos e Geodinâmicos.

DISSERTAÇÃO APROVADA EM 04/10/2023

Conforme deliberação número 248/2022 do CEPE, as defesas de dissertação de mestrado podem ser realizadas de modo remoto e, portanto, nas versões finais das teses e dissertações, as assinaturas originais dos membros da banca examinadora poderão ser substituídas por documento(s) com assinaturas eletrônicas. Estas devem ser feitas na própria folha de assinaturas, através do SIPAC, ou do Sistema Eletrônico de Informações (SEI) e neste caso a folha com a assinatura deve constar como anexo ao final da tese / dissertação. O referido procedimento já vinha sendo adotado conforme deliberação número 001/2020 da PROPPG, de 30/06/2020, tendo em vista a implementação de trabalho remoto e durante a vigência do período de suspensão das atividades acadêmicas presenciais, em virtude das medidas adotadas para reduzir a propagação da pandemia de Covid-19.

Coordenador do PPGMEG: Artur Corval Vieira.

BANCA EXAMINADORA:

Dra. Helena Saraiva Koenow Pinheiro (orientadora e presidente da banca; Departamento de Solos-UFRRJ).
Dr. Francisco José Fonseca Ferreira (co-orientador; Departamento de Geologia - UFRRJ).
Dr. Suze Nei Pereira Guimarães (membro titular interno; Departamento de Petrologia e Geotectônica -UFRRJ).
Dr. Waldir de Carvalho Junior (membro titular externo, EMBRAPA).

(Assinado digitalmente em 04/10/2023 13:31)
ARTUR CORVAL VIEIRA
PROFESSOR DO MAGISTERIO SUPERIOR
DPETRO (11.39.40)
Matricula: 1648928

(Assinado digitalmente em 04/10/2023 13:59)
HELENA SARAIVA KOENOW PINHEIRO
CHEFE DE DEPARTAMENTO
DeptS (12.28.01.00.00.00.33)
Matricula: 2223668

(Assinado digitalmente em 04/10/2023 14:15)
SUZE NEI PEREIRA GUIMARAES
ASSINANTE EXTERNO
CPF: 856.780.741-72

(Assinado digitalmente em 04/10/2023 17:31)
FRANCISCO JOSÉ FONSECA FERREIRA
ASSINANTE EXTERNO
CPF: 037.395.384-49

(Assinado digitalmente em 04/10/2023 14:53)
WALDIR DE CARVALHO JUNIOR
ASSINANTE EXTERNO
CPF: 583.287.696-20

AGRADECIMENTOS

Primeiramente, agradeço à minha família, em especial meus pais Marcos e Sandra, que sempre me apoiaram incondicionalmente. Seus ensinamentos de perseverança, determinação e responsabilidade foram e são fundamentais para o meu sucesso. Agradeço também à minha madrinha Livia, pelo incentivo e por ser minha primeira referência da ciência na família.

Agradeço à minha querida orientadora, amiga e irmã, Helena. Sua orientação e apoio foram inestimáveis ao longo de toda minha trajetória acadêmica desde 2018 (do PIBIC para vida). Sou muito grata por sua confiança, paciência, compreensão e amizade.

Agradeço também ao meu coorientador, Francisco, por sua valiosa contribuição. Sua experiência e conhecimento foram fundamentais para o sucesso deste trabalho.

Agradeço aos meus colegas Waldir, Nilson e Pedro, por todo conhecimento e bons momentos compartilhados em campanhas de campo. Agradeço também à Lúcia, pelo suporte durante a minha trajetória.

Agradeço também à minha família ruralina: Alfano, Aline, Aysla, Babi, Bia, Brenda, Bruno, Cadu, Caroly, Dri, Elis, Guerart, Gustavo, Helena, Iago, Jéssica, Lara, Larissa, Lili, Marcellin, Marie, Matheus, Nattasha, Nicole, Renan, Talyson, Thiago, Vitória, Vivian, Vivi, Zé e toda a galera do “rolê”. Obrigada por cada momento compartilhado.

Por fim, gostaria de agradecer à Universidade Federal Rural do Rio de Janeiro, em especial ao PPGMEG pelo suporte e estrutura. Agradeço também aos órgãos Faperj e CNPq pelo financiamento dos projetos que permitiram o desenvolvimento da minha pesquisa.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

RESUMO

Os dados aerogeofísicos possuem grande potencial no Mapeamento Digital de Solos (MDS) como covariáveis na predição de atributos físicos e químicos do solo, pois representam fatores de formação como, por exemplo, material de origem e relevo. O objetivo geral desse estudo foi avaliar a importância e o desempenho dos dados aerogeofísicos na predição de atributos do solo, a saber: Saturação por Alumínio (ASat), Saturação por Bases (BS), Capacidade de Troca Catiônica (CEC), Argila (Clay) e Carbono Orgânico (OC). As covariáveis aerogeofísicas incluem contagem total (TC), potássio (K), urânio equivalente (eU), tório equivalente (eTh), razões entre os radioelementos (eTh/K, eU/K e eU/eTh), Fator F, potássio anômalo (Kd), urânio anômalo (Ud), campo magnético anômalo (AMF), derivada vertical (GZ), derivadas horizontais (GX e GY) e índice máfico (MI). A abordagem utilizada para verificar a importância dos dados aerogeofísicos na predição dos atributos do solo foi baseada na aplicação de técnicas de modelagem preditiva utilizando dois conjuntos distintos de covariáveis: (1) covariáveis derivadas do Modelo Digital de Elevação (MDE), imagens de satélite da missão Sentinel-2 e dados aerogeofísicos; e (2) covariáveis derivadas do MDE e imagens de satélite da missão Sentinel-2, excluindo os dados aerogeofísicos. O estudo foi realizado no município de Bom Jardim, Rio de Janeiro, Brasil, com uma área de 382,430 km² e banco de dados legados de solo composto por 208 amostras (harmonizados a uma profundidade predefinida de 0-30cm). As covariáveis não explicativas para os atributos de solo selecionados foram excluídas usando as funções *nearZeroVar*, *findCorrelation* e *Recursive Feature Elimination (RFE)*. Por meio das covariáveis selecionadas, os modelos *Random Forest (RF)* e *Support Vector Machine (SVM)* foram aplicados com amostras separadas para treinamento (75%) e validação (25%). O desempenho dos modelos foi avaliado quantitativamente por meio dos índices *R-squared (R²)*, *Root Mean Square Error (RMSE)*, e *Mean Absolute Error (MAE)*, bem como valores nulos do modelo (*NULL RMSE* e *NULL MAE*) e *Coefficiente de Variação (CV%)*. As etapas *RFE*, treinamento, validação e predição para cada propriedade do solo foram repetidas 100 vezes para garantir robustez dos modelos. O modelo *RF* mostrou melhor performance na modelagem com os dados aerogeofísicos (valores de *R²* variando entre 0.15 e 0.23), assim como o modelo *SVM* (valores de *R²* variando entre 0.08 e 0.23), quando comparados aos modelos *RF* (valores de *R²* variando entre 0.10 e 0.20) e *SVM* (valores de *R²* variando entre 0.04 e 0.10), sem os dados aerogeofísicos. Os resultados sugerem que a modelagem com dados aerogeofísicos pode ser uma ferramenta útil para o mapeamento de solos e seus atributos, especialmente em regiões onde as técnicas convencionais de levantamento de solos são impraticáveis ou economicamente inviáveis. No entanto, é crucial reconhecer que a precisão dos dados aerogeofísicos na previsão das propriedades do solo pode variar dependendo de vários fatores (a qualidade e resolução dos dados, e amostragem em escala apropriada, por exemplo) e mais pesquisas são necessárias para explorar esses fatores e determinar a abordagem ideal para utilizá-los no MDS.

Palavras-chave: mapeamento digital de solos, dados gamaespectrométricos, dados magnetométricos, material de origem

ABSTRACT

Airborne geophysical data (AGD) have great potential as covariates in Digital Soil Mapping (DSM) once they represent soil-forming factors, as parental material and relief, for example. The study main goal was to evaluate the importance of AGD in predicting soil attributes, such as: aluminum saturation (ASat), base saturation (BS), cation exchange capacity (CEC), clay, and organic carbon (OC). The AGD predictor variables include total count (TC), potassium (K), uranium equivalent (eU), and thorium equivalent (eTh), ratios between these elements (eTh/K, eU/K, and eU/eTh), factor F or F-parameter, anomalous potassium (Kd), anomalous uranium (Ud), anomalous magnetic field (AMF), vertical derivative (GZ), horizontal derivatives (GX and GY), and mafic index (MI). The approach to assess the importance of AGD data was based on apply predictive modeling techniques using two input datasets: (1) digital elevation model (DEM) covariates and Sentinel-2 images with AGD; and (2) DEM covariates and Sentinel-2 images, without the AGD. The study was conducted in Bom Jardim, a county in Rio de Janeiro-Brazil with an area of 382,430 km², with a database of 208 soil samples to a predefined depth (0–30 cm). Non-explanatory covariates for the selected soil attributes were excluded through nearZeroVar, findCorrelation and Recursive Feature Elimination (RFE) functions. Through the selected covariables, the Random Forest (RF) and Support Vector Machine (SVM) models were applied with separate samples for training (75%) and validation (25%). The model's performance was evaluated through the R-squared (R²), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE), as well as null model values and Coefficient of Variation (CV%). The RF algorithm showed better performance with AGD (R² values ranging from 0.15 to 0.23), as well as the SVM model (R² values ranging from 0.08 to 0.23), when compared to RF (R² values ranging from 0.10 to 0.20) and SVM (R² values ranging from 0.04 to 0.10) models, without AGD. The results suggest that AGD can be helpful for soil mapping. Nevertheless, it is crucial to acknowledge that the accuracy of AGD in predicting soil properties could vary depending on various common factors in DSM, such as the quality and resolution of the covariates and available soil data. Further research is needed to determine the optimal approach for using AGD in soil mapping.

Keywords: machine learning; digital soil mapping; gamma-ray spectrometry data; magnetic data; parent material

SUMÁRIO

1. Introdução	12
2. Revisão bibliográfica	15
2.1 Dados geofísicos: um breve resumo.....	15
2.1.1 Gamaespectrometria.....	15
2.1.2 Magnetometria	20
2.2 Produtos aerogeofísicos.....	26
3. Caracterização da área de estudo	31
3.1 Contexto Geológico.....	34
3.2 Contexto Geomorfológico.....	38
3.3 Contexto pedológico	42
4. Artigo: Could Airborne Geophysical Data Be Used to Improve Predictive Modeling of Agronomic Soil Properties in Tropical Hillslope Area?	47
4.1 Introduction.....	47
4.2 Materials and Methods.....	49
4.2.1 Study Area and Soil Data.....	49
4.2.2 Covariate Acquisition and Processing.....	52
4.2.3 Modeling Strategy	55
4.3 Results	59
4.3.1 Covariates Selection.....	59
4.3.2 Models' Performance	60
4.3.3 Map Prediction and uncertainty	62
4.3.4 AGD Data Importance	67
4.4 Discussion	69
4.4.1 AGD Importance to Predict Soil Attributes through DSM.....	69
4.4.2 Soil Properties and AGD Relationships	72
4.4.3 Precautions and Challenges.....	75
4.5 Conclusions	76
5. Considerações finais	77
6. Referências	78
7. Anexo A	87
8. Anexo B	89

LISTA DE FIGURAS

Figura 1: Espectros de radiação gama mostrando as posições da janela de energia para cada elemento e da contagem total (Ribeiro et al., 2014).....	16
Figura 2: Variação das concentrações de K, U e Th de rochas ígneas com o incremento de sílica. Modificado de Dickson e Scott (1997) (Ribeiro et al., 2014).....	17
Figura 3: Perfil topográfico simplificado ilustrando a influência dos processos geomorfológicos com relação as respostas gamaespectrométricas. Modificado de Wilford et al. (1997) (Ferreira et al., 2016).....	20
Figura 4: Aproximação do campo geomagnético por um dipolo inclinado ao eixo de rotação da Terra. Os polos magnéticos reais não coincidem com o polo geomagnético do dipolo ou os polos geográficos. GNP - polo norte geomagnético, GSP - polo sul geomagnético, ME - equador magnético, NP - polo norte geográfico e SP - polo sul geográfico, VMI - intensidade magnética vertical, F - campo geomagnético (F), Inc – inclinação, e Dec - declinação (adaptado de Dentith e Mudge, 2014).....	21
Figura 5: Ilustração esquemática do alinhamento dos dipolos magnéticos de diferentes tipos de magnetismo (Dentith e Mudge, 2014).....	22
Figura 6: Suscetibilidade magnética em diferentes tipos de rochas (Dentith e Mudge, 2014).....	24
Figura 7: Suscetibilidade magnética de granitoides relacionados a depósitos de cobre porfírico, molibdênio e estanho-tungstênio. Ton – tonalita, Qmd – quartzo monzodiorito, Grd – granodiorito, e Gra – granito (principalmente monzogranito) (Dentith e Mudge, 2014).....	25
Figura 8: Representação das derivadas horizontais em pontos de aquisição de uma grade regular (Dentith e Mudge, 2014).....	30
Figura 9: Mapa de localização do município de Bom Jardim - RJ.....	32
Figura 10: Mapa de Clima em escala 1:5.000.000 representa as diferentes zonas climáticas do território brasileiro agrupadas pela temperatura e umidade (IBGE, 2002); Rede de drenagem e Modelo Digital de Elevação (IBGE, 2018).....	33
Figura 11: Mapa geológico da região de Bom Jardim – RJ (escala original 1:400.000). Adaptado de CPRM (2016a).....	35
Figura 12: Padrão de relevo do município de Bom Jardim - RJ (escala original 1:100.000) (Adaptado de CPRM, 2017).....	40
Figura 13: Solos do Médio Alto Curso do Rio Grande, Região Serrana do Estado do Rio de Janeiro (escala original 1:100.000) (adaptado de Calderano Filho et al. (2012)). Mapa com padronização das cores das classes de 1º e 2º níveis categóricos de acordo com o Sistema Brasileiro de Classificação de Solos SiBCS – Santos et al. (2018).....	42

Figure 14: Location map of the Bom Jardim County and soil samples described by Calderano Filho (2012). On the left below, the digital elevation model (DEM) derived from Rio de Janeiro cartographic database (original scale 1:25,000) (IBGE, 2018).....	50
Figure 15: (a) Soil map adapted from (Calderano Filho et al., 2012), SiBCS: Brazilian Soil Classification System (Santos et al., 2018), WRB: World Reference Base for Soil Resources (FAO, 2014); (b) geological map (original scale 1:400,000) (CPRM, 2016a).	51
Figure 16: Modeling strategy flowchart.....	56
Figure 17: Boxplots with the number of predictors selected by the models for each soil property throughout the 100 runs.....	59
Figure 18: Performance of the models' RF and SVM on the prediction of the ASat, BS, CEC, Clay, and OC with AGD and without AGD assessed by the R-squared (R^2) metric.	61
Figure 19: Final mean maps of the RF model with AGD: (a) ASat (%), (b) BS (%), (c) CEC (cmolc kg ⁻¹), (d) Clay (g kg ⁻¹), and (e) OC (g kg ⁻¹).	63
Figure 20: Coefficient of variation maps of the RF model with AGD: (a) ASat, (b) BS, (c) CEC, (d) Clay, and (e) OC.	64
Figure 21: Final mean maps of the SVM model with AGD: (a) ASat (%), (b) BS (%), (c) CEC (cmolc kg ⁻¹), (d) Clay (g kg ⁻¹), and (e) OC (g kg ⁻¹).	64
Figure 22: Coefficient of variation maps of the SVM model with AGD: (a) ASat, (b) BS, (c) CEC, (d) Clay, and (e) OC.	65
Figure 23: Top 10 predictors' frequency of the models' RF and SVM on the ASat, BS, and CEC prediction with AGD.	67
Figure 24: Top 10 predictors' frequency of the models' RF and SVM on the Clay and OC prediction with AGD.	68
Figure 25: Spearman's correlation matrix between AGD and soil attributes. "X" represents not statistically significant values at the 0.2 significance level.	69

LISTA DE TABELAS

Tabela 1: Variação da concentração média dos radioelementos em rochas e solos na Austrália. Modificada de Dickson e Scott (1997).	18
Tabela 2: Resumo das litologias mapeadas no município de Bom Jardim – RJ (CPRM, 2016a; CPRM, 2016b).	35
Tabela 3: Padrões de relevo do município de Bom Jardim – RJ adaptado de CPRM (2017).	41
Tabela 4: Descrição das unidades de mapeamento no município de Bom Jardim de acordo com SiBCS – Santos et al. (2018) adaptado de Calderano Filho et al. (2012).	42
Table 5: Statistics of soil attributes used in the prediction.	52
Table 6: DEM covariates used for prediction and their respective references.	53
Table 7: Sentinel-2 covariates used for prediction and their respective references.	54
Table 8: Survey characteristics of Rio de Janeiro Aerogeophysical Project (CPRM, 2012).	54
Table 9: AGD covariates used for prediction and their respective references.	55
Table 10: Models' performance with and without AGD for all soil attributes.	60
Table 11: Number of times the model reached values $R^2 \geq 0.20$	62
Table 12: The mean values of CV% maps for modeling with and without AGD.	63
Table 13: Comparison of the results of this study with those obtained by Mello et al. (2022).	70

LISTA DE SIGLAS E ABREVIATURAS

AGD	Airborne Geophysical Data
DSM	Digital Soil Mapping
AMF	Anomalous Magnetic Field
Asat	Aluminum Saturation
BS	Base Saturation
CatchmentA	Catchment Area
CatchmentS	Catchment Slope
CEC	Cation Exchange Capacity
Convergenc	Convergence Index
CV%	Coefficient of Variation
DEM	Digital Elevation Model
eTh	Thorium equivalent
eU	Uranium equivalent
FactorF	Factor F or F-parameter
FerricIron	Ferric Iron
FerricOxi	Ferric Oxides
FerrousSilic	Ferrous silicates (Biotite, chlorite, amphibole)
Gradient	Downslope Distance Gradient
GSI	Grain Size Index
GX e GY	Horizontal derivatives
GZ	Vertical derivative
K, Kperc	Potassium
Kd	Anomalous potassium
Longitudin	Longitudinal Curvature
MAE	Mean Absolute Error
MaximumCurv	Maximum Curvature
MI	Mafic Index
MidSlope	Mid-Slope Position
NDVI	Normalized Difference Vegetation Index
Normalized	Normalized Height
OC	Organic Carbon
PlanCurv	Plan Curvature
ProfileCurv	Profile Curvature
R²	R-squared
RF	Random Forest
RFE	Recursive Feature Elimination
RMSE	Root Mean Square Error

SlopeHeig	Slope Height
Standardiz	Standardized Height
SVM	Support Vector Machine
TC	Total Count
TPI	Topographic Position Index
TRI	Terrain Ruggedness Index
TWI	Topographic Wetness Index
Ud	Anomalous uranium
ValleyDep	Valley Depth

1. Introdução

O Mapeamento Digital de Solos (MDS), segundo Lagacherie e McBratney (2006), busca a integração entre sistemas de informações espaciais convencionais, que englobam conhecimentos especializados em levantamento de solo e sensoriamento remoto, e sistemas de inferência espacial de solo baseados em modelos preditivos, visando aprimorar a compreensão da distribuição espacial das propriedades/atributos do solo. Em outras palavras, o objetivo é gerar novas informações (ou novos dados de solo) a partir dos dados já disponíveis, estudar suas propriedades e variáveis ambientais relacionadas e produzir mapas mais informativos, com maior resolução espacial, precisão e estimativas de incerteza (Lagacherie e McBratney, 2006; Lagacherie, 2008; McBratney et al., 2003).

Para alcançar esse objetivo, o MDS depende de variáveis ambientais adequadas para representar fatores de formação do solo (Jenny, 1994) como preditores. De acordo com Chen et al. (2022), relevo, organismos e clima são as três covariáveis ambientais mais utilizadas no MDS, e a frequência de outros fatores, como material de origem, é limitada. Sabendo que a maioria dos solos estudados são solos minerais e, conseqüentemente, suas propriedades são influenciadas por características intrínsecas da rocha “mãe”, é de amplo interesse atualizar as informações sobre material de origem na ciência do solo (Wilson, 2019). Wilson (2019) demonstra que o material de origem pode ser um fator importante, destacando as diferenças entre solos herdados de granitos e basaltos em algumas propriedades importantes do solo como, por exemplo, teores de argila, retenção de água, acidez, saturação por bases, porosidade, erodibilidade, entre outros.

A utilização de dados gamaespectrométricos como covariável representante do material de origem no MDS tem se tornado cada vez mais frequente (por exemplo, Loiseau et al., 2020; Loiseau et al., 2021; Mello et al., 2021; Mello et al., 2022; Maino et al., 2022). Esses dados também são úteis para estabelecer associações com processos de denudação do relevo, bem como as taxas relativas de formação do solo e erosão (Minty, 1997; Wilford e Minty, 2006), conforme também apontado por Reinhardt e Herrmann (2019) em uma revisão crítica sobre o uso dos dados gamaespectrométricos na ciência do solo. A gamaespectrometria é um método geofísico que mede a radiação natural emitida por rochas e solos. Esse método permite medir as concentrações de K (potássio), U (urânio) e Th (tório) a partir de suas séries de decaimento radioativo (^{40}K , ^{238}U e

²³²Th), geralmente utilizando detectores de NaI(Tl) em levantamentos *in situ* ou em levantamentos aéreos (Minty, 1997). Uma característica relevante desse método é o aumento das concentrações de potássio, urânio e tório em rochas ígneas, à medida que o teor de sílica aumenta, especialmente em rochas ácidas como granitos e gnaisses (Wilford et al., 2012; Dickson e Scott, 1997). Outra aplicação relevante na ciência do solo consiste no mapeamento da distribuição desses elementos em perfis de solo (Dickson e Scott, 1997; Mello et al., 2021; Reinhardt e Herrmann, 2019). Tal procedimento se deve ao fato de que distintos horizontes do solo podem apresentar concentrações variadas desses elementos radioativos, possibilitando, assim, o mapeamento vertical desses elementos e o entendimento da relação com as propriedades do solo em profundidade.

A magnetometria é mais um exemplo de método geofísico que pode ser utilizado no MDS. Esse método permite medir a intensidade e direção do campo geomagnético na superfície terrestre. Ao mapear as anomalias magnéticas, é possível obter informações relevantes acerca das propriedades magnéticas das rochas e estruturas geológicas (Dentith e Mudge, 2014). No caso dos dados magnéticos, as leituras resultantes são principalmente influenciadas pela presença de minerais magnéticos. Geralmente, os minerais magnéticos estão associados a rochas mais básicas e com teor elevado de minerais ferromagnesianos, como os basaltos. No entanto, dentro de uma mesma área, é possível fazer algumas generalizações, uma vez que rochas da mesma região podem apresentar um aumento na susceptibilidade magnética conforme aumenta a maficidade (Dentith e Mudge, 2014).

Assim como a gamaespectrometria, a magnetometria pode ser medida *in situ* ou através de levantamentos aéreos. Os estudos que envolvem dados magnéticos e solos geralmente são realizados *in situ* (por exemplo, Valaee et al., 2016; Sarmast et al., 2017; Mello et al., 2020; Mello et al., 2022). No entanto, sua aplicabilidade pode ser válida no MDS devido à possibilidade de contrastar com os dados gamaespectrométricos em relação ao tipo de material de origem enfatizado por esses métodos. Além disso, os dados gamaespectrométricos e magnéticos aéreos são disponibilizados gratuitamente pelo Serviço Geológico do Brasil (Projeto Aerogeofísico do Rio de Janeiro – CPRM (2012)), o que incentiva pesquisadores e cientistas a promoverem avanços em diversas áreas, impulsionando setores como ciência de dados, inteligência artificial, tecnologia geoespacial, entre outros. O Índice Máfico é um exemplo de covariável com potencial para o MDS,

uma vez que estabelece uma relação entre dados gamaespectrométricos e magnéticos. Conforme mencionado por Iza et al. (2018), essa técnica pode ser utilizada para mapear solos lateríticos.

Portanto, a hipótese do presente estudo é formulada na premissa de que os dados aerogeofísicos podem aprimorar a modelagem preditiva das propriedades do solo como Saturação por Alumínio (ASat), Saturação por Bases (BS), Capacidade de Troca Catiônica (CEC), Argila (Clay) e Carbono Orgânico (OC). Para avaliar essa hipótese, a pesquisa utilizou dados legados de solo das propriedades mencionadas em mapeamento realizado por Calderano Filho et al. (2012) no município de Bom Jardim, Estado do Rio de Janeiro; e técnicas de modelagem preditiva para comparar a eficácia de dois diferentes grupos de covariáveis que representam os fatores de formação do solo: (1) covariáveis derivadas do Modelo Digital de Elevação (MDE) e imagens de satélite da missão Sentinel-2 com dados aerogeofísicos; e (2) covariáveis derivadas do MDE e imagens de satélite da missão Sentinel-2, sem os dados aerogeofísicos.

As covariáveis preditoras candidatas dos dados aerogeofísicos incluem contagem total ($\mu\text{R/h}$), K (potássio), eU (equivalente de urânio) e eTh (equivalente de tório), bem como as razões entre esses elementos (eTh/K, eU/K e eU/eTh), o Parâmetro F ou Fator F, potássio anômalo (Kd), urânio anômalo (Ud), campo magnético anômalo (AMF), derivada vertical (GZ), derivadas horizontais (GX e GY) e índice máfico (MI). A significância dessas covariáveis foi avaliada através da análise de dois modelos preditivos Random Forest (RF) (Breiman, 2001) e Support Vector Machine (SVM) (Cortes e Vannik, 1995), métricas como R-squared (R^2), Root Mean Square Error (RMSE), e Mean Absolute Error (MAE), bem como valores nulos do modelo (NULL RMSE e NULL MAE), Coeficiente de Variação (CV%), frequência das covariáveis e correlações de Spearman, metodologia de modelagem baseada em trabalhos recentes como Gomes et al. (2019) Mello et al. (2022) e Siqueira et al. (2023).

A dissertação foi estruturada da seguinte forma: inicialmente, foi realizada uma breve revisão bibliográfica sobre os dados aerogeofísicos e exemplos de produtos associados; em seguida, foi feita uma revisão da área de estudo, incluindo localização, contexto geológico, geomorfológico e pedológico; seguindo com a apresentação do artigo publicado: “*Could Airborne Geophysical Data Be Used to Improve Predictive Modeling of Agronomic Soil Properties in Tropical Hillslope Area?*” que inclui descrição metodológica, apresentação dos resultados, discussão, conclusão e anexo associado ao presente estudo; encerrando, também em anexo, com o

artigo publicado: “*Clustering airborne gamma-ray spectrometry data in Nova Friburgo, State of Rio de Janeiro, southeastern Brazil*”. Produto também associado aos projetos de pesquisas “Aerogamaespectrometria Aplicada ao Mapeamento Digital Geopedológico na Região Serrana do Estado do Rio de Janeiro”, aprovado pelo edital MCTIC/CNPq N° 28/2018, e “Propriedades e funções do solo como suporte ao manejo de terras: uso de dados legados e de campo, sensoriamento remoto e abordagem pedométrica no mapeamento de solos”, aprovado pelo edital FAPERJ N° 11/2018.

2. Revisão bibliográfica

2.1 Dados geofísicos: um breve resumo

2.1.1 Gamaespectrometria

A gamaespectrometria é um método científico que tem como propósito a determinação das concentrações dos radioelementos potássio (%), urânio (ppm) e tório (ppm) por meio da detecção da radiação gama emitida durante o processo de decaimento radioativo, conforme destacado por Wilford et al. (1997) e Minty (1997). Um gamaespectrômetro mede a contagem total num intervalo energético entre 0,41-2,81 MeV (milhões de elétrons-volt), no qual compreende os fotopicos característicos de cada elemento (Figura 1).

A concentração do potássio é medida através do decaimento do isótopo ^{40}K para ^{40}Ar com energia de 1.46 MeV. Já o urânio e tório possuem séries de decaimento radioativo mais complexas. Urânio ocorre naturalmente nas formas dos radioisótopos ^{238}U e ^{235}U e possui uma série de decaimento que termina nos isótopos estáveis ^{206}Pb e ^{207}Pb , respectivamente. Em levantamentos aéreos, é utilizada a série de decaimento da forma ^{238}U , visto que ^{235}U contribui apenas com 0.72% da ocorrência natural de urânio e a energia associada à sua série de decaimento é baixa para ser diagnosticada (Minty, 1997). O tório ocorre naturalmente como ^{232}Th e possui série de decaimento que termina no isótopo estável ^{208}Pb . Ambos não emitem radiação gama e suas concentrações são estimadas através de seus radiogênicos, ^{208}Tl (2.61 MeV) para tório e ^{214}Bi (1.76 MeV) para urânio (IAEA, 2003).

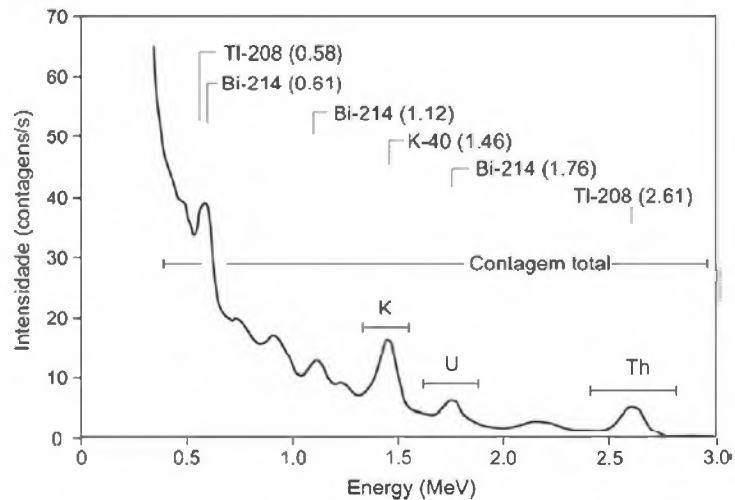


Figura 1: Espectros de radiação gama mostrando as posições da janela de energia para cada elemento e da contagem total (Ribeiro et al., 2014).

No levantamento aéreo, a fonte dos dados se encontra entre 30 e 40 cm de profundidade, e relacionam-se com a mineralogia e geoquímica da rocha fonte e materiais intemperizados como, por exemplo, solos e saprólitos. O potássio é um dos elementos mais abundantes da crosta terrestre com média de 2.3%, urânio e tório são menos comuns e possuem médias de 3 e 12 ppm, respectivamente (Wilford e Minty, 2006). O potássio ocorre principalmente nos álcalis-feldspatos, micas e illita e pode ocorrer em menores quantidades em argilas como montmorilonita e caulinita. Por outro lado, as principais ocorrências do urânio e tório são em minerais acessórios como zircão, monazita, esfeno e rutilo, por exemplo (Wilford e Minty, 2006; Dentith e Mudge, 2014).

Em relação as suas propriedades cristalóquímicas, o potássio possui raio iônico de 1.33 Å (ângström), formando com outros cátions (Cs, Rb, Sr, Eu bivalente, Pb e Ba) o grupo dos cátions LFS (Low Field Strength) e LIL (Large Ion Lithophile). Já o urânio e o tório constituem cátions de tamanhos médios com facilidade mútua de substituição em estruturas cristalinas com elementos do grupo HFS (High Field Strength), incluindo ETR (Elementos Terras Raras) trivalentes, além de Ce^{4+} , Y, Pb^{4+} , Zr, Hf, Cs, Ti, Nb e Ta (Ulbrich et al., 2009).

Minerais formadores de rocha podem conter traços de urânio e tório, como quartzo, feldspatos e piroxênio por substituição de cátions de raio iônico e carga similares. Em condições redutoras, urânio e tório ocorrem no estado tetravalente e são relativamente insolúveis. No estado U^{+6} , o urânio é solúvel no meio aquoso e pode formar complexos com uma variedade de íons. Sua

mobilidade pode ser afetada pela adsorção em óxidos de ferro hidratados, argilominerais e coloides. Em condições neutras ou ácidas, compostos orgânicos podem aumentar a solubilidade do tório, mas, de maneira geral, a baixa solubilidade em águas naturais faz com que esse elemento seja amplamente transportado em material particulado (Dentith e Mudge, 2014).

Uma importante característica desses elementos é que eles tendem a estar correlacionados com a maioria dos litotipos. O conteúdo de sílica exerce influência na abundância de K, U e Th de maneira proporcional em rochas ígneas, ou seja, quanto maior o conteúdo de sílica, maior o conteúdo desses elementos (Figura 2). Consequentemente, rochas félsicas possuem maiores teores desses radioelementos, seguidas por rochas intermediárias e rochas máficas e ultramáficas com teores mais baixos (Dickson e Scott, 1997).

Nas rochas metamórficas o conteúdo desses elementos, de maneira geral, depende do principalmente do protólito e, no caso das rochas sedimentares, da rocha-mãe e do ciclo sedimentar. Gnaisses félsicos estão entre as rochas mais radioativas, assim como os migmatitos. Já os gnaisses máficos e anfibolitos são menos radioativos. Metassedimentos, incluindo xistos, pelitos, filitos e ardósias, possuem conteúdo semelhante aos gnaisses. Nas rochas sedimentares, a resposta radioativa é determinada pelo conteúdo de feldspatos, micas e argilominerais e minerais pesados contendo urânio e tório (Dentith e Mudge, 2014).

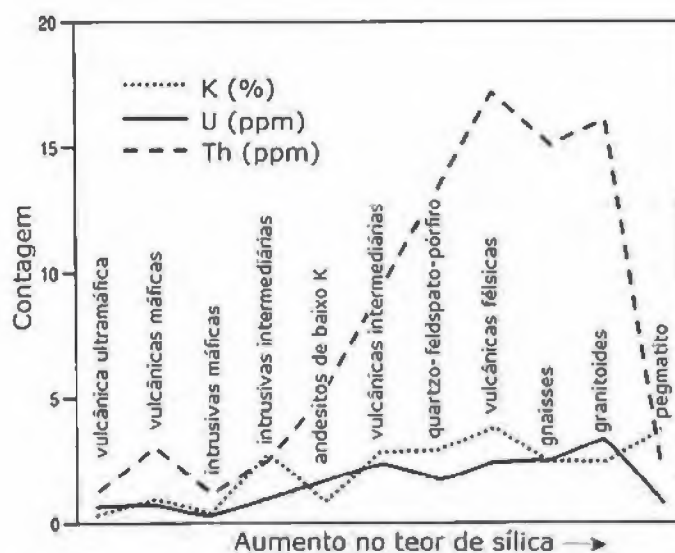


Figura 2: Variação das concentrações de K, U e Th de rochas ígneas com o incremento de sílica. Modificado de Dickson e Scott (1997) (Ribeiro et al., 2014).

Na Tabela 1 abaixo estão listados os teores dos três elementos em rochas e solos em estudos realizados na Austrália por Dickson e Scott (1997).

Tabela 1: Variação da concentração média dos radioelementos em rochas e solos na Austrália. Modificada de Dickson e Scott (1997).

Tipo de rocha	Rocha			Solo		
	K (%)	U (ppm)	Th (ppm)	K (%)	U (ppm)	Th (ppm)
Intrusivas						
Granitóides	0.3-4.5	0.4-7.8	2.3-45.0	0.4-3.9	0.5-7.8	2.0-37.0
Gnaisses	2.4-3.8	2.1-3.6	18.0-55.0	0.7-1.9	1.6-3.8	6.0-19.0
Pegmatito	2.6-5.5	0.3-1.0	0.3-9.6			
Aplitos	0.6-4.0	1.0-8.0	3.0-20.0			
Quartzo-feldspática porfírica	1.0-5.0	1.3-2.9	6.0-14.0			
Intermediárias	0.7-5.6	0.1-1.2	0.8-6.1	0.7-3.4	1.5-2.3	1.5-2.3
Máficas	0.1-0.8	0.0-1.1	0.0-3.1			
Extrusivas						
Félsicas	2.0-4.4	1.4-13	13.0-28.0	1.8-3.2	1.3-2.4	10.0-18.0
Intermediárias	1.8-4.1	0.9-5.6	1.5-15.0	1.0-2.7	1.2-3.6	4.0-17.0
Andesitos baixo-K	0.7-0.9	1.0-2.5	3.0-8.0	0.8-1.5	1.2-1.5	4.0-6.0
Máficas	0.3-1.3	0.3-1.3	2.0-5.0	0.2-1.4	0.6-2.5	3.3-13.0
Ultramáficas	0.2-0.9	0.3-0.9	0.0-4.0	0.6	2.0	6.0
Sedimentares						
Xistos arqueanos	0.4-1.6	0.3-1.3	1.0-5.0	0.8	1.2	3.0
Outros xistos	0.1-4.0	1.6-3.8	10.0-55.0	0.7-3.0	1.2-5.0	6.0-19.0
Arenitos	0.0-5.5	0.7-5.1	4.0-22.0	0.1-2.4	1.2-4.4	7.0-18.0
Carbonatos	0.0-0.5	0.4-2.9	0.0-2.9			

Segundo Wilford et al. (1997) a resposta gamaespectrométrica do solo pode ser descrita como primária, onde as concentrações estão relacionadas diretamente a geoquímica da rocha, ou seja, solos *in situ* pouco evoluídos que conservam as características mineralógicas da fonte; e secundária, relacionada a redistribuição dos radioelementos por processos de intemperismo e pedogênese. A segunda é mais complexa, envolvendo fatores como grau de evolução do intemperismo, incorporação de solos transportados, relevo, erosão e clima.

De maneira geral, por ser solúvel na maioria dos ambientes, o conteúdo de potássio diminui com o avanço do intemperismo, logo, tende a ser lixiviado nos perfis de solo. Exceções ocorrem quando o potássio é incorporado a argilas potássicas ou quando está associado a grandes fenocristais de k-feldspatos ou micas que levam maior tempo para intemperizar. Em contraste, urânio e tório são mais estáveis em perfis do solo. Quando liberados da rocha, tendem a ser adsorvidos por minerais argilosos, oxihidróxidos de ferro e alumínio e matéria orgânica, além de estarem associados a minerais resistentes que persistem no solo (Ulbrich et al., 2009).

Portanto, a resposta radioativa depende da história evolutiva da paisagem, como as taxas relativas de erosão e intemperismo, desnudação do relevo ou preservação de antigos perfis de intemperismo. Paisagens com superfícies estáveis, menos ativas, irão preservar o manto de intemperismo e sua consequente evolução, refletindo a resposta gamaespectrométrica de materiais secundários, como argilas e óxidos. Já em áreas mais ativas, de topografia íngreme, a taxa de erosão excede a formação de regolito, formando solos menos espessos ou expondo totalmente a rocha fonte (Wilford et al., 1997).

Nos levantamentos aéreos, esses fatores influenciam a intensidade da resposta radioativa. Em termos gerais, altos valores indicam regiões de solo menos espessos e/ou exposição da rocha. Por outro lado, valores mais baixos são característicos de áreas onde ocorre acumulação e desenvolvimento da camada de intemperismo, conforme ilustrado na Figura 3 a seguir.

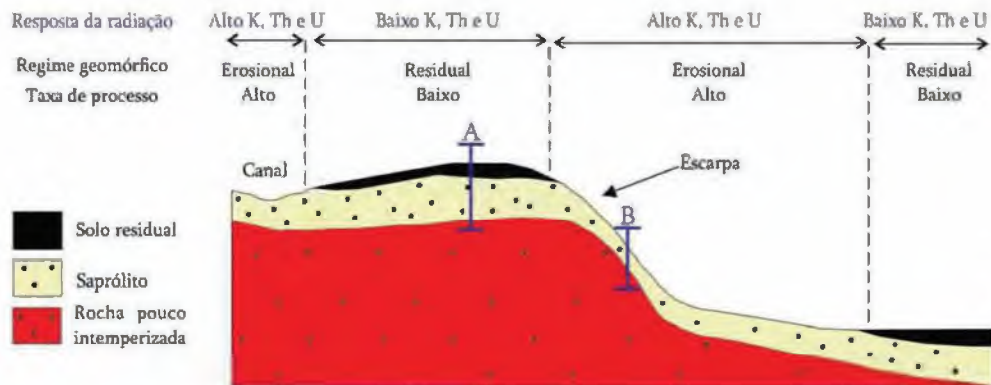


Figura 3: Perfil topográfico simplificado ilustrando a influência dos processos geomorfológicos com relação às respostas gamaespectrométricas. Modificado de Wilford et al. (1997) (Ferreira et al., 2016).

2.1.2 Magnetometria

A susceptibilidade magnética no mapeamento geológico é uma medida de resposta da rocha ao campo magnético da Terra, sendo essa rocha magneticamente suscetível. Em um levantamento magnético são detectados, além das anomalias magnética crustais, a intensidade do campo magnético da Terra, que inclui o campo geomagnético principal estável associado ao núcleo, e uma componente instável associada ao campo magnético externo. O campo externo variável representa um ruído na interpretação desses dados, portanto deve ser removido (Dentith e Mudge, 2014).

O campo magnético interno originado do núcleo da Terra é responsável pelo magnetismo das rochas. O campo geomagnético em um determinado local é definido por sua intensidade e direção (descrita pela sua inclinação e declinação). Locais onde as linhas do campo são paralelas à superfície da Terra, ou seja, onde a sua inclinação magnética é zero, definem o Equador magnético. Ao norte do Equador magnético, a inclinação é considerada positiva e ao sul do Equador magnético, a inclinação é considerada negativa (Dentith e Mudge, 2014) (Figura 4).

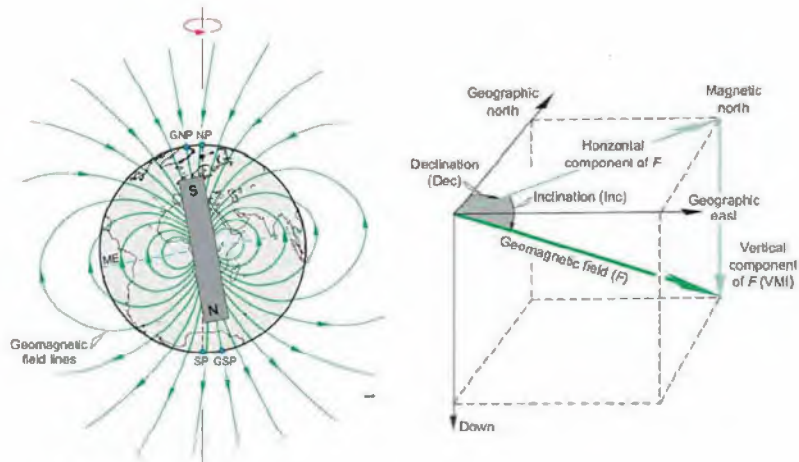


Figura 4: Aproximação do campo geomagnético por um dipolo inclinado ao eixo de rotação da Terra. Os polos magnéticos reais não coincidem com o polo geomagnético do dipolo ou os polos geográficos. GNP - geomagnetic north pole (polo norte geomagnético), GSP - geomagnetic south pole (polo sul geomagnético), ME - magnetic equator (equador magnético), NP - north geographic pole (polo norte geográfico), SP - south geographic pole (polo sul geográfico), VMI - vertical magnetic intensity (intensidade magnética vertical), F - geomagnetic field (campo geomagnético), Inc - inclination (inclinação), e Dec - declination (declinação) (adaptado de Dentith e Mudge, 2014).

O campo interno da Terra é descrito por um modelo matemático baseado em observações realizadas ao longo de anos em diversas localidades ao redor do mundo, conhecido como o Campo de Referência Geomagnética Internacional (*International Geomagnetic Reference Field - IGRF*). O IGRF fornece a direção e a intensidade do campo geomagnético em qualquer localização, sendo atualizado a cada 5 anos para considerar a atividade magnética atual observada (Dentith e Mudge, 2014). Mudanças no magnetismo das rochas se sobrepõem às variações originadas no interior do núcleo, portanto, em levantamentos magnéticos, o IGRF é reduzido, de forma a garantir que as medições magnéticas reflitam unicamente os elementos da crosta (CPRM, 2012).

Segundo Kearey et al. (2009), todo material é magnético em escala atômica, onde cada átomo se comporta como um dipolo em razão do comportamento do spin de seus elétrons e trajetória orbital dos elétrons ao redor do núcleo. A forma como os elétrons estão dispostos e como se comportam quando há a influência de um campo externo, classificam os materiais em: diamagnético, paramagnético, ferromagnético, ferrimagnético ou antiferromagnético.

O material diamagnético possui todos os elétrons pareados e susceptibilidade magnética negativa muito pequena. Minerais com essa característica raramente são detectados em

levantamentos aéreos, como por exemplo fases puras de quartzo, calcita e feldspato. Uma exceção é a halita que, dependendo da quantidade e do ambiente, como em bacias sedimentares, pode exibir anomalias. Minerais classificados como paramagnéticos possuem orbitais com elétrons não pareados, resultando em um campo magnético gerado pelos seus spins. Possuem baixa susceptibilidade positiva e são representados por minerais como olivina, piroxênios, anfibólios, granada, micas e carbonatos de ferro e manganês. Assim como os minerais diamagnéticos, minerais paramagnéticos exibem anomalias em levantamentos aéreos quando ocorrem em ambientes “magneticamente silenciosos”, sendo assim destacados (Isles e Rankin, 2013; Dentith e Mudge, 2014).

Substâncias paramétricas podem constituir um domínio magnético acoplado a outras substâncias paramétricas através da interação de seus elétrons não pareados, resultando em materiais com características ferromagnéticas, ferrimagnéticas ou antiferromagnéticas (Kearey et al., 2009). Sendo esses diferenciados pela intensidade e orientação de seus dipolos, podendo ser paralelos ou antiparalelos (Figura 5). Os minerais ferromagnéticos, ferrimagnéticos e antiferromagnéticos são, dependendo da bibliografia, classificados como apenas um grupo: os ferromagnéticos. Sendo o ferrimagnetismo e antiferromagnetismo classificados como diferentes tipos de ferromagnetismo (Isles e Rankin, 2013).

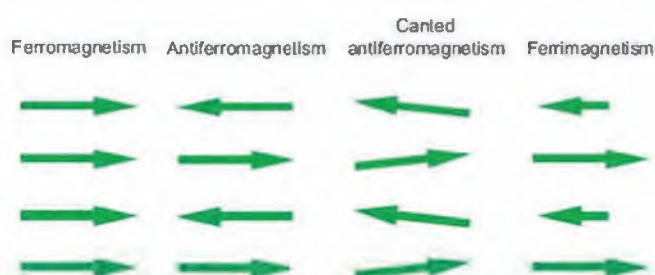


Figura 5: Ilustração esquemática do alinhamento dos dipolos magnéticos de diferentes tipos de magnetismo: ferromagnetismo (ferromagnetism), antiferromagnetismo (antiferromagnetism), antiferromagnetismo inclinado (canted antiferromagnetism) e ferrimagnetismo (ferrimagnetism) (Dentith e Mudge, 2014).

Os minerais ferromagnéticos possuem dipolos paralelos na mesma direção e isso resulta numa magnetização forte e uma susceptibilidade muito alta (Kearey et al., 2009). Em materiais antiferromagnéticos ocorre acoplamento antiparalelo de mesma intensidade que se anulam. Porém, o antiparalelismo imperfeito dos dipolos, ou seja, o antiferromagnetismo inclinado, pode causar um pequeno magnetismo intrínseco, como por exemplo na hematita (Dentith e Mudge, 2014). Nos

materiais ferrimagnéticos, como magnetita, pirrotita monoclínica, maghemita e ilmenita, o acoplamento é antiparalelo com diferentes intensidades, assumindo forte magnetização espontânea a alta susceptibilidade (Kearey et al., 2009; Dentith e Mudge, 2014).

Sob a influência de um campo magnético externo, como o da Terra, os domínios magnetizados crescem na orientação do campo do aplicado. Dependendo da intensidade, o material pode possuir magnetismo remanescente ou induzido. Em condições de menor intensidade do campo aplicado, a magnetização é do tipo induzida e é perdida após a remoção do campo. Por outro lado, em condições de maior intensidade, o material retém a magnetização após a remoção do campo, sendo chamada de remanescente ou remanente (Kearey et al., 2009). Alguns materiais exibem ambos os tipos de magnetismo, ou são predominantemente ou exclusivamente de um tipo, ou são incapazes de se tornarem magnéticos. Portanto, o magnetismo geral de um corpo é a soma vetorial das magnetizações induzida e remanente podem ser expressas por suas intensidades pela razão de Königsberger (Königsberger ratio) (Dentith e Mudge, 2014).

Conforme foi descrito acima, os minerais podem ser categorizados de acordo com suas características químicas e físicas em resposta a um campo magnético externo. Sendo o grupo ferromagnético (aqui colocado na forma generalizada) representados por propriedades magnéticas mais fortes, exercendo maior influência nas propriedades magnéticas das rochas (Dentith e Mudge, 2014). Consequentemente, apesar da susceptibilidade magnética variar amplamente em todos os principais grupos de rochas, algumas generalizações podem ser feitas como mostra a Figura 6. Onde, na maioria dos casos, a distribuição de susceptibilidade magnética é bimodal, apresentando membros com altas e baixas susceptibilidades, refletindo a presença ou ausência, respectivamente, de minerais ferromagnéticos (Dentith e Mudge, 2014).

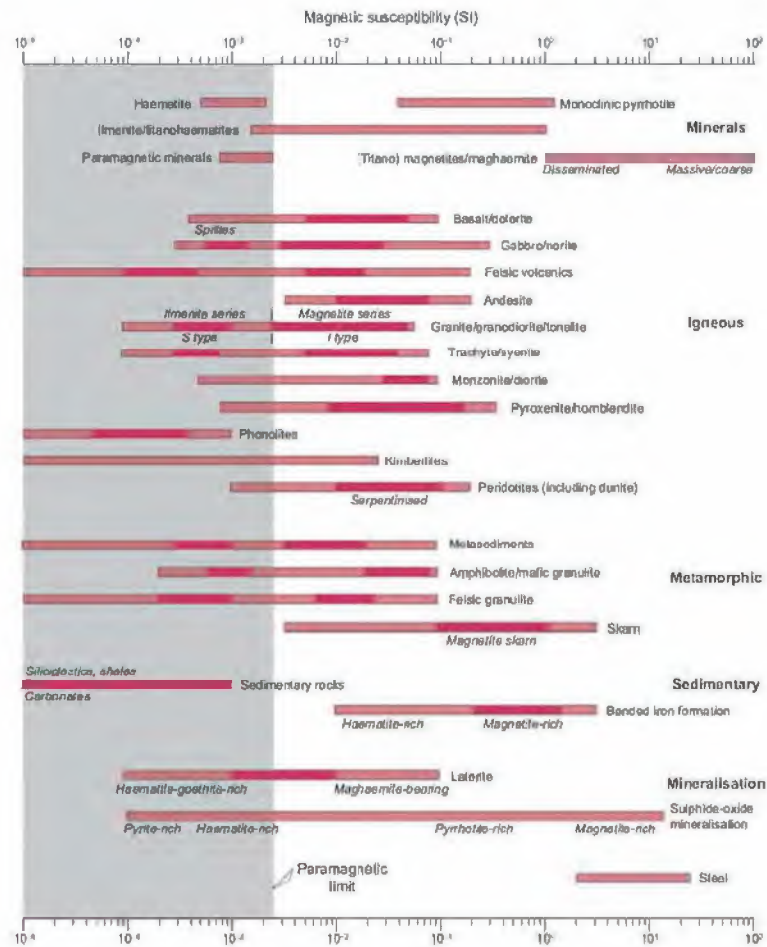


Figura 6: Susceptibilidade magnética em diferentes tipos de rochas (Dentith e Mudge, 2014).

Portanto, relações magnéticas importantes podem ser identificadas entre diferentes classes de rochas de um ambiente geológico. Em rochas ígneas, por exemplo, observa-se que a suscetibilidade magnética tende a aumentar com a maficidade em rochas de uma mesma área, o que possibilita a análise de contrastes magnéticos (Dentith e Mudge, 2014).

De acordo com a Figura 6, em granitoides, a suscetibilidade magnética bimodal reflete duas séries distintas que podem ser associadas aos granitos do tipo S e I de Chappell e White (1974) (citado por Dentith e Mudge, 2014), conforme também mostra a Figura 7. Foi observado que a série ilmenita, que corresponde aproximadamente aos granitos do tipo S, possui níveis mais baixos de suscetibilidade magnética, enquanto a série magnetita, que corresponde aproximadamente aos

granitos do tipo I, apresentam maior magnetismo. Em ambos os casos, a susceptibilidade diminui à medida que o teor de sílica aumenta (Dentith e Mudge, 2014).

As propriedades magnéticas dos granitoides também podem ser analisadas a partir da presença de minerais que indicam a condição de oxidação do ambiente. Por exemplo, os granodioritos com anfíbólio e biotita são predominantemente ferromagnéticos e apresentam susceptibilidade moderada, enquanto os granitoides com muscovita e biotita geralmente são paramagnéticos e fracamente magnéticos (Dentith e Mudge, 2014).

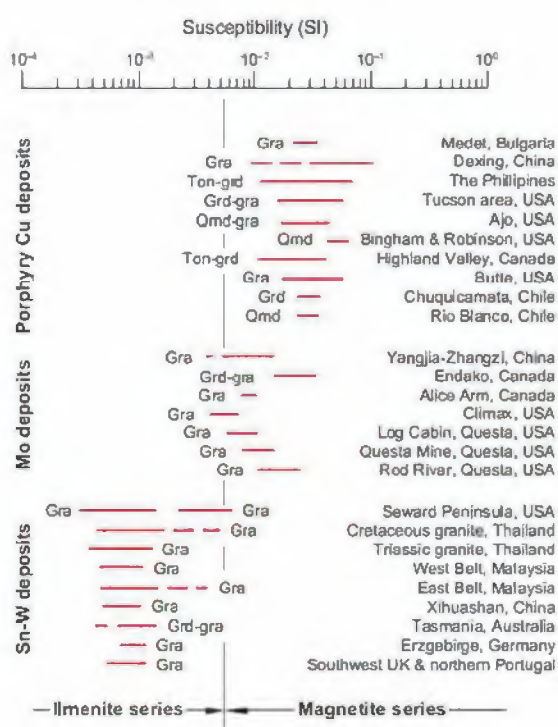


Figura 7: Suscetibilidade magnética de granitoides relacionados a depósitos de cobre porfírico, molibdênio e estanho-tungstênio. Ton – tonalita, Qmd – quartzo monzodiorito, Grd – granodiorito, e Gra – granito (principalmente monzogranito) (Dentith e Mudge, 2014).

A susceptibilidade magnética em rochas sedimentares é geralmente baixa, exceto no caso da formação ferrífera bandada. Porém, pode ser influenciada pela mineralogia da rocha fonte ou estado de oxidação do ambiente. Como por exemplo, unidades com alto teor de quartzo, originárias de áreas mais evoluídas, apresentam susceptibilidades mais baixas do que os sedimentos provenientes de terrenos vulcânicos imaturos (Dentith e Mudge, 2014). O estado de oxidação reflete na estabilidade do ferro associado a argilas em rochas sedimentares. Em um ambiente de natureza oxidante, onde a água do sistema está em contato com a atmosfera, minerais contendo

ferro ferroso são transformados em minerais de ferro férrico (como por exemplo, magnetita transformada em hematita). Por outro lado, em um ambiente redutor estagnado, especialmente rico em matéria orgânica, a magnetita é estável (Dentith e Mudge, 2014).

Rochas metamórficas também são influenciadas pela mineralogia de sua rocha fonte. Porém, o metamorfismo também pode alterar significativamente as propriedades magnéticas de uma rocha, através da destruição e criação de minerais magnéticos, aquecimento e consequente exsolução de fases minerais, deformação e recristalização de grãos, alinhamento preferencial de grãos etc. Ou seja, o efeito do metamorfismo nas propriedades magnéticas é complexo e existe uma gama quase infinita de possibilidades (Dentith e Mudge, 2014).

2.2 Produtos aerogeofísicos

Após o tratamento dos dados brutos, incluindo procedimentos de controle de qualidade e a aplicação de técnicas de correção (mais detalhes em CPRM (2012)), algumas técnicas de aprimoramento e exibição dos dados aerogeofísicos podem ser aplicadas, como operações matemáticas e combinações entre os canais.

Em relação aos dados aerogamaespectrométricos, além dos mapas de concentração de K (cm %), cU (urânio equivalente em ppm) e cTh (tório equivalente em ppm), estão disponíveis:

- Mapa de Contagem Total ($\mu\text{R/h}$ - micro Röntgen por hora):

O mapa de Contagem Total registra a quantidade geral de radiação gama detectada durante o levantamento aerogeofísico. É utilizado para realçar anomalias da radioatividade total que podem representar contatos entre unidades litológicas e estruturas (IAEA, 2003).

- Mapa de composição ternária

O mapa ternário é gerado a partir do padrão de cores RGB (Red, Green, Blue) ou CMY (Cyan, Magenta, Yellow) proporcionais aos valores de concentração do potássio, do urânio e do tório. O RGB associa o vermelho ao potássio, verde ao tório e azul ao urânio. Já o CMY considera as cores ciano, magenta e amarelo. Em ambas as formas de representação, o branco corresponde a alta concentração dos três elementos e o preto baixa concentração dos mesmos (IAEA, 2003).

A composição RGB, normalmente, é eficaz na distinção de unidades geológicas, mas pode levar a ambiguidades no processo de interpretação. Por exemplo, a predominância do vermelho, que normalmente representa altos valores de potássio e baixos valores de tório e urânio, pode estar associada tanto a granitos quanto a sedimentos imaturos. Da mesma forma, o branco pode indicar a presença de sienitos ou rochas vulcânicas ácidas, sem necessariamente apresentar altos valores dos três elementos. Em outras palavras, a interpretação desse mapa deve ser sempre realizada com o auxílio de outras informações para garantir sua consistência (Sordi, 2020).

- Mapas de razões entre os elementos:

Refere-se à relação entre as concentrações dos elementos potássio, urânio e tório. As razões são calculadas para destacar a diferença relativa entre as concentrações desses elementos em rochas de uma mesma região ou identificar áreas de alteração hidrotermal do tipo potássica ou fílica. A variação é resultado da geologia local, processos de enriquecimento e/ou empobrecimento e grau de intemperismo (Ribeiro et al., 2014). Alguns exemplos de razões utilizadas:

$$\text{Razão entre } eTh \text{ e } K = \frac{eTh}{K} \quad \text{Eq. 1}$$

$$\text{Razão entre } eU \text{ e } K = \frac{eU}{K} \quad \text{Eq. 2}$$

$$\text{Razão entre } eU \text{ e } eTh = \frac{eU}{eTh} \quad \text{Eq. 3}$$

- Mapa de Fator F:

O mapa de Fator F é uma relação matemática que combina as concentrações dos três radioelementos, realçando o enriquecimento de potássio e urânio em relação ao tório. Baixos valores do Fator F representam áreas com taxas de intemperismo e lixiviação elevadas, enquanto áreas enriquecidas em potássio e urânio por alteração hidrotermal apresentam altos valores (Ribeiro et al., 2014). O mapa de Fator F é calculado pela seguinte equação:

$$\text{Fator } F = \frac{K \times eU}{eTh} \quad \text{Eq. 4}$$

- Mapas anômalos (Kd e Ud)

Os mapas de potássio anômalo e urânio anômalo podem ser utilizados na interpretação de depósitos minerais associados a processos hidrotermais, por exemplo. Na definição desses mapas, a concentração de tório é utilizada como um controle litológico para estabelecer um valor "ideal" de potássio e urânio, com base na hipótese proposta por Saunders et al. (1993). Os autores sugerem que os fatores que influenciam a concentração e distribuição do tório também afetam de maneira previsível a concentração do urânio e potássio. Em outras palavras, as concentrações desses radioelementos variam de forma paralela, e a normalização pelo tório é empregada para eliminar esse efeito, destacando possíveis anomalias nos mapas.

O método descrito considera o K ideal (K_i) em relação ao eTh calculado pela equação:

$$K_i = \left(\frac{\text{média de } K}{\text{média de } eTh} \right) \times eTh \quad \text{Eq. 5}$$

Em seguida, os desvios dos valores reais a partir dos valores ideais são obtidos pela equação:

$$Kd = \frac{K - K_i}{K_i} \quad \text{Eq. 6}$$

Em relação aos dados aerogeofísicos magnéticos, primeiro é gerado o mapa do Campo Magnético Anômalo (nT – Nanotesla). A partir deste, é realizada a redução ao polo (Baranov, 1957), se necessário, de acordo com os parâmetros calculados para a data de aquisição. A bipolaridade do campo magnético terrestre gera diferentes padrões de anomalias conforme varia a latitude, tornando complexa a análise dos dados. A redução ao polo consiste na transformação dos dados, coletados em qualquer latitude, para a latitude onde a inclinação do campo é 90°, transformando anomalias dipolares em monopolares, tornando a análise mais viável. Em dados próximos ao Equador magnético, ou seja, baixas latitudes, a redução não é recomendada devido à complexidade das características das amplitudes magnéticas (Dentith e Mudge, 2014). Feita a redução ao polo, algumas técnicas de realce podem ser aplicadas como por exemplo:

- Continuações ascendente e descendente

Filtros de continuação ascendente e descendente envolvem a modificação dos dados magnéticos ao aumentar ou diminuir a altitude de aquisição. A continuação ascendente é

equivalente a aumentar a altura do levantamento, enquanto a continuação descendente é equivalente a diminuí-la (Dentith e Mudge, 2014).

A continuação ascendente é uma técnica de filtragem que se assemelha a um filtro passa-baixa. Nesse processo, os comprimentos de onda mais curtos (altas frequências) associados a fontes magnéticas próximas à superfície são atenuados, enquanto as fontes com comprimentos de onda mais longos ou mais profundas (baixas frequências) são destacadas. Isso permite uma melhor identificação das fontes magnéticas localizadas em profundidades maiores. Por outro lado, a continuação descendente amplifica o espectro com a diminuição da altura, sendo destacados os comprimentos de onda mais curtos, o que resulta no realce das fontes magnéticas mais rasas próximas à superfície (Dentith e Mudge, 2014).

- Derivada Vertical (GZ) e Derivadas Horizontais (GX e GY):

Os gradientes ou derivadas de um parâmetro P adquirido, ao longo de linhas paralelas, são conhecidos como derivada X ($\partial P/\partial x$), medida na direção ao longo da linha, e derivada Y ($\partial P/\partial y$) medida na direção transversal à linha (Figura 8). Esses gradientes representam as variações horizontais. Já o gradiente vertical, ou seja, a derivada Z ($\partial P/\partial z$), indica como o parâmetro medido se altera à medida que a distância para a fonte (altura da pesquisa) varia (Dentith e Mudge, 2014).

As derivadas horizontais realçam mudanças no parâmetro medido na direção perpendicular em que foram calculadas, ou seja, a derivada GX realça estruturas orientadas na direção Y e a derivada GY realça estruturas orientadas na direção X. A derivada vertical GZ, ao contrário das derivadas horizontais, apresenta a vantagem de responder a mudanças independentemente da orientação de sua fonte em relação à linha de pesquisa (Dentith e Mudge, 2014).

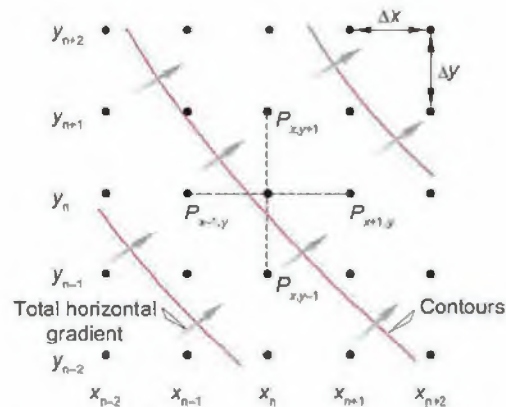


Figura 8: Representação das derivadas horizontais em pontos de aquisição de uma grade regular (Dentith e Mudge, 2014).

As derivadas horizontais e vertical também funcionam como filtros passa-alta e, portanto, acentuam as bordas de corpos magnéticos rasos, sendo chamadas de "detectores de borda", úteis no mapeamento de contatos e estruturas geológicas (Dentith e Mudge, 2014).

- Gradiente Horizontal Total (GHT):

As derivadas horizontais podem ser combinadas para gerar o GHT de P ($\partial P/\partial r$) na localização (x_n, y_n) na direção resultante (r) de seus componentes horizontais a partir da seguinte fórmula:

$$\frac{\partial P}{\partial r}(x_n, y_n) = \sqrt{\left(\frac{\partial P}{\partial x}\right)^2 + \left(\frac{\partial P}{\partial y}\right)^2} \quad \text{Eq. 7}$$

- Amplitude do Sinal Analítico (ASA)

Da mesma forma, o gradiente total ou Amplitude do Sinal Analítico (ASA) também pode ser calculado a partir das derivadas x, y e z do campo magnético pela seguinte fórmula:

$$\frac{\partial P}{\partial r}(x_n, y_n) = \sqrt{\left(\frac{\partial P}{\partial x}\right)^2 + \left(\frac{\partial P}{\partial y}\right)^2 + \left(\frac{\partial P}{\partial z}\right)^2} \quad \text{Eq. 8}$$

A ASA independe da direção de magnetização da fonte e é uma técnica eficiente para determinar localização de limites geológicos e estruturais, além da profundidade de corpos (Dentith e Mudge, 2014).

- Índice Máfico (IM):

O Índice Máfico (MI) é calculado a partir da combinação de dois métodos e tem como objetivo delinear assinaturas magnéticas, excluindo a influência de materiais superficiais ricos em ferro, como solos ou lateritas (Barbosa et al., 2013). Esse índice mostrou eficiência na diferenciação de rochas máficas e ultramáficas, eliminando a resposta magnética resultante de solos alterados, conforme constatado por Barbosa et al. (2013). Além disso, o IM também permite, com o auxílio de outros métodos, a diferenciação de crostas lateríticas desenvolvidas em diferentes rochas fontes (máficas e félsicas) como foi demonstrado por Iza et al. (2018).

O MI é gerado utilizando os canais de concentração de potássio, urânio e tório e a ASA pela seguinte fórmula:

$$IM = \frac{ASA}{(K \times eU \times eTh)} \quad \text{Eq. 8}$$

3. Caracterização da área de estudo

De acordo com a Divisão Territorial Brasileira (DTB) (IBGE, 2021a), o município de Bom Jardim está localizado na mesorregião Centro Fluminense do Estado do Rio de Janeiro (Figura 9), com área de 382.430 km² e população estimada de 27.779 habitantes (IBGE, 2021b). A área faz limites com os municípios Nova Friburgo, Duas Barras, Cordeiro e Trajano de Moraes que, com os municípios Cantagalo, Carmo, Macuco, Petrópolis, Santa Maria Madalena, São José do Vale do Rio Preto, São Sebastião do Alto, Sumidouro e Teresópolis, integram a região administrativa denominada Região Serrana.

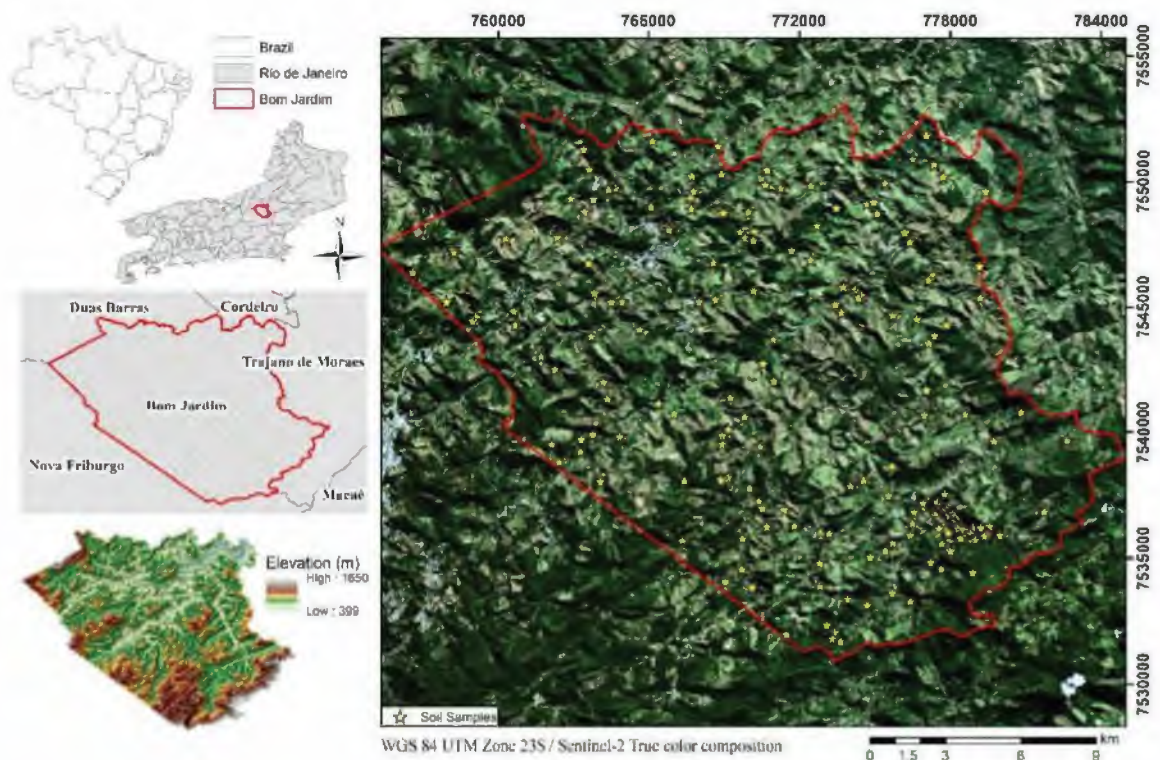


Figura 9: Mapa de localização do município de Bom Jardim - RJ.

Segundo IBGE (2002), Bom Jardim integra a região de clima classificado como Tropical Brasil Central, subdividido em subquente com temperatura média de 15 a 18 °C; e mesotérmico brando com temperatura média entre 10 e 15 °C. A pluviosidade média da série de 1941 a 2020 é de 1.413,63 mm, sendo dezembro o mês mais úmido e o mais seco setembro (ANA, 2020). De acordo com a base cartográfica do Rio de Janeiro (IBGE, 2018), possui como principais rios o Córrego do Sertão, Rio Grande, Córrego Rosário, Ribeirão São José, Ribeirão do Capitão, Córrego Novo, Córrego Frio, Ribeirão do Klein, Córrego do Barro e Córrego do Humaitá, inclusos nas bacias do Rio Grande e Ribeirão São Domingos (Figura 10).

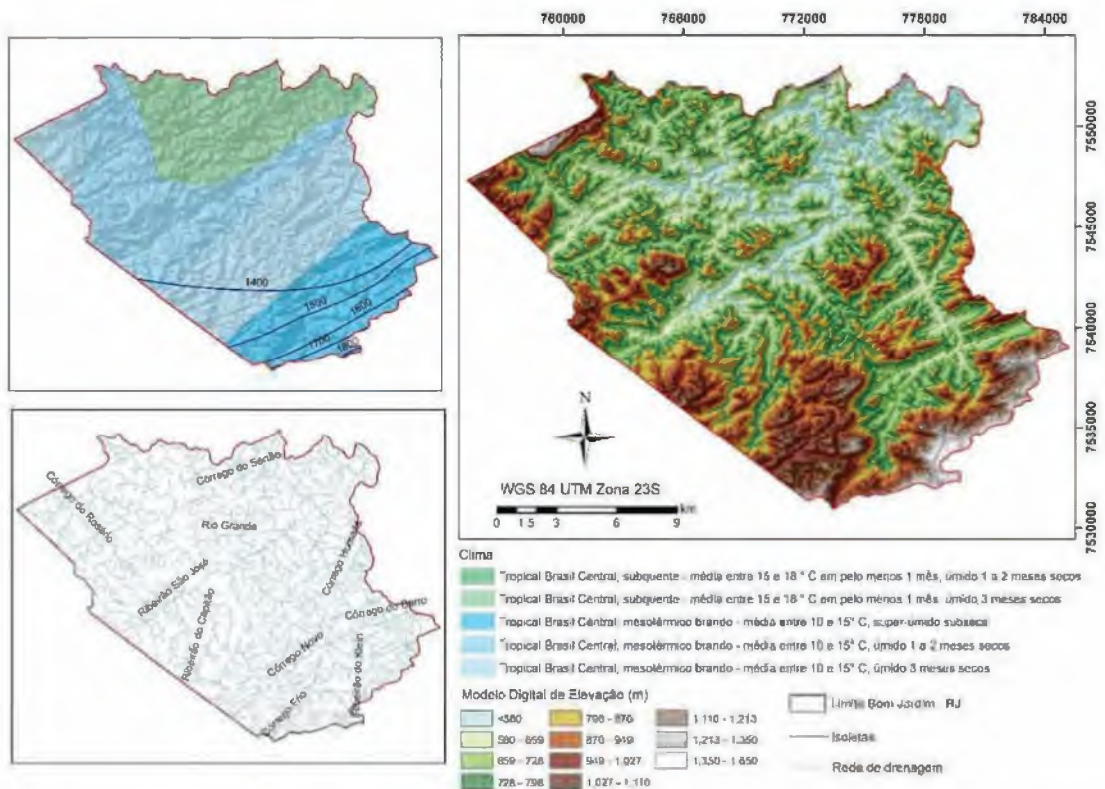


Figura 10: Mapa de Clima em escala 1:5.000.000 representa as diferentes zonas climáticas do território brasileiro agrupadas pela temperatura e umidade (IBGE, 2002); Rede de drenagem e Modelo Digital de Elevação (IBGE, 2018).

Em relação a cobertura vegetal, a região de Bom Jardim é composta predominantemente pela floresta tropical perenifólia e subperenifólia, caracterizada por apresentar vegetação densa e espécies arbóreas de grande porte. Apesar das intervenções antrópicas, principalmente por atividades agrícolas, a vegetação original da Mata Atlântica remanescente é de 51%, constituindo matas secundárias e matas nativas preservadas em áreas de maior altitude (Calderano Filho, 2012).

O restante, 49% da área, é composta predominantemente por pastagens naturais, culturas perenes, como o café, e cultivos de olerícolas e de flores ornamentais, como rosas e reflorestamento de eucalipto. A agricultura é regida em pequenos módulos rurais (agricultura familiar) nas encostas dos vales até as cabeceiras de drenagem (Calderano Filho et al., 2012). As interferências antrópicas associadas ao gradiente elevado do relevo montanhoso dominante, caracterizam um alto potencial de vulnerabilidade a eventos de erosão e movimentos de massa na região. Esse potencial é reduzido onde os terrenos são mais íngremes e elevados devido a preservação da cobertura florestal (Calderano Filho, 2012).

3.1 Contexto Geológico

Geologicamente, a região de Bom Jardim está inserida no contexto de evolução tectônica e magmática da porção oriental da Faixa Ribeira. Esse processo inclui rochas plutônicas pré a tardi-colisionais deformadas, coberturas sedimentares inclusas no processo colisional, gerando paragneisses e metassedimentos de alto grau metamórfico e corpos graníticos não deformados pós-colisionais (Tupinambá et al., 2012).

A evolução tectônica teve como início a instalação do arco magmático do Complexo Rio Negro e sedimentação das rochas do grupo São Fidelis (Criogeniano). Seguido pela colisão Ediacarana deste arco com a margem passiva a oeste, onde foram gerados os gnaisses da Suíte Cordeiro e colisão tardi-tectônica que gerou o magmatismo calci-alcalino da Suíte Serra dos Órgãos. Por fim, ocorreu exumação tectônica e colapso tectônico, gerando a intrusão dos granitos pós-colisionais da Suíte Nova Friburgo (Conselheiro Paulino, Granitos Sana e São José do Ribeirão) durante o Cambro-Ordoviciano (Tupinambá et al., 2013). A Figura 11 e a Tabela 2 exemplificam as litologias mapeadas na área de acordo com o Mapa Geológico do Estado do Rio de Janeiro (CPRM, 2016a; CPRM, 2016b).

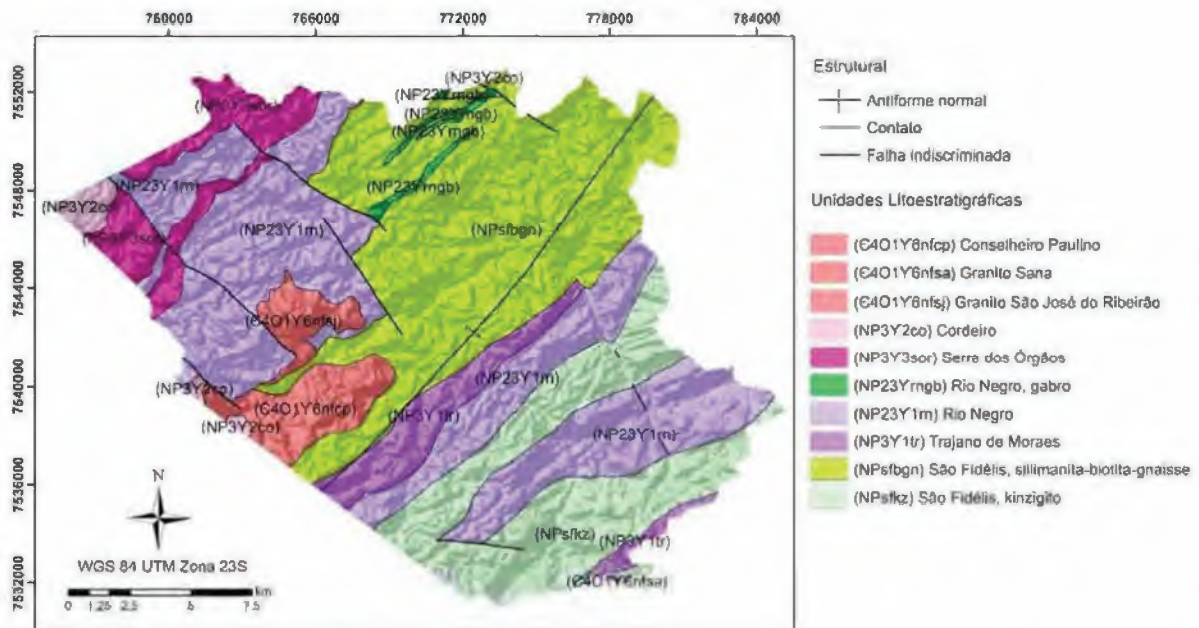


Figura 11: Mapa geológico da região de Bom Jardim – RJ (escala original 1:400.000). Adaptado de CPRM (2016a).

Tabela 2: Resumo das litologias mapeadas no município de Bom Jardim – RJ (CPRM, 2016a; CPRM, 2016b).

MAGMATISMO PÓS-COLISIONAL	
Suíte Nova Friburgo	Granito Conselheiro Paulino (E4O1Y6nfc)
	Granito São José do Ribeirão (E4O1Y6nfsj)
	Granito Sana (E4O1Y6nfsa)
MAGMATISMO SIN A TARDI-COLISIONAL	
Granitoides tardicolisionais	Suíte Serra dos Órgãos (NP3Y3sor)
Granitoides sin-colisionais anatóticos	Suíte Cordeiro (NP3Y2co)
MAGMATISMO NEOPROTEROZÓICO	
Complexo Rio Negro (NP23Y1rn)	
Complexo Rio Negro, gabro (NP23Y1ngb)	
Complexo Trajano de Moraes (NP3Y1tr)	
UNIDADES METASSEDIMENTARES NEOPROTEROZÓICAS	
Grupo São Fidélis (NPsf)	Kinzigito (NPsfkz)
	Sillimanita-biotita gnaiss (NPsfbgn)

As rochas do Grupo São Fidélis (NPsf) são regionalmente constituídas por biotita gnaisses granatíferos com sillimanita e, de maneira menos expressiva, cordierita. Apresentam também lentes de rochas calcissilicáticas, anfíbolitos e quartzitos feldspáticos (Tupinambá et al., 2007). Dentro desse grupo ocorrem as Unidades Kinzigito, Sillimanita-biotita gnaiss, Anfíbolito e Quartzito, estando presentes na área de estudo as duas primeiras. A rocha principal da Unidade Kinzigito (NPsfkz) é descrita por Tupinambá et al. (2012) como de difícil caracterização, devido a sua variedade textural causada por deformação cisalhante e diferentes graus de fusão parcial. A litologia mais frequente é o granada biotita gnaiss migmatítico mesocrático e de grão fino. Essa rocha possui matriz composta por quartzo, granada, sillimanita, microclina, plagioclásio (oligoclásio), biotita e biotita alterando para muscovita, e bandamento contendo lentes leucocráticas formadas por porfiroclastos milimétricos de feldspato e fitas de quartzo.

A Unidade Sillimanita-biotita gnaiss (NPsfbgn) é caracterizada principalmente por um (muscovita)-granada-sillimanita-biotita gnaiss com estrutura bandada (espessura milimétrica a centimétrica) e arranjos de porfiroblastos com cristais de feldspato. Possui composição tonalítica, migmatítica, mesocrática, de grão médio e com níveis xistosos ricos em biotita e granada (Tupinambá et al., 2012).

As rochas do Complexo Rio Negro (NP23Y1rn) são produtos do metamorfismo de rochas ígneas mais antigas aflorantes na região (630-600 Ma) (Heilbron e Machado, 2003), constituídas por gnaisses cinzentos bandados, de composição tonalítica e trondhjemítica (TTG) (Geraldes et al., 2012). Na área ocorrem na porção meridional do Batólito Serra dos Órgãos com direção NE-SW, com gnaisses apresentando forte foliação de baixo a médio ângulo. Também ocorrem intrusões e apófises de granitoides e de leucogranitos S, intercalações de metagabros, metaquartzo-dioritos e paleodiques anfíbolíticos (Geraldes et al., 2012). Os protólitos definem uma série expandida com biotita, hornblenda e titanita, variando de gabro a álcali-granito, de médio a baixo-K. As rochas com características de magmatismo de arco pré-colisional dessa série possuem elevado conteúdo em Ca e pouca distribuição de elementos terras raras (Tupinambá et al., 2012). O corpo gabróico (NP3β1rn gb) ocorre associado a litologias descritas por Geraldes et al. (2012) como leucogranodiorito gnaiss, rocha quartzo feldspática, brecha magmática com enclaves de rocha gabróicas e hornblendenditos ultramáficos; e é descrito como afloramento de soleira composto por piroxênio hornblenda gabronorito (Geraldes et al., 2012).

O Complexo Trajano de Moraes (NP3Y1tr) encerra um grupo magmático metamórfico migmatizado com porções migmatizadas distribuídas em uma longa faixa de direção N50E. Apresentam uma série de rochas plutônicas de composição básica a intermediária, classificadas como dioritos e quartzo dioritos, hornblenda granodioritos, ortognaisses tonalíticos, allanita sienogranitos e allanita monzogranitos. Na área de estudo, seu contato com a Unidade São Fidélis está relacionado a falhas de empurrão, apresentando plútons de composição granítica associados (Geraldés et al., 2012).

As rochas da Suíte Cordeiro (NP3Y2co) ocorrem de três formas de acordo com a encaixante: (1) como grandes soleiras ou pequenos diques junto às rochas do Complexo Rio Negro, apresentando composição granodiorítica à granítica e com raros termos tonalíticos; (2) como stocks em metassedimentos, predominando tipos hololeucocráticos de composição variando de granito à álcali-feldspato granito; e (3) como soleiras em metassedimentos com bandas pegmatóides concordantes. A litologia principal é um gnaiss leucocrático a hololeucocrático com alta susceptibilidade à erosão, poucos afloramentos rochosos e espessos mantos eluviais quartzo-feldspáticos. Apresenta predominância de feldspato potássico e quartzo, com plagioclásio de maneira menos expressiva. Sua composição varia de sienogranito a granodiorito, com maior concentração de sienogranito (Tupinambá et al., 2012).

A Suíte Serra dos Órgãos (NP3Y3sor) apresentam série petrográfica contínua à hornblenda e à titanita, com variações tonalíticas a graníticas com predomínio granodiorítico. Apresentam textura semelhante aos gnaisses do Complexo Rio Negro, diferenciando-se pela grande variação composicional do plagioclásio (de oligoclásio a labradorita, predominando andesina). Possui biotita, hornblenda, granada, quartzo, plagioclásio e feldspato potássico como minerais principais, sendo a proporção de feldspato potássico maior em relação ao plagioclásio (razão $\geq 2:1$) (Tupinambá et al., 2012). Segundo Tupinambá (1999), o magmatismo é descrito como uma série expandida, metaluminosa, calci-alcalina de médio a alto-K.

O magmatismo pós-colisional (Cambro-Ordoviciano) da Suíte Nova Friburgo é representado nessa área por três corpos intrusivos de granitos não deformados (Conselheiro Paulino, Sana e São José do Ribeirão). Esses são intrusivos nos ortognaisses da Suíte Serra dos Órgãos, do Complexo Rio Negro e paragnaisses do Grupo São Fidélis na forma de grandes corpos e diques. O magmatismo granítico pós-colisional agrupa uma série petrográfica descontínua à

biotita, titanita e allanita, com termos álcali-feldspato graníticos a granodioríticos (Tupinambá, 1999). A Suíte Nova Friburgo é representada por uma série calci-alcálica a álcali-cálcica metaluminosa de alto-K com magmatismo datado em dois pulsos (511 Ma e 486 Ma) (Valeriano et al., 2016).

O Corpo Conselheiro Paulino (€4O1Y6nfc) possui composição sienogranítica de carácter bimodal, representado por texturas e estruturas indicando mistura magmática. Possui como litotipo predominante o microgranito equigranular a microporfirítico, composto por biotita, feldspato, quartzo e plagioclásio. De forma menos expressiva, ocorre microdiorito assimilado a microgranito, sugerindo mistura magmática mecânica (Tupinambá et al., 2012b).

O Corpo Sana (€4O1Y6nfsa) é intrusivo nos paragneisses do Grupo São Fidélis e é cortado por pegmatitos graníticos e diques de diabásio, de espessura métrica e extensão quilométrica. Possui quartzo, microclina e plagioclásio como minerais essenciais e biotita, allanita, apatita, magnetita, titanita e zircão como acessórios. Possui composição ácida com elevados teores de álcalis, Ba, Sr e Zr, assinatura calcioalcalina de alto K e correlações negativas para Al_2O_3 , CaO, FeO, MgO, TiO_2 e P_2O_5 , sugerindo fracionamento de plagioclásio, biotita/óxidos de Fe e Ti e apatita na evolução magmática (CPRM, 2016b).

O Granito São José do Ribeirão (€4O1Y6nfsj) ocorre na área intrudindo os gnaisses migmatíticos do Complexo Rio Negro na forma de um extenso corpo tabular, bem como na forma de diques e sills de espessura pequena a média. Sua rocha principal é descrita como um granito de cor cinza, isotrópico, leucocrático e de textura inequigranular hipidiomórfica a porfírica com fenocristais de microclima de até 1 cm. Em relação a composição química, são rochas classificadas como metaluminosas, variando entre 68,68% e 71,28% em peso de conteúdo de SiO_2 , de natureza álcali-cálcica, com considerável conteúdo de K_2O e Ba e baixas de MgO (CPRM, 2016b).

3.2 Contexto Geomorfológico

O município de Bom Jardim está localizado na Unidade Morfoestrutural Cinturão Orogênico do Atlântico, caracterizado como macroestruturas pertencentes ao arcabouço tectônico da Faixa Ribeira submetidas a diferentes ciclos orogênicos. Essa unidade encerra as Unidades Morfoesculturais: Maciços Costeiros e Interiores, Maciços Alcalinos Intrusivos, Superfícies

Aplainadas nas Baixadas Litorâneas, Escarpas Serranas, Planaltos Residuais, Depressões Interplanálticas e Depressões Interplanálticas com Alinhamentos Serranos Escalonados (CPRM, 2020).

A área de estudo está situada na Unidade Morfoescultural Planaltos Residuais e na Unidade Geomorfológica Planalto Reverso da Região Serrana. A primeira é caracterizada por superfícies residuais soerguidas por tectônica, resistentes aos processos erosivos do Cenozóico Superior. Possui características morfológicas diferenciadas, incluindo relevos colinosos de baixa amplitude alternados com tipos de relevos mais movimentados e mais elevados, alta densidade de drenagem e suas vertentes são geralmente cobertas por colúvios, com gradientes variando entre suaves a médios, e elevados nas áreas de relevo acidentado (Dantas, 2000).

A segunda trata-se de um extenso planalto situado no reverso da Serra do Mar, prolongando até a escarpa reversa do planalto da Região Serrana. Na área se destaca relevo de colinas pouco dissecadas com predomínio de amplitudes topográficas inferiores a 100 m, vertentes convexo-côncavas e topos arredondados ou alongados, com sedimentação de colúvios e alúvios e densidade de drenagem média. Os padrões de relevo são do tipo montanhoso com amplitudes topográficas superiores a 500 m, muito acidentado, aguçados ou levemente arredondados com densidade de drenagem muito alta (Dantas, 2000).

Seguindo a hierarquia de classificação taxonômica (IBGE, 2009), a Unidade Geomorfológica Planalto Reverso da Região Serrana é subdividida na área em Padrões de Relevo como mostra a Figura 12 e Tabela 3.

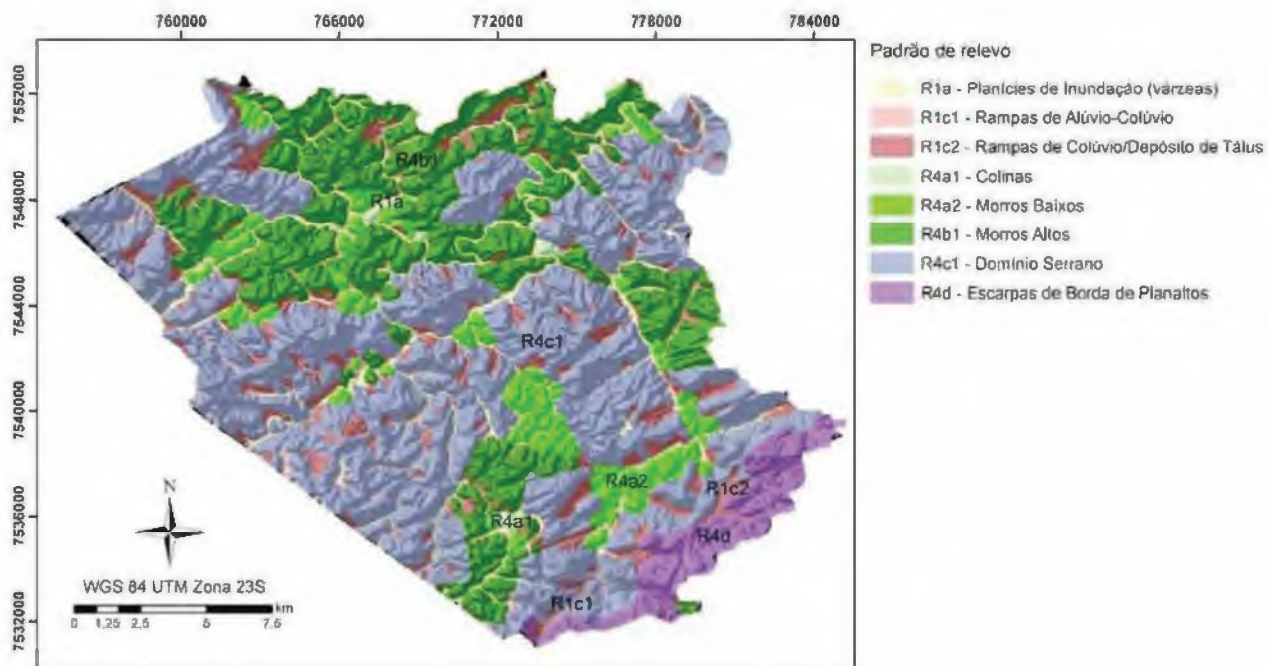


Figura 12: Padrão de relevo do município de Bom Jardim - RJ (escala original 1:100.000) (Adaptado de CPRM, 2017).

Tabela 3: Padrões de relevo do município de Bom Jardim – RJ adaptado de CPRM (2017).

Padrão de Relevo	Características predominantes
Escarpas de borda de Planaltos	Relevo montanhoso e extremamente acidentado com vertentes muito íngremes e dissecadas, retilíneas a côncavas, paredões rochosos e topos de cristas alinhadas ou aguçadas. Alta densidade de drenagem com a presença de tálus e colúvios nas baixas vertentes.
Domínio Serrano	Relevo montanhoso, muito acidentado, com vertentes retilíneas a côncavas e topos de cristas alinhadas, aguçados ou levemente arredondados. Sedimentação de colúvios e tálus e alta densidade de drenagem. Predomínio de vertentes com alto gradiente e eventual ocorrência de paredões rochosos subverticais e pães-de-açúcar.
Morros Altos	Relevo de morros de geometria convexo-côncava, francamente dissecados, movimentado com vertentes de gradientes médios a elevados e topos arredondados a aguçados. Drenagem moderada a alta com padrão subdendrítico a treliça.
Morros Baixos	Relevo típico do domínio de “mares-de-morros” com colinas dissecadas, vertentes convexo-côncavas com gradiente suave a moderado e topos arredondados. Moderada densidade de drenagem com padrão dendrítico ou subdendrítico.
Colinas	Relevo com colinas pouco dissecadas, vertentes convexas ou convexo-côncavas de gradiente suave, topos amplos, de morfologia alongada ou arredondada e baixas amplitudes de relevo. Em geral, baixa densidade de drenagem com padrão dendrítico.
Rampas de Colúvio/Depósitos de Tálus	Superfícies deposicionais fortemente inclinadas com presença de depósitos de encosta, matriz areno-argilosa a argilo-arenosa, rica em blocos mal selecionados e depósitos suavemente inclinados de rampas de alúvio-colúvio. Ocorrem nos sopés das vertentes íngremes das serras e escarpas.
Rampas de Alúvio-Colúvio	Superfícies deposicionais inclinadas com depósitos de encosta, arenoargilosos a argilo-arenosos, mal selecionados e depósitos praticamente planos das planícies fluviais. Ocorrem em meio ao domínio de mar-de-morro.
Planícies de Inundação (várzeas)	Superfícies sub-horizontais com depósitos arenosos ou areno-argilosos a argilosos, bem selecionados e situados nos fundos de vales. Gradientes extremamente suaves e convergentes aos cursos d’água principais. Terrenos mal drenados, periodicamente inundáveis.

3.3 Contexto pedológico

De acordo com o mapeamento realizado por Calderano Filho et al. (2012) de escala 1:100.000, o município de Bom Jardim inclui associações de Cambissolos Háplicos, Latossolos Amarelos, Latossolos Vermelho-amarelos, Latossolos Vermelhos, Argissolos Vermelho-amarelos, Argissolos Vermelhos como mostra a Figura 13 e a Tabela 4.

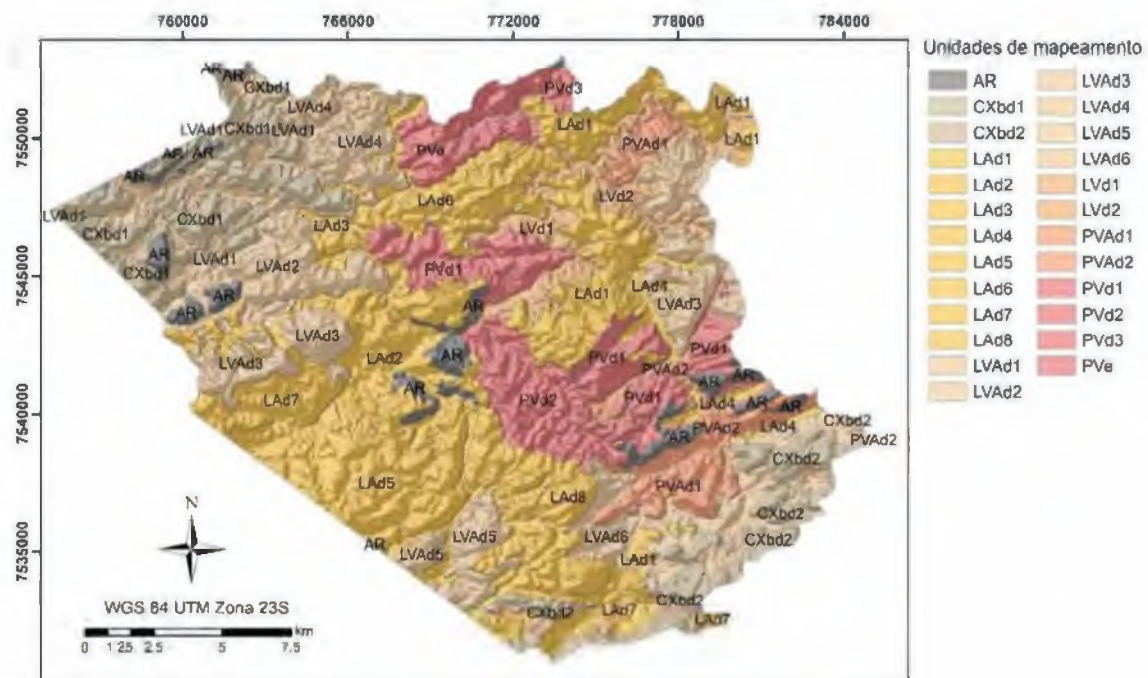


Figura 13: Solos do Médio Alto Curso do Rio Grande, Região Serrana do Estado do Rio de Janeiro (escala original 1:100.000) (adaptado de Calderano Filho et al. (2012)). Mapa com padronização das cores das classes de 1º e 2º níveis categóricos de acordo com o Sistema Brasileiro de Classificação de Solos SiBCS – Santos et al. (2018).

Tabela 4: Descrição das unidades de mapeamento no município de Bom Jardim de acordo com SiBCS – Santos et al. (2018) adaptado de Calderano Filho et al. (2012).

Legenda	Classificação SiBCS (2018)
AR	AFLORAMENTOS DE ROCHA + CAMBISSOLO HÁPLICO Tb Distrófico típico + NEOSSOLO LITÓLICO Tb Distrófico típico
CXbd1	CAMBISSOLO HÁPLICO Tb Distrófico típico + LATOSSOLO VERMELHO-AMARELO Distrófico típico + Afloramento de Rocha
CXbd2	CAMBISSOLO HÁPLICO Tb Distrófico úmbrico ou típico + Afloramento de Rocha + ARGISSOLO VERMELHO-AMARELO Distrófico típico
LAd1	LATOSSOLO AMARELO Distrófico típico + LATOSSOLO VERMELHO Distrófico típico + Afloramento de Rocha + inclusão CAMBISSOLO HÁPLICO Tb Distrófico típico + ARGISSOLO VERMELHO Distrófico latossólico

LAd2	LATOSSOLO AMARELO Distrófico típico + LATOSSOLO VERMELHO-AMARELO Distrófico típico + CAMBISSOLO HÁPLICO Tb Distrófico típico ou úmbrico
LAd3	LATOSSOLO AMARELO Distrófico típico + LATOSSOLO VERMELHO Distrófico típico + CAMBISSOLO HÁPLICO Tb Distrófico típico ou úmbrico
LAd4	LATOSSOLO AMARELO Distrófico típico + NEOSSOLO FLÚVICO Tb Distrófico típico + CAMBISSOLO FLÚVICO Tb Distrófico típico
LAd5	LATOSSOLO AMARELO Distrófico típico + ARGISSOLO VERMELHO - AMARELO Distrófico Latossólico + CAMBISSOLO HÁPLICO Tb Distrófico típico ou úmbrico + inclusão LATOSSOLO AMARELO Distrófico húmico
LAd6	LATOSSOLO AMARELO Distrófico típico + LATOSSOLO VERMELHO Distrófico típico + ARGISSOLO VERMELHO - AMARELO Distrófico típico
LAd7	LATOSSOLO AMARELO Distrófico húmico + LATOSSOLO VERMELHO-AMARELO Distrófico típico + CAMBISSOLO HÁPLICO Tb Distrófico úmbrico ou típico + Inclusão de LATOSSOLO VERMELHO Distrófico típico
LAd8	LATOSSOLO AMARELO Distrófico húmico + LATOSSOLO AMARELO Ácrico húmico + LATOSSOLO VERMELHO-AMARELO Distrófico típico + CAMBISSOLO HÁPLICO Tb Distrófico típico + Inclusão Afloramento de Rocha
LVAd1	LATOSSOLO VERMELHO-AMARELO Distrófico típico + ARGISSOLO VERMELHO-AMARELO Distrófico Latossólico + CAMBISSOLO HÁPLICO Tb Distrófico típico
LVAd2	LATOSSOLO VERMELHO - AMARELO Distrófico típico + LATOSSOLO AMARELO Distrófico húmico + LATOSSOLO VERMELHO Distrófico típico
LVAd3	LATOSSOLO VERMELHO-AMARELO Distrófico típico + LATOSSOLO VERMELHO Distrófico típico+ CAMBISSOLO HÁPLICO Tb Distrófico típico + inclusão ARGISSOLO VERMELHO-AMARELO Distrófico Latossólico + Afloramento de Rocha. 20%
LVAd4	LATOSSOLO VERMELHO-AMARELO Distrófico típico + Afloramento de Rocha + CAMBISSOLO HÁPLICO Tb Distrófico típico+ Inclusão LATOSSOLO Vermelho Distrófico típico + ARGISSOLO VERMELHO-AMARELO Distrófico típico
LVAd5	LATOSSOLO VERMELHO-AMARELO Distrófico típico + ARGISSOLO VERMELHO-AMARELO Distrófico Latossólico + CAMBISSOLO HÁPLICO Tb Distrófico típico + Afloramento de Rocha
LVAd6	LATOSSOLO VERMELHO - AMARELO Distrófico típico + CAMBISSOLO HÁPLICO Tb Distrófico + LATOSSOLO AMARELO Distrófico típico e húmico + Inclusão de ARGISSOLO VERMELHO-AMARELO Distrófico típico
LVd1	LATOSSOLO VERMELHO Distrófico típico + ARGISSOLO VERMELHO-AMARELO Distrófico típico ou abruptico + LATOSSOLO VERMELHO-AMARELO Distrófico típico
LVd2	LATOSSOLO VERMELHO Distrófico típico + LATOSSOLO VERMELHOAMARELO, Distrófico típico + CAMBISSOLO HÁPLICO Tb Distrófico típico + Inclusão de LATOSSOLO AMARELO Distrófico húmico + ARGISSOLO VERMELHO-AMARELO Distrófico Latossólico ou abruptico
PVAd1	ARGISSOLO VERMELHO-AMARELO Distrófico típico + ARGISSOLO AMARELO Distrófico úmbrico ou latossólico + CAMBISSOLO HÁPLICO Tb Distrófico típico + Afloramento de Rocha (30, 20, 20 e 20%) + inclusão de LATOSSOLO AMARELO e LATOSSOLO VERMELHO, ambos Distróficos típicos
PVAd2	ARGISSOLO VERMELHO-AMARELO Distrófico típico ou latossólico + ARGISSOLO VERMELHO Distrófico típico + LATOSSOLO AMARELO Distrófico húmico
PVd1	ARGISSOLO VERMELHO Distrófico típico + LATOSSOLO AMARELO Distrófico típico + CAMBISSOLO HÁPLICO Tb Distrófico típico + Inclusão de Afloramento de Rocha. (10%).
PVd2	ARGISSOLO VERMELHO Distrófico típico ou latossólico + ARGISSOLO VERMELHO AMARELO Distrófico típico ou latossólico
PVd3	ARGISSOLO VERMELHO Distrófico típico + ARGISSOLO AMARELO Distrófico Úmbrico + CAMBISSOLO HÁPLICO Tb Distrófico típico
PVe	ARGISSOLO VERMELHO Eutrófico típico + ARGISSOLO VERMELHO-AMARELO Distrófico típico + CAMBISSOLO HÁPLICO Tb Distrófico típico + Inclusão ARGISSOLO AMARELO Distrófico típico + LATOSSOLO AMARELO Distrófico típico

ARGISSOLO AMARELO (PA)

Na área de estudo essa classe ocorre de forma dispersa: como segundo componente das unidades de mapeamento PVAd1 e PVd3 como Argissolos Amarelos Distróficos úmbricos ou latossólicos; e como inclusão na unidade de mapeamento PVe como Argissolo Amarelo Distrófico típico. De maneira geral, são solos de perfis profundos, com argila de atividade baixa, cerosidade variando de moderada e comum a forte e abundante, com baixos teores de ferro e drenagem variando de bem a moderadamente drenado. De acordo com a posição fisiográfica, ocorrem em relevos suaves e em topos amplos e ondulados (Calderano Filho et al., 2012).

ARGISSOLO VERMELHO (PV)

A classe PV ocorre na área como primeiro componente em quatro unidades de mapeamento (PVd1, PVd2, PVd3 e PVe), sendo predominante os Argissolos Vermelhos Distróficos típicos. São caracterizados como solos minerais, normalmente profundos, não hidromórficos, com argila de atividade baixa, baixos teores de ferro, saturação por bases de baixa a alta, bem drenados e com cerosidade variando de moderada e comum a forte e abundante. Possuem textura argilosa a muito argilosa com predomínio de caulinita na fração argila. Situam-se em áreas de relevo ondulado a forte ondulado com cobertura vegetal de floresta tropical subperenifólia ou em áreas de plantio de café, pastagem natural e pastagem plantada (Calderano Filho et al., 2012).

ARGISSOLO VERMELHO-AMARELO (PVA)

A classe PVA ocorre na área como primeiro componente nas unidades PVAd1 e PVAd2, predominando os Argissolos Vermelho-Amarelos Distróficos típicos. De maneira geral, são solos de fertilidade natural baixa com argila de atividade baixa, bem drenados, baixos teores de ferro, cerosidade variando de moderada e comum a forte e abundante e textura muito argilosa em profundidade. Ocorrem em relevo ondulado a montanhoso com declives de 15% a 65%, em áreas de floresta tropical subperenifólia, pastagem natural e pastagem plantada (Calderano Filho et al., 2012).

LATOSSOLO AMARELO (LA)

Entre as unidades definidas como LA como primeiro componente, seis compreendem solos com horizonte A moderado ou proeminente, de textura argilosa, distróficos típicos ou úmbricos, situadas em relevo suave, ondulado e forte ondulado (LAd1, LAd2, LAd3, LAd4, LAd5 e LAd6).

E duas compreendem solos com horizonte A húmico de textura argilosa, distróficos húmicos, situados nas posições de relevo forte ondulado e montanhoso (LAd7 e LAd8). São solos muito profundos e de baixa fertilidade natural, bem drenados, de alta permeabilidade e muito intemperizados, definidos como distróficos. Apesar da baixa fertilidade, são utilizados no plantio de café, olerícolas, pastagens naturais e plantadas, ocorrendo também em áreas expressivas de floresta (Calderano Filho et al., 2012).

LATOSSOLO VERMELHO (LV)

Duas unidades de mapeamento foram definidas tendo a classe LV como primeiro componente (LVd1 e LVd2). Esses solos são definidos como Latossolos Vermelhos Distróficos típicos e apresentam horizonte A moderado com textura argilosa, situados em áreas de relevo forte ondulado. De maneira geral, são solos profundos, bem drenados, de alta permeabilidade, muito porosos e muito intemperizados. Seu uso na área está relacionado a pequenos plantios de café, plantios de eucalipto e pastagem (Calderano Filho et al., 2012).

LATOSSOLO VERMELHO-AMARELO (LVA)

Os Latossolos Vermelho-Amarelos são definidos como primeiros componentes nas unidades LVAd1, LVAd2, LVAd3, LVAd4, LVAd5 e LVAd6. Compreendem solos de baixa fertilidade natural, muito profundos, bem drenados, muito intemperizados, argilosos e com alta permeabilidade. De acordo com a posição fisiográfica, ocorrem nas cotas mais elevadas do relevo, onde o processo erosivo é favorecido, sob cobertura vegetal remanescente do tipo floresta tropical subperenifólia, pastagens e áreas expressivas de floresta (Calderano Filho et al., 2012).

CAMBISSOLO HÁPLICO (CX)

A classe CX ocorre como primeiro componente na unidade de mapeamento CXbd1 como CAMBISSOLO HÁPLICO Tb Distrófico típico e na unidade CXbd2 como CAMBISSOLO HÁPLICO Tb Distrófico úmbrico ou típico. Na área apresentam argila de atividade baixa, baixa saturação por bases, boa drenagem e variações em relação a espessura dos perfis. Podendo ser profundos ou de espessura menor que 80 cm onde ocorre contato com material saprolítico. A posição fisiográfica também é diversificada, ocorrendo em morros, morrotes e montanhas, de relevo ondulado até montanhoso e declives de 15% a 75%. Possui forte susceptibilidade à erosão

em alguns casos e, quando o relevo é montanhoso, apresentam rochiosidade e estão associados a solos rasos (Calderano Filho et al., 2012).

CAMBISSOLO FLÚVICO (CY)

Constitui uma classe dispersa na área, definido como terceiro componente da unidade de mapeamento LAd4. São solos o caráter flúvico dentro de 120 cm a partir da superfície de solo e que ocupam posições de baixada ou ambientes de várzea, em relevo suave ondulado e associados aos Neossolos Flúvicos Calderano Filho et al., 2012).

NEOSSOLO LITÓLICO (RL)

Essa classe ocorre na área associada a afloramento de rocha e Cambissolos Háplicos Tb Distróficos. Por definição, são solos rasos assentados diretamente sobre rocha ou sobre horizontes C ou Cr, ou sobre material com 90% (por volume) ou mais de sua massa constituída por fragmentos de rocha (cascalhos, calhaus e matacões) com contato lítico dentro de 50 cm da superfície do solo. Ocorrem em relevo movimentado e são muito susceptíveis à erosão, apresentando limitações para atividade agrícola Calderano Filho et al., 2012).

NEOSSOLO FLÚVICO (RY)

Essa classe está associada aos Cambissolos Flúvicos, sob áreas de vegetação remanescente do tipo floresta tropical subperenifólia de várzea e em áreas de relevo plano a suave ondulado com declives entre 0 a 5%, em estreitas planícies e terraços fluviais. São solos derivados de sedimentos aluviais e que apresentam o caráter flúvico, moderadamente a imperfeitamente drenados e com evidências de gleização em subsuperfície. No geral, são solos profundos e com potencial agrícola, dependendo do nível do lençol freático Calderano Filho et al., 2012).

4. Artigo: Could Airborne Geophysical Data Be Used to Improve Predictive Modeling of Agronomic Soil Properties in Tropical Hillslope Area?

Este tópico se refere ao artigo de título “*Could Airborne Geophysical Data Be Used to Improve Predictive Modeling of Agronomic Soil Properties in Tropical Hillslope Area*” (Bastos et al., 2023), fruto desta dissertação de mestrado e publicado na revista *Remote Sensing* com coautoria de Helena S. K. Pinheiro, Francisco J. F. Ferreira, Waldir de Carvalho Junior e Lúcia Helena C. dos Anjos.

4.1 Introduction

According to Lagacherie and Mcbratney (2006), the purpose of digital soil mapping (DSM) is to extend spatial soil information system functionalities (conventional soil survey expertise and remote/proximal sensing) by combining with the spatial soil inference systems (predictive models) to increase the understanding of the spatial distribution of soil properties. In other words, the objective is to infer new soil data from the one already available for predicting soil classes and attributes, studying their properties and related environmental variables, and producing more informative and cost-efficient maps with higher spatial resolution, accuracy, and uncertainty estimates (Lagacherie, 2008; Lagacherie e Mcbratney, 2006; Mcbratney et al., 2003).

The DSM depends on adequate environmental variables to represent soil formation factors (Jenny, 1994) as predictors to achieve this purpose. According to Chen et al. (2022), relief, organisms, and climate were the three most frequently used environmental covariates, and the frequency of other factors, such as parental material, are poorly related in the literature. Knowing that most of the world’s soils are “mineral soils” and their properties are influenced by their mineral composition, it is our interest to update the parent material information in soil science (Wilson, 2019). Parent material can be an essential soil-forming factor where granitic and basaltic inherited soils occur, for example. As suggested Wilson (2019), highlighting the differences between those soils in some important soil properties (e.g., sand contents, clay contents, water retention, porosity, erodibility, and clay minerals of young soils).

In addressing this issue, gamma-ray spectrometry data have increased as a covariate in DSM to represent parental material information (e.g.: Mello et al. (2022); Loiseau et al. (2020 and 2021); Maino et al. (2022); Mello et al. (2021)). It is possible to make associations between relief denudation processes and the relative rates of soil formation and erosion (Minty, 1997; Wilford and Minty, 2006), as Reinhardt and Herrmann (2019) also pointed out in a critical review of gamma-ray spectrometry as a tool in soil science. Among the geophysical data, gamma-ray spectrometry is a remote sensing technique used to measure the natural radiation emitted by rocks and soil. This method measures the concentrations of K (potassium), U (uranium), and Th (thorium) from their radioactive decay series (^{40}K , ^{238}U , and ^{232}Th) generally through NaI(Tl) detectors in the portable spectrometer or in airborne surveys (Minty, 1997). One important characteristic is a general increase in its concentration with increasing silica content in igneous rocks, mainly correlated with acid rocks such as granites and gneisses (Wilford, 2012). Another application in soil science is mapping the distribution of these elements in soil profiles (Mello et al., 2021; Dickson e Scott, 1997; Reinhardt and Herrmann, 2019). Different soil horizons may have different concentrations of radioactive elements, which can be mapped to understand the vertical distribution of these elements and how they correlate with soil properties and soil genesis.

The magnetic method is another example of geophysical data that can be used to represent parent material. This technique measures the strength and direction of the magnetic field at the Earth's surface. Magnetic anomalies (i.e., deviations from the expected magnetic field) can be mapped and provide information about the magnetic properties of the rocks and geological structures (Dentith and Mudge, 2014). Magnetic minerals and their characteristics, such as grain size, shape, and orientation, control the magnetic survey responses. Magnetic data are usually related to more basic rocks such as basalts. However, some generalizations may be made within the same area since rocks from the same site can exhibit increased magnetic susceptibility with maficity (Dentith and Mudge, 2014).

As well as gamma-ray spectrometry, the magnetic method can be measured in situ or based on aerial surveys. Studies involving magnetic data and soils are usually in situ (Mello et al., 2020; Mello et al., 2022; Sarmast et al., 2017; Valaee et al., 2016). However, its applicability may be valid due to the possibility of contrasting with gamma-ray data in the type of parent material highlighted by these methods in DSM. Furthermore, airborne gamma-ray and magnetic data are

freely available by the Geological Survey of Brazil—CPRM (CPRM, 2012), encouraging researchers and scientists to promote advances in several areas, boosting sectors such as data science, artificial intelligence, and geospatial technology. The mafic index is an example of a covariable with the potential for digital soil mapping that integrates both data types. According to (Iza et al., 2018), that is a helpful technique that can be used to analyze the influence of lateritic soils on the surface (Fe-rich material).

Therefore, the hypothesis was that AGD could enhance the predictive modeling of aluminum saturation (ASat), base saturation (BS), cation exchange capacity (CEC), clay, and organic carbon (OC) in Bom Jardim County, Rio de Janeiro State. To evaluate this hypothesis, the research aimed to employ predictive modeling techniques and compare the effectiveness of two different groups of covariates representing the soil-forming factors: the first group consisted of covariates derived from DEM and Sentinel-2 images with AGD; the second group included covariates derived from DEM and Sentinel-2 images without AGD.

The AGD candidate predictor variables include total count ($\mu\text{R/h}$), K (potassium), eU (uranium equivalent), and eTh (thorium equivalent), ratios between these elements (eTh/K, eU/K, and eU/eTh), factor F or F-parameter, anomalous potassium (Kd), anomalous uranium (Ud), anomalous magnetic field (AMF), vertical derivative (GZ), horizontal derivatives (GX and GY), and mafic index (MI). The significance of these covariates was assessed through the evaluation of two predictive models (RF and SVM), model performance criteria (R^2 , RMSE, and MAE), null model and coefficient of variation (CV%) values, covariate frequency, and Spearman's correlations.

4.2 Materials and Methods

4.2.1 Study Area and Soil Data

Bom Jardim County is in the Centro Fluminense mesoregion of the State of Rio de Janeiro, as indicated by the Brazilian Territorial Division from (IBGE, 2021a) (Figure 14). Situated between the valleys and escarpments of Serra do Mar, the county comprises an area of 382,430 km². The climate is subtropical, categorized by Cwa per Köppen's criteria. The average rainfall from 1941

to 2020 is 1413.63 mm, with December being the wettest month and September being the driest (ANA, 2020).

Despite anthropic interventions, mainly through agricultural activities, the original vegetation of the remaining Atlantic Forest is 51%, constituting secondary forests and native forests preserved in areas of higher altitude. The remaining 49% of the area predominantly comprises natural pastures, perennial crops such as coffee, vegetable crops, and ornamental flowers such as roses and eucalyptus reforestation (Calderano Filho et al., 2012). The anthropic interferences associated with the dominant mountainous relief characterize a high potential vulnerability to the region's erosion and mass movement events. This potential is reduced when the terrain is still covered with the original tropical rainforest (Calderano Filho, 2012).

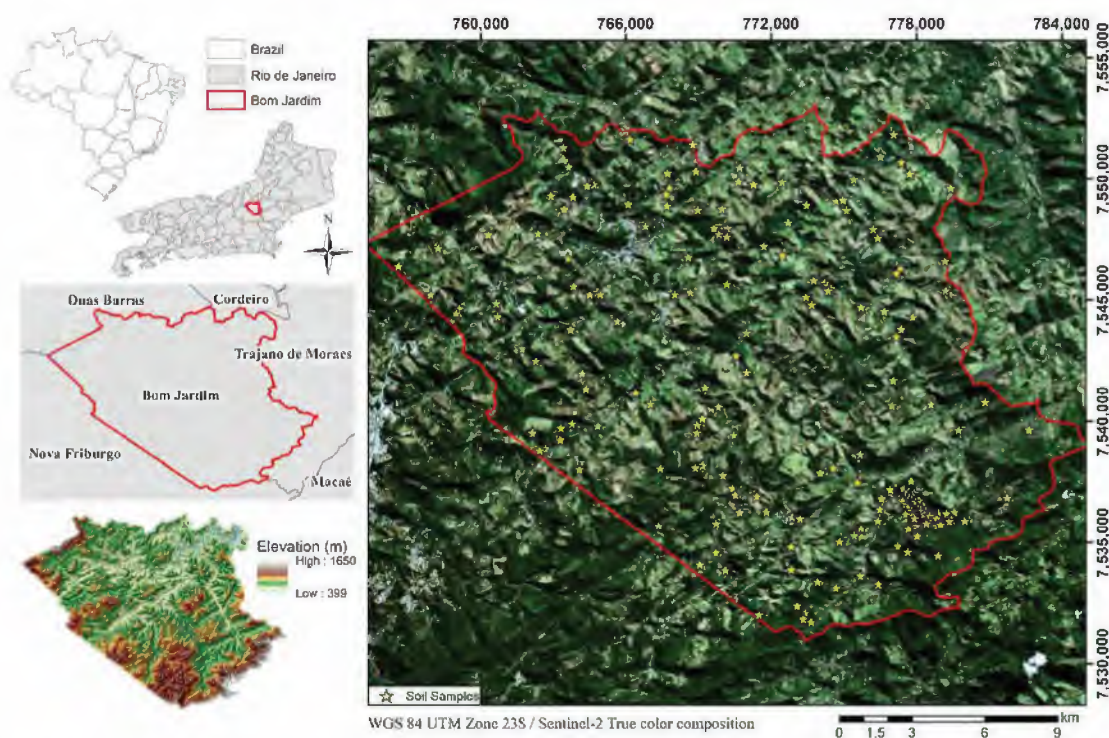


Figure 14: Location map of the Bom Jardim County and soil samples described by Calderano Filho (2012). On the left below, the digital elevation model (DEM) derived from Rio de Janeiro cartographic database (original scale 1:25,000) (IBGE, 2018).

According to the World Reference Base for Soil Resources Classification System (FAO, 2014), the region has as principal soil classes Cambisols, Ferralsols, Acrisols, and Fluvisols (Figure 15a) and soil associations detailed by (Calderano Filho et al., 2012). Regarding geology, Bom

Jardim is inserted in the Oriental Terrane of the Ribeira Belt. These include plutonic rock intrusion and deformation during pre- and late-collisional periods, paragneiss and high-grade metamorphic metasediments, and non-deformed granitic bodies (CPRM, 2016b; Tupinamba et al., 2013) (Figure 15b).

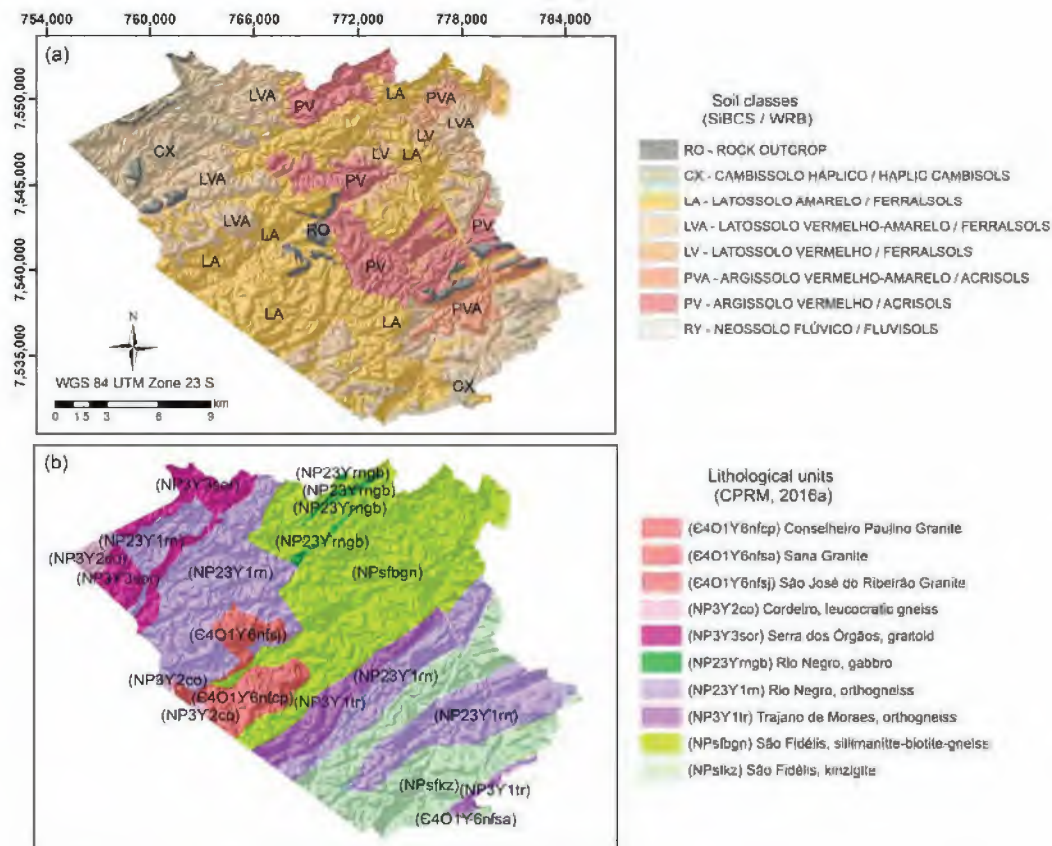


Figure 15: (a) Soil map adapted from (Calderano Filho et al., 2012), SiBCS: Brazilian Soil Classification System (Santos et al., 2018), WRB: World Reference Base for Soil Resources (FAO, 2014); (b) geological map (original scale 1:400,000) (CPRM, 2016a).

The legacy soil data were from (Calderano Filho et al., 2012)’s investigation “Soils of the Medium Upper Course of the Rio Grande, Mountainous Region of the State of Rio de Janeiro”, collected between 2009 and 2011. The study aimed to classify and map different soil types on a 1:100,000 scale using the Brazilian Soil Classification System criteria and norms adopted by Embrapa (EMBRAPA, 2006). The soil dataset consists of 208 samples, including 74 soil profiles, 44 complementary soil profiles, and 90 surface-horizon samples. The samples are divided into 97 Ferralsols, 35 Cambisols, 62 Acrisols, three Leptosols, five Fluvisols, five Gleysols, and one

Nitisol, detailed by (Calderano Filho et al., 2012). The analytical results of aluminum saturation (ASat), base saturation (BS), cation exchange capacity (CEC), clay, and organic carbon (OC) were selected for the present study.

Given the wide variety of soil profile thicknesses, the slice-wise aggregation algorithm of the AQP package developed by Beaudette et al. (2013) was applied. This function aims to interpolate values for soil properties at predefined depth intervals. The slice-wise function assumes that soil properties along the profile are continuous, respecting their average values. In the present work, the topsoil layer corresponding to 0–30 cm was used to compound the input soil dataset. The procedure was done in R and RStudio software version 4.3.0 (R Core Team, 2023). The basic statistics for these attributes are presented in Table 5.

Table 5: Statistics of soil attributes used in the prediction.

	ASat (%)	BS (%)	CEC	Clay (g/kg)	OC (g/kg)
Min	0	2	3.39	102	4.8
Max	89	100	35.07	556	76.2
Median	20.25	26	9.06	344.5	15.6
Mean	30.20	30.25	9.70	336.80	16.43
Standard deviation	28.37	20.91	3.70	96.31	7.26

Aluminum saturation = ASat, base saturation = BS, cation exchange capacity = CEC, and organic carbon = OC.

4.2.2 Covariate Acquisition and Processing

The DEM was generated using vector data containing primary elevation information, equidistant contour lines at 10 m intervals, hydrography, and boundary of the study area, all projected in UTM/WGS84 datum (Universal Transverse Mercator/World Geodetic System 1984), Zone 23 S, EPSG 32723. These data comprise the cartographic base of the State of Rio de Janeiro, available by the Brazilian Institute of Geography and Statistics, with an original scale of 1:25,000 (IBGE, 2018). The vector data were interpolated into a regular 20 m grid by applying the Topo To Raster tool in ArcGIS Desktop software (v. 10.6). Interpolation errors, such as spurious depressions, were filled. Additionally, the Morphometry and Hydrology modules of the SAGA-GIS software (Conrad et al., 2015) were used to generate the covariates derived from DEM. DEM covariates used for pre-diction and their respective references are represented in Table 6.

Table 6: DEM covariates used for prediction and their respective references.

Covariate	Abbreviation	Reference
Digital Elevation Model	DEM	-
Convergence Index	Convergen	Köthe and Lehmeier (1996)
Downslope Distance Gradient	Gradient	Hjerdt et al. (2004)
Slope	-	
Aspect	-	
Profile Curvature	ProfileCurv	Wood (2009)
Plan Curvature	PlanCurv	
Longitudinal Curvature	Longitudin	
Maximum Curvature	MaximumCurv	
Topographic Position Index	TPI	Guisan et al. (1999); Weiss (2001); Wilson and Gallant (2000)
Slope Height	SlopeHeig	Böhner and Selige (2006)
Valley Depth	ValleyDep	
Normalized Height	Normalized	
Standardized Height	Standardiz	
Mid-Slope Position	MidSlope	
Terrain Ruggedness Index	TRI	Riley et al. (1999)
Topographic Wetness Index	TWI	
Catchment Area	CatchmentA	Böhner and Selige (2006)
Catchment Slope	CatchmentS	

The Sentinel-2 mission's Multispectral Instrument (MSI) was carried out by two satellites, Sentinel 2A, and 2B, launched in June 2015 and March 2017. The European Space Agency (ESA) provided these sensor images, including 13 spectral bands ranging from visible and near-infrared (VNIR) to shortwave infrared (SWIR), to users in Level 1C of processing, which consists of TOA (top-of-atmosphere) images. The images are geometrically corrected and orthorectified, with 100 km × 100 km dimensions, projected on UTM/WGS84 (Drusch et al., 2012). The ESA also provides additional data for the processing of images from Level 1C to Level 2A, which refers to the application of atmospheric correction, transforming TOA images into BOA (bottom-of-atmosphere). The procedure was performed using the Sen2Cor processor. More details about processing can be seen in the user manual written by (Mueller-Wilm et al., 2017).

Cloud-free images were selected at processing Level 1C and collected on 7 December 2021 for this study. The atmospheric correction method described in the previous paragraph was applied, resulting in 20 m resolution BOA images. Calculations to generate covariates were performed using mathematical operations between bands in the QGIS software v. 3.24.1 (QGIS Development Team, 2023).

The Sentinel-2 covariates used for prediction and their respective references are represented in Table 7.

Table 7: Sentinel-2 covariates used for prediction and their respective references.

Covariate	Abbreviation	Reference
Grain Size Index	GSI	Perera et al. (2005); Xiao et al. (2006)
Normalized Difference Vegetation Index	NDVI	Rouse et al. (1973)
Alteration Ferric Iron Ferric Oxides Ferrous silicates (Biotite, chloride, amphibole)	- FerricIron FerricOxi FerrousSilic	Van Der Meer et al. (2014)

The airborne geophysical data were obtained by the Geological Survey of Brazil (Rio de Janeiro Project (CPRM, 2012)). This project was carried out during 2011 and 2012 and covered an area of 32,202 km², totaling 66,111.40 km of high-resolution geophysical profiles. The summary of survey characteristics is described in Table 8.

Table 8: Survey characteristics of Rio de Janeiro Aerogeophysical Project (CPRM, 2012).

Flight line direction	N-S
Flight line spacing	500 m
Control line direction	E-W
Control line spacing	10 km
Measurement intervals (gamma-ray spectrometer)	1.0 s
Measurement intervals (magnetometer)	0.1 s
Flight height average	100 m
Approximate flight speed	270 km/h

All the AGDs (Table 9) were processed using Oasis Montaj software (Educational software edition v. 9.8) at a spatial resolution of 100 m. First, minimum curvature (Briggs, 1974) interpolation was performed to generate primary gamma-ray variables: total count ($\mu\text{R/h}$), K (potassium), eU (uranium equivalent), and eTh (thorium equivalent). The ratios between elements, Factor F or F-parameter, anomalous potassium (Kd), and anomalous uranium (Ud) were generated from those variables. The anomalous magnetic field (AMF), which represents the magnetic

susceptibility of the rocks in the area, was generated by performing magnetic data interpolation using the bidirectional method (GEOSOFT, 2023). The data were then reduced to the pole (Baranov, 1957) according to the parameters calculated for the date of acquisition. From AMF, its derivatives were generated: analytic signal amplitude (ASA), vertical derivative (GZ), and horizontal derivatives (GX and GY). The ASA (Li, 2006) was used with K, eU, and eTh channels to generate the mafic index, the calculated consistency in the division between the ASA and the product between the K, eU, and eTh channels (Iza et al., 2018).

After processing, all geophysical rasters were resampled from 100 m to 20 m of spatial resolution in the Oasis Montaj software, using the Regrid tool to adapt them to the finer resolution of the DEM and Sentinel-2 covariates since morphologically complex areas are affected by the pixel size, preferring finer resolutions as observed by (Cavazzi et al., 2013).

Table 9: AGD covariates used for prediction and their respective references.

Covariate	Abbreviation	Reference
Total count	TC	
Potassium, uranium equivalent, and thorium equivalent	Kperc, eU, and eTh	CPRM (2012); IAEA (2003)
Ratios between elements	eTh/K, eU/K, and eU/eTh	
Factor F or F-parameter	FactorF	Gnojek and Prichystal (1985)
Anomalous potassium and anomalous uranium	Kd and Ud	Saunders et al. (1993)
Anomalous magnetic field	AMF	CPRM (2012)
Vertical derivative and horizontal derivatives	GZ, GX and GY	Dentith and Mudge (2014)
Mafic index	MI	Iza et al. (2018)

4.2.3 Modeling Strategy

The following steps were applied to predict the selected soil attributes using variables derived from DEM and Sentinel-2 imagens with AGD and without AGD (Figure 16):

- (1) Dataset preparation;

- (2) Removal of non-significant covariates with nearZeroVar and findCorrelation functions;
- (3) Data splitting in training (75%) and validation (25%) datasets;
- (4) Removal of covariates by importance using recursive feature elimination (RFE);
- (5) Training of predictive models using the selected covariates;
- (6) Model performance evaluation using validation samples;
- (7) Computation of frequency of top 10 RFE predictors;
- (8) Generation of average maps for each soil property;
- (9) Generation of coefficient of variation (CV%) maps;
- (10) Interpretation of final results.

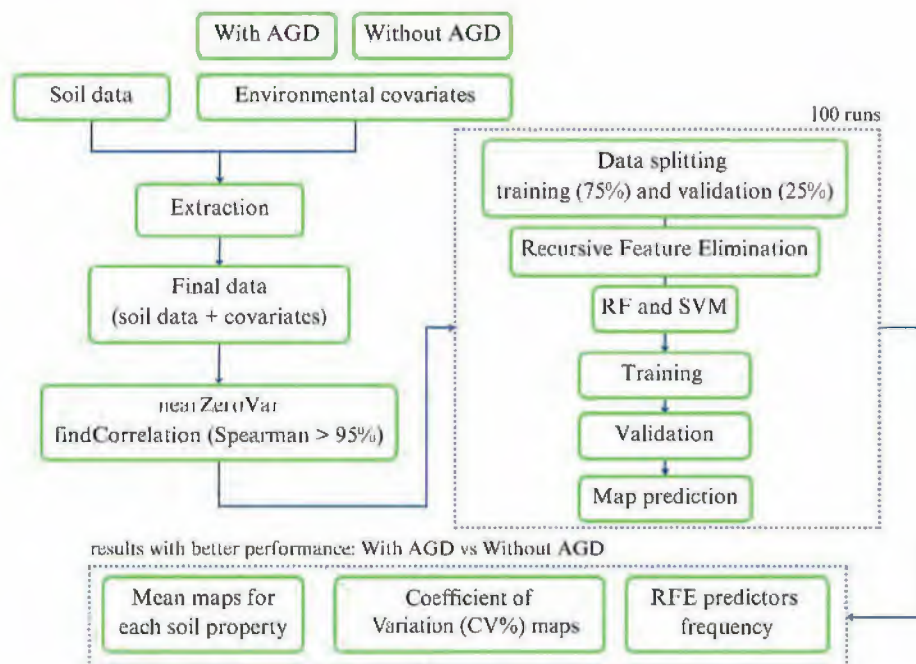


Figure 16: Modeling strategy flowchart.

According to Kuhn and Johnson (2013), there are benefits to eliminating predictors in the pre-processing step: optimize the computational time required to build the models, remove covariables with degenerate distributions, and remove one of highly correlated two predictors, which can measure the same underlying information. Eliminating non-significant covariates should

not harm the model's performance and may result in a more concise and understandable model, helping maximize accuracy.

The elimination of the non-significant covariate process was carried out through `nearZeroVar` and `findCorrelation` functions, both available in the `caret` package (Kuhn, 2022) for R software. `NearZeroVar` function was applied to eliminate variables with zero or nearly zero variance that will not contribute to the model's performance. The `findCorrelation` function was used to identify and remove the highly correlated covariates that could jeopardize the model's performance. The approach involves computing correlation coefficients for all possible pairs of covariates. When the correlation coefficient between two covariates exceeds a user-defined threshold, the correlation coefficients between each covariate and all other covariates are calculated and summed. The covariate with the highest sum of correlation coefficients with the other covariates is eliminated (Kuhn and Johnson, 2013). In this study, the `findCorrelation` function was applied with Spearman's correlation method with a critical value of 95% as adopted by Mello et al. (2022) and Siqueira et al. (2023).

Before performing recursive feature elimination (RFE), the samples were randomly divided into 75% for training and 25% for validation with the function `createDataPartition`. The RFE was performed in the training dataset through `rfe` and `rfeControl` functions, also from the `caret` package. This backward selection procedure evaluates multiple models by removing covariates in order of importance and is utilized in recent studies such as Mello et al. (2022), Siqueira et al. (2023) and Kaya et al. (2023). The initial model contains all predictors. At each step, the less significant predictors are iteratively eliminated before rebuilding the model. The final subset of covariates corresponds to the best value of the defined decision metric (Kuhn and Johnson, 2013). In this study, the RFE was performed through the ancillary functions `rfFuncs` for the RF model and `caretFuncs` for the SVM model, with 10-fold cross-validation using the `repeatedcv` method and evaluated by the R-squared accuracy metric. The predictor's subsets were used following each approach (with AGD and without AGD) and the result of the previous step. For modeling with AGD, we used 5 to 34 predictors subsets and for modeling without AGD, we used 5 to 22 predictors subsets. The ideal set of covariates generated by the RFE for each algorithm was used in the model training step.

The training process was performed through `train` and `trainControl` functions from the `caret` package. The procedure was performed with 10-fold cross-validation using the `repeatedcv` method and ten possible values of tuning hyperparameters evaluated by the R-squared accuracy metric (`mtry` for RF and `sigma` and `cost` for SVM (Siqueira et al., 2023)). The hyperparameters of each algorithm are described in the `caret` package manual, as cited by Mello et al. (2022). In the 10-fold cross-validation, also used in the RFE step, the training samples were partitioned into 10 near-equally sized folds. The models are trained by repeatedly excluding one of the folds, and the training performance is evaluated by making predictions on the excluded fold (Meyer et al., 2019; Siqueira et al., 2023).

The fitted models were applied to the validation data to assess their accuracy. Their performance was evaluated using three commonly used metrics in DSM: R-squared (R^2), root Mean square error (RMSE), and mean absolute error (MAE). Additionally, null model values (NULL RMSE and NULL MAE) were also calculated for comparison purposes. Utilizing null models can be a practical approach for setting thresholds and assessing the quality of models because it enables evaluating the tested models against a model with parameters set to zero (null) (Siqueira et al., 2023).

As shown in Figure 16, the RFE, training, validation, and prediction for the entire study area were repeated 100 times to ensure robustness. So, the final maps for each soil property were generated by computing the average value of the products of the 100 model runs, and the frequency of the top 10 RFE predictors. Additionally, the coefficient of variation ($CV\% = (\text{standard deviation}/\text{mean}) \times 100$) was calculated through the products of the 100 runs for each soil property following Gomes et al. (2019) and Siqueira et al. (2023)'s proposal.

A high CV % indicates a large amount of variability in the data, which can result in higher uncertainty in the map predictions. Conversely, a low CV % shows relatively little variability in the data, which can result in lower uncertainty in the map predictions. In addition, Spearman's correlation between AGD and soil properties was performed to assist in discussing their relationships. As the study focuses on the AGD's importance in predicting the select soil properties, the results and discussion will be directed toward them.

4.3 Results

4.3.1 Covariates Selection

In modeling with AGD data, none of the total 40 covariates were removed by the nearZeroVar function. The selection by Spearman's correlation considered six variables highly correlated with other covariates, which were removed from the input set of the modeling (eTh, eU/eTh, Kd, Longitudinal Curvature, Alteration, and Gz), resulting in a final set of 34 covariates. In modeling without AGD data, of the 25 covariates, none were removed by the nearZeroVar function. The selection by Spearman's correlation considered two variables highly correlated with other covariates, which were removed from the input set of the modeling (Longitudinal Curvature, Alteration), resulting in a final set of 22 covariates. The subsets of covariates in the RFE procedure varied in size according to the model and the soil property. The average number of predictors selected by the models for each property is presented in Figure 17.

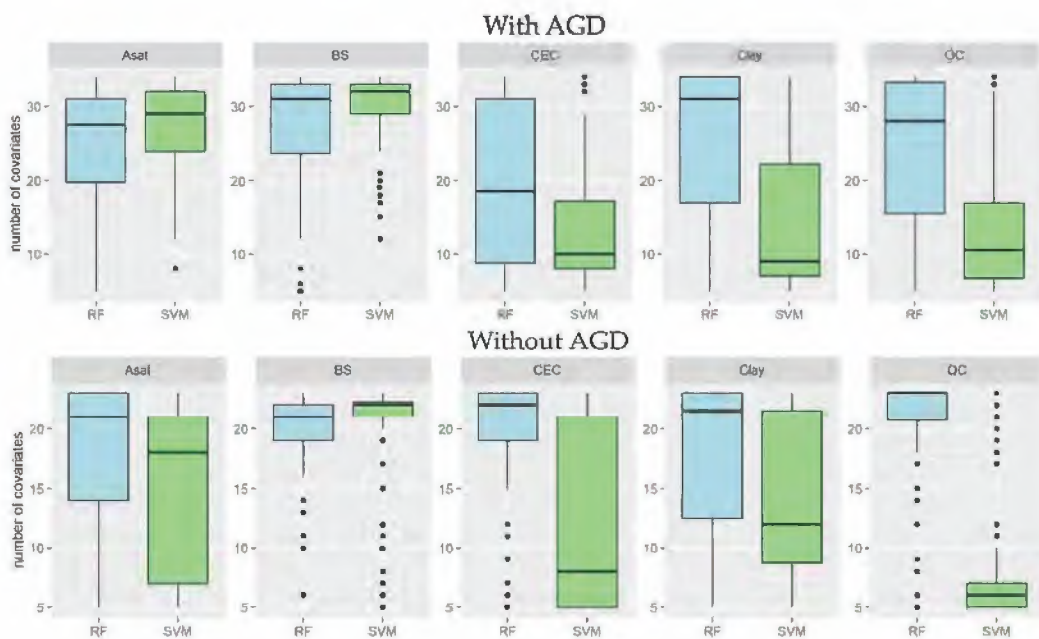


Figure 17: Boxplots with the number of predictors selected by the models for each soil property throughout the 100 runs.

In modeling with AGD, considering the 100 runs, the mean number of predictors selected for the RF model was 25 for ASat, 28 for BS, 20 for CEC, 26 for Clay, and 24 for OC. The mean number of predictors selected for the SVM model was 27 for ASat, 30 for BS, 14 for CEC, 14 for Clay, and 13 for OC. In modeling without AGD, the mean number of predictors selected for the RF model was 18 for ASat, 19 for BS, 19 for CEC, 18 for Clay, and 20 for OC. The mean number of predictors selected for the SVM model was 15 for ASat, 20 for BS, 13 for CEC, 14 for Clay, and 8 for OC, indicating a tendency to construct simpler models, mainly for the SVM model and CEC, Clay, and OC properties.

4.3.2 Models' Performance

The worst performances were seen in the modeling of soil attributes, excluding the AGD for both models (RF and SVM) (Table 10).

Table 10: Models' performance with and without AGD for all soil attributes.

		With AGD				
		ASat	BS	CEC	Clay	OC
R ²	RF	0.20	0.22	0.23	0.15	0.16
	SVM	0.19	0.23	0.11	0.11	0.08
RMSE	RF	25.66	18.47	3.20	91.08	6.66
	SVM	26.78	18.64	3.47	93.23	6.86
	NULL	28.32	20.64	3.57	97.11	6.97
MAE	RF	22.32	15.14	2.24	75.05	4.48
	SVM	21.97	14.79	2.20	77.57	4.39
	NULL	25.85	17.61	2.35	80.26	4.55
		Without AGD				
		ASat	BS	CEC	Clay	OC
R ²	RF	0.10	0.13	0.20	0.12	0.11
	SVM	0.06	0.10	0.07	0.10	0.04
RMSE	RF	27.19	19.54	3.27	92.72	6.80
	SVM	29.29	20.02	3.53	93.46	6.98
	NULL	28.32	20.64	3.57	97.11	6.96
MAE	RF	23.99	16.21	2.29	76.65	4.64
	SVM	24.04	16.28	2.23	78.33	4.53
	NULL	25.85	17.61	2.35	80.26	4.55

The RF algorithm showed the best performance to predict all soil attributes with AGD based on R², except for BS: ASat (R² = 0.20), CEC (R² = 0.23), Clay (R² = 0.15), and OC (R² = 0.16). Based on RMSE, the RF algorithm performed best for all soil attributes. On the other hand, based

on MAE, the SVM showed the best performance for ASat (MAE = 21.97), BS (MAE = 14.79), CEC (MAE = 2.20), and OC (MAE = 4.39).

The reliability of these observed results was confirmed through a comparison with null values, where all NULL RMSE and NULL MAE values were higher for all soil attributes in modeling with AGD. Conversely, in modeling without AGD, the SVM model showed higher RMSE values than NULL RMSE for ASat (RMSE = 29.29, NULL RMSE = 28.32) and for OC (RMSE = 6.98, NULL RMSE = 6.96). The RF model showed a higher MAE value than NULL MAE for OC (MAE = 4.64, NULL MAE = 4.55).

Figure 18 shows the performance results of the 100 model runs for each soil property in a boxplot based on R^2 .

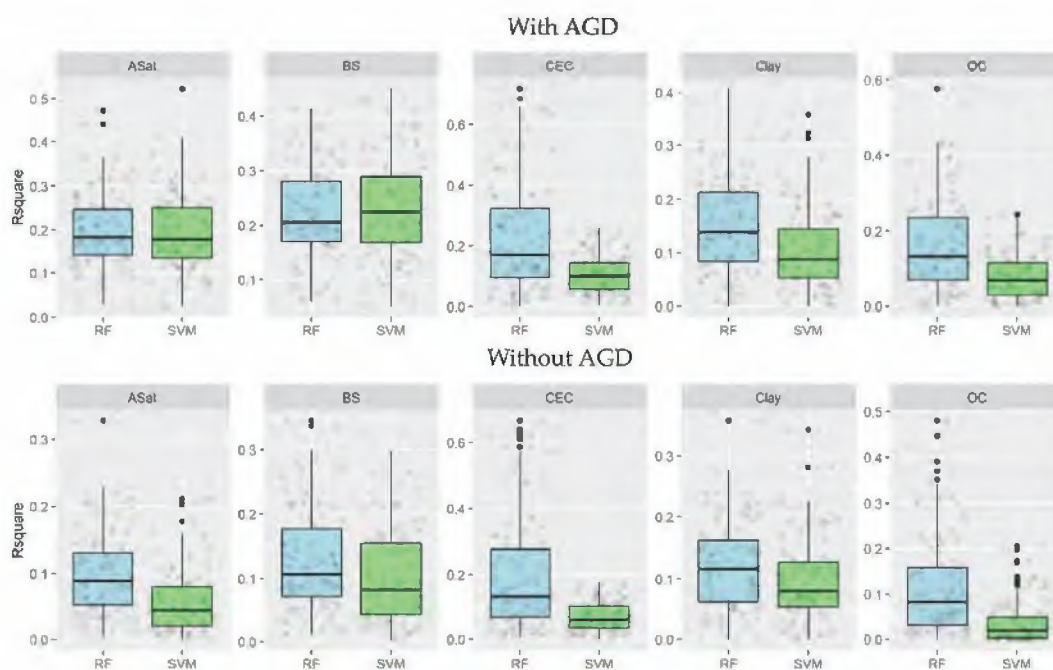


Figure 18: Performance of the models' RF and SVM on the prediction of the ASat, BS, CEC, Clay, and OC with AGD and without AGD assessed by the R-squared (R^2) metric.

Overall, the results exhibit substantial variation. The models with the poorest performance had the slightest deviations, as seen in the SVM model for CEC, Clay, and OC in modeling with AGD and SVM model for ASat, CEC, and OC in modeling without AGD. However, after thoroughly examining all 100 runs, it was found that using AGD for modeling was beneficial. The

advantage was evidenced by soil attribute predictions with AGD showing more R^2 values equal to or higher than 0.2, as shown in Table 11, which is considered satisfactory for machine learning algorithms regarding soil properties (Mello et al., 2022).

Table 11: Number of times the model reached values $R^2 \geq 0.20$.

With AGD					
	ASat	BS	CEC	Clay	OC
RF	44	58	44	32	33
SVM	43	63	14	12	3
Without AGD					
	ASat	BS	CEC	Clay	OC
RF	11	21	35	18	21
SVM	3	11	0	11	2

The most noteworthy differences between the models (RF and SVM) when utilizing AGD versus without AGD were observed in the ASat and BS attributes. The RF model achieved R^2 values of at least 0.20 in 44 instances for ASat and 58 instances for BS, compared to only 11 for ASat and 21 for BS instances without AGD. Similarly, the SVM model produced R^2 values of at least 0.20 in 43 instances for ASat and 63 instances for BS, in contrast to just 3 and 11 instances without AGD. For CEC, there was also an improvement in the results. The RF model with AGD achieved R^2 values of at least 0.20 in 44 instances versus 35 in modeling without AGD. For the SVM model, there were 14 instances with AGD against zero without AGD. While the models' performances for Clay and OC did not attain satisfactory values, using AGD slightly improved their overall performance.

4.3.3 Map Prediction and uncertainty

The mean values of CV% maps for modeling with and without AGD are represented in Table 12 below. The minimum, maximum, and median of CV% maps are attached in Appendix A, Table A1.

Table 12: The mean values of CV% maps for modeling with and without AGD.

With AGD					
	ASat	BS	CEC	Clay	OC
RF	13.86	9.29	8.30	4.56	9.29
SVM	14.35	8.31	4.73	6.22	5.43
Without AGD					
	ASat	BS	CEC	Clay	OC
RF	14.85	10.95	7.53	4.93	9.07
SVM	22.08	8.41	5.50	4.89	5.50

Comparing the RF values for modeling with and without AGD, the RF results with AGD reached a lower variability in the data for ASat, BS, and Clay. For SVM, the modeling with AGD reached a lower variability in the data for ASat, BS, CEC, and OC. Comparing the RF and the SVM models for modeling with AGD, the SVM showed better results for BS, CEC, and OC. The final prediction and CV% maps of the RF model are represented in Figures 19 and 20, and the final prediction and CV% maps of the SVM model are described in Figures 21 and 22 to analyze the results of modeling with AGD in more detail.

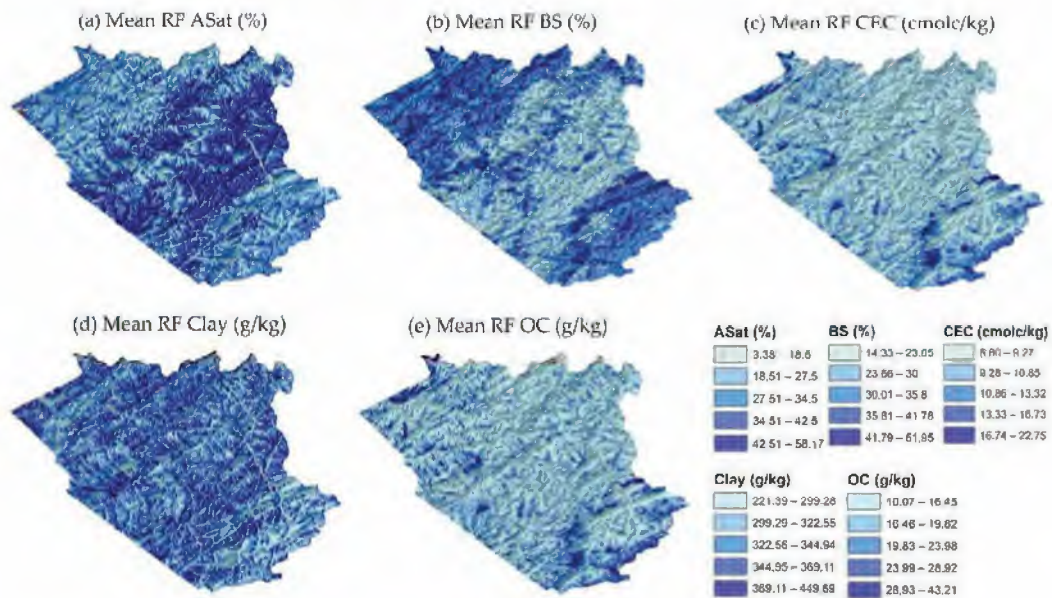


Figure 19: Final mean maps of the RF model with AGD: (a) ASat (%), (b) BS (%), (c) CEC (cmolc kg⁻¹), (d) Clay (g kg⁻¹), and (e) OC (g kg⁻¹).

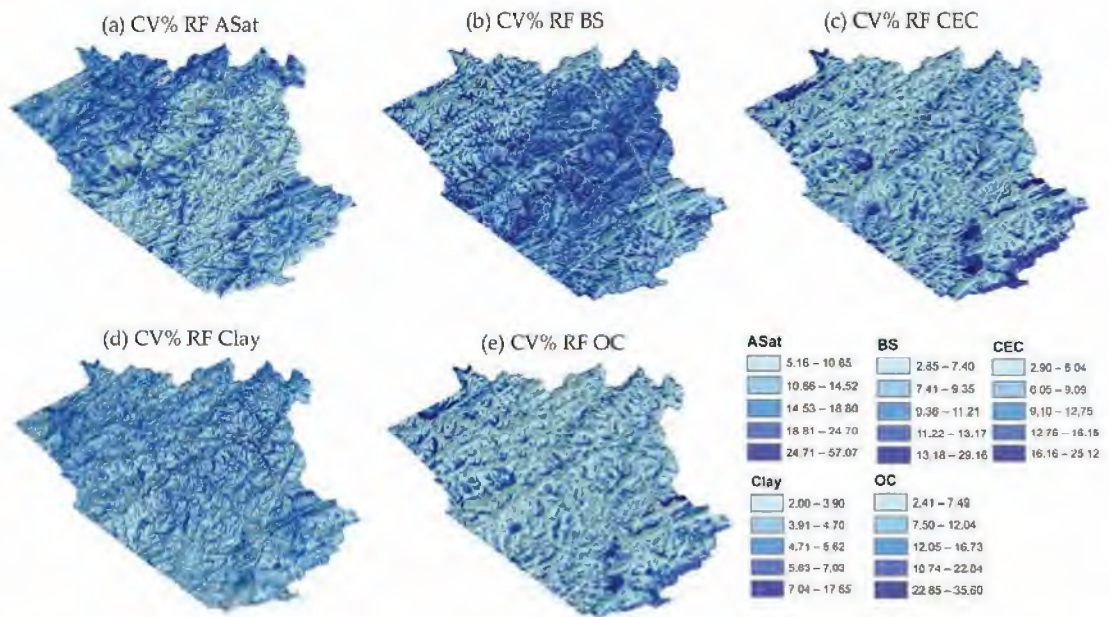


Figure 20: Coefficient of variation maps of the RF model with AGD: (a) ASat, (b) BS, (c) CEC, (d) Clay, and (e) OC.

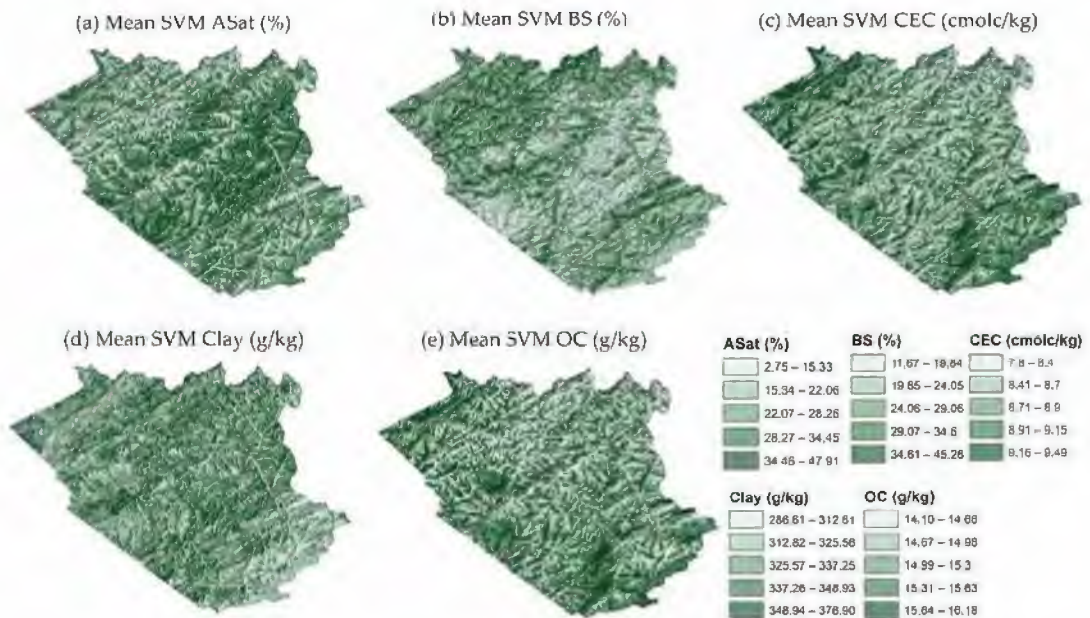


Figure 21: Final mean maps of the SVM model with AGD: (a) ASat (%), (b) BS (%), (c) CEC (cmolc kg⁻¹), (d) Clay (g kg⁻¹), and (e) OC (g kg⁻¹).

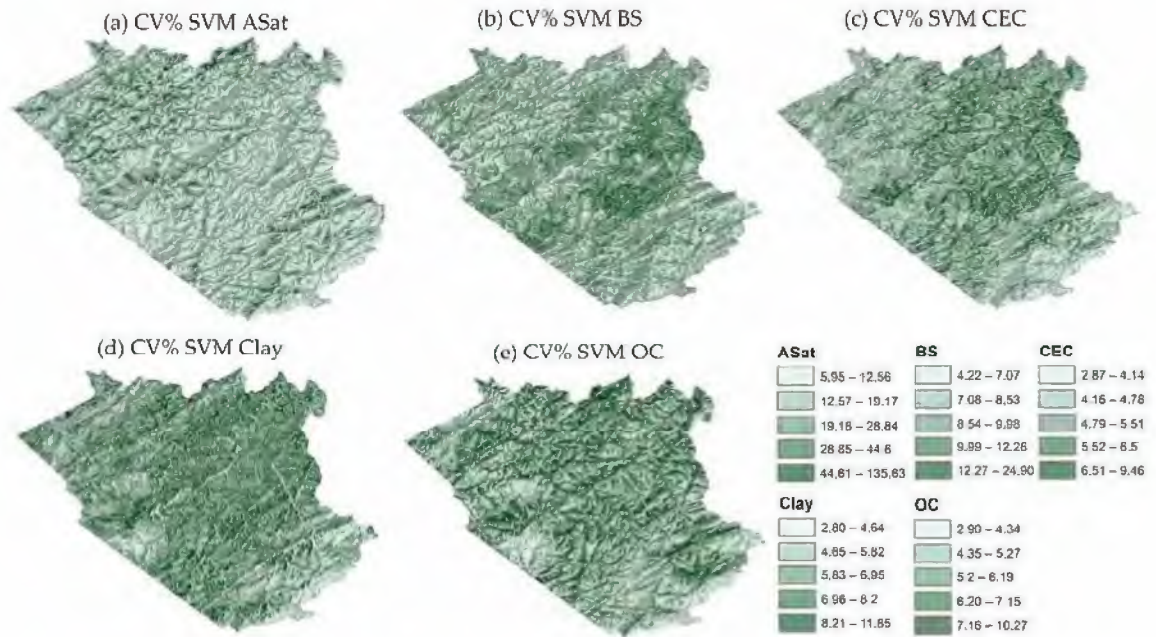


Figure 22: Coefficient of variation maps of the SVM model with AGD: (a) ASat, (b) BS, (c) CEC, (d) Clay, and (e) OC.

Figure 19a presents the ASat map, which clearly depicts a central region with high ASat values ranging from 42.51% to 58.17%. This region is dominated by Ferralsols and Acrisols, as shown in Figure 15a. It is characterized by the lithological units of the Rio Negro Complex, Trajano de Moraes, São Fidélis Group (sillimanite–biotite–gneiss), and Nova Friburgo Suite (Conselheiro Paulino and Sana granites), as indicated in Figure 15b. Interestingly, this area exhibits a lower CV% ranging from 5.16% to 10.65%, demonstrating greater reliability in the data in this region, as illustrated in Figure 20a.

In contrast, the BS map in Figure 19b displays the opposite trend, with the regions with the highest BS values located in the NW and SE regions ranging from 41.79% to 61.95%. These areas are the only ones with Cambisols as the dominant soil in this mapping unit (Figure 15a) and are also characterized by Ferralsols and Acrisols. Geologically, these high BS values are associated with the Rio Negro Complex, Serra dos Órgãos, Cordeiro, and São Fidélis (Kinzigite), as indicated in Figure 15b. This area also exhibits a lower CV% represented by classes ranging from 2.85% to 7.40% and 7.41% to 9.35%, as depicted in Figure 20b.

Figure 19c displays the CEC map, which seems to have its values influenced by the topography, with high CEC values (6.60–9.27 cmolc kg⁻¹) matching the outcrop area with higher altitudes (Figure 14). This region is dominated by Cambisols, Ferralsols, and rock outcrop, as shown in Figure 15a. It is characterized by the lithological units of the Nova Friburgo Suite, Rio Negro Complex, Serra dos Órgãos, and Kinzigite, as indicated in Figure 2b. In this case, the CV% is also higher, ranging from 16.16% to 25.12%, demonstrating a higher level of uncertainty in the predicted data in this region (Figure 20c).

For Clay (Figure 19d), it is more challenging to observe a pattern. However, relief also influences its distribution, where low values (221.39–299.28 and 299.29–322.55 g kg⁻¹) correspond to valleys. The medium and higher values (other classes ranging from 322.56 to 449.69 g kg⁻¹) are distributed in areas with higher altitudes (Figure 14). The CV% map for Clay concentrates its higher values mainly in the valleys, ranging from 7.04% to 17.65% (Figure 7d). The OC mean map (Figure 19e) has a distribution similar to that observed in the CEC mean map (Figure 6c). The higher values (28.93–43.21 g kg⁻¹) match with outcrop areas. The same occurs for the CV% map, showing higher values, ranging from 22.85% to 35.60%, as illustrated in Figure 7e.

The visual distribution for SVM (Figure 21) is generally similar for all mean maps compared to the RF model (Figure 19). However, the RF model seems to have the most defined distribution patterns, with higher contrast between the highest and lowest values, mainly for ASat, BS, and Clay.

Comparing the two models, a difference worth highlighting is the CV% values. Although the RF model performs better according to the presented metrics (Table 10), the SVM model's CV% values for all soil attributes showed a smaller amplitude, except for ASat. For SVM, the CV% ranges from 4.22% to 24.9% for BS, 2.87% to 9.46% for CEC, 2.80% to 11.85% for Clay, and 2.90% to 10.27% for OC (Figure 22b–e, respectively). For ASat, the RF model showed a smaller amplitude ranging from 5.16% to 57.07% (Figure 20a), while the SVM model showed an amplitude ranging from 5.95% to 135.63% (Figure 22a).

Another difference that can be highlighted is that the highest values in the average map (Figure 19c,e) correspond to the highest CV% values (Figure 20c,e) for the RF model for CEC and OC. In contrast, for the SVM model, the highest values in the average map (Figure 21c,e)

correspond to the lower CV% values (Figure 22c,e) for CEC and OC. For Clay, the highest values in the average map (Figure 19d) correspond to the RF model's lower CV% values (Figure 20d). In contrast, for the SVM model, the lower values in the average map (Figure 21d) correspond to the lower CV% values (Figure 22d) (valleys). The analysis suggests that while the RF model performs better, the SVM model may be more consistent in its predictions (lower uncertainty), as indicated by the smaller CV% values in these cases.

4.3.4 AGD Data Importance

Figure 23 shows the top 10 covariates' importance frequency to predict ASat, BS, and CEC with AGD. The top 10 covariates' importance frequency in predicting Clay and OC contents are represented in Figure 24.

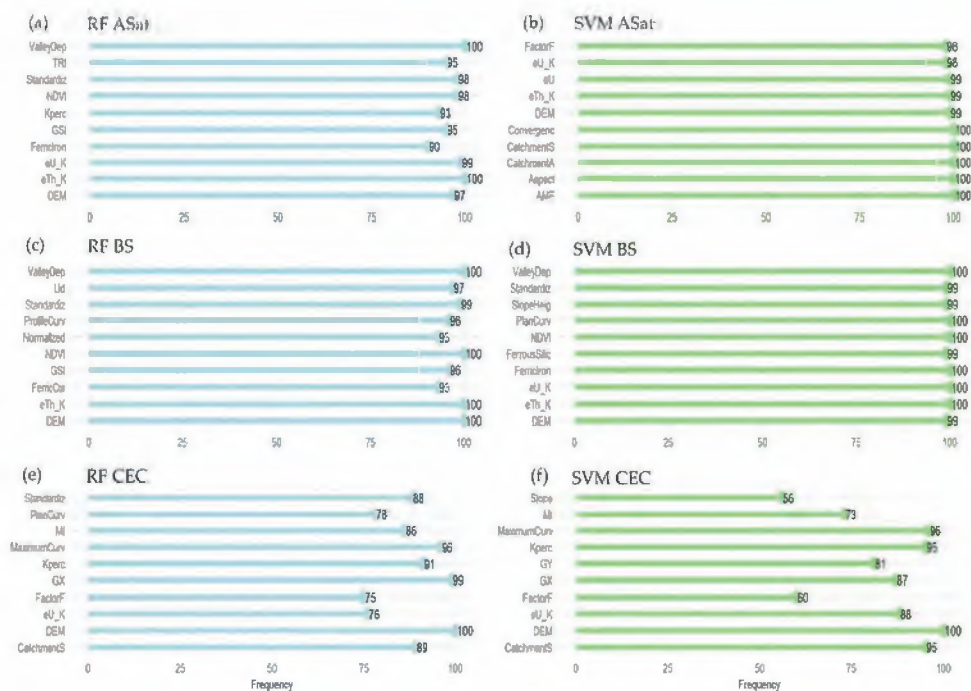


Figure 23: Top 10 predictors' frequency of the models' RF and SVM on the ASat, BS, and CEC prediction with AGD.

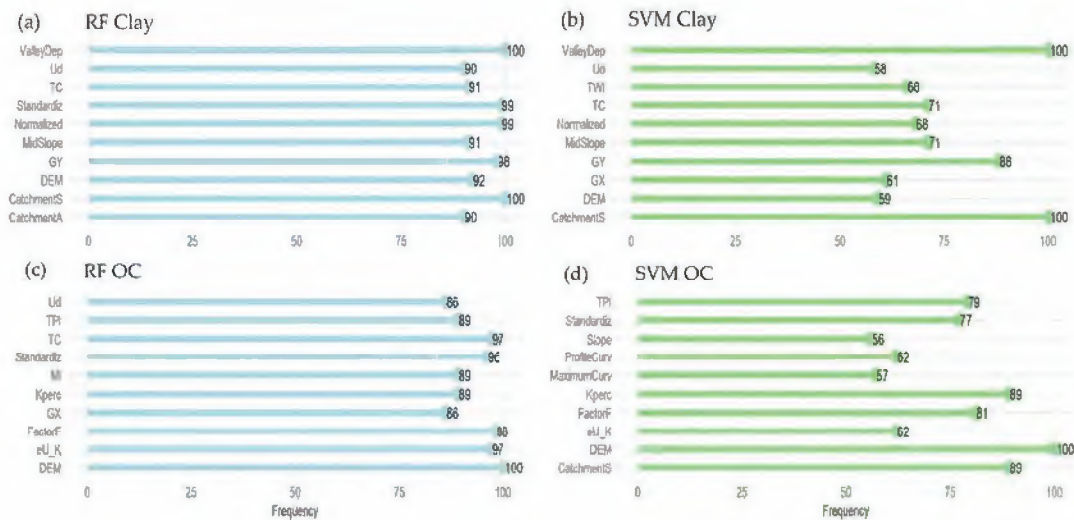


Figure 24: Top 10 predictors' frequency of the models' RF and SVM on the Clay and OC prediction with AGD.

Overall, AGD frequently appeared as significant for covariates in all models tested. In RF modeling for ASat, three of the ten most essential predictors were AGD (Kperc, eU/K, eTh/K), while in SVM modeling, five of the top ten predictors were AGD (Factor F, eU/K, eU, eTh, AMF). The terrain attributes were dominant in both RF and SVM modeling for BS, with two AGDs appearing among the top predictors (Ud and eTh/K, eU/K and eTh/K, respectively). In contrast, in RF modeling for CEC, five out of the top ten predictors were AGD (MI, Kperc, GX, FactorF, eU/K), while in SVM modeling, six of the top ten predictors were AGD (MI, Kperc, GY, GX, FactorF, eU/K).

In RF modeling for Clay content, three of the 10 most essential predictors were AGD (Ud, TC, GY), while in SVM modeling, 4 of the top 10 predictors were AGD (Ud, TC, GY, GX). In RF modeling for OC, 7 out of the top 10 predictors were AGD (Ud, TC, MI, Kperc, GX, FactorF, eU/K), while in SVM modeling, three of the top 10 predictors were AGD (Kperc, FactorF, eU/K). For Clay and OC, few covariates reached a frequency of 100. Only Valley Depth and Catchment Slope for Clay reached a frequency of 100 in both RF and SVM models, and DEM for OC in both RF and SVM models.

Figure 25 shows Spearman's correlation matrix between AGD and soil attributes.

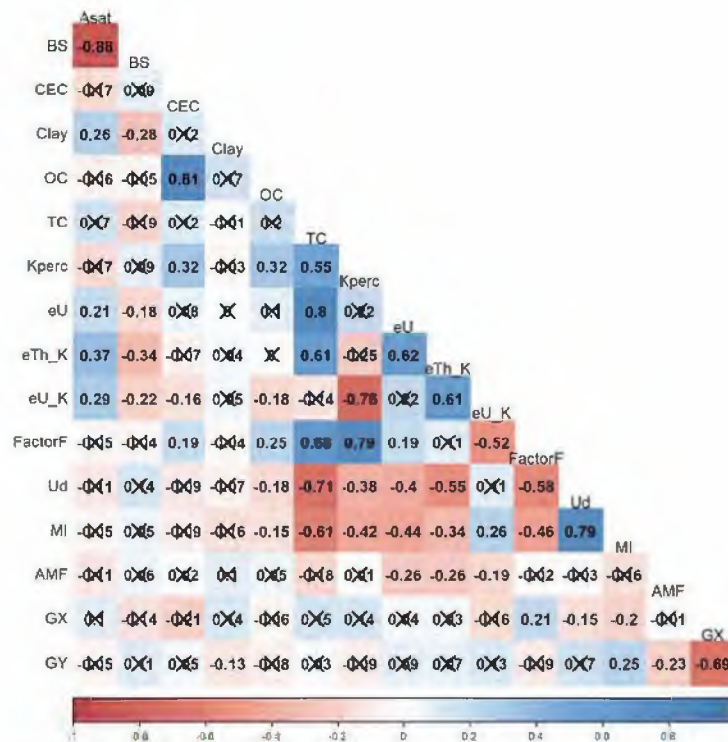


Figure 25: Spearman's correlation matrix between AGD and soil attributes. "X" represents not statistically significant values at the 0.2 significance level.

The AGD covariates eTh/K and eU/K ratios exhibited moderate correlations with ASat and BS properties, with inverse relationships. The eTh/K correlation was 0.37 with ASat and -0.34 with BS, while eU/K correlated 0.29 with ASat and -0.22 with BS. Despite the low values, another predictor that showed a correlation with ASat and BS was eU, with values of 0.21 and -0.18, respectively. For CEC, Kperc (0.32) exhibited moderate correlation, while Factor F (0.19) and eU/K (-0.16) showed weak correlations. For Clay, only GY (-0.13) showed weak correlations. Finally, for OC, Kperc (0.32) exhibited moderate correlation, while eU/K (-0.18), FactorF (0.25), Ud (-0.18), and MI (-0.15) showed weak correlations.

4.4 Discussion

4.4.1 AGD Importance to Predict Soil Attributes through DSM

From the observed results, the modeling with the AGD obtained better performance in terms of R^2 , RMSE, and MAE for both prediction algorithms used for modeling the selected soil

attributes. Mello et al. (2022) performed a similar study by applying different geophysical sensors combination (measured in situ) for modeling soil properties, and, in general, the modeling without using geophysical sensors also showed the poorest results. According to Mello et al. (2022), gamma-ray spectrometry and magnetic susceptibility were the best combinations of geophysical data. The comparison of the results of this study with those obtained by Mello et al. (2022) is represented in Table 13, except for ASat.

Table 13: Comparison of the results of this study with those obtained by Mello et al. (2022).

		R² Results			
		BS	CEC	Clay	OC
Mello et al. (2022)	RF	0.17	0.14	0.46	0.05
	SVM	0.11	0.29	0.49	0.03
This study	RF	0.22	0.23	0.15	0.16
	SVM	0.23	0.11	0.11	0.08

The comparison between the studies shows that the results obtained in the present study were satisfactory. The most significant discrepancy is in the performance of the Clay models with lower R² values than those observed by Mello et al. (2022). For BS, the performance of this study was better for both models: 0.22 versus 0.17 for the RF model and 0.23 versus 0.11 for the SVM model. For CEC, the performance of this study was better for the RF model (0.23 versus 0.14). For OC, both models also presented a better performance, highlighting the difference in the values observed for the RF model with values 0.16 in this study in contrast with 0.05 presented in Mello et al. (2022).

Loiseau et al. (2020) also analyzed the different models to predict topsoil particle-size distribution, with and without gamma-ray spectrometry, to replace lithology maps. According to the authors, a significant increase in models' performance was observed across all particle sizes when gamma-ray spectrometry was used instead of lithology, permitting the creation of more pedologically meaningful maps. Another example is presented by Vasques et al. (2020). Soil properties were measured in situ using several sensors to test the performance of the individual sensors and their combination to enhance soil property predictions (organic carbon, sum of bases, CEC, clay content, volumetric moisture, and bulk density). In this case, the X-ray fluorescence spectrometer sensor was superior. However, the gamma-ray sensor was the second best among individual sensors for predicting all those soil properties and the best for predicting CEC values.

The better performance of modeling with AGD was also confirmed through a comparison with NULL RMSE and NULL MAE values. In modeling with AGD, all NULL RMSE and NULL MAE values were higher for all soil attributes. Algorithms that exhibit RMSE and MAE values exceeding those of the NULL method are inferior and perform less than the use of mean value for the entire area (Mello et al., 2022).

Carvalho Junior et al. (2014) used the same database of the same area but with a different approach to predict soil properties. In their research, the modeling was performed only once for each attribute, without AGD, using the cross-validation method to evaluate the models' performance. The results in terms of R^2 were 0.19 for Clay 0–5 cm using ordinary kriging, 0.19 for Clay 5–15 cm using ordinary kriging, and 0.18 for Clay 15–30 cm using regression kriging. For OC, the results were 0.06 for 0–5 cm using linear regression, 0.07 for 5–15 cm using linear regression, and 0.11 for 15–30 cm using regression tree.

Although the average accuracy of the model's performance to predict Clay content has a lower R^2 value in this study (0.15 for RF and 0.11 for SVM in modeling with AGD), the number of times the models reached values $R^2 \geq 0.20$ was 32 for RF and 12 for SVM in Clay modeling, as shown in Table 11. For OC, the result presented by Carvalho Junior et al. (2014) ($R^2 = 0.11$) is a lower value than observed for the RF model in this study (0.16). Additionally, the number of times the models reached $R^2 \geq 0.20$ was 33 for RF and three for SVM in OC modeling.

In this case, another point that can be analyzed is that generalized harmonization at a depth of 0–30 in the present study seems not to have affected the performance of the models. Since the different depth-interval results (0–5, 5–15, and 15–30) shown by Carvalho Junior et al. (2014) for Clay and OC contents did not show a significant performance increase, and considering that the results of the present study were obtained through 100 models and that the validation was made from a set of unknown samples, it can be said that the results are more reliable and showed better performance than those presented by Carvalho Junior et al. (2014) to predict soil properties in the same study area.

4.4.2 Soil Properties and AGD Relationships

The results show that AGD combined with terrain and Sentinel-2 covariates played an essential role in predicting soil properties in the study area. AGDs were commonly observed as essential predictors for ASat, BS, CEC, Clay, and OC for both models (Figures 23 and 24). However, some covariates stand out considering Spearman's correlation analyses (Figure 25). The ratios eTh/K and eU/K are examples, exhibiting positive correlations with ASat (0.37 and 0.29, respectively) and negative with BS (-0.34 and -0.22 , respectively).

Interpreting the ratios between the elements' concentrations helps characterize different lithotypes and highlights zones of radioelement enrichment and alteration (Dentith and Mudge, 2014). Assuming that K is more mobile and tends to be leached from the weathering profile in tropical and subtropical climates, while eTh and eU are generally retained in the weathering profile and associated with clays, oxides, and resistant minerals, it is possible to establish relationships between weathering and erosion rates (Wilford, 2012). These relationships also agree with the observed results, where less-weathered soils have relatively higher values of BS (low values of these ratios). In contrast, more evolved soils are depleted in bases and are more acidic, and consequently have higher values of ASat (high values of these ratios). Therefore, the radioactive response largely depends on the evolutionary history of the landscape (Wilford, 2012).

The "Map Prediction and Uncertainty" analysis supports this hypothesis. The central region, which has high ASat values, mainly comprises Ferralsols and Acrisols. On the other hand, the highest BS values in the NW and SE regions correspond to the locations where Cambisol prevails. The eU correlations with ASat (0.21) and BS (-0.18) follow the same pattern. Mello et al. (2021) also demonstrated a negative correlation between BS and gamma-ray uranium. In Appendix A, Figure A1, attached to this paper, the airborne gamma-ray spectrometry maps of eTh/K and eU/K also support this idea. The areas with the highest values correspond to the high values observed in the average map for ASat (Figure 19a), and the lower values correspond to the high values observed in the average map for BS (Figure 19b), as well as the CV% maps demonstrating lower values and greater reliability in the data in these regions (Figure 20a,b). This relationship is essential as it can be a good indicator of soil fertility since ASat and BS are used for this purpose (Ker et al., 2015).

GY highlights anomalies perpendicular to its direction (in the x direction). It highlights superficial magnetic anomalies (Figure A2 in Appendix A), helping map geological contacts and shallow features, such as lineament distribution (Dentith and Mudge, 2014). Since lineament distribution influences groundwater flow (Falebita and Ayua, 2023; Hung et al., 2005), the tested hypothesis was to find a correlation between the drainage system and valley regions. In this sense, the GY correlated with Clay content (-0.13), where areas with low GY values present Clay deposits. According to Brubaker et al. (1993), clay values, among other soil properties, are significantly related to landscape position and tend to decrease downslope.

The negative correlation with GY corroborates with the hypothesis. However, the weak correlation value and the lack of significant correlation with other AGDs did not conform to what was expected. Furthermore, Clay content demonstrated an inverse relationship with GX (Figure A3 in Appendix A), suggesting that Clay values would change according to the directions of the applied filters, which does not make sense from a pedological point of view—demonstrating that these covariates may not be reliable for predicting soil properties. Despite the low performance observed in the present study, the use of AGD data to predict Clay contents showed satisfactory results, as reported by Mello et al. (2020), Mello et al. (2021), Mello et al. (2022) and Loiseau et al. (2020).

Factor F (or F-parameter) is a valuable tool for maximizing and distinguishing areas characterized by potassium enrichment resulting from hydrothermal alterations (Gnojek and Prichystal, 1985). This covariate showed a positive correlation with CEC (0.19), suggesting that high potassium values are related to high CEC values, confirmed by the positive correlations with Kperc (0.32) and negative with eU/K (-0.16). These results agree with Mello et al. (2021), showing a positive correlation between CEC and gamma-ray potassium (0.42). However, the CV% maps do not support the idea. Although the CV% maps for the SVM model demonstrate good reliability for the areas corresponding to the highest CEC values (Figures 22c and 21c, respectively), the RF model does not follow the same pattern (Figures 20c and 19c, respectively), demonstrating duality in the results and, consequently, low reliability.

In Appendix A, Figure A1, the Kperc and Factor F maps mainly match with higher elevation areas, as shown in Figure 1. The areas of occurrence of Conselheiro Paulino and Sana granites, for example, highlight high potassium values as they represent undeformed granites resistant to

weathering (CPRM, 2016b; Tupinambá and Heilbron, 2013). So, in this case, the high potassium values are related to the parent material, not the soil potassium content, explaining why BS did not correlate well with K_{perc} and factor F. Furthermore, knowing that CEC is the sum of bases plus H^+ and Al^+ (Ker et al., 2015), it makes sense that CEC is related to potassium concentrations, but since it is also associated with aluminum, it may have confused the models.

The K_{perc} covariate showed a positive correlation with OC (0.32) as well factor F (0.25). According to a study conducted by (Rawlins et al., 2009), it was discovered that radiometric potassium plays a crucial role in predicting soil carbon in Northern Ireland. Nevertheless, the relationship is inversed in Rawlins et al. (2009)'s case, with a correlation value of -0.51 . Soils with high organic carbon content significantly diminish gamma rays' intensity, as confirmed by the negative correlation with radiometric thorium (-0.36) and radiometric uranium (-0.29).

According to Powers et al. (2011), soil carbon has a high spatial variation, mainly where the land cover was altered for different purposes in tropical areas. To obtain an accurate picture of the carbon content in tropical regions, gathering data from a wide range of locations to account for this variation is essential. Mello et al. (2021) found the same pattern observed in the present study, showing a positive correlation between OC and radiometric potassium (0.17) in a study area in southeastern Brazil and climate classified as Cwa. The difference between the results can be explained mainly by the climatic conditions. In contrast, in the case of subtropical and tropical climates, the positive correlation values between OC and K could be related to topsoil erosion.

However, as observed for CEC, the CV% maps for OC demonstrated duality in the data. Although the CV% maps for the SVM model show good reliability for the areas corresponding to the highest OC values (Figures 22e and 21e, respectively), the RF model did not follow the same pattern (Figures 20e and 19e, respectively). So, further studies are needed to understand the relationship between the distribution of radioelements and organic carbon contents.

According to Mello et al. (2020), Mello et al. (2021) and Mello et al. (2022), gamma-ray and magnetic susceptibility can be associated with soil attributes. However, there is still a gap in understanding the optimal covariates and their potential combinations to investigate further soil weathering, pedogenesis, and their relationship with soil attributes, especially when using airborne geophysical data, where appropriate scale can be an issue.

4.4.3 Precautions and Challenges

Based on the discussed results, some considerations should be made. First, data availability and quality are essential for more reliable modeling. Regrettably, the current soil databases lack the necessary comprehensiveness and precision to support the utilization of soil information (Lagacherie and Mcbratney, 2006). Digital soil modeling relies on accurate and comprehensive soil data. However, available soil data are generally insufficient or of varying quality, making it challenging to build reliable models based on a representative input soil dataset. Improving data collection methods, standardizing data formats, and enhancing data-sharing practices are crucial for better results (Chen et al., 2022). For example, in this study, although the analysis of the models and some possible relationships with soil properties suggest that AGD can be a helpful tool for soil mapping, more reliable results could be achieved if the sample density were representative. Of the 208 samples, 97 are Ferralsols, 35 are Cambisols, and 62 Acrisols, lacking representation for classes such as Leptosols (3), Fluvisols (5), Gleysols (5), and one Nitisol which were also reported in the study area.

The number of samples also affects the model's validation and evaluation. Although the proportion for training (75%) and validation (25%) is suitable for DSM, only 156 samples were used for training and 52 for validation. Appropriate scale sampling and sampling design are other issues. Appendix A shows the correlation table of the studied properties and all the proposed covariates (Figure A3). The observed correlations were well below expectations, mainly for terrain covariates widely used in soil mapping (Chen et al., 2022). This problem may be related to the density of samples collected, as shown in the SE of the map (Figure 1). According to Lagacherie et al. (2019), the statistical parameters are sensitive to the number and the locations of the soil observations. In this sense, sample distribution according to scale and number of samples is essential to ensure the accuracy and reliability of the models.

Spatial and temporal variability are also a challenge since soils exhibit considerable variability, and most DSM studies typically concentrate on predicting soil properties for a specific period (Chen et al., 2022). Topography, land use, climate, and geological processes are examples of factors that influence soil properties, and capturing this variability accurately in digital soil models is challenging. Bom Jardim County is a good example of an area that undergoes variations

mainly through anthropic interferences associated with agricultural activities and the dominant mountainous relief (Calderano Filho et al., 2012), which may explain the low significance of the Sentinel-2 based indices since the soil collection was carried out between the years 2009 and 2011, and the sentinel images were from 2021.

DSM often faces several challenges since soil observations are scarce and costly. However, achieving satisfactory results with AGD is still possible, as reported by Adler et al. (2023), Chen et al. (2021), Loiseau et al. (2020), Loiseau et al. (2021), Maino et al. (2022) and Ng et al. (2023). After extensive research over the past years, DSM has made significant progress in producing soil maps, a credible alternative to fulfill the increasing worldwide demand for spatial soil information (Lagacherie, 2008).

4.5 Conclusions

The RF algorithm showed the best performance in terms of data usage with AGD to predict all the selected soil attributes (ASat, BS, CEC, Clay, and OC). The SVM model also performed better regarding data usage with AGD than without AGD. Moreover, the comparison with null values revealed that using AGD for modeling was beneficial in terms of improving the accuracy of soil attribute predictions. The models with AGD showed more R^2 values equal to or higher than 0.2, which is considered satisfactory for machine learning algorithms to predict soil attributes that present a wide spatial variability. The most significant improvements were observed in the ASat and BS attributes.

In summary, using AGD for modeling showed its benefits in improving the accuracy of soil attribute predictions. Although the models' performances for CEC, Clay, and OC contents did not attain satisfactory values, using AGD led to a slightly improved overall performance. The AGD covariate with the best correlations with soil properties was eTh/K, showing that the ratio between the element's concentrations can be a helpful tool for highlighting zones of weathering and an indicator of soil fertility. Spearman's correlation showed associations mainly with the gamma-ray spectrometry data, while the magnetic data did not show satisfactory results. Magnetic data may provide good results combined with gamma-ray spectrometry data in areas where chemical contrasts between lithological units are evident.

Overall, the results suggest that AGD can be a helpful tool for soil mapping, particularly in areas where traditional soil survey methods are impractical or cost-prohibitive. However, it is important to note that the effectiveness of AGD in predicting soil properties may depend on several factors, including the quality and resolution of the covariates, number and representativeness of the soil samples, appropriate scale sampling and design, land cover spatial and temporal variability, and the statistical models used for analysis. Further research is needed to better understand the factors that rule models' performance variability and to determine the optimal approach for using AGD for soil mapping.

5. Considerações finais

O presente estudo ressalta a contribuição positiva dos dados aerogeofísicos no âmbito do MDS, exemplificado principalmente pelos dados gamaespectrométricos. Embora os dados magnéticos não tenham apresentado resultados promissores, sua aplicação ainda pode ser relevante em áreas onde as unidades litológicas possuem contrastes químicos mais acentuados, demandando estudos adicionais. De toda forma, os dados aerogeofísicos desempenham um papel complementar em relação às variáveis ambientais tradicionalmente utilizadas. Esses dados oferecem uma perspectiva mais abrangente ao considerar covariáveis representativas do material de origem, fator de formação do solo menos utilizado no MDS. Portanto, a realização de estudos adicionais se faz necessária, a fim de promover discussões mais aprofundadas, o que por sua vez contribuirá para uma compreensão mais aprimorada dessa abordagem e uma apreciação mais completa de suas limitações, especialmente no que se refere à correlação desses dados com as propriedades do solo.

6. Referências

- ADLER, K. et al. Digital Soil Mapping of Cadmium: Identifying Arable Land for Producing Winter Wheat with Low Concentrations of Cadmium. *Agronomy*, v. 13, n. 2, p. 317, 2023.
- ANA - AGÊNCIA NACIONAL DE ÁGUAS. Agência Nacional de Águas e Saneamento Básico. Índices e Estatísticas das Estações Pluviométricas e Fluviométricas. 2020. Disponível em: <<https://dadosabertos.ana.gov.br/>>.
- BANNARI, A.; ASALHI, H.; TEILLET, P. M. Transformed Difference Vegetation Index (TDVI) for Vegetation Cover Mapping. In: *Proceedings of the Geoscience and Remote Sensing Symposium, IGARSS '02, IEEE International*, v. 5, 2002.
- BARANOV, V. A new method for interpretation of aeromagnetic maps: pseudo-gravimetric anomalies. *Geophysics*, v. 22, n. 2, p. 359-382, 1957.
- BARBOSA, I. O. et al. Geology, airborne geophysics, geomorphology and soils in the individualization of the Niquelândia mafic-ultramafic complex, Goiás state, Brazil. *Brazilian Journal of Geophysics*, v. 31, n. 3, p. 463-481, 2013.
- BASTOS, B. P. et al. Could Airborne Geophysical Data Be Used to Improve Predictive Modeling of Agronomic Soil Properties in Tropical Hillslope Area? *Remote Sensing*, v. 15, n. 15, p. 3719, 2023.
- BEAUDETTE, D. E.; ROUDIER, P.; O'GEEN, A. T. Algorithms for quantitative pedology: A toolkit for soil scientists. *Computers & Geosciences*, v. 52, p. 258-268, 2013.
- BÖHNER, J.; SELIGE, T. Spatial prediction of soil attributes using terrain analysis and climate regionalisation. In: *SAGA-Analyses and modelling applications*. Goltze, 2006.
- BOUSBIH, S. et al. Soil Texture Estimation Using Radar and Optical Data from Sentinel-1 and Sentinel-2. *Remote Sensing*, v. 11, n. 13, p. 1520, 2019.
- BREIMAN, L. Random forests. *Machine learning*, v. 45, p. 5-32, 2001.
- BRIGGS, I. C. Machine contouring using minimum curvature. *Geophysics*, v. 39, n. 1, p. 39-48, 1974.

BRUBAKER, S. C. et al. Soil Properties Associated with Landscape Position. Soil Science Society of America Journal, v. 57, n. 1, p. 235–239, 1993.

CALDERANO FILHO, B. Análise geoambiental de Paisagens Rurais Montanhosas da Serra do Mar utilizando Redes Neurais Artificiais. Subsídios a sustentabilidade ambiental de ecossistemas frágeis e fragmentados sob interferência antrópica, 2012.

CALDERANO FILHO, B. et al. Solos do médio alto curso do Rio Grande, região serrana do Estado do Rio de Janeiro. 2012.

CARVALHO JUNIOR, W. D. et al. Evaluation of statistical and geostatistical models of digital soil properties mapping in tropical mountain regions. Revista Brasileira de Ciência do Solo, v. 38, n. 3, p. 706–717, 2014.

CAVAZZI, S. et al. Are fine resolution digital elevation models always the best choice in digital soil mapping? Geoderma, v. 195–196, p. 111–121, 2013.

CHEN, S. et al. Digital mapping of GlobalSoilMap soil properties at a broad scale: A review. Geoderma, v. 409, p. 115567, 2022.

CHEN, S. et al. Digital mapping of the soil thickness of loess deposits over a calcareous bedrock in central France. CATENA, v. 198, p. 105062, 2021.

CONRAD, O. et al. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geoscientific Model Development, v. 8, n. 7, p. 1991–2007, 2015.

CORTES, C.; VAPNIK, V. Support-vector networks. Machine learning, v. 20, p. 273-297, 1995.

CPRM - COMPANHIA DE PESQUISA DE RECURSOS MINERAIS. Serviço Geológico do Brasil. Geodiversidade do Estado do Rio de Janeiro. Programa Geologia do Brasil levantamento da Geodiversidade. 2020.

CPRM - COMPANHIA DE PESQUISA DE RECURSOS MINERAIS. Serviço Geológico do Brasil. Geologia e Recursos Minerais do Estado do Rio de Janeiro: texto explicativo do mapa geológico e de recursos minerais. 2016b.

CPRM - COMPANHIA DE PESQUISA DE RECURSOS MINERAIS. Serviço Geológico do Brasil. Mapa Geológico e de recursos minerais do Estado do Rio de Janeiro. Escala 1:400.000. 2016a. Disponível em: <<http://rigeo.cprm.gov.br/jspui/handle/doc/18458>>.

CPRM - COMPANHIA DE PESQUISA DE RECURSOS MINERAIS. Serviço Geológico do Brasil. Carta Geomorfológica do Município de Bom Jardim, escala 1:100.000. 2017. Disponível em: <<https://rigeo.cprm.gov.br/handle/doc/18187>>.

CPRM - COMPANHIA DE PESQUISA DE RECURSOS MINERAIS. Serviço Geológico do Brasil. Projeto Aerogeofísico Rio de Janeiro (Projeto 1.117). Relatório final do levantamento e processamento dos dados magnetométricos e gamaespectrométricos. 219p, 2012.

DANTAS, M. E. Geomorfologia do estado do Rio de Janeiro. COMPANHIA DE PESQUISA DE RECURSOS MINERAIS-CPRM. Serviço Geológico do Brasil. Estudo Geoambiental do Estado do Rio de Janeiro. Rio de Janeiro, CPRM/Embrapa Solos, 2000.

DENTITH, M. C.; MUDGE, S. T. Geophysics for the mineral exploration geoscientist. Cambridge, United Kingdom: Cambridge University Press, 2014.

DRUSCH, M. et al. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. Remote Sensing of Environment, v. 120, p. 25–36, 2012.

EMBRAPA - EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. Sistema Brasileiro de Classificação de Solos. Embrapa Solos, 2006. Disponível em <<https://www.infoteca.cnptia.embrapa.br/handle/doc/338818>>

FALEBITA, D. E.; AYUA, K. J. Appraisal of lineaments for groundwater prognosis in the Middle Benue Trough, Nigeria: a case study. Sustainable Water Resources Management, v. 9, n. 1, p. 12, 2023.

FAO - FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. World Reference Base for Soil Resources 2014. International Soil Classification System for Naming Soils and Creating Legends for Soil Maps; World Soil Resources Report 106, 2014

FERREIRA, F. J. F. et al. Gamaespectrometria Aérea e sua Aplicação na Cartografia Geológica. Tópicos Especiais em Cartografia Geológica. 2ª edição, Universidade Federal do Paraná, Curitiba, 2016.

FERREIRA, R. G. et al. Machine learning models for streamflow regionalization in a tropical watershed. *Journal of Environmental Management*, v. 280, p. 111713, 2021.

FRIEDMAN, J. H. Stochastic gradient boosting. *Computational Statistics & Data Analysis*, v. 38, n. 4, p. 367–378, 2002.

GALLANT, J. C.; DOWLING, T. I. A multiresolution index of valley bottom flatness for mapping depositional areas: MULTIREOLUTION VALLEY BOTTOM FLATNESS. *Water Resources Research*, v. 39, n. 12, 2003.

GEOSOFT. *Tutorials Oasis Montaj: Bi-Directional Gridding*. Geosoft Inc. 2023. Disponível em: <<https://www.seequent.com>>

GERALDES, M. C. et al. *Geologia e recursos minerais da folha Casimiro de Abreu SF. 23-ZB-III*. CPRM, 2012.

GNOJEK, I.; PŘICHYSTAL, A. A new zinc mineralization detected by airborne gamma-ray spectrometry in northern Moravia (Czechoslovakia). *Geoexploration*, v. 23, n. 4, p. 491-502, 1985.

GOMES, L. C. et al. Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma*, v. 340, p. 337–350, 2019.

GUISAN, A.; WEISS, S. B.; WEISS, A. D. GLM versus CCA spatial modeling of plant species distribution. *Plant ecology*, v. 143, p. 107-122, 1999.

HEILBRON, M.; MACHADO, N. Timing of terrane accretion in the Neoproterozoic–Eopaleozoic Ribeira orogen (SE Brazil). *Precambrian Research*, v. 125, n. 1-2, p. 87-112, 2003.

HJERDT, K. N. et al. A new topographic index to quantify downslope controls on local drainage: Technical Note. *Water Resources Research*, v. 40, n. 5, 2004.

HUNG, L. Q.; BATELAAN, Okke; DE SMEDT, Florimond. Lineament extraction and analysis, comparison of LANDSAT ETM and ASTER imagery. Case study: Suoimuoi tropical karst catchment, Vietnam. In: *Remote sensing for environmental monitoring, GIS applications, and geology V*. SPIE, 2005. p. 182-193.

IAEA - INTERNATIONAL ATOMIC ENERGY AGENCY. *Guidelines for Radioelement Mapping Using Gamma Ray Spectrometry Data*; IAEA: Vienna, Austria, 2003.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTIA. Mapa de Clima do território brasileiro. Escala 1:5.000.000. 2002. Disponível em: <<https://www.ibge.gov.br>>

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTIA. Cidades e Estados. 2021b. Disponível em: <<https://www.ibge.gov.br/cidades-e-estados/rj/bom-jardim.html>>

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Manual Técnico de Geomorfologia. Coordenação de Recursos Naturais e Estudos Ambientais. Rio de Janeiro, RJ: IBGE, 2009.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Base cartográfica vetorial contínua do Estado do Rio de Janeiro na escala 1:25.000. Projeto RJ-25, 2018. Disponível em: <<https://www.ibge.gov.br/>>

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Divisão Territorial Brasileira. 2021a. Disponível em: <<https://www.ibge.gov.br/>>

ISLES, D. J.; RANKIN, L. R. Geological interpretation of aeromagnetic data. Society of Exploration Geophysicists and Australian Society of Exploration Geophysicists, 2013.

IZA, E. R. H. F. et al. Integration of Geochemical and Geophysical Data to Characterize and Map Lateritic Regolith: An Example in the Brazilian Amazon. *Geochemistry, Geophysics, Geosystems*, v. 19, n. 9, p. 3254–3271, 2018.

JENNY, H. Factors of soil formation: a system of quantitative pedology. New York: Dover, 1994.

KAYA, F. et al. Combining Digital Covariates and Machine Learning Models to Predict the Spatial Variation of Soil Cation Exchange Capacity. *Land*, v. 12, n. 4, p. 819, 2023.

KEAREY, P.; BROOKS, M.; HILL, I. Geofísica de exploração. São Paulo: Oficina de Textos, 2009.

KER, J.C.; CURI, N.; SCHAEFER, C.E.G.R.; VIDAL-TORRADO, P. Pedologia: Fundamentos; Sociedade Brasileira de Ciência do Solo: Viçosa, MG, Brasil, 2015; ISBN 978-85-86504-09-9.

KÖTHE, R.; LEHMEIER, F. SARA—System zur Automatischen Relief-Analyse: User Manual. Tool Convergence Index available in SAGA-GIS Tool Library Documentation. 1996. Disponível em: <<https://saga-gis.sourceforge.io/>>

KUHN, M. Package caret: classification and regression training. R package version 6.0-92; 2022. Disponível em: <<https://CRAN.R-project.org/package=caret>>

KUHN, M.; JOHNSON, K. Applied Predictive Modeling; Springer: New York, NY, USA, 2013; ISBN 978-1-4614-6848-6.

LAGACHERIE, P. Digital soil mapping: a state of the art. Digital soil mapping with limited data, p. 3-14, 2008.

LAGACHERIE, P. et al. How far can the uncertainty on a Digital Soil Map be known?: A numerical experiment using pseudo values of clay content obtained from Vis-SWIR hyperspectral imagery. Geoderma, v. 337, p. 1320-1328, 2019.

LAGACHERIE, P. et al. Analysing the impact of soil spatial sampling on the performances of Digital Soil Mapping models and their evaluation: A numerical experiment on Quantile Random Forest using clay contents obtained from Vis-NIR-SWIR hyperspectral imagery. Geoderma, v. 375, p. 114503, 2020.

LAGACHERIE, P.; MCBRATNEY, A. B. Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. Developments in soil science, v. 31, p. 3-22, 2006.

LI, X. Understanding 3D analytic signal amplitude. Geophysics, v. 71, n. 2, p. L13-L16, 2006.

LOISEAU, T. et al. Could airborne gamma-spectrometric data replace lithological maps as covariates for digital soil mapping of topsoil particle-size distribution? A case study in Western France. Geoderma Regional, v. 22, p. e00295, 2020.

LOISEAU, T. et al. Density of soil observations in digital soil mapping: A study in the Mayenne region, France. Geoderma Regional, v. 24, p. e00358, 2021.

MAINO, A. et al. Airborne Radiometric Surveys and Machine Learning Algorithms for Revealing Soil Texture. Remote Sensing, v. 14, n. 15, p. 3814, 2022.

MAYER, B.; KYLLING, A. The libRadtran software package for radiative transfer calculations-description and examples of use. Atmospheric Chemistry and Physics, v. 5, n. 7, p. 1855-1877, 2005.

MCBRATNEY, A. B.; MENDONÇA SANTOS, M. L.; MINASNY, B. On digital soil mapping. *Geoderma*, v. 117, n. 1–2, p. 3–52, 2003.

MELLO, D. C. D. et al. A new methodological framework by geophysical sensors combinations associated with machine learning algorithms to understand soil attributes. *Geoscientific Model Development Discussions*, v. 2021, p. 1-42, 2022.

MELLO, D. C. D. et al. Applied gamma-ray spectrometry for evaluating tropical soil processes and attributes. *Geoderma*, v. 381, p. 114736, 2021.

MELLO, D. C. D. et al. Soil magnetic susceptibility and its relationship with naturally occurring processes and soil attributes in pedosphere, in a tropical environment. *Geoderma*, v. 372, p. 114364, 2020.

MEYER, H. et al. Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. *Ecological Modelling*, v. 411, p. 108815, 2019.

MINTY, B. R. S. Fundamentals of airborne gamma-ray spectrometry. *AGSO Journal of Australian Geology and Geophysics*, v. 17, p. 39-50, 1997.

MUELLER-WILM U.; DEVIGNOT O.; PESSIOT L. S2 MPC Sen2Cor Configuration and User Manual. European Space Agency. 2017.

NG, W. et al. Digital soil mapping of lithium in Australia. *Earth System Science Data*, v. 15, n. 6, p. 2465-2482, 2023.

PERERA, Y. Y. et al. Prediction of the soil-water characteristic curve based on grain-size-distribution and index properties. In: *Advances in pavement engineering*, p. 1-12, 2005.

POWERS, J. S. et al. Geographic bias of field observations of soil carbon stocks with tropical land-use changes precludes spatial extrapolation. *Proceedings of the National Academy of Sciences*, v. 108, n. 15, p. 6318-6322, 2011.

QGIS DEVELOPMENT TEAM. QGIS Geographic Information System. Open Source Geospatial Foundation Project. 2023. Disponível em: <<http://qgis.osgeo.org>>

R CORE TEAM. A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Austria, 2023.

RAWLINS, B. G. et al. Airborne radiometric survey data and a DTM as covariates for regional scale mapping of soil organic carbon across Northern Ireland. *European Journal of Soil Science*, v. 60, n. 1, p. 44-54, 2009.

REINHARDT, N.; HERRMANN, L. Gamma-ray spectrometry as versatile tool in soil science: A critical review. *Journal of Plant Nutrition and Soil Science*, v. 182, n. 1, p. 9–27, 2019.

RIBEIRO, V. B.; MANTOVANI, M.; LOURO, V. H. A. Aerogamaespectrometria e suas aplicações no mapeamento geológico. *Terrae Didatica*, v. 10, n. 1, p. 29, 2014.

RILEY, S. J.; DEGLORIA, S. D.; ELLIOT, R. Index that quantifies topographic heterogeneity. *intermountain Journal of sciences*, v. 5, n. 1-4, p. 23-27, 1999.

ROUSE, J. W. et al. Monitoring vegetation systems in the Great Plains with ERTS. *NASA Spec. Publ*, v. 351, n. 1, p. 309, 1973.

SANTOS, H. G. dos et al. *Sistema Brasileiro de Classificação de Solos*. Brasília, DF: Embrapa, 2018.

SARMAST, M.; FARPOOR, M. II.; BOROUJENI, I. E. Magnetic susceptibility of soils along a lithotoposequence in southeast Iran. *Catena*, v. 156, p. 252-262, 2017.

SAUNDERS, D. F. et al. Relation of thorium-normalized surface and aerial radiometric data to subsurface petroleum accumulations. *Geophysics*, v. 58, n. 10, p. 1417-1427, 1993.

SIQUEIRA, R. G. et al. Machine learning applied for Antarctic soil mapping: Spatial prediction of soil texture for Maritime Antarctica and Northern Antarctic Peninsula. *Geoderma*, v. 432, p. 116405, abr. 2023.

SORDI, D. A. de. *Atlas aerogeofísico: estado do Rio de Janeiro*. 2020.

TUPINAMBÁ, M. et al. *Geologia e recursos minerais da folha Nova Friburgo SF. 23-ZB-II, estado do Rio de Janeiro*. CPRM, 2012.

TUPINAMBÁ, M. et al. *Geologia da Faixa Ribeira Setentrional: estado da arte e conexões com a Faixa Araçuaí*. Geonomos, 2007.

TUPINAMBÁ, M.; TEIXEIRA, W.; HEILBRON, M. Evolução tectónica e magmática da faixa ribeira entre o neoproterozoico e o paleozoico inferior na região serrana do estado do Rio de Janeiro, Brasil. *Anuário do Instituto de Geociências*, v. 35, n. 2, p. 140-151, 2013.

ULBRICH, H. H. G. J. et al. Levantamentos gamaespectrométricos em granitos diferenciados. I: revisão da metodologia e do comportamento geoquímico dos elementos K, Th e U. *Geologia USP. Série Científica*, v. 9, n. 1, p. 33-53, 2009.

VALAEE, M. et al. Using magnetic susceptibility to discriminate between soil moisture regimes in selected loess and loess-like soils in northern Iran. *Journal of Applied Geophysics*, v. 127, p. 23-30, 2016.

VALERIANO, C. M., MENDES, J. C., TUPINAMBÁ, M., BONGIOLO, E., HEILBRON, M., & JUNHO, M. D. C. B. (2016) - Cambro-Ordovician post-collisional granites of the Ribeira belt, SE-Brazil: A case of terminal magmatism of a hot orogen. *Journal of South American Earth Sciences*, 68, p. 269-281.

MORISSON VALERIANO, C. de et al. Cambro-Ordovician post-collisional granites of the Ribeira belt, SE-Brazil: a case of terminal magmatism of a hot orogen. *Journal of South American Earth Sciences*, v. 68, p. 269-281, 2016.

VAN DER MEER, F. D.; VAN DER WERFF, H. M. A.; VAN RUITENBEEK, F. J. A. Potential of ESA's Sentinel-2 for geological applications. *Remote sensing of environment*, v. 148, p. 124-133, 2014.

VASQUES, G. M. et al. Field proximal soil sensor fusion for improving high-resolution soil property maps. *Soil Systems*, v. 4, n. 3, p. 52, 2020.

WEISS, A. Topographic position and landforms analysis. In: Poster presentation, ESRI user conference, San Diego, CA. 2001.

WILFORD, J. A weathering intensity index for the Australian continent using airborne gamma-ray spectrometry and digital terrain analysis. *Geoderma*, v. 183-184, p. 124-142, 2012.

WILFORD, J. R.; BIERWIRTH, P. E.; CRAIG, M. A. Application of airborne gamma-ray spectrometry in soil/regolith mapping and applied geomorphology. *AGSO Journal of Australian Geology and Geophysics*, v. 17, n. 2, p. 201-216, 1997.

WILFORD, J.; MINTY, B. The use of airborne gamma-ray imagery for mapping soils and understanding landscape processes. *Developments in soil science*, v. 31, p. 207-610, 2006.

WILSON, J. P.; GALLANT, J. C. *Terrain analysis: principles and applications*. John Wiley & Sons, 2000.

WILSON, M. J. The importance of parent material in soil classification: A review in a historical context. *CATENA*, v. 182, p. 104131, 2019.

WOOD, J. Geomorphometry in landserf. *Developments in soil science*, v. 33, p. 333-349, 2009.

XIAO, J. et al. Development of topsoil grain size index for monitoring desertification in arid land using remote sensing. *International Journal of Remote Sensing*, v. 27, n. 12, p. 2411-2422, 2006.

7. Anexo A

Table A1: Basic statistics of CV% maps for modeling with and without AGD.

		With AGD				
		ASat	BS	CEC	Clay	OC
RF	minimum	5.16	2.85	2.90	2.00	2.41
	mean	13.86	9.29	8.30	4.56	9.29
	median	12.86	9.85	6.38	4.43	6.40
	maximum	57.07	29.16	25.12	17.65	35.60
SVM	minimum	5.95	4.23	2.87	2.80	2.90
	mean	14.35	8.31	4.73	6.22	5.43
	median	11.84	8.23	4.57	6.09	5.34
	maximum	135.63	24.91	9.46	11.85	10.27
		Without AGD				
		ASat	BS	CEC	Clay	OC
RF	minimum	5.79	2.31	2.64	2.20	2.31
	mean	14.85	10.95	7.53	4.93	9.07
	median	13.80	10.30	5.33	4.78	5.99
	maximum	56.24	28.15	24.91	17.28	33.23
SVM	minimum	10.13	5.61	2.47	2.68	3.04
	mean	22.08	8.41	5.50	4.89	5.50
	median	20.08	8.21	5.41	4.80	5.43
	maximum	103.84	18.41	10.85	8.25	9.56

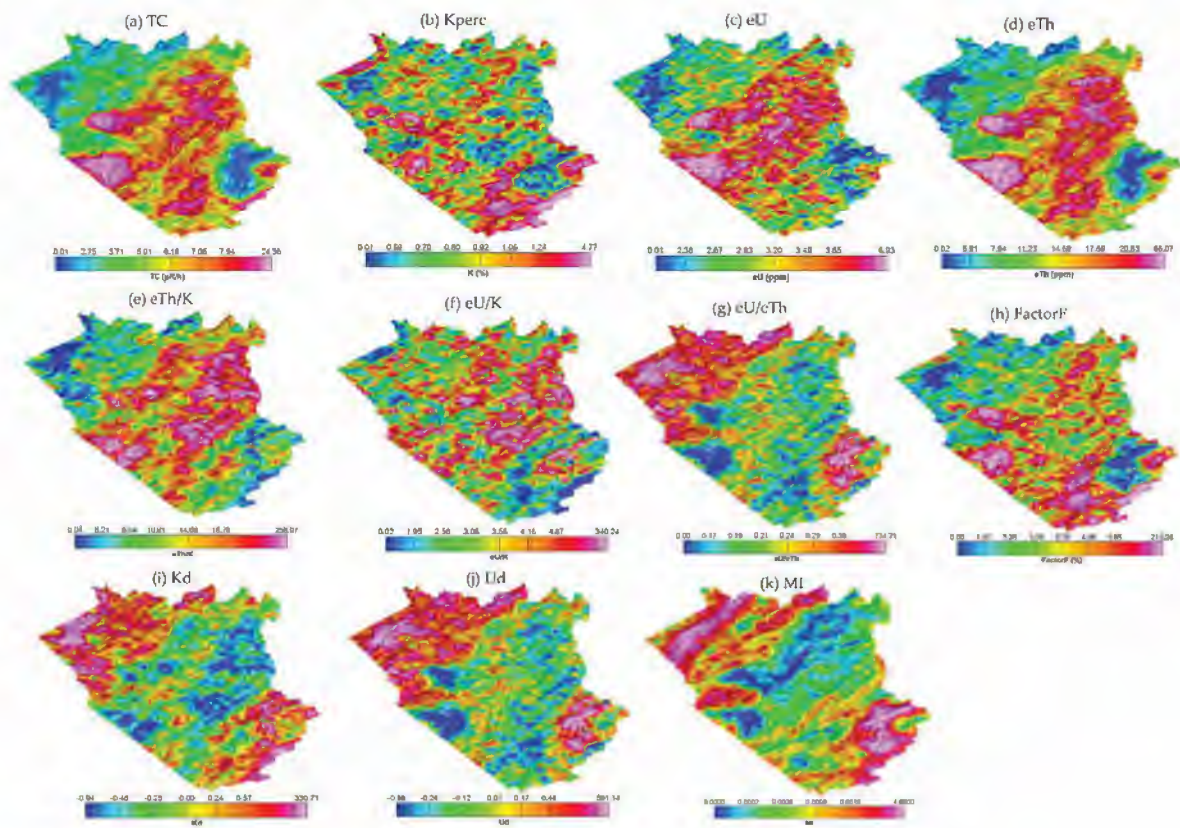


Figure A1: Airborne gamma-ray spectrometry maps. (a) Total Count, (b) potassium, (c) uranium, (d) thorium, (e) eTh/K ratio, (f) eU/K ratio, (g) eU/eTh ratio, (h) Factor F, (i) anomalous potassium, (j) anomalous uranium, and (k) Mafic Index.

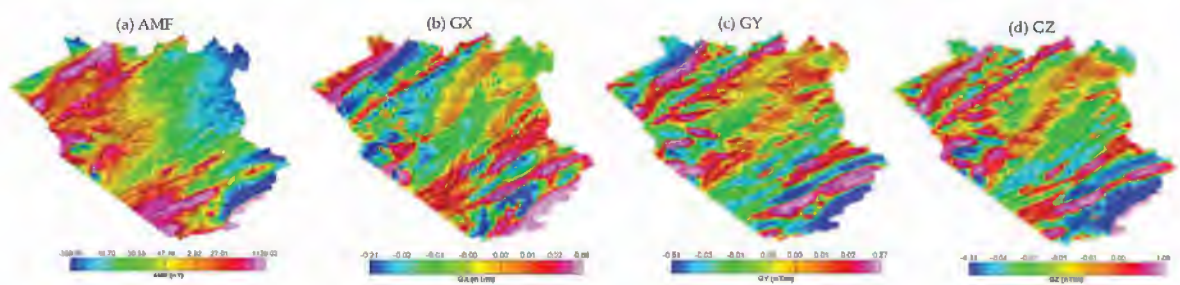


Figure A2: Airborne magnetic maps. (a) AMF, (b) GX, (c) GY and (d) GZ.



Clustering airborne gamma-ray spectrometry data in Nova Friburgo, State of Rio de Janeiro, southeastern Brazil

Blenda Pereira Bastos^{a,*}, Helena Saraiva Koenow Pinheiro^b, Waldir Carvalho Junior^c, Lúcia Helena Cunha dos Anjos^b, Francisco José Fonseca Ferreira^d

^a Federal Rural University of Rio de Janeiro (UFRRJ), Geoscience Institute, Postgraduate Program in Modeling and Geological Evolution, Seropédica, Rio de Janeiro, Brazil

^b Federal Rural University of Rio de Janeiro (UFRRJ), Agronomy Institute, Soils Department, Seropédica, Rio de Janeiro, Brazil

^c Brazilian Agricultural Research Corporation, Embrapa Soils, Jardim Botânico, Rio de Janeiro, Brazil

^d Federal University of Paraná (UFPR), Laboratory for Research in Applied Geophysics, Department of Geology, Curitiba, Paraná, Brazil

ARTICLE INFO

Keywords:

Data-driven classification
Knowledge-driven classification
Gamma-ray spectrometry
Geological mapping

ABSTRACT

The goal of this study was to test different image clustering techniques, using airborne gamma-ray spectrometry data to optimize the interpretation of areas with similar properties in large scale. The methodology applied was based on the comparison of two techniques of data-driven classification (K-means and Gaussian Mixture Models) and a technique of knowledge-driven classification (Simplified RGB) to discriminate domains from the primary variables potassium (K), uranium (eU), and thorium (eTh) obtained from airborne gamma-ray spectrometry surveys. The performance of these three methods was evaluated through pre-processing techniques and post-processing including the best number of clusters/classes, visual interpretation, internal validation, and boxplot analysis. The clustering performance was considered satisfactory through the visual interpretation and comparison with the geological map and the DEM, for all three methods. For a quantitative analysis, the simplest model of unsupervised clustering Gaussian Mixture Models prevailed.

1. Introduction

Technological development has provided a great advance in the techniques of acquiring information from the earth's surface, and airborne gamma-ray spectrometry data from Rio de Janeiro Aero-geophysical Project carried out by Geological Survey of Brazil (CPRM, 2012), is an example. This superficial method can measure the concentrations of K (potassium), U (uranium) and Th (thorium) about 30–40 cm of surface depth. These concentrations are obtained by capturing gamma-ray radiation from the radioactive decay series (^{40}K , ^{238}U and ^{232}Th) and through the calibrations and corrections of these data, it is possible to associate to the lithological variations and their respective surface cover (Minty, 1997; Wilford et al., 1997). The characteristic of covering large areas in a short time interval, with high spatial resolution and sampling uniformity, motivates the incorporation of these data into the routine of geological and soil mapping (e.g., Bedini and Rasmussen, 2018; Becegato et al., 2019; Grochowski et al., 2019;

Loiseau et al., 2020; Mello et al., 2021; Hassan et al., 2022; Mamouch et al., 2022; and Mello et al., 2022).

Due to this wide applicability, several studies use different techniques for the interpretation of these data, trying to reach more reliable products, whether using a qualitative or quantitative approaches. In general, qualitative interpretations are based on the visual analysis of the radioelement concentration maps integrated with thematic maps (e.g., geological maps: Ribeiro et al., 2014; Grochowski et al., 2019) or are based on knowledge-driven classification (e.g., Weihermann et al., 2019), so these approaches mostly depend on the analyst's experience. Whereas the quantitative approach uses geoprocessing and remote sensing techniques to reduce the subjectivity and optimize the interpretation (e.g., Eberle et al., 2013; Lima and Marfurt, 2018; Weihermann et al., 2021; and Xu et al., 2022).

Clustering is a spectral pattern recognition technique used to group similar classes in a multidimensional dataset. No training data is needed to perform unsupervised clustering and the interpreter's responsibility

* Corresponding author at: Federal Rural University of Rio de Janeiro (UFRRJ), Geoscience Institute, Postgraduate Program in Modeling and Geological Evolution, Seropédica, Rio de Janeiro, Brazil.

E-mail addresses: blendabastos@ufrrj.br (B.P. Bastos), koenow@ufrrj.br (H.S.K. Pinheiro), waldir.carvalho@embrapa.br (W.C. Junior), lanjos@ufrrj.br (L.H.C. dos Anjos), francisco.ferreira@ufpr.br (F.J.F. Ferreira).

<https://doi.org/10.1016/j.jappgeo.2022.104900>

Received 29 September 2022; Received in revised form 22 November 2022; Accepted 2 December 2022

Available online 5 December 2022

0926-9851/© 2022 Published by Elsevier B.V.

consists in specifying the input parameters and in the post-classification analysis, guaranteeing the meaning of the created classes (Kassambara, 2017; Ghosa et al., 2020).

Considering this, the objective of this work was to test two different types of data-driven clustering (K-means and Gaussian Mixture Model) and knowledge-driven classification (Simplified RGB) to compare their performance in defining airborne gamma-ray spectrometry domains. To reach this proposal, we use pre-processing techniques to choose the best number of clusters and post-processing techniques to evaluate the accuracy of each method.

2. Site description

According to the Brazilian Territorial Division from Brazilian Institute of Geography and Statistics (IBGE, 2021), the Nova Friburgo County is in the Centro Fluminense mesoregion of the State of Rio de Janeiro (Fig. 1), between valleys and escarpments of Serra do Mar, with an area of 935,429 km². The climate is Cwb (subtropical highland climate) according to Köppen classification and has an average temperature of 18.1 °C with a large annual variation of rainfall, defined in two distinct periods: a wet one, from November to March, and a dry one, from April to October (Marafon, 2019).

The region is inserted in the tectonic context and magmatic evolution of the Oriental Terrane from Ribeira Belt. This process includes pre- to late-collisional deformed plutonic rocks, sedimentary covers included in the collision process, generating paragneiss and metasediments of high metamorphic grade, and post-collisional non deformed granitic bodies. The tectonic evolution starts with the installation of the magmatic arc of the Rio Negro Complex and São Fidelis Group sediments (late Cryogenian). Between 590 and 570 Ma, collision apex, occurred the formation of the Cordeiro Suite. Posteriorly, late-collisional tectonic generated the calc-alkaline magmatism of the Serra dos Órgãos Suite.

And finally, tectonic exhumation and tectonic collapse generated the intrusion of the post-collisional granites of the Nova Friburgo Suite during the Cambrian-Ordovician (Tupinamba et al., 2012). Fig. 2 exemplifies the lithologies mapped in the study area according to the Geological Map of the State of Rio de Janeiro (CPRM, 2016a) and its report (CPRM, 2016b).

3. Materials

3.1. Airborne gamma-ray spectrometry data

The airborne gamma-ray spectrometric data were obtained by Geological Survey of Brazil – CPRM (Rio de Janeiro Aerogeophysical Project). This acquisition was performed from perpendicular flight lines with a flight height average of 100 m, constituting NS flight lines spaced at 500 m and EW control lines spaced at 10000 m. The differences in the crossover points between the flight lines and the control lines are used to level the survey data (CPRM, 2012). Radiometric data were acquired with RS-500 gamma-ray spectrometer from Radiation Solution Inc., with sodium iodide (NaI) crystals downward-looking and upward-looking detectors. Sample collection was performed at 1 s intervals which corresponds approximately 75 m on the ground (CPRM, 2012).

The data were processed in Oasis Montaj software (Educational software edition v. 9.8) with minimum curvature interpolation method (Briggs, 1974), adopting a spatial resolution of 100 m. The interpolation procedure created regular grids representing K (%) (Fig. 3a), eU (ppm) (Fig. 3b), and eTh (ppm) (Fig. 3c) contents. The RGB-ternary image (Red = K; Green = eTh; Blue = eU) was obtained in Oasis Montaj software by combining these grids (Fig. 3d).

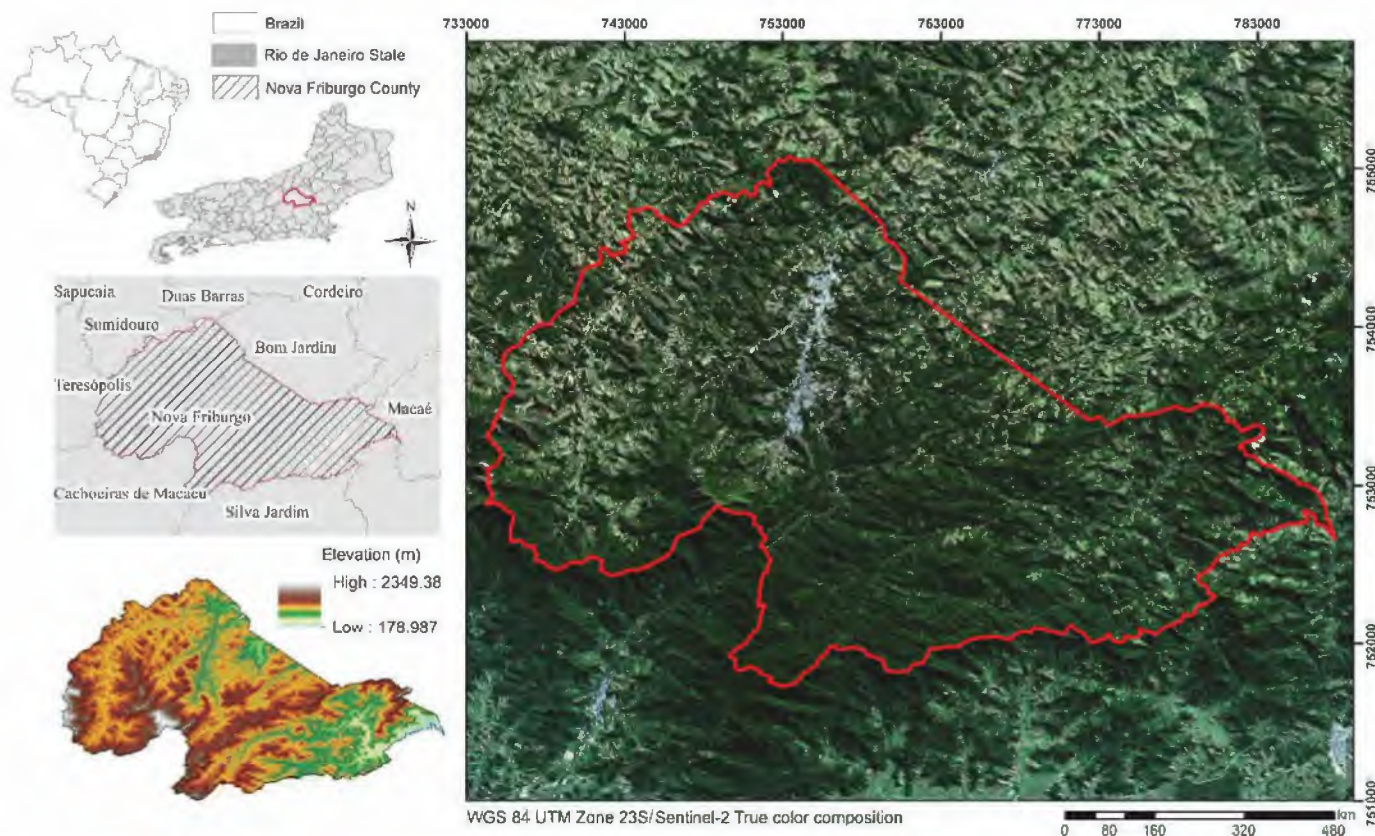


Fig. 1. Location of the study area. On the left below, the Digital Elevation Model (DEM) derived from Rio de Janeiro cartographic database (original scale 1:25,000) (IBGE, 2018).

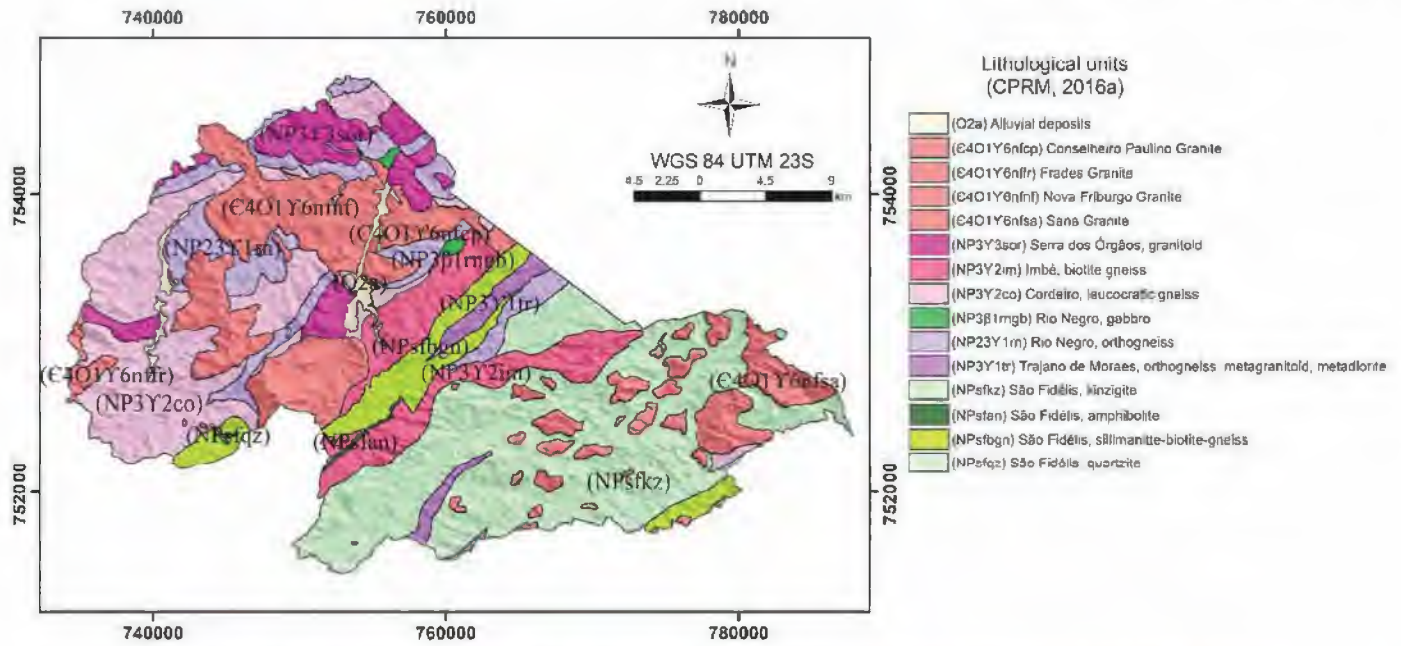


Fig. 2. Geological map of Nova Friburgo – RJ (original scale 1:400.000). Adapted from CPRM - Serviço Geológico do Brasil (2016a).

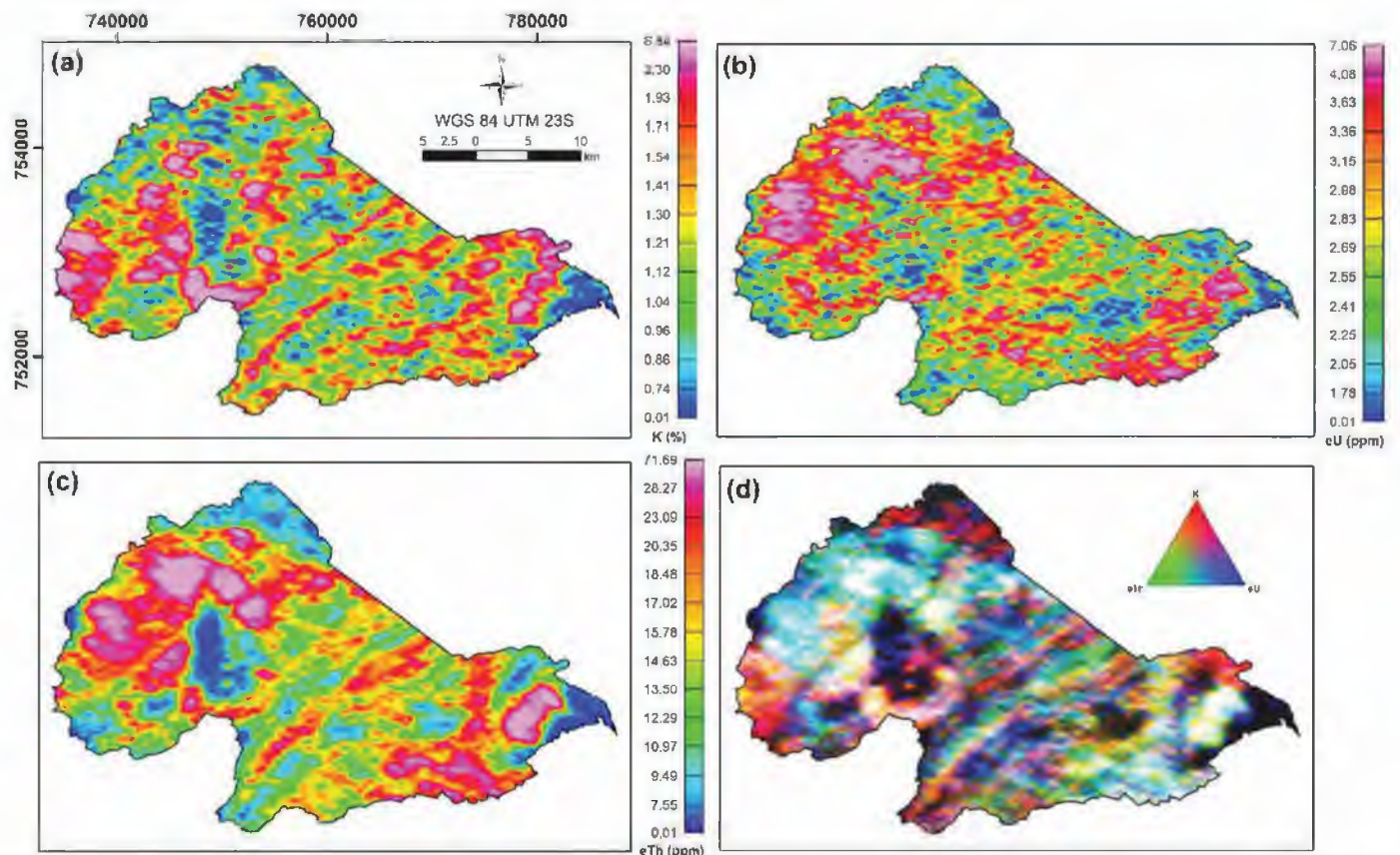


Fig. 3. Airborne gamma-ray spectrometry maps: (a) Potassium – K (%), (b) Uranium – eU (ppm), (c) Thorium – eTh (ppm) and (d) Ternary image – (Red = K; Green = eTh; Blue = eU). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Methods

The methodology can be divided into four steps, summarized at the Fig. 4, comprising:

- I. Unsupervised cluster classification (GMM and K-means) using as input dataset the K, eU, and eTh grids; considering the optimal number of clusters for each method.

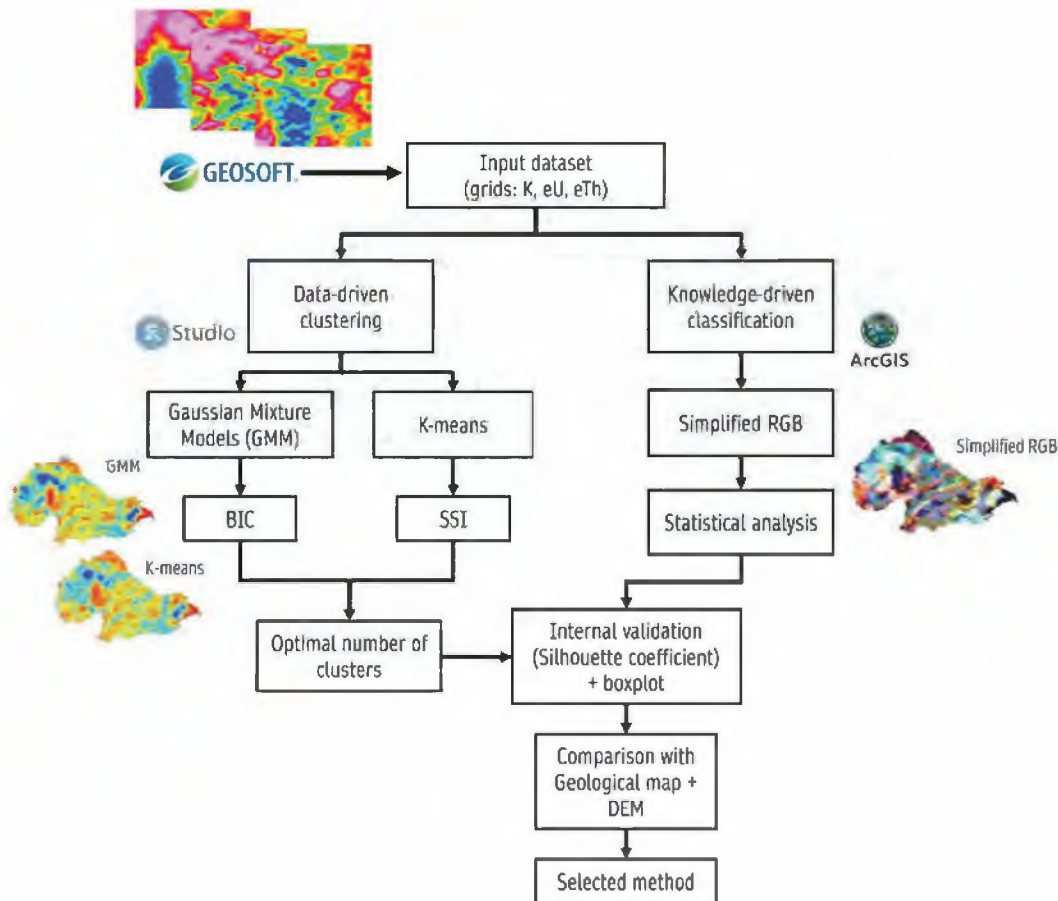


Fig. 4. Flowchart of the methodological procedure.

- II. Simplified RGB classification according to the statistical distribution of K, eU, and eTh concentrations in the study area.
- III. Internal validation considering the silhouette coefficient and a boxplot graphs to evaluate the methods performance.
- IV. Visual comparison of the final maps created by each method with the geological map and the DEM of the area.

All these procedures aided the selection of the best method to represent the gamma-ray domains at the area, and they are described in detail at the subsection below.

4.1. Data-driven clustering: K-means

K-means, proposed by MacQueen et al. (1967), is the most widely used partitional clustering algorithm. This method finds similarity among points (intra-cluster) according to their distances from the cluster-centroid. Two considerations are taking into account to build clusters: (1) there is not any cluster without at least one object/point and (2) each object must be into to at least one cluster (Fahad et al., 2014; Ghosal et al., 2020). The data set is partitioned into K groups, where K represents the number of clusters specified by the analyst, where the center corresponds to the mean of points assigned to the cluster (Kasambara, 2017).

In this work, K-means was performed in R software through the cascadeKM function of the Vegan package (Oksanen et al., 2019). This function allows to test several numbers of clusters, forming a cascade from small to large number of groups formed by K-means. The best partition is evaluated through the Simple Structure Index (SSI). The SSI considers the maximum difference between clusters, the most contrasting clusters sizes, and the variable deviation in the cluster centers

compared to its overall mean. All these parameters are normalized between 0 and 1. Then, the best number of clusters is indicated by the highest SSI value (Dimitriadou et al., 2002). The parameters used include:

- The minimum and maximum numbers of groups to be displayed: 2 to 27;
- The number of random starting configurations for each value of K: 100;
- The criterion that will be used to select the best partition: SSI.

4.2. Data-driven clustering: Gaussian Mixture Models (GMM)

Gaussian Mixture Models (GMM) is a probabilistic clustering approach which, as well as K-means, attempts to minimize variance within the cluster by adjusting cluster-centroids, most commonly, through the means. The difference among the data-driven models is that GMM aims to seek solutions that maximize the likelihood function (for example, components covariances), allowing more flexibility to fit the data (Bishop and Nasrabadi, 2006; Jones et al., 2019). Because of that, the GMM usually works with two algorithms. First, it is common to run the K-means algorithm to find a suitable initialization for the model and, subsequently, to run the expectation-maximization algorithm or EM algorithm (Bishop and Nasrabadi, 2006). More details about the EM algorithm can be seen in Dempster et al. (1977).

Another technique commonly used with the GMM model is the Bayesian Information Criterion (BIC) introduced by Schwarz (1978) (e.g., Bouayad et al., 2021; Weihermann et al., 2021; and Whiteley et al., 2021). BIC can be used as a pre-processing technique. From BIC, comparisons between models can be based on Bayes factors and the optimal

number of clusters can be determined. The optimal number of clusters can be considered as the lowest BIC between the tested classes (Weihermann et al., 2021; Whiteley et al., 2021).

To perform GMM, we use ClusterR package (Mouselimis, 2022) in R software. First, we centered the data by using the *center_scale* function, posteriorly, we performed the optimal number of clusters through *Optimal_Clusters_GMM* function. The parameters used in this stage were:

- Max number of clusters: 27;
- Criterion to choose the best model: Bayesian Information Criterion (BIC).
- Distance used during the seeding of initial means clustering: Euclidean distance.
- How the initial means are seeded prior to running the algorithm: random subset.
- The number of iterations of the K-means and EM algorithms: 100.

After choosing the best model, we applied the GMM and GMM prediction functions. The GMM returns the centroids, covariance matrix, weights, and log-likelihoods for each gaussian component. The GMM prediction returns the cluster probabilities and cluster labels of the data (Mouselimis, 2022). In this work, the GMM function parameters include:

- Number of clusters according to BIC.
- Distance used during the seeding of initial means clustering: Euclidean distance.
- How the initial means are seeded prior to running the algorithm: random subset.
- The number of iterations of the K-means and EM algorithms: 100.

4.3. Knowledge-driven classification: Simplified RGB

According to Medeiros and Rosa-Costa (2020), the interpretation of gamma-ray spectrometric domains must be carried out through a discriminatory assessment between concentrations of each radioelement in three classes of low, medium, and high relative content. The Simplified RGB delimits the classes based on statistical values of the data to simplify RGB composition from $256 \times 256 \times 256$ possible color tones (Fig. 3d) to $3 \times 3 \times 3$ possible color tones, so the concentration values of each radioelement will be divided into: K (high, medium and low), eTh (high, medium and low) and eU (high, medium and low), resulting in a combination of 27 classes (Weihermann et al., 2019). This approach can be considered Knowledge-drive once the human expertise is used to provide classes limits (Todrovski and Džeroski, 2006).

In this study, each grid (K, eTh and eU) has been reclassified according to the Table 1. To achieve this proposal, we use the ArcGIS Desktop software (v. 10.6). The grids were joined according to CPRM classification (Medeiros and Rosa-Costa, 2020).

4.4. Internal validation: Silhouette coefficient

The cluster validation is a post-processing technique, which consists in measuring the clustering performance. As we know, in clustering, the objects of interest are grouped into similar classes. Then, a way to achieve a good result is by maximizing the intra-cluster similarity and minimizing the inter-cluster similarity (Halkidi et al., 2001). Internal

Table 1

Radioelements statistical classification of the study area (X = grid value; M = median; SD = standard deviation).

	Low (3)	Medium (2)	High (1)
Radioelement	$X \leq M - \frac{1}{2}(SD)$	$M - \frac{1}{2}(SD) < X < M + \frac{1}{2}(SD)$	$X \geq M + \frac{1}{2}(SD)$
K (%)	$X \leq 1.00$	$1.00 < X < 1.59$	$X \geq 1.59$
eU (ppm)	$X \leq 2.40$	$2.40 < X < 3.22$	$X \geq 3.22$
eTh (ppm)	$X \leq 11.70$	$11.70 < X < 20.46$	$X \geq 20.46$

validation is used when there is no external information available, so the accuracy is measured by using the same input dataset (Palacio-Niño and Berzal, 2019).

The validation process adopted the Silhouette coefficient (Rousseeuw, 1987) to compare the performance of each grouping method (K-means, GMM, and Simplified RGB). The Silhouette coefficient combines the metrics of cohesion and separation in a value and is defined in the interval [-1; 1]. Positive values indicate high segregation between clusters, and negative values indicate overlapping. When the silhouette coefficient is zero indicates uniformity in data distribution throughout the Euclidean space (Palacio-Niño and Berzal, 2019). The Silhouette coefficient was obtained through clusterCrit package (Desgraupes and Desgraupes, 2018) in R software.

5. Results and discussion

5.1. Optimizing the number of clusters with data-driven models

The optimal number of clusters was evaluated by SSI for K-means and by BIC for GMM. As a result, the best partition for K-means clustering was 22 groups. The SSI value showed an increasing trend until 21 number of clusters, reached the maximum in 22, followed by a decrease from 22 to 23 groups, as observed at the Fig. 5a. On the other hand, the BIC value for GMM pointed out no significant improvement in the range of 10 to 27 groups (Fig. 5b).

According to Weakliem (1999), in large data, Bayesian estimation often makes little difference between the models, making the choice of the best distribution difficult. Similar behavior in BIC values was presented by Weihermann et al. (2021) also using airborne gamma-ray spectrometry data and, through an analysis of the statistical separability of classes, it was possible to choose the best number of clusters. Addressing this issue and to test a different approach, we decided to add another step for choosing the best number of clusters for the GMM model. Using the set of models that, according to Fig. 5b, present similar values (10 to 27 groups), we perform the Silhouette coefficient. According to Kassambara (2017), the Silhouette Index is also a useful approach to determine the optimal number of clusters. The BIC and Silhouette values were presented in the Fig. 6.

The model with the best performance pointed by the BIC (the lowest value with 24 clusters) does not show a significant difference between the models with 21 to 27 clusters. Coincidentally, the Silhouette coefficient values also showed no significant difference. The same occurs for the models with 15, 16, 18, 19, and 20 clusters. The model with 17 cluster has the highest BIC value compared to models with 15 to 27 clusters, also presented the worst Silhouette value.

The Silhouette coefficient showed that the best partitions are represented by the simplest models with the lowest number of clusters (10 to 14). And despite the model with 10 clusters having the worst performance regarding the BIC, it showed the best performance according to the Silhouette. Aiming to simplify the interpretation of the classes and based on the performance of Silhouette the simplest model, with 10 clusters, was chosen to represent GMM classification.

5.2. Comparing data-driven and knowledge-driven models

As a result of the steps described above, the following maps were obtained: Simplified RGB with 26 classes (Fig. 7a), GMM with 10 classes (Fig. 7b) and K-means with 22 classes (Fig. 7c). These figures show, in general, visual similarity between methods. The Simplified RGB (Fig. 7a) as well as the ternary image (Fig. 3d) indicates high values with white color (in the legend: 333) and low values with black color (in the legend: 111) in all three radionuclides. To facilitate the interpretation, the models GMM10 and K-means22 were also organized in a similar way, in other words, with the clusters organized in an ascending way.

All methods highlighted the high values represented by Nova Friburgo Suite. These rocks are represented by undeformed post-collisional

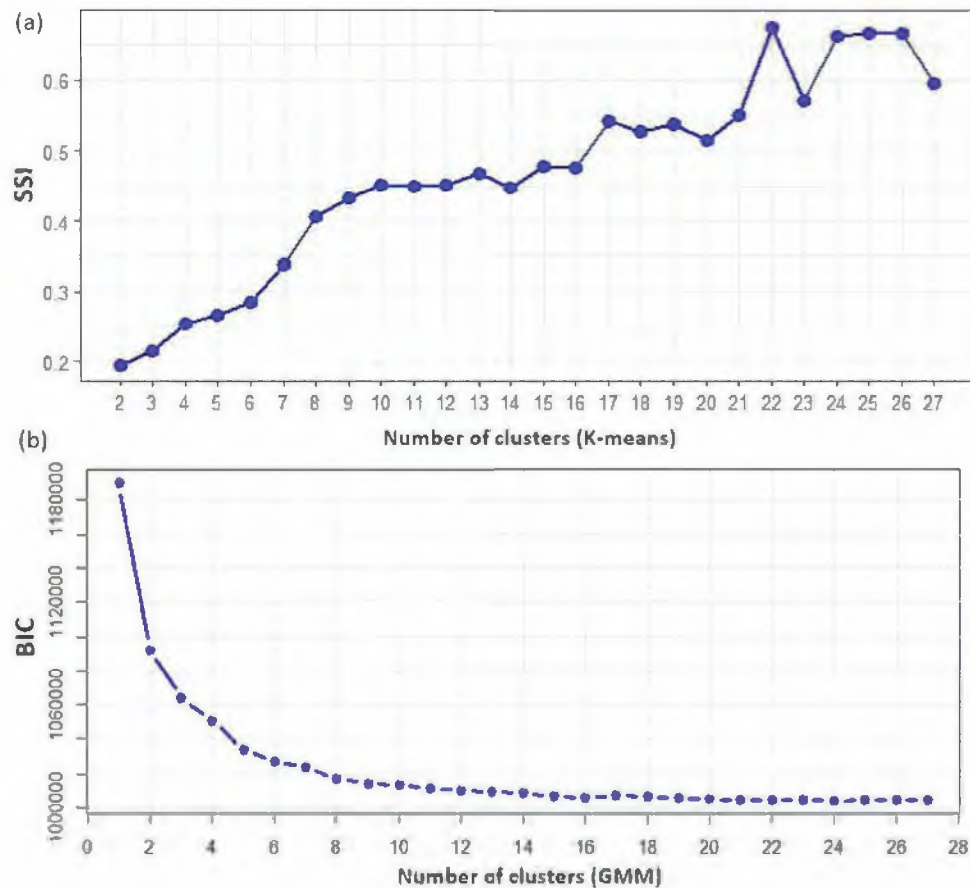


Fig. 5. (a) Simple Structure Index (SSI) computed for 1 to 27 classes for K-means; (b) Bayesian Information Criterion (BIC) computed for 1 to 27 classes for GMM.



Fig. 6. Silhouette coefficient and Bayesian information criterion (BIC) comparison for eighteen GMM possible models (10 to 27).

magmatism with high-K calcalkaline to alkali-calcic predominantly metaluminous granites and quartz monzonites (Valeriano et al., 2016). According to the same authors, based on geochemical studies, the collisional magmatism of the central portion of the Ribeira Belt is marked by a wide distribution of SiO₂ (about 71%). As cited by Ulbrich et al. (2009), granitogenesis is marked by an important enrichment of K, eU and eTh and, as mentioned by Dickson and Scott (1997), the concentrations of these radioelements are proportional to the increase in

silica. This corroborates the highlighted correlation of these granites outcrops with class 333 from Simplified RGB (Fig. 7a), cluster 10 from GMM10 (Fig. 7b) and cluster 22 from K-means22 (Fig. 7c).

To better understand the correlation among the models, the Fig. 8 presents a detailed image of the Nova Friburgo Granite (upper portion), that shows the GMM representing this feature with smaller number of clusters, simplifying the interpretation.

It is interesting to observe that the high values of the methods are

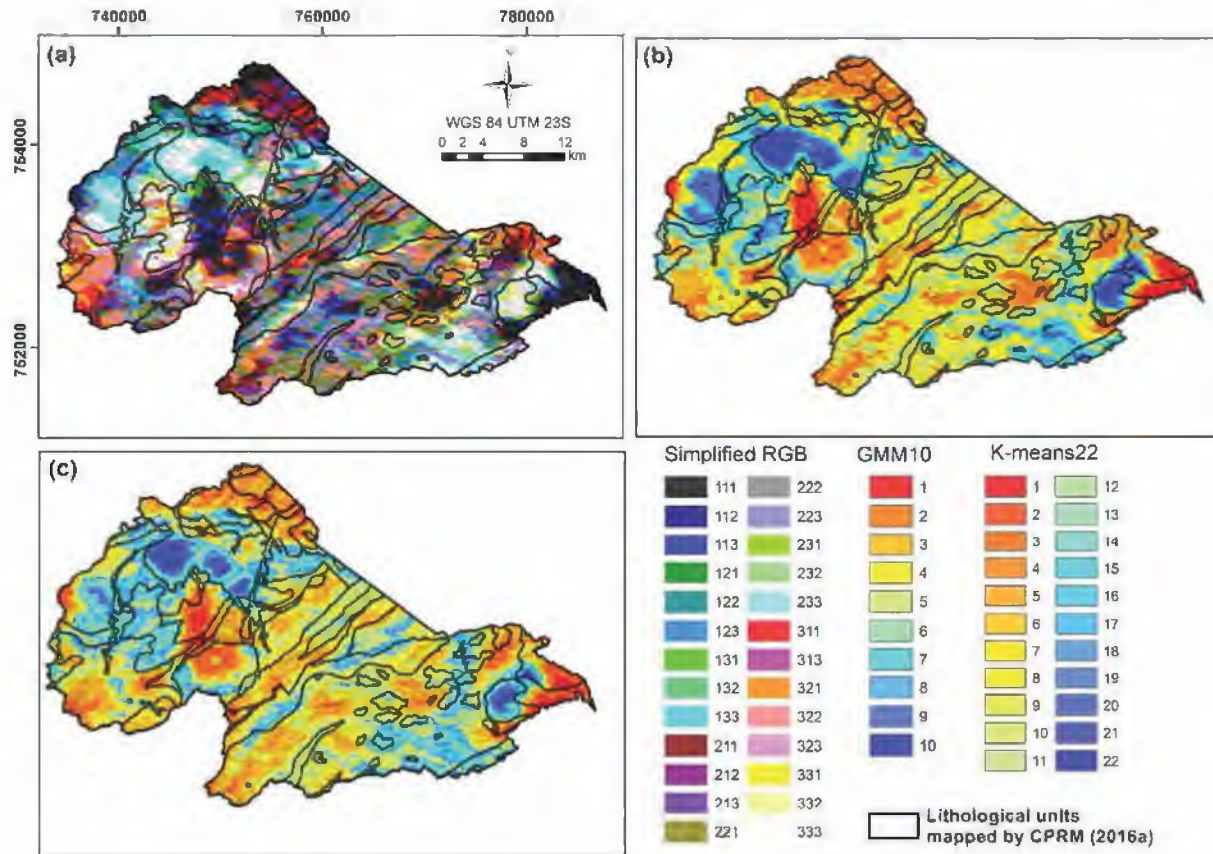


Fig. 7. Results maps from each method: (a) Simplified RGB with 26 classes; (b) GMM with 10 classes; and (c) K-means with 22 classes.

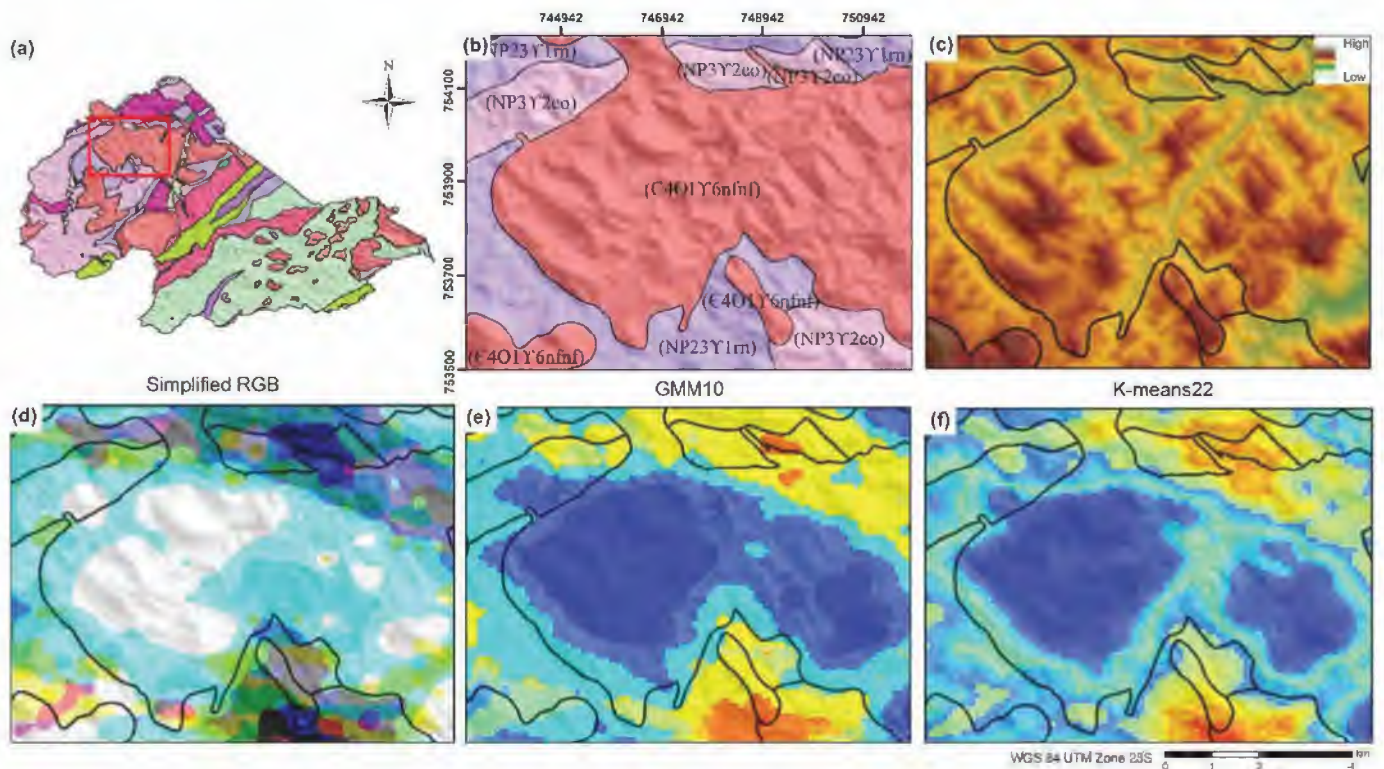


Fig. 8. (a) and (b) Geological map with the location zoom area marked in red; (c) zoom area in DEM; (d) zoom area in Simplified RGB; (e) zoom area in GMM10; (f) zoom area in K-means22. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

correlated with the high topography as we can also see in Fig. 9. The highlighted area presented at the Fig. 9 represents the lower portion of the Nova Friburgo Granite occurrence area. All methods (Fig. 9d–f) highlighted granite outcrop area that has higher altitude according to DEM (Fig. 9c).

Another example is presented at the Fig. 10 representing the Sana Granite in contrast to an area of low altitude according to DEM (Fig. 10a–c). The Fig. 10 also presents visual similarity between methods (Fig. 10d–f), showing the efficiency of three methods to highlight the high values for all radionuclides matching the region with the higher altitude and the low values represented by the region with the lower altitude.

This means that the spatial variation of radioelement concentrations according to topography was well represented by the three methods. According to Wilford et al. (1997), weathering modifies the distribution and concentration of radioelements. In areas with steep topography, the erosion rate exceeds regolith formation, forming thinner soils or completely exposing the rock (outcrops). In areas with more stable and smoothed relief, there is the accumulation of material and its consequent evolution, differing geochemically from the underlying rock. In airborne surveys, these factors interfere with the intensity of the gamma-ray response, where higher values will be observed coinciding with rock outcrops and lower values where the accumulation of sediments and soil evolution prevails (Wilford et al., 1997).

The visual comparison with Geology and DEM examples showed good correspondence with the three classification methods. However, according to the values obtained in the internal validation, the best method was K-means22 with a Silhouette coefficient of 0.33, followed by GMM10 with 0.14 and, Simplified RGB with -0.16. This means that K-means and GMM10 have a good segregation between clusters (positive values), while the negative value of Simplified RGB shows overlapping between classes (Palacio-Niño and Berzal, 2019). A good way to observe this behavior is through the analysis of the boxplot diagram that summarizes the cluster/class statistics of each method. The boxplot

diagram consists of the minimum and maximum range values, the upper and lower quartiles, the median, and the outliers of the dataset distribution. From the Simplified RGB boxlots (Fig. 11a–c) we can observed the classes overlapping for the three radioelements.

In general, the distribution of the three elements per class seems to follow a “ladder standard”, forming groups with almost the same range. For potassium boxplot Fig. 11a, the classes 111, 112, 113, 121, 122, 123, 131, 132, 133 are distributed in practically the same range. As well as the classes 211, 212, 213, 221, 222, 223, 231, 232, 233 and 311, 313, 321, 322, 323, 331, 332, 333 forming in total 3 big groups with almost no significant difference between them. In addition, it presents outliers mainly in the last group (classes 311 to 333). In uranium boxplot (Fig. 11b), the classes 111, 121, 131, 211, 221, 231, 311, 321, 331, for example, have very close median values. The same behavior occurs in thorium boxplot (Fig. 11c) with the classes 111, 112, 113, 211, 212, 213, 311, 313, for example.

In GMM10 boxplots (Fig. 11d–f) we can observe a better division per clusters. Mainly of the thorium element (Fig. 11f) that, despite the outliers, no cluster median values overlaps. In potassium boxplot (Fig. 11d), the cluster 2 and cluster 4 have practically the same distribution and, in the uranium boxplot (Fig. 11e) showed similarity between clusters 2 and 3. On the other hand, the good result in Silhouette coefficient do not reflects in boxplots analysis for K-means model (Fig. 11g–i). The only element whose distribution is well divided by cluster is thorium (Fig. 11i), with considerably better performance when compared to potassium and uranium boxplots (Fig. 11g and h) in which the clusters are poorly divided. According to Palacio-Niño and Berzal (2019), cohesion and separation metrics (like Silhouette coefficient) perform well for partitional clustering techniques as K-means, therefore the result can be misinterpreted. As we can see GMM10 presents better performance than K-means22 for all the three elements.

Even with the data normalization in the step before the unsupervised clustering, thorium seems to be the “clusters definer”, represented by a better division of the data by cluster, showing less overlaps, for both

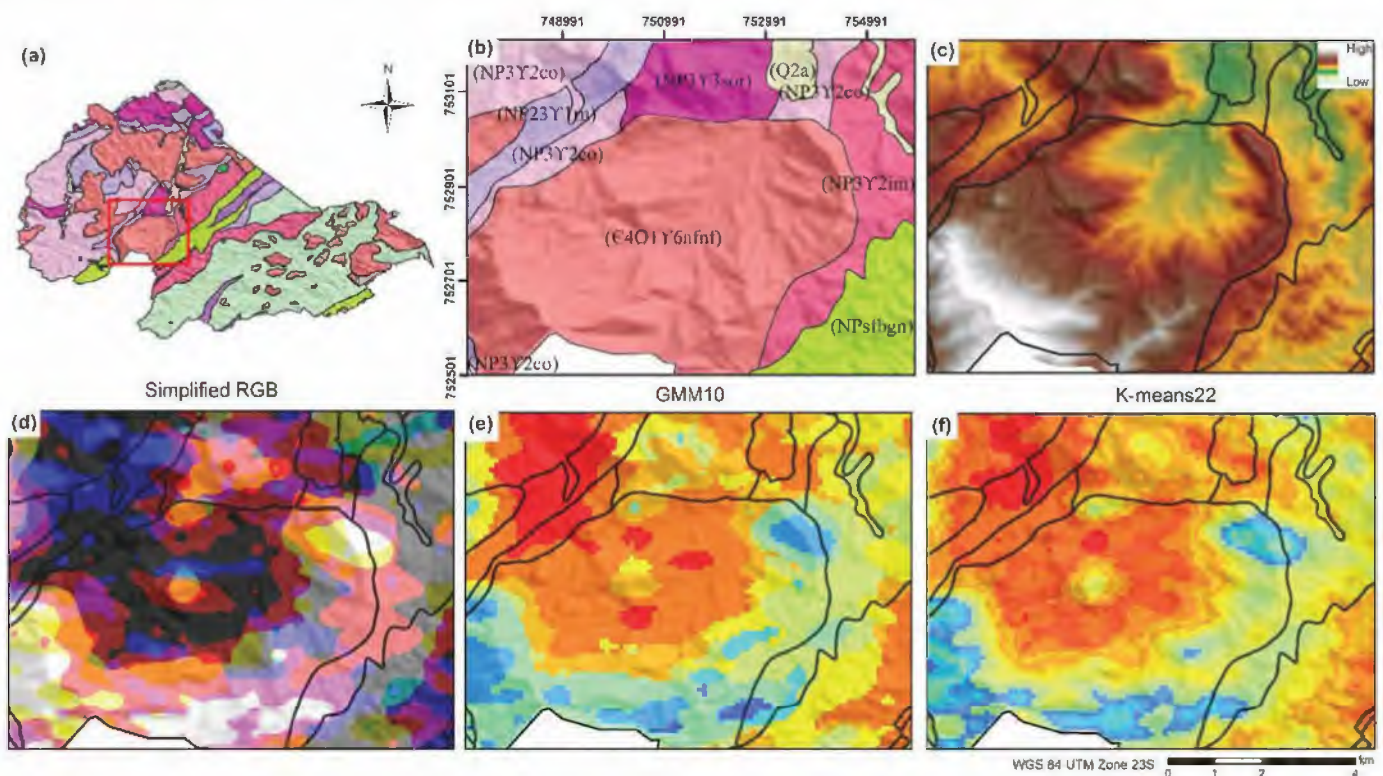


Fig. 9. (a) and (b) Geological map with the location zoom area marked in red; (c) zoom area in DEM; (d) zoom area in Simplified RGB; (e) zoom area in GMM10; (f) zoom area in K-means22. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

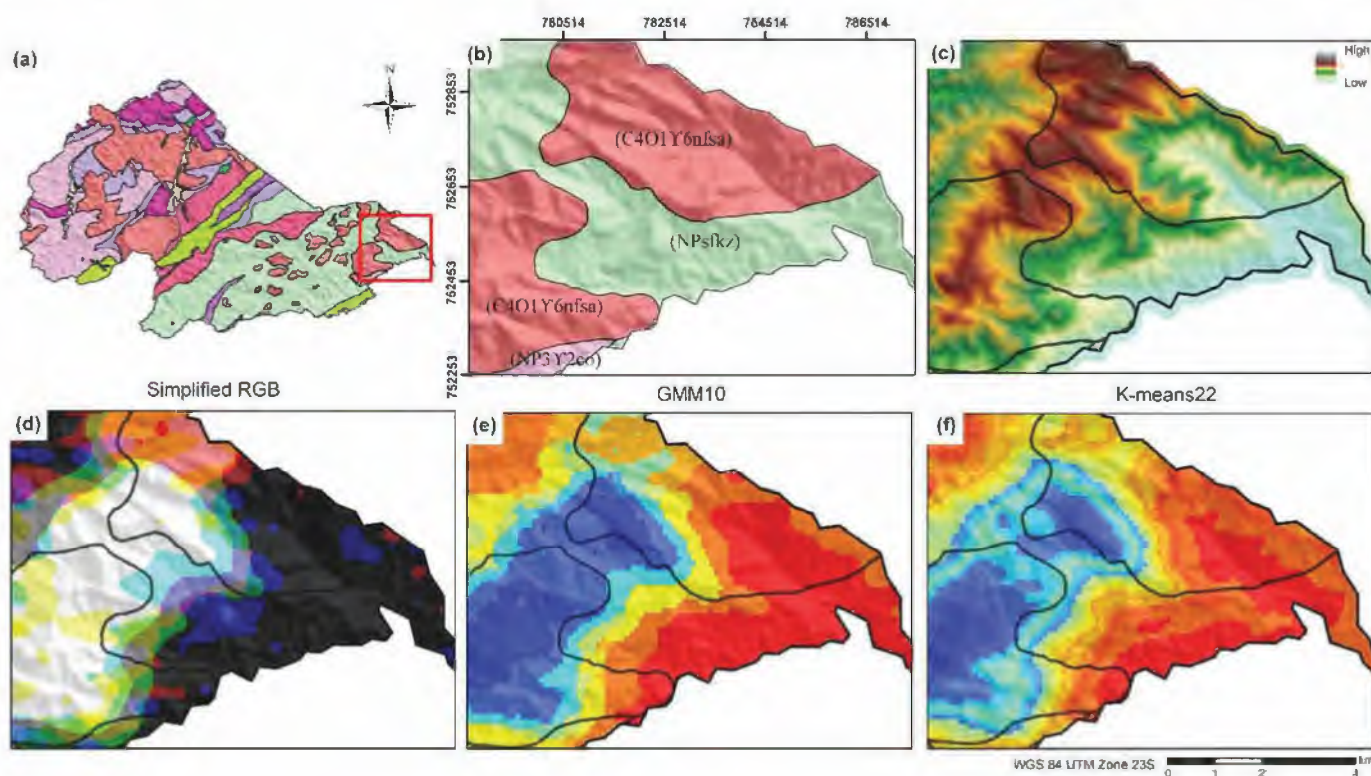


Fig. 10. (a) and (b) Geological map with the location zoom area marked in red; (c) zoom area in DEM; (d) zoom area in Simplified RGB; (e) zoom area in GMM10; (f) zoom area in K-means22. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

data-driven methods (GMM and K-means). This behavior may be linked to the natural characteristics of these radioelements. Among the three elements, thorium is the least soluble, therefore it is the most resistant to processes such as weathering (Ulbrich et al., 2009), and consequently, its spatial distribution is more restricted. Potassium is extremely mobile in the hydrothermal alteration process and during weathering, especially in tropical and subtropical climates (Ulbrich et al., 2009). The surface concentration of uranium is usually associated with resistant minerals, such as thorium. Uranium can be mobile in the oxidized form or in hydrothermal alteration environments (Ulbrich et al., 2009; Dickson and Scott, 1997). However, probability, the more erratic variation of uranium may be related to the characteristics of the survey itself, as pointed out by Minty (1997), which highlights that in airborne survey, disequilibrium in the uranium decay series is a serious source of error. Furthermore, the Radon element (Rn) and its derived products, are the major contributors to the background since are detected in the same energy window as equivalent uranium (1.66–1.86 MeV) (Minty, 1997), what can explain this erratic distribution.

6. Conclusions

The BIC criterion presented the worst performance to define the number of clusters for the GMM. Addressing this issue, further analysis testing several models through Silhouette coefficient was applied, pointing out the importance of a detailed analysis to define the optimal number of clusters in unsupervised classification.

The visual interpretation showed that, regardless of the number of clusters, in regional scale, the three methods performed well to explain the spatial variation of the radioelements. Demonstrating the importance of the dynamics of these elements in relation to topography. Therefore, weathering processes and landscape forms are relevant, and should be considered in the data interpretation, especially in tropical climates, which affects the leaching and/or concentration of these radioelements.

The Silhouette coefficient, in general, was important to define the best cluster number for the GMM and to differentiate the performance between Knowledge-driven and Data-driven classification. However, it must be used with caution. The negative result found to SSI, and confirmed by boxplot analysis, for the Knowledge-driven classification (Simplified RGB) showed overlapping among several classes. The best result found for SSI by using k-means22 classification did not agree with the arrangement of the concentrations of the three radioelements in the boxplot analysis. The result was influenced only by the thorium, which could lead to misinterpretations, if the data distribution had not been analyzed.

The best model was the GMM with 10 clusters, once its performance was considered satisfactory in the visual interpretation with the geological map and the DEM, as well as the intra-cluster cohesion confirmed by the Silhouette coefficient and boxplot analysis. Finally, we conclude that for a regional qualitative interpretation all three classification methods can be used. However, for a quantitative analysis, the simplest model of unsupervised clustering prevailed.

CRedit authorship contribution statement

Blenda Pereira Bastos: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Helena Saraiva Koenow Pinheiro:** Methodology, Supervision, Resources, Funding acquisition, Writing – original draft, Writing – review & editing. **Waldir Carvalho Junior:** Methodology, Writing – review & editing. **Lúcia Helena Cunha dos Anjos:** Resources, Funding acquisition, Writing – review & editing. **Francisco José Fonseca Ferreira:** Methodology, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

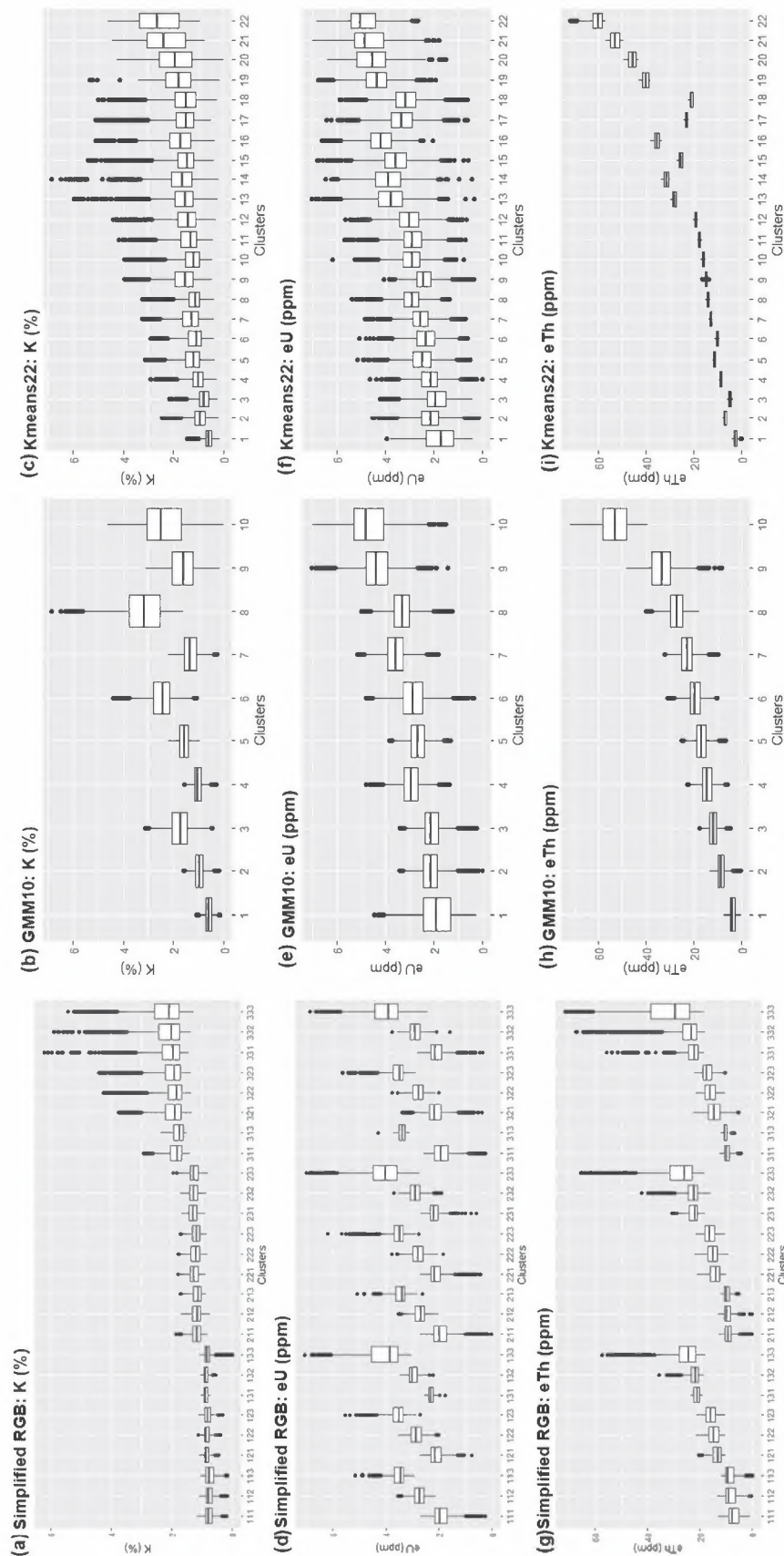


Fig. 11. Simplified RGB, GMM10, and K-means22 boxplots of the primary variables: (a), (b) and (c) Potassium – K (%) for Simplified RGB, GMM10 and K-means22 respectively; (d), (e) and (f) Uranium – eU (ppm) for Simplified RGB, GMM10 and K-means22 respectively; (g), (h) and (i) Thorium – eTh (ppm) for Simplified RGB, GMM10 and K-means22 respectively.

the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research was supported by Federal Rural University of Rio de Janeiro (UFRRJ, Brazil), Postgraduate Program in Modeling and Geological Evolution (PPGMEG, Brazil), Federal University of Paraná (UFPR, Brazil), Embrapa Soils (Brazil), Coordination of Improvement of Higher Level Personnel (CAPES, Brazil), Scientific and Technological Development (CNPq, Brazil), and Research Support Foundation of the State of Rio de Janeiro (FAPERJ, Brazil), providing infrastructure, software licenses, and funding.

References

- Becegado, V.A., Becegado, V.R., Baum, C.A., Lavnitcki, L., Paulino, A.T., 2019. Multivariate statistical analysis correlating 238 U, 232 Th, and 40 K equivalent activities in soil to geochemical data from an agricultural area. *J. Soils Sediments* 19 (4), 1901–1910.
- Bedini, E., Rasmussen, T.M., 2018. Use of airborne hyperspectral and gamma-ray spectroscopy data for mineral exploration at the Sarfartoq carbonate complex, southern West Greenland. *Geosci. J.* 22 (4), 641–651.
- Bishop, C.M., Nasrabadi, N.M., 2006. *Pattern Recognition and Machine Learning*, 4. Springer, New York, p. 738.
- Bouayad, D., Baroth, J., Dano, C., 2021. Gaussian mixture model based soil classification using multiple cone penetration tests. In: *IOP Conference Series: Earth and Environmental Science*, vol. 696. IOP Publishing, p. 012034, 1.
- Briggs, I.C., 1974. Machine contouring using minimum curvature. *Geophysics* 39 (1), 39–48.
- CPRM - Serviço Geológico do Brasil, 2012. Projeto Aerogeofísico Rio de Janeiro (Projeto 1.117). Relatório final do levantamento processamento dos dados magnetométricos e gamaespectrométricos. In: *Prospeções Aerolevantaamentos e Sistemas Ltda*, Vol. 1 (219p).
- CPRM - Serviço Geológico do Brasil, 2016a. Mapa Geológico e de recursos minerais do Estado do Rio de Janeiro. Belo Horizonte. Escala 1:400.000. Programa geológica do Brasil. Mapas geológicos estaduais. <https://geosgbc.cprm.gov.br/> (accessed in 01 April 2022).
- CPRM - Serviço Geológico do Brasil, 2016b. Geologia e Recursos Minerais do Estado do Rio de Janeiro: texto explicativo do mapa geológico e de recursos minerais. Belo Horizonte, MG (182p).
- Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B Methodol.* 39 (1), 1–22.
- Desgraupes, B., Desgraupes, M.B., 2018. Package 'clusterCrit'.
- Dickson, B.L., Scott, K.M., 1997. Interpretation of aerial gamma-ray surveys-adding the geochemical factors. *AGSO J. Aust. Geol. Geophys.* 17, 187–200.
- Dimitriadou, E., Dolničar, S., Weingessel, A., 2002. An examination of indexes for determining the number of clusters in binary data sets. *Psychometrika* 67 (1), 137–159.
- Eberle, D.G., Hutchins, D., Titus, N., 2013. Unsupervised soft clustering of high resolution airborne geophysical and satellite data suites from the Sperrgebiet, Karas region, Southern Namibia, to enhance lithology mapping. *ASEG Extended Abstracts*. 2013 (1), 1–5.
- Falad, A., Alshatri, N., Tari, Z., Alamri, A., Khalil, I., Zomaya, A.Y., et al., 2014. A survey of clustering algorithms for big data: taxonomy and empirical analysis. *IEEE Trans. Emerg. Top. Comput.* 2, 267–279.
- Ghosal, A., Nandy, A., Das, A.K., Goswami, S., Panday, M., 2020. A short review on different clustering techniques and their applications. *Emerging technology In modelling and graphics*. 69–83.
- Grochowski, J., Kuchenbecker, M., Barbuena, D., Novo, T.A., 2019. Integrating geological and airborne geophysical data to review the cartography of Rio Itanguá Batholith, Araçuaí Orogen, Brazil. *Brazilian J. Geol.* 49.
- Halkidi, M., Batistakis, Y., Vazirgiannis, M., 2001. On clustering validation techniques. *J. Intell. Inf. Syst.* 17 (2), 107–145.
- Hassan, S.M., Youssef, M.A., Gabr, S.S., Sadek, M.F., 2022. Radioactive mineralization detection using remote sensing and airborne gamma-ray spectrometry at Wadi Al-Miyah area, Central Eastern Desert, Egypt. *The Egyptian Journal of Remote Sensing and Space Science*. 25 (1), 37–53.
- IBGE - Instituto Brasileiro de Geografia e Estatística, 2018. Base cartográfica vetorial contínua do Estado do Rio de Janeiro na escala 1:25.000. Diretoria de Geociências. Departamento de Cartografia. Projeto RJ-25. ftp://geofp.ibge.gov.br/cartas_e_mapas/bases_cartograficas_continuas/bc25/rj/ (accessed in 01 April 2022).
- IBGE - Instituto Brasileiro de Geografia e Estatística, 2021. Divisão Territorial Brasileira. <https://www.ibge.gov.br> (accessed in 01 April 2022).
- Jones, D.C., Holt, H.J., Meijers, A.J., Shuckburgh, E., 2019. Unsupervised clustering of Southern Ocean Argo float temperature profiles. *Journal of Geophysical Research: Oceans* 124 (1), 390–402.
- Kassambara, A., 2017. *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning*, vol. 1. Sthda.
- Lima, R.P., Marfurt, K.J., 2018. Principal component analysis and K-means analysis of airborne gamma-ray spectrometry surveys. In: *SEG Technical Program Expanded Abstracts 2018*. Society of Exploration Geophysicists, pp. 2277–2281.
- Loiseau, T., Richer-de-forges, A.C., Martelet, G., Bialkowski, A., Nehlig, P., Arrouays, D., 2020. Could airborne gamma-spectrometric data replace lithological maps as covariates for digital soil mapping of topsoil particle-size distribution? A case study in Western France. *Geoderma Regional*. 22, e00295.
- MacQueen, J., et al., 1967. Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 281–297.
- Mamouch, Y., Attou, A., Miffah, A., Onchchen, M., Dadi, B., Achkouch, L., et al., 2022. Mapping of Hydrothermal Alteration zones in the Kelâat M'Gouna Region using Airborne Gamma-Ray Spectrometry and Remote Sensing Data: Mining Implications (Eastern Anti-Atlas, Morocco). *Appl. Sci.* 12 (3), 957.
- Marafon, G.J., 2019. Atlas Geográfico do Município de Nova Friburgo. Editora da Universidade do Estado do Rio de Janeiro, Rio de Janeiro (83p).
- Medeiros, V.C., Rosa-Costa, L.T., 2020. Guia de procedimentos técnicos: volume 1 - cartografia geológica. CPRM, Brasília (170 p).
- Mello, D.C., Demattê, J.A.M., De Oliveira Mello, F.A., Poppiel, R.R., Silvero, N.E., Safanelli, J.L., et al., 2021. Applied gamma-ray spectrometry for evaluating tropical soil processes and attributes. *Geoderma*. 381, 114736.
- Mello, D.C.D., Veloso, G.V., Lana, M.G.D., Mello, F.A.D.O., Poppiel, R.R., Cabrero, D.R. O., et al., 2022. New methodological framework by geophysical sensors combinations associated with machine learning algorithms to understand soil attributes. *Geosci. Model Dev.* 15 (3), 1219–1246.
- Minty, B.R.S., 1997. Fundamentals of airborne gamma-ray spectrometry. *AGSO J. Aust. Geol. Geophys.* 17, 39–50.
- Mouselimis, L., 2022. ClusterR: Gaussian Mixture Models, K-Means, Mini-Batch-Kmeans, K-Medoids and Affinity Propagation Clustering. R package version 1.2.6. <https://CRAN.R-project.org/package=ClusterR>.
- Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., Minchin, P.R., O'hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szöcs, E., Wagner, H., 2019. Package 'vegan' v. 2.5-6: Community Ecology Package. Available at: <https://CRAN.R-project.org/package=vegan>.
- Palacio-Niño, J.O., Berzal, F., 2019. Evaluation metrics for unsupervised learning algorithms. *arXiv preprint 2, 1–9* (arXiv:1905.05667).
- Ribeiro, V.B., Mantovani, M.S.M., Louro, V.H.A., 2014. Airborne gamma-ray spectrometry and applications in geological mapping (In Portuguese). *Terra Didática*. 10 (1), 29–51.
- Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20, 53–65.
- Schwarz, G., 1978. Estimating the dimension of a model. *Ann. Stat.* 461–464.
- Todorovski, L., Dzeroski, S., 2006. Integrating knowledge-driven and data-driven approaches in modeling. *Ecol. Model.* 194 (1–3), 3–13.
- Tupinambá, M., Teixeira, W., Heilbron, M., 2012. Evolução tectônica e magmática da faixa ribeira entre o neoproterozoico e o paleozoico inferior na região serrana do estado do Rio de Janeiro, Brasil. *Anuário do Instituto de Geociências*. 35 (2), 140–151.
- Ulbrich, H.H.G.J., Ulbrich, M.N.C., Ferreira, F.J.F., Alves, L.S., Guimarães, G.B., Fruchting, A., 2009. Gamma-ray spectrometry surveys in differentiated granites. I: review of the methodology and geochemical behavior of the elements K, Th and U (In Portuguese). *Revista do Instituto de Geociências - USP, São Paulo* 9 (1), 33–53.
- Valeriano, C.M., Mendes, J.C., Tupinambá, M., Bongioio, E., Heilbron, M., Junho, M.D.C. B., 2016. Cambro-Ordovician post-collisional granites of the Ribeira belt, SE-Brazil: a case of terminal magmatism of a hot orogen. *J. S. Am. Earth Sci.* 68, 269–281.
- Weakliem, D.L., 1999. A critique of the Bayesian information criterion for model selection. *Sociol. Methods Res.* 27 (3), 359–397.
- Weiermann, J.D., Castro, F.R., de Castro, L.G., Ferreira, F.J.F., Oliveira, S.P., Silva, A. M., 2019. Statistical analysis for lithogeophysical maps using airborne gamma-ray spectrometry. In: *Sixteenth International Congress of the Brazilian Geophysical Society*.
- Weiermann, J.D., Ferreira, M.P., de Castro, L.G., Ferreira, F.J.F., Silva, A.M., 2021. Retrieving geological units with unsupervised clustering of gamma-ray spectrometry data. *J. Appl. Geophys.* 184, 104225.
- Whiteley, J.S., Wadet, A., Uhlemann, S., Wilkinson, P., Boyd, J.P., Jordan, C., Chambers, J.E., 2021. Rapid characterisation of landslide heterogeneity using unsupervised classification of electrical resistivity and seismic refraction surveys. *Eng. Geol.* 290, 106189.
- Wilford, J.R., Bierwirth, P.E., Craig, M.A., 1997. Application of airborne gamma-ray spectrometry in soil/regolith mapping and applied geomorphology. *AGSO J. Aust. Geol. Geophys.* 17 (2), 201–216.
- Xu, S., Zhang, G., Dong, G., Sun, W., Wei, D., Li, H., et al., 2022. Radiological, geochemical, and environmental assessment in Xiucheng, China: the airborne gamma-ray spectrometric view. *J. Geochem. Explor.* 106980.