



## Article

# Could Airborne Geophysical Data Be Used to Improve Predictive Modeling of Agronomic Soil Properties in Tropical Hillslope Area?

Blenda P. Bastos <sup>1,\*</sup>, Helena S. K. Pinheiro <sup>2</sup>, Francisco J. F. Ferreira <sup>3</sup>, Waldir de Carvalho Junior <sup>4</sup>  
and Lúcia Helena C. dos Anjos <sup>2</sup>

- <sup>1</sup> Postgraduate Program in Modeling and Geological Evolution, Geoscience Institute, Federal Rural University of Rio de Janeiro (UFRRJ), Seropédica 23890-000, RJ, Brazil
- <sup>2</sup> Soils Department, Agronomy Institute, Federal Rural University of Rio de Janeiro (UFRRJ), Seropédica 23890-000, RJ, Brazil; lenask@gmail.com (H.S.K.P.); lanjosh@ufrrj.br (L.H.C.d.A.)
- <sup>3</sup> Laboratory for Research in Applied Geophysics, Department of Geology, Federal University of Paraná (UFPR), Curitiba 81530-000, PR, Brazil; francisco.ferreira@ufpr.br or francisco.lpga@gmail.com
- <sup>4</sup> Empresa Brasileira de Pesquisa Agropecuária (Embrapa Solos), 1.024 Jardim Botânico Street, Rio de Janeiro 22460-000, RJ, Brazil; waldir.carvalho@embrapa.br
- \* Correspondence: blendabastos@ufrrj.br or blenda.bastos@hotmail.com; Tel.: +55-21-98847-3056

**Abstract:** Airborne geophysical data (AGD) have great potential to represent soil-forming factors. Because of that, the objective of this study 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 ( $\mu\text{R}/\text{h}$ ), 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 approach was based on applying predictive modeling techniques using (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 km<sup>2</sup>, 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 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 ( $R^2$ ), 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 ( $R^2$  values ranging from 0.15 to 0.23), as well as the SVM model ( $R^2$  values ranging from 0.08 to 0.23) when compared to RF ( $R^2$  values ranging from 0.10 to 0.20) and SVM ( $R^2$  values ranging from 0.04 to 0.10) models without AGD. Overall, 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; hillslope areas; parent material



**Citation:** Bastos, B.P.; Pinheiro, H.S.K.; Ferreira, F.J.F.; Carvalho Junior, W.d.; dos Anjos, L.H.C. Could Airborne Geophysical Data Be Used to Improve Predictive Modeling of Agronomic Soil Properties in Tropical Hillslope Area? *Remote Sens.* **2023**, *15*, 3719. <https://doi.org/10.3390/rs15153719>

Received: 2 June 2023

Revised: 3 July 2023

Accepted: 6 July 2023

Published: 25 July 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

According to [1], 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 [1–3].

The DSM depends on adequate environmental variables to represent soil formation factors [4] as predictors to achieve this purpose. According to [5], 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 [6]. Parent material can be an essential soil-forming factor where granitic and basaltic inherited soils occur, for example. As suggested [6], 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., [7–11]). It is possible to make associations between relief denudation processes and the relative rates of soil formation and erosion [12,13], as [14] 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 ( $^{40}\text{K}$ ,  $^{238}\text{U}$ , and  $^{232}\text{Th}$ ) generally through NaI(Tl) detectors in the portable spectrometer or in airborne surveys [13]. 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 [15]. Another application in soil science is mapping the distribution of these elements in soil profiles [11,14,16]. 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 [17]. 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 [17].

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 [10,18–20]. 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 [21], 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 [22], 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.

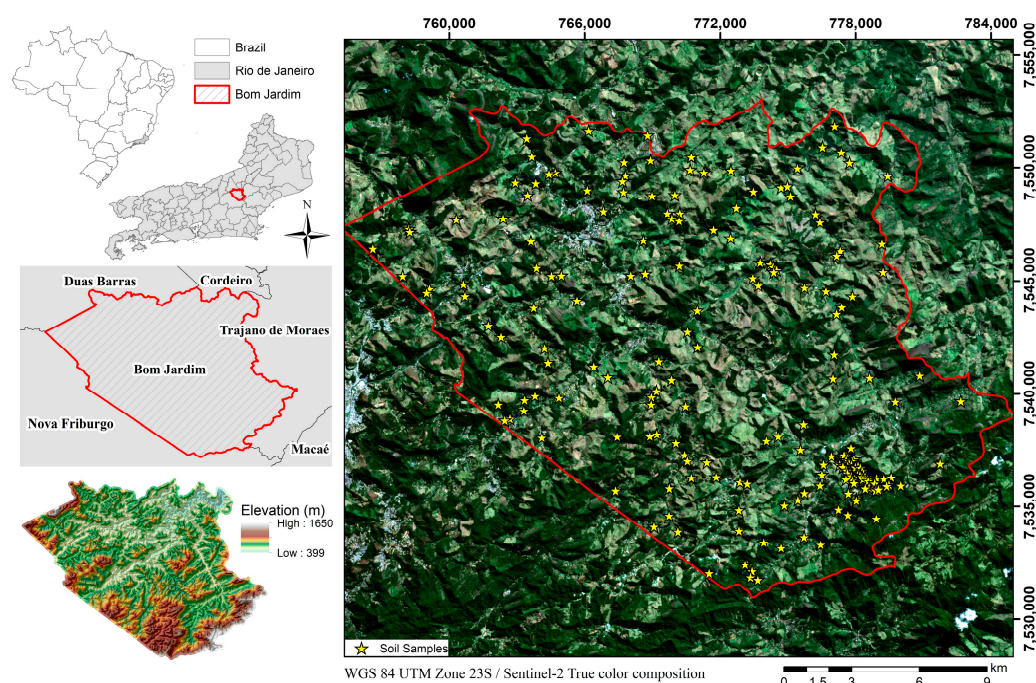
The AGD candidate predictor variables include total count ( $\mu\text{R}/\text{h}$ ), K (potassium), eU (uranium equivalent), and eTh (thorium equivalent), ratios between these elements (eTh/K, eU/K, and eU/eTh), factor F or F-parameter, anomalous potassium (Kd), anomalous uranium (Ud), anomalous magnetic 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.

## 2. Materials and Methods

### 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 [23] (Figure 1). Situated between the valleys and escarpments of Serra do Mar, the county comprises an area of 382,430 km<sup>2</sup>. 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 [24].

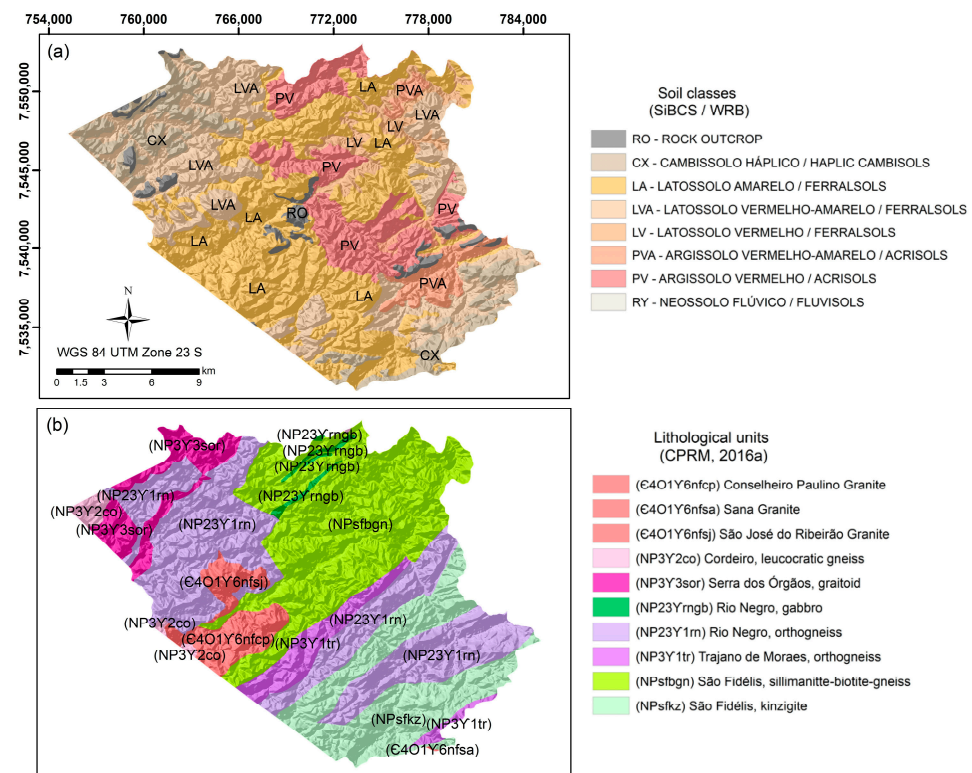
Despite anthropic interventions, mainly through agricultural activities, the original vegetation of the remaining Atlantic Forest is 51%, constituting secondary forests and native forests preserved in areas of higher altitude. The remaining 49% of the area predominantly comprises natural pastures, perennial crops such as coffee, vegetable crops, and ornamental flowers such as roses and eucalyptus reforestation [25]. 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 [26].



**Figure 1.** Location map of the Bom Jardim County and soil samples described by [26]. On the left below, the digital elevation model (DEM) derived from Rio de Janeiro cartographic database (original scale 1:25,000) [27].

According to the World Reference Base for Soil Resources Classification System [28], the region has as principal soil classes Cambisols, Ferralsols, Acrisols, and Fluvisols

(Figure 2a) and soil associations detailed by [25]. 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 [29,30] (Figure 2b).



**Figure 2.** (a) Soil map adapted from [25], SiBCS: Brazilian Soil Classification System [31], WRB: World Reference Base for Soil Resources [28]; (b) geological map (original scale 1:400,000) [32].

The legacy soil data were from [25]’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 [33]. 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 [25]. 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 [34] 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 [35]. The basic statistics for these attributes are presented in Table 1.

**Table 1.** Statistics of soil attributes used in the prediction.

	ASat (%)	BS (%)	CEC (cmol <sub>c</sub> kg <sup>-1</sup> )	Clay (g kg <sup>-1</sup> )	OC (g kg <sup>-1</sup> )
Minimum	0	2	3.39	102	4.8
Maximum	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.

## 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 [27]. 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 [36] were used to generate the covariates derived from DEM. DEM covariates used for prediction and their respective references are represented in Table 2.

**Table 2.** DEM covariates used for prediction and their respective references.

Covariate	Abbreviation	Reference
Digital Elevation Model	DEM	-
Convergence Index	Convergen	[37]
Downslope Distance Gradient	Gradient	[38]
Slope	-	
Aspect	-	
Profile Curvature	ProfileCurv	
Plan Curvature	PlanCurv	[39]
Longitudinal Curvature	Longitudin	
Maximum Curvature	MaximumCurv	
Topographic Position Index	TPI	[40–42]
Slope Height	SlopeHeig	
Valley Depth	ValleyDep	
Normalized Height	Normalized	[43]
Standardized Height	Standardiz	
Mid-Slope Position	MidSlope	
Terrain Ruggedness Index	TRI	[44]
Topographic Wetness Index	TWI	
Catchment Area	CatchmentA	[43]
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 [45]. 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 [46].

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 [47]. The Sentinel-2 covariates used for prediction and their respective references are represented in Table 3.

**Table 3.** Sentinel-2 covariates used for prediction and their respective references.

Covariate	Abbreviation	Reference
Grain Size Index	GSI	[48,49]
Normalized Difference Vegetation Index	NDVI	[50]
Alteration	-	
Ferric Iron	FerricIron	[51]
Ferric Oxides	FerricOxi	
Ferrous silicates (Biotite, chloride, amphibole)	FerrousSilic	

The airborne geophysical data were obtained by the Geological Survey of Brazil—CPRM (Rio de Janeiro Project: [21]). This project was carried out during 2011 and 2012 and covered an area of 32,202 km<sup>2</sup>, totaling 66,111.40 km of high-resolution geophysical profiles. The summary of survey characteristics is described in Table 4.

**Table 4.** 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 5) were processed using Oasis Montaj software (Educational software edition v. 9.8) at a spatial resolution of 100 m. First, minimum curvature [52] 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 [53]. The data were then reduced to the pole [54] 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 [55] 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 [22].

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 [56].



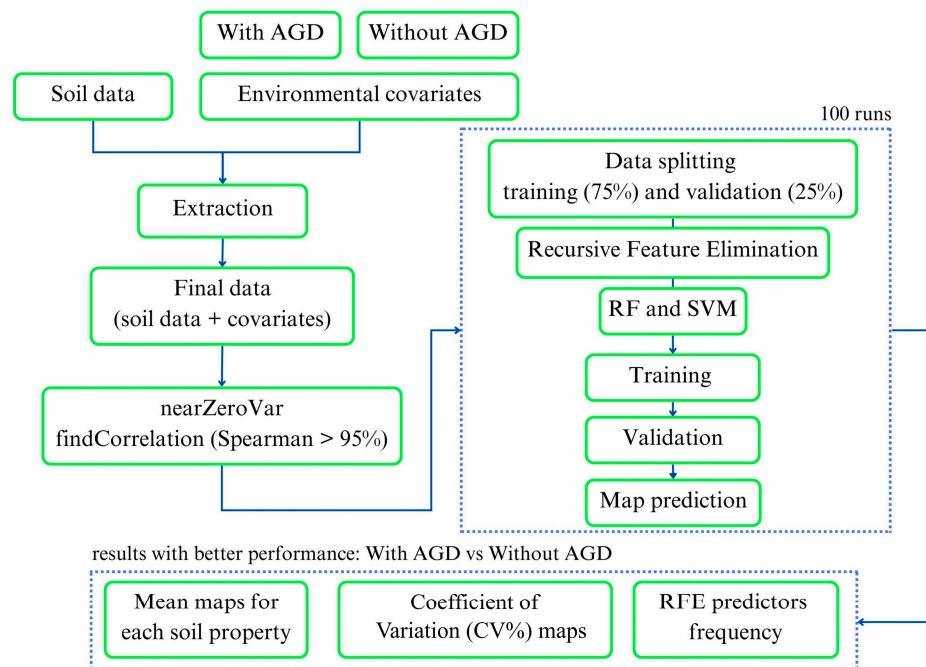
**Table 5.** AGD covariates used for prediction and their respective references.

Covariate	Abbreviation	Reference
Total count	TC	
Potassium, uranium equivalent, and thorium equivalent	Kperc, eU, and eTh	[21,57]
Ratios between elements	eTh/K, eU/K, and eU/eTh	
Factor F or F-parameter	FactorF	[58]
Anomalous potassium and anomalous uranium	Kd and Ud	[59]
Anomalous magnetic field	AMF	[21]
Vertical derivative and horizontal derivatives	GZ, GX and GY	[17]
Mafic index	MI	[22]

### 2.3. Modeling Strategy

The following steps were applied to predict the selected soil attributes using variables derived from DEM and Sentinel-2 images with AGD and without AGD (Figure 3):

- (1) Dataset preparation;
- (2) Removal of non-significant covariates with nearZeroVar and findCorrelation functions [60];
- (3) Data splitting in training (75%) and validation (25%) datasets;
- (4) Removal of covariates by importance using recursive feature elimination (RFE) [60];
- (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 3.** Modeling strategy flowchart.

According to [61], 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 [60] 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 [61]. In this study, the `findCorrelation` function was applied with Spearman's correlation method with a critical value of 95% (as adopted by [10,62]).

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 [10,62,63]. 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 [61]. 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 [62]). The hyperparameters of each algorithm are described in the `caret` package manual, as cited by [10]. 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 [62,64].

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) [62].

As shown in Figure 3, 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 [62,65]'s proposal.

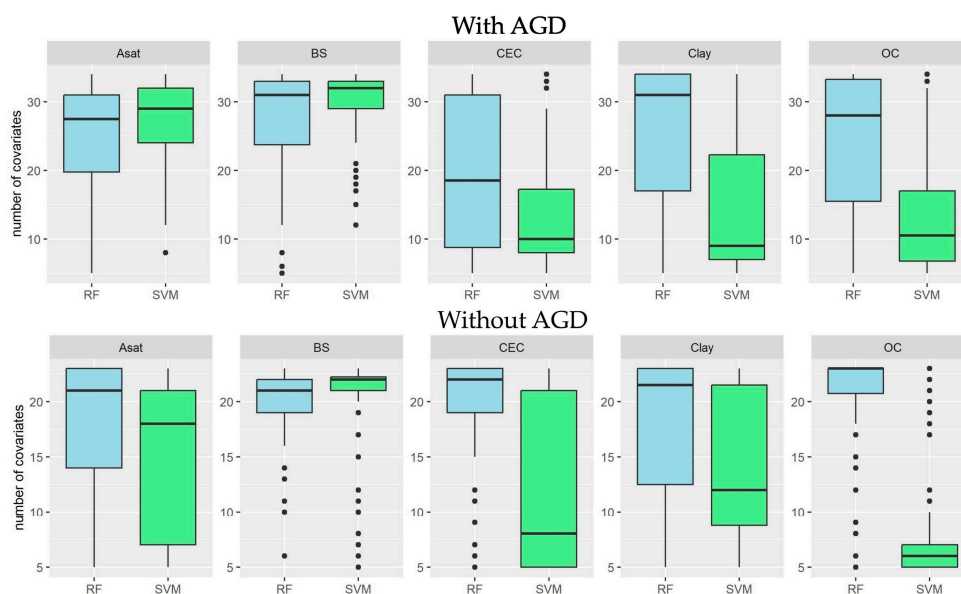
A high CV % indicates a large amount of variability in the data, which can result in higher uncertainty in the map predictions. Conversely, a low CV % shows relatively little variability in the data, which can result in lower uncertainty in the map predictions.

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.

### 3. Results

#### 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 4.



**Figure 4.** The average 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.

#### 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 6).

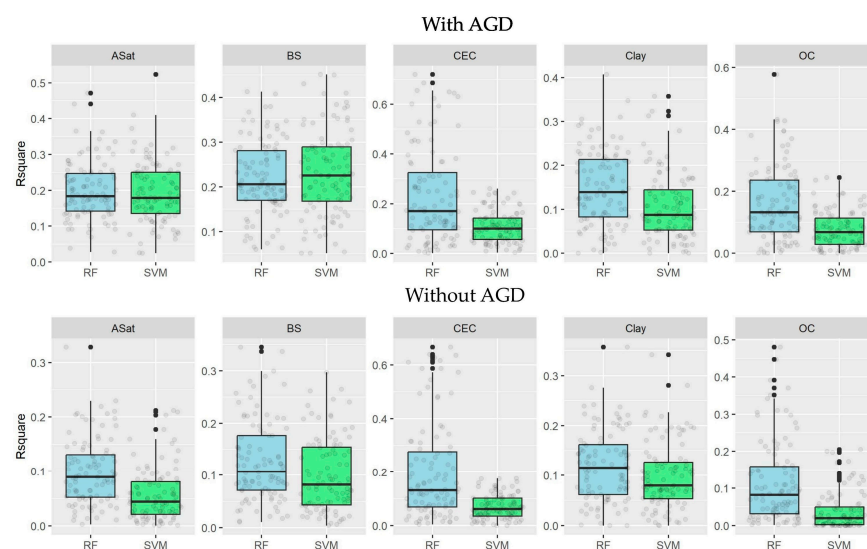
**Table 6.** Models' performance with and without AGD for all soil attributes.

		With AGD				
		ASat	BS	CEC	Clay	OC
R <sup>2</sup>	RF	0.20	0.22	0.23	0.15	0.16
	SVM	0.19	0.23	0.11	0.11	0.08
RSME	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 <sup>2</sup>	RF	0.10	0.13	0.20	0.12	0.11
	SVM	0.06	0.10	0.07	0.10	0.04
RSME	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 R<sup>2</sup>, except for BS: ASat (R<sup>2</sup> = 0.20), CEC (R<sup>2</sup> = 0.23), Clay (R<sup>2</sup> = 0.15), and OC (R<sup>2</sup> = 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 5 shows the performance results of the 100 model runs for each soil property in a boxplot based on R<sup>2</sup>.

**Figure 5.** 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<sup>2</sup>) 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 7, which is considered satisfactory for machine learning algorithms regarding soil properties [10].

**Table 7.** 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.

### 3.3. Map Prediction and Uncertainty

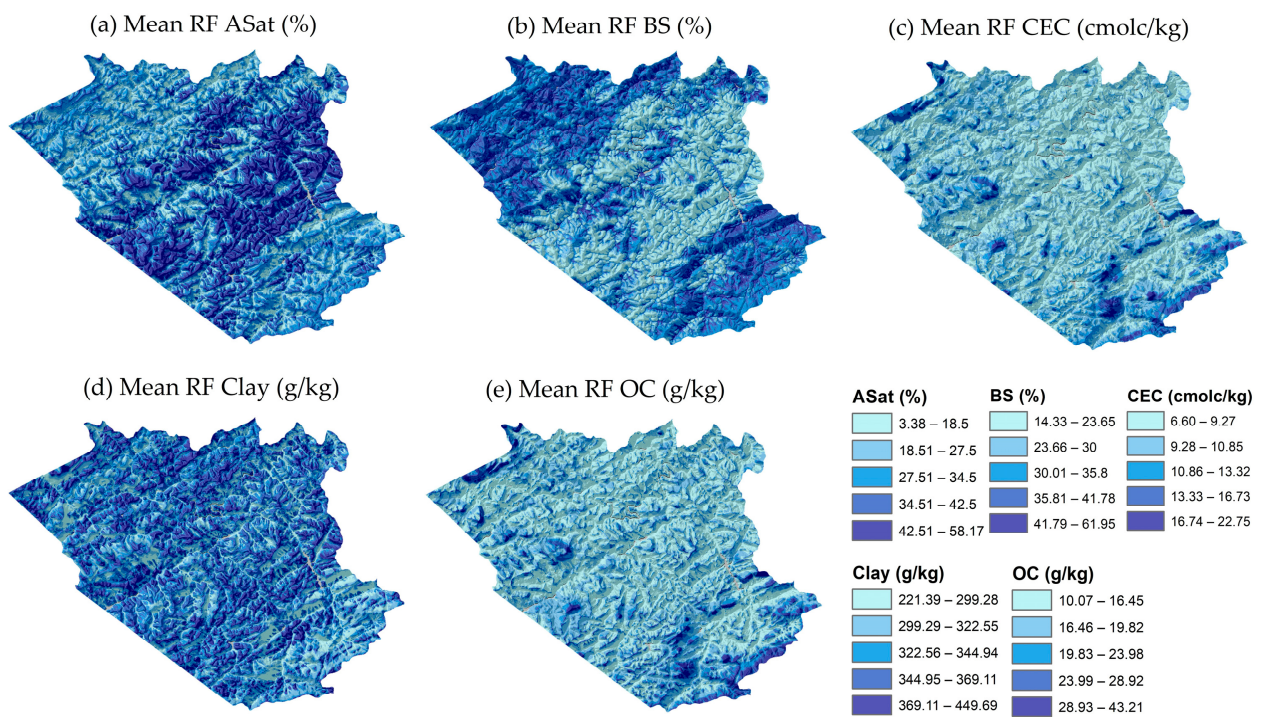
The mean values of CV% maps for modeling with and without AGD are represented in Table 8 below. The minimum, maximum, and median of CV% maps are attached in Appendix A, Table A1.

**Table 8.** The mean values of CV% maps for modeling with and without AGD.

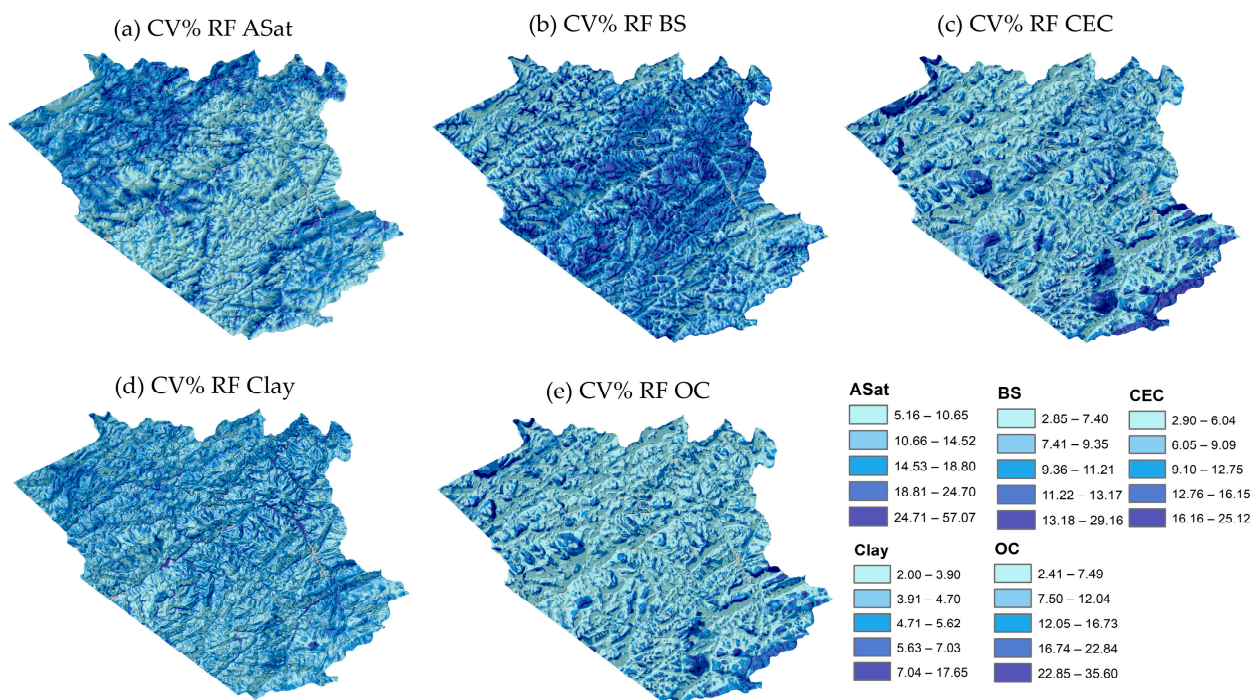
With AGD					
	ASat	BS	CEC	Clay	OC
RF	13.86	9.29	8.30	4.56	9.29
SVM	14.35	8.31	4.73	6.22	5.43
Without AGD					
	ASat	BS	CEC	Clay	OC
RF	14.85	10.95	7.53	4.93	9.07
SVM	22.08	8.41	5.50	4.89	5.50

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 6 and 7, and the final prediction and CV% maps of the SVM model are described in Figures 8 and 9 to analyze the results of modeling with AGD in more detail.



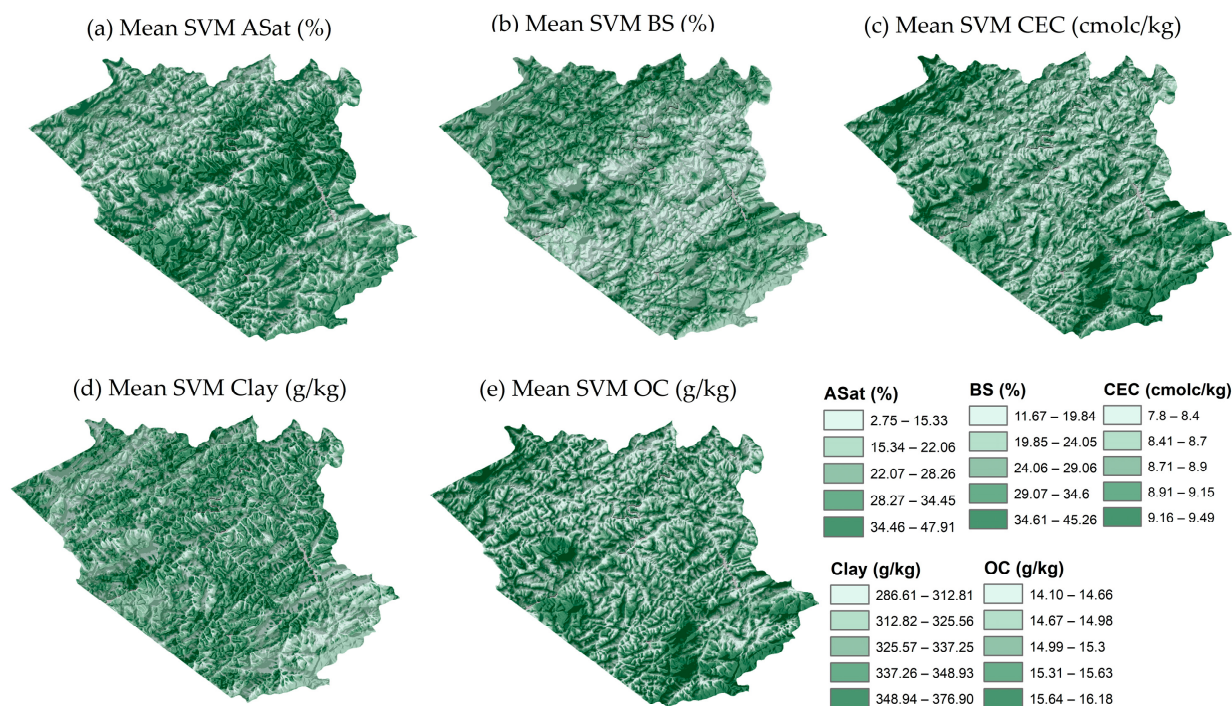


**Figure 6.** Final mean maps of the RF model with AGD: (a) ASat (%), (b) BS (%), (c) CEC (cmol<sub>c</sub> kg<sup>-1</sup>), (d) Clay (g kg<sup>-1</sup>), and (e) OC (g kg<sup>-1</sup>).

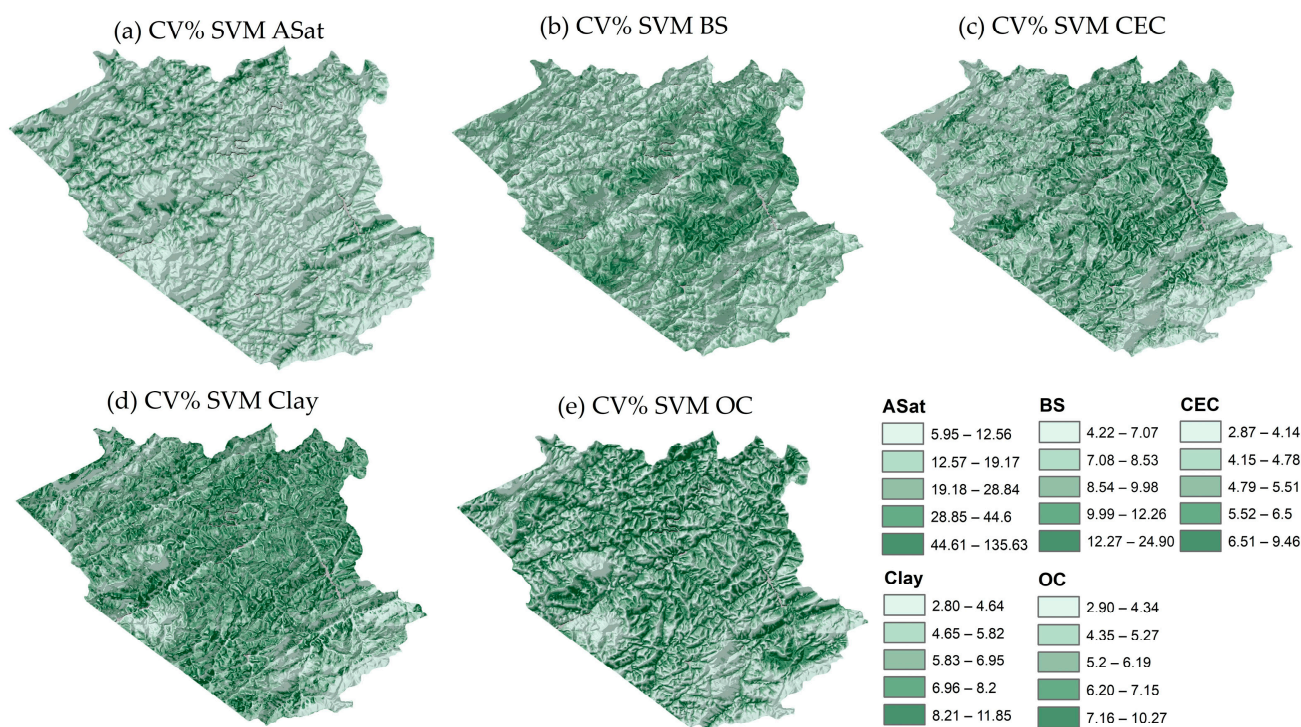


**Figure 7.** Coefficient of variation maps of the RF model with AGD: (a) ASat, (b) BS, (c) CEC, (d) Clay, and (e) OC.





**Figure 8.** Final mean maps of the SVM model with AGD: (a) ASat (%), (b) BS (%), (c) CEC (cmolc kg<sup>-1</sup>), (d) Clay (g kg<sup>-1</sup>), and (e) OC (g kg<sup>-1</sup>).



**Figure 9.** Coefficient of variation maps of the SVM model with AGD: (a) ASat, (b) BS, (c) CEC, (d) Clay, and (e) OC.

Figure 6a 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 2a. 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 2b. 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 7a.

In contrast, the BS map in Figure 6b 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 2a) 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 2b. 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 7b.

Figure 6c displays the CEC map, which seems to have its values influenced by the topography, with high CEC values ( $6.60\text{--}9.27\text{ cmol}_c\text{ kg}^{-1}$ ) matching the outcrop area with higher altitudes (Figure 1). This region is dominated by Cambisols, Ferralsols, and rock outcrop, as shown in Figure 2a. 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 7c).

For Clay (Figure 6d), it is more challenging to observe a pattern. However, relief also influences its distribution, where low values ( $221.39\text{--}299.28$  and  $299.29\text{--}322.55\text{ g kg}^{-1}$ ) correspond to valleys. The medium and higher values (other classes ranging from  $322.56$  to  $449.69\text{ g kg}^{-1}$ ) are distributed in areas with higher altitudes (Figure 1). 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 6e) has a distribution similar to that observed in the CEC mean map (Figure 6c). The higher values ( $28.93\text{--}43.21\text{ g kg}^{-1}$ ) 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.

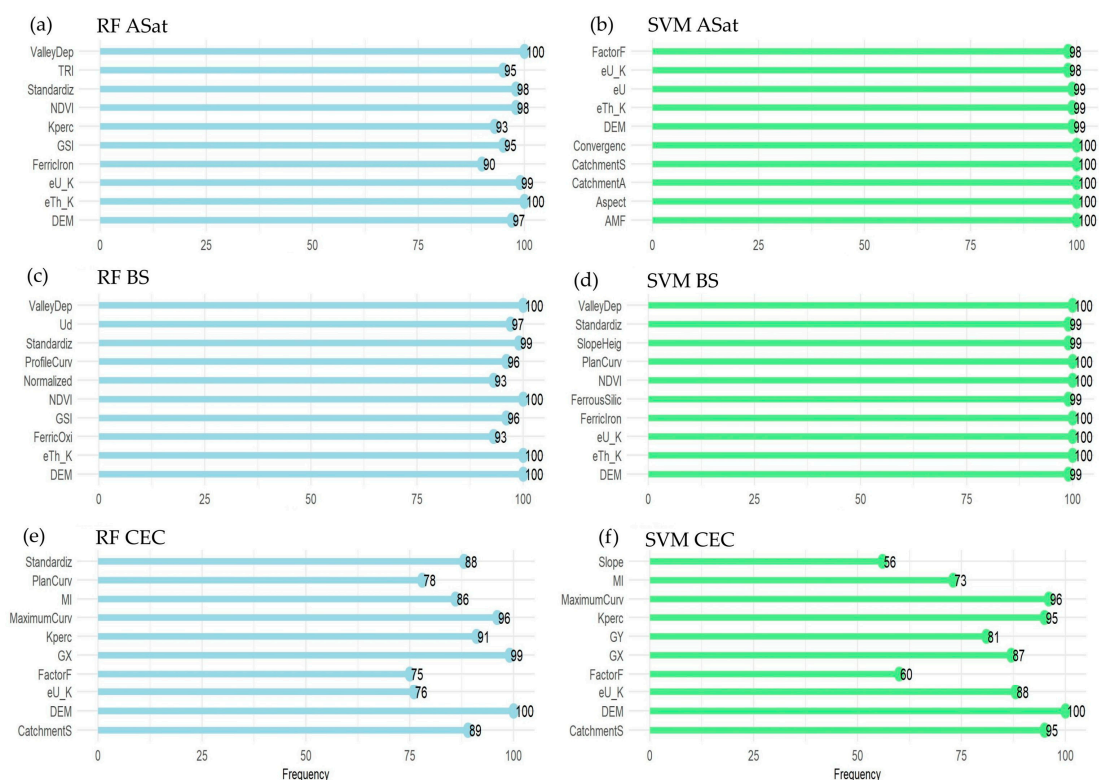
The visual distribution for SVM (Figure 8) is generally similar for all mean maps compared to the RF model (Figure 6). However, the RF model seems to have the most defined distribution patterns, with higher contrast between the highest and lowest values, mainly for ASat, BS, and Clay.

Comparing the two models, a difference worth highlighting is the CV% values. Although the RF model performs better according to the presented metrics (Table 6), 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 9b–e, respectively). For ASat, the RF model showed a smaller amplitude ranging from 5.16% to 57.07% (Figure 7a), while the SVM model showed an amplitude ranging from 5.95% to 135.63% (Figure 9a).

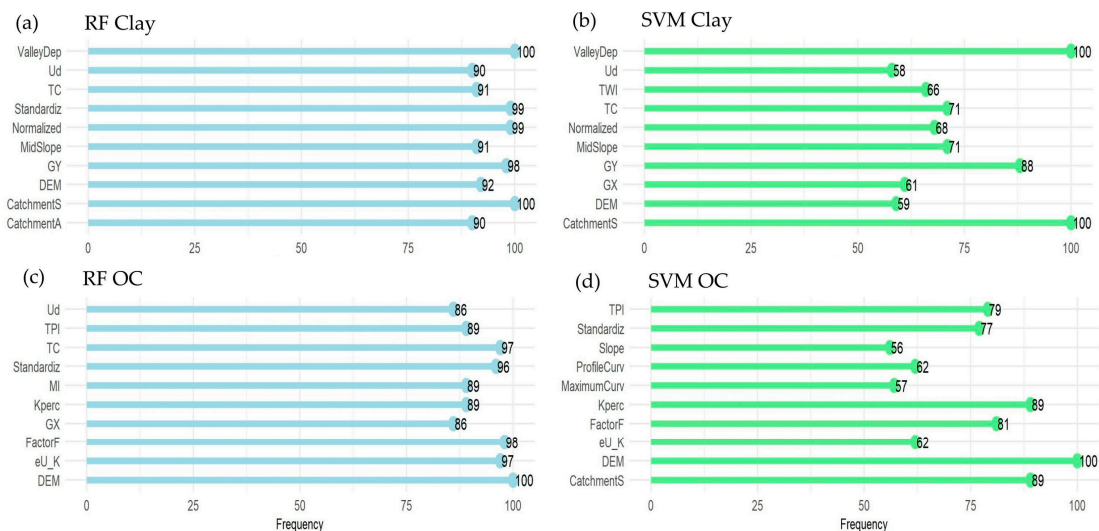
Another difference that can be highlighted is that the highest values in the average map (Figure 6c,e) correspond to the highest CV% values (Figure 7c,e) for the RF model for CEC and OC. In contrast, for the SVM model, the highest values in the average map (Figure 8c,e) correspond to the lower CV% values (Figure 9c,e) for CEC and OC. For Clay, the highest values in the average map (Figure 6d) correspond to the RF model's lower CV% values (Figure 7d). In contrast, for the SVM model, the lower values in the average map (Figure 8d) correspond to the lower CV% values (Figure 9d) (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.

### 3.4. AGD Data Importance

Figure 10 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 11.



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



**Figure 11.** 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 12 shows Spearman’s correlation matrix between AGD and soil attributes.

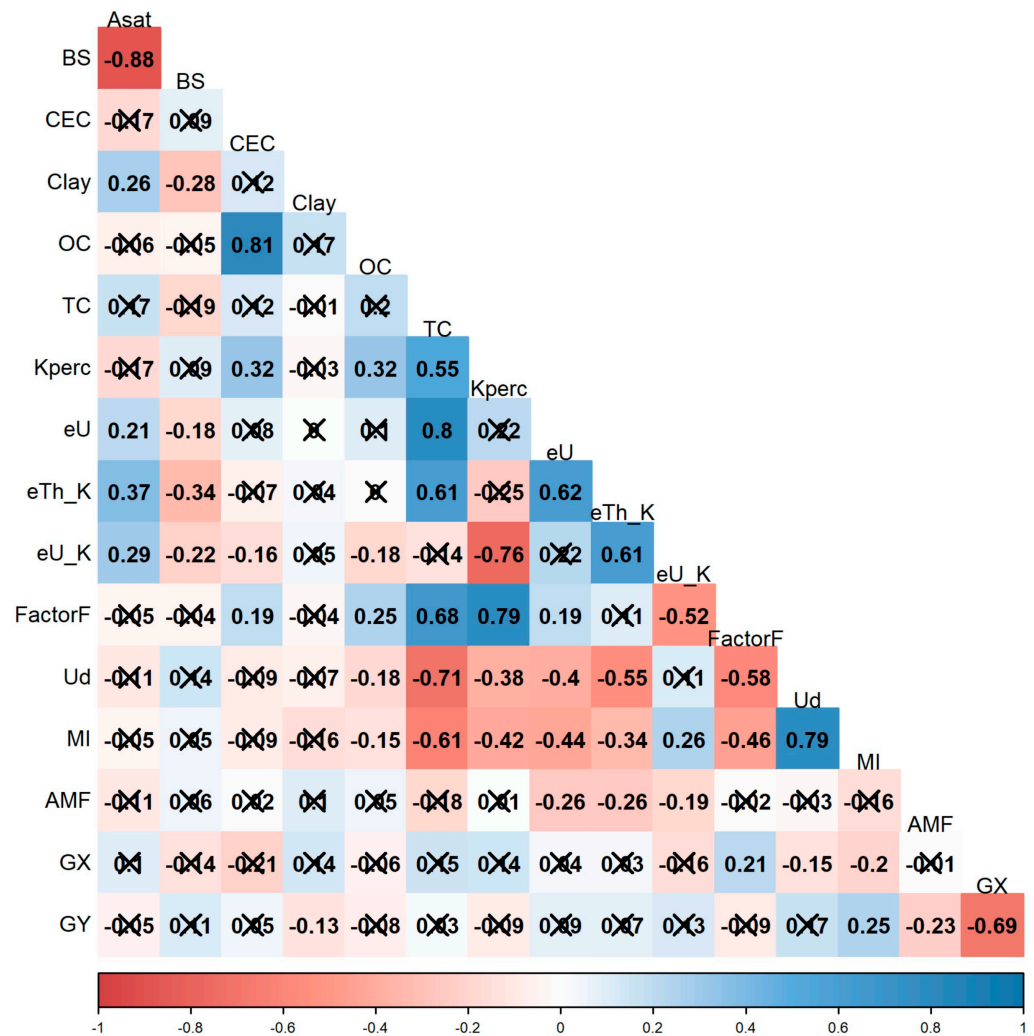


Figure 12. Spearman’s correlation matrix between AGD and soil attributes. “x” represents not statistically significant values at the 0.2 significance level.

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. Discussion

### 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. Ref. [10] 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 [10], 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 [10] is represented in Table 9, except for ASat.

**Table 9.** Comparison of the results of this study with those obtained by [10].

		$R^2$ Results			
		BS	CEC	Clay	OC
[10]	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  $R^2$  values than those observed by [10]. 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 [10].

Ref. [9] 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 [66]. 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 [10].

Ref. [67] 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 7. For OC, the result presented by [67] ( $R^2 = 0.11$ ) 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 [67] 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 [67] to predict soil properties in the same study area.

#### 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 10 and 11). However, some covariates stand out considering Spearman's correlation analyses (Figure 12). 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 [17]. 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 [15]. 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 [15].

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. Ref. [11] 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 6a), and the lower values correspond to the high values observed in the average map for BS (Figure 6b), as well as the CV% maps demonstrating lower values and greater reliability in the data in these regions (Figure 7a,b). This relationship is essential as it can be a good indicator of soil fertility since ASat and BS are used for this purpose [68].

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 [17]. Since lineament distribution influences groundwater flow [69,70], 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 [71], 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 [9–11,18].

Factor F (or F-parameter) is a valuable tool for maximizing and distinguishing areas characterized by potassium enrichment resulting from hydrothermal alterations [58]. 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 [11], 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 9c and 8c, respectively), the RF model does not follow the same pattern (Figures 7c and 6c, 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 [29,32]. 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 Kperc and factor F. Furthermore, knowing that CEC is the sum of bases plus H<sup>+</sup> and Al<sup>+</sup> [68], 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 Kperc covariate showed a positive correlation with OC (0.32) as well factor F (0.25). According to a study conducted by [72], it was discovered that radiometric potassium plays a crucial role in predicting soil carbon in Northern Ireland. Nevertheless, the relationship is inversed in [72]'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 [73], 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. Ref. [11] 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 9e and 8e, respectively), the RF model did not follow the same pattern (Figures 7e and 6e, respectively). So, further studies are needed to understand the relationship between the distribution of radioelements and organic carbon contents.

According to [10,11,18], 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.

#### 4.3. 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 [1]. 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 [5]. 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 [5]. This problem may be related to the density of samples collected, as shown in the SE of the map (Figure 1). According to [74], 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 [5]. 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 [25], which may explain the low significance of the Sentinel 2 MSI-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 [7–9,75–77]. 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 [2].

## 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.

**Author Contributions:** Conceptualization, B.P.B.; methodology, B.P.B., H.S.K.P. and W.d.C.J.; software, B.P.B.; validation, B.P.B. and H.S.K.P.; formal analysis, B.P.B. and H.S.K.P.; investigation, B.P.B.; resources, H.S.K.P. and L.H.C.d.A.; writing—original draft preparation, B.P.B. and H.S.K.P.; writing—review and editing, B.P.B., H.S.K.P., F.J.F.F., W.d.C.J. and L.H.C.d.A.; supervision, H.S.K.P., F.J.F.F., W.d.C.J. and L.H.C.d.A.; project administration, H.S.K.P. and L.H.C.d.A.; funding acquisition, H.S.K.P. and L.H.C.d.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Scientific and Technological Development (CNPq, Brazil) (MCTIC/CNPq No. 28/2018 and No. 312121/2021-0), and Research Support Foundation of the State of Rio de Janeiro (FAPERJ, Brazil) (FAPERJ No. 11/2018).

**Data Availability Statement:** Airborne gamma-ray and magnetic data are available from the Geological Survey of Brazil at <https://geosgb.cprm.gov.br/> (accessed on 1 March 2021).

**Acknowledgments:** This research was supported by the Federal Rural University of Rio de Janeiro (UFRRJ, Brazil), Postgraduate Program in Modeling and Geological Evolution (PPGMEG, Brazil), Coordination of Improvement of Higher Level Personnel (CAPES, Brazil), Federal University of Parana (UFPR, Brazil), Embrapa Soils (Brazil), providing infrastructure and software licenses.

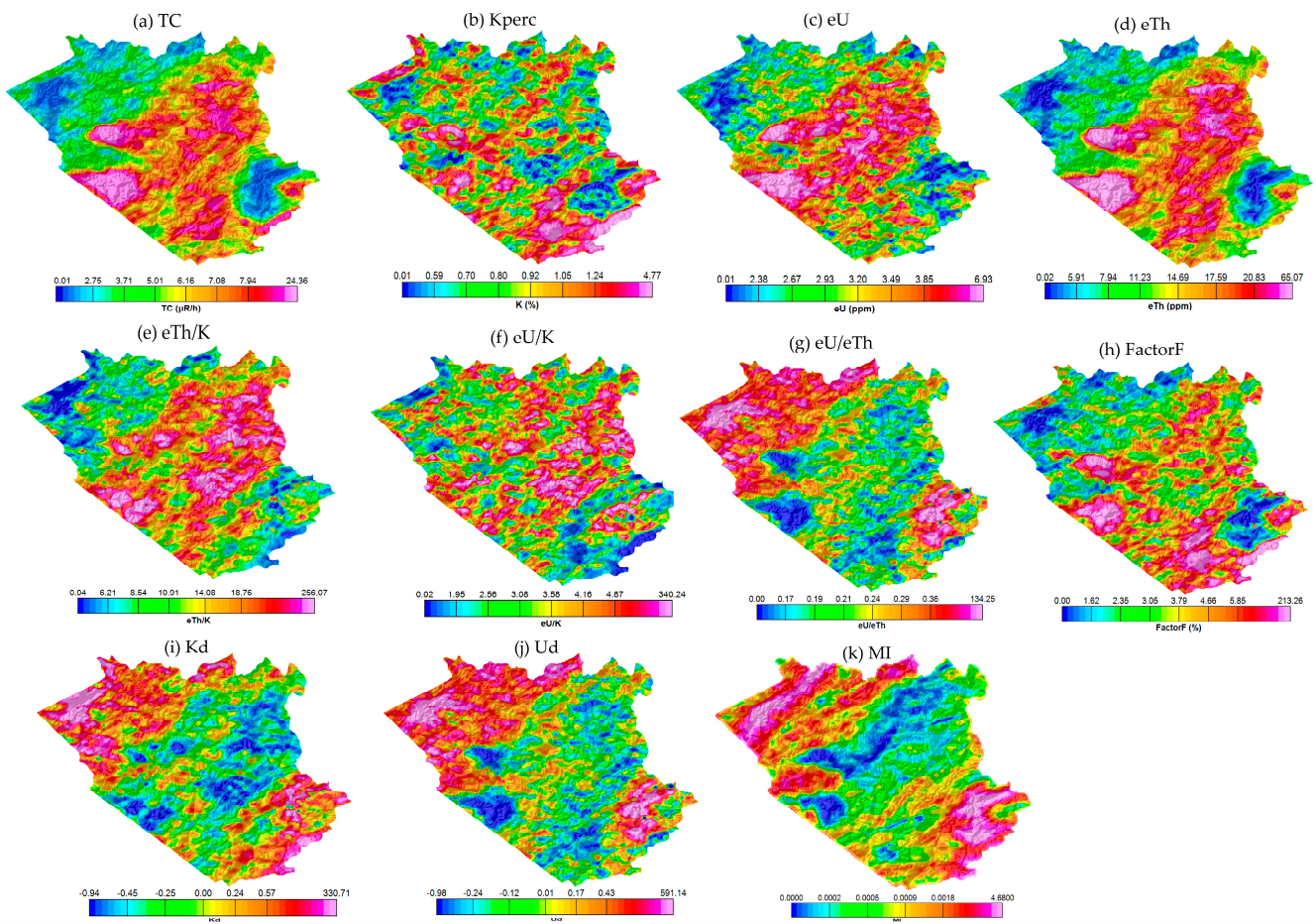
**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

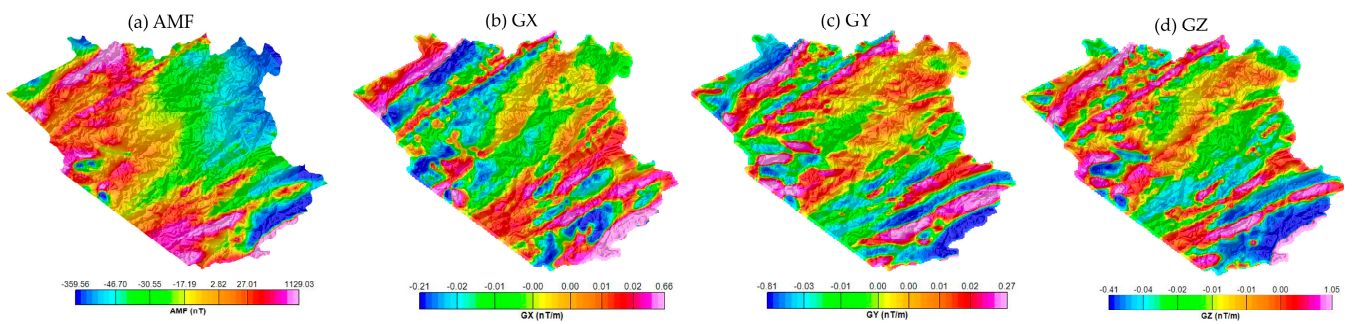
**Table A1.** Basic statistics of CV% maps for modeling with and without AGD.

		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.



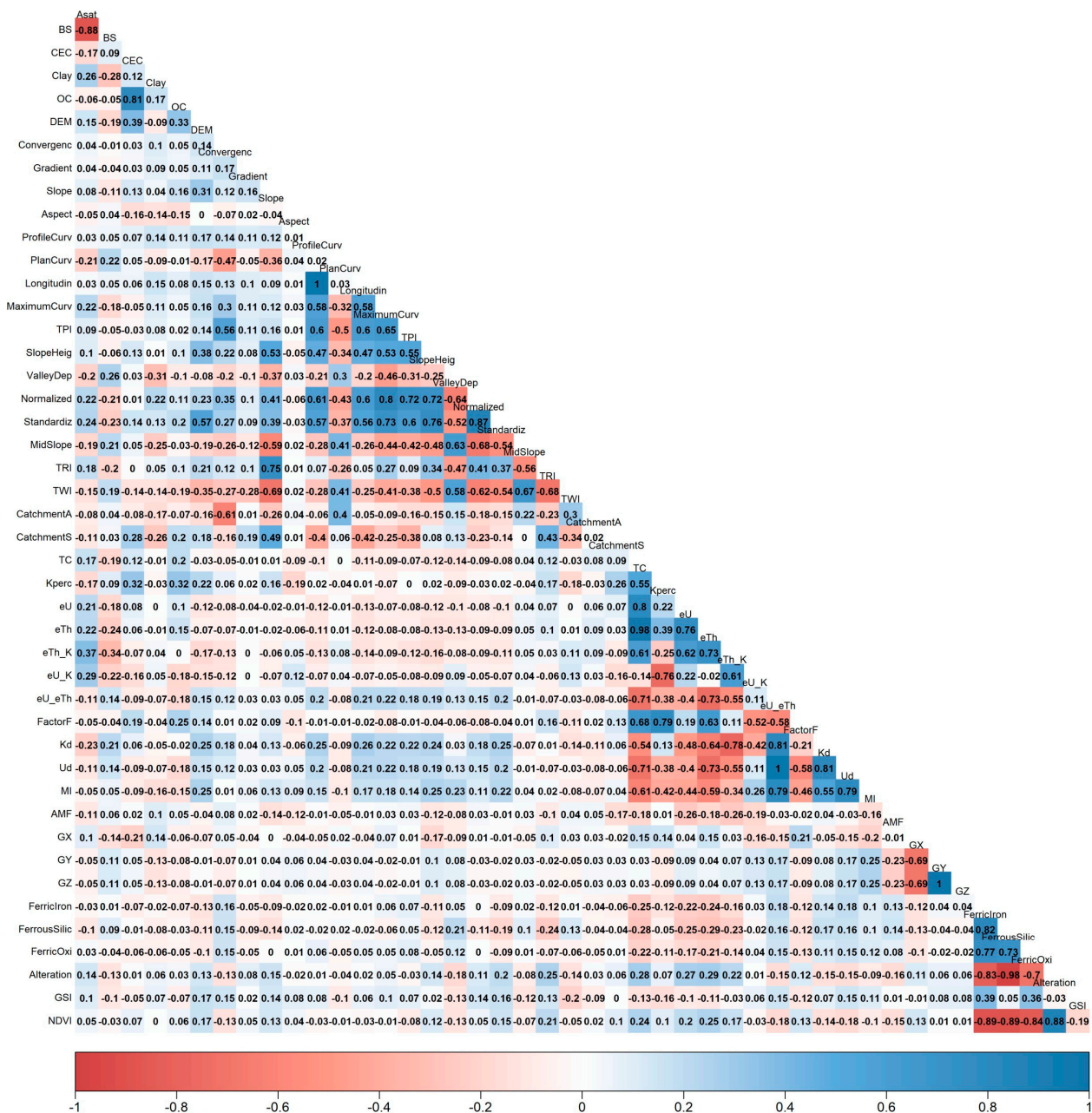


Figure A3. Spearman’s correlation matrix between all the proposed covariates and soil attributes.

References

1. Lagacherie, P.; McBratney, A.B. Chapter 1 Spatial Soil Information Systems and Spatial Soil Inference Systems: Perspectives for Digital Soil Mapping. In *Developments in Soil Science*; Elsevier: Amsterdam, The Netherlands, 2006; Volume 31, pp. 3–22, ISBN 978-0-444-52958-9.
2. Lagacherie, P. Digital Soil Mapping: A State of the Art. In *Digital Soil Mapping with Limited Data*; Hartemink, A.E., McBratney, A., Mendonça-Santos, M.D.L., Eds.; Springer: Dordrecht, The Netherlands, 2008; pp. 3–14, ISBN 978-1-4020-8591-8.
3. McBratney, A.B.; Mendonça Santos, M.L.; Minasny, B. On Digital Soil Mapping. *Geoderma* **2003**, *117*, 3–52. [\[CrossRef\]](#)
4. Jenny, H. *Factors of Soil Formation: A System of Quantitative Pedology*; Dover: New York, NY, USA, 1994; ISBN 978-0-486-68128-3.
5. Chen, S.; Arrouays, D.; Leatitia Mulder, V.; Poggio, L.; Minasny, B.; Roudier, P.; Libohova, Z.; Lagacherie, P.; Shi, Z.; Hannam, J.; et al. Digital Mapping of GlobalSoilMap Soil Properties at a Broad Scale: A Review. *Geoderma* **2022**, *409*, 115567. [\[CrossRef\]](#)
6. Wilson, M.J. The Importance of Parent Material in Soil Classification: A Review in a Historical Context. *Catena* **2019**, *182*, 104131. [\[CrossRef\]](#)

7. Maino, A.; Alberi, M.; Anceschi, E.; Chiarelli, E.; Cicala, L.; Colonna, T.; De Cesare, M.; Guastaldi, E.; Lopane, N.; Mantovani, F.; et al. Airborne Radiometric Surveys and Machine Learning Algorithms for Revealing Soil Texture. *Remote Sens.* **2022**, *14*, 3814. [[CrossRef](#)]
8. Loiseau, T.; Richer-de-Forges, A.C.; Martelet, G.; Bialkowski, A.; Nehlig, P.; Arrouays, D. Could Airborne Gamma-Spectrometric Data Replace Lithological Maps as Co-Variates for Digital Soil Mapping of Topsoil Particle-Size Distribution? A Case Study in Western France. *Geoderma Reg.* **2020**, *22*, e00295. [[CrossRef](#)]
9. Loiseau, T.; Arrouays, D.; Richer-de-Forges, A.C.; Lagacherie, P.; Ducommun, C.; Minasny, B. Density of Soil Observations in Digital Soil Mapping: A Study in the Mayenne Region, France. *Geoderma Reg.* **2021**, *24*, e00358. [[CrossRef](#)]
10. Mello, D.C.D.; Veloso, G.V.; Lana, M.G.D.; Mello, F.A.O.; Poppiel, R.R.; Cabrero, D.R.O.; Di Raimo, L.A.D.L.; Schaefer, C.E.G.R.; Filho, E.I.F.; Leite, E.P.; et al. A New Methodological Framework by Geophysical Sensors Combinations Associated with Machine Learning Algorithms to Understand Soil Attributes. *Earth Space Sci. Inform.* **2022**, *15*, 1219–1246. [[CrossRef](#)]
11. Mello, D.C.D.; Demattê, J.A.M.; Mello, F.A.D.O.; Roberto Poppiel, R.; Elizabet Quiñonez Silvero, N.; Lucas Safanelli, J.; Barros E Souza, A.; Augusto Di Loreto Di Raimo, L.; Rizzo, R.; Eduarda Bispo Resende, M.; et al. Applied Gamma-Ray Spectrometry for Evaluating Tropical Soil Processes and Attributes. *Geoderma* **2021**, *381*, 114736. [[CrossRef](#)]
12. Wilford, J.; Minty, B. Chapter 16 The Use of Airborne Gamma-ray Imagery for Mapping Soils and Understanding Landscape Processes. In *Developments in Soil Science*; Elsevier: Amsterdam, The Netherlands, 2006; pp. 207–610. [[CrossRef](#)]
13. Minty, B.R.S. Fundamentals of airborne gamma-ray spectrometry. *AGSO J. Aust. Geol. Geophys.* **1997**, *17*, 39–50.
14. Reinhardt, N.; Herrmann, L. Gamma-ray Spectrometry as Versatile Tool in Soil Science: A Critical Review. *J. Plant Nutr. Soil Sci.* **2019**, *182*, 9–27. [[CrossRef](#)]
15. Wilford, J. A Weathering Intensity Index for the Australian Continent Using Airborne Gamma-Ray Spectrometry and Digital Terrain Analysis. *Geoderma* **2012**, *183–184*, 124–142. [[CrossRef](#)]
16. Dickson, B.L.; Scott, K.M. Interpretation of aerial gamma-ray surveys-adding the geochemical factors. *AGSO J. Aust. Geol. Geophys.* **1997**, *17*, 187–200.
17. Dentith, M.C.; Mudge, S.T. *Geophysics for the Mineral Exploration Geoscientist*; Cambridge University Press: Cambridge, UK, 2014.
18. Mello, D.C.D.; Demattê, J.A.M.; Silvero, N.E.Q.; Di Raimo, L.A.D.L.; Poppiel, R.R.; Mello, F.A.O.; Souza, A.B.; Safanelli, J.L.; Resende, M.E.B.; Rizzo, R. Soil Magnetic Susceptibility and Its Relationship with Naturally Occurring Processes and Soil Attributes in Pedosphere, in a Tropical Environment. *Geoderma* **2020**, *372*, 114364. [[CrossRef](#)]
19. Sarmast, M.; Farpoor, M.H.; Esfandiarpour Boroujeni, I. Magnetic Susceptibility of Soils along a Lithotoposequence in Southeast Iran. *Catena* **2017**, *156*, 252–262. [[CrossRef](#)]
20. Valaee, M.; Ayoubi, S.; Khormali, F.; Lu, S.G.; Karimzadeh, H.R. Using Magnetic Susceptibility to Discriminate between Soil Moisture Regimes in Selected Loess and Loess-like Soils in Northern Iran. *J. Appl. Geophys.* **2016**, *127*, 23–30. [[CrossRef](#)]
21. CPRM—Serviço Geológico do Brasil. *Relatório Final do Levantamento Processamento dos Dados Magnetométricos e Gamaespectrométricos; Projeto Aerogeofísico Rio de Janeiro (Projeto 1.117); Prospectores Aerolevantamentos e Sistemas Ltda: Rio de Janeiro, Brazil, 2012; 219p.*
22. Iza, E.R.H.F.; Horbe, A.M.C.; Castro, C.C.; Herrera, I.L.I.E. Integration of Geochemical and Geophysical Data to Characterize and Map Lateritic Regolith: An Example in the Brazilian Amazon. *Geochem. Geophys. Geosyst.* **2018**, *19*, 3254–3271. [[CrossRef](#)]
23. IBGE—Instituto Brasileiro de Geografia e Estatística. Divisão Territorial Brasileira. 2021. Available online: <https://www.ibge.gov.br> (accessed on 1 March 2021).
24. ANA—Agência Nacional de Águas e Saneamento Básico. Índices e Estatísticas das Estações Pluviométricas e Fluviométricas. 2020. Available online: <https://dadosabertos.ana.gov.br> (accessed on 1 March 2021).
25. Calderano Filho, B.; Polivanov, H.; Chagas, C.S.; de Carvalho Junior, W.; Calderano, S.B.; Guerra, A.J.T.; Donagemma, G.K.; Bhering, S.B.; Aglio, M.L.D. *Solos do Médio alto Curso do Rio Grande, Região Serrana do Estado do Rio de Janeiro*; Embrapa: Rio de Janeiro, Brazil, 2012.
26. 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. Tese de Doutorado, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil, 2012.*
27. 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. Available online: <https://www.ibge.gov.br> (accessed on 1 March 2021).
28. 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; FAO: Rome, Italy, 2014.
29. Tupinambá, M.; Teixeira, W.; Heilbron, M. 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 IGEO-UFRJ* **2013**, *35*, 140–151. [[CrossRef](#)]
30. CPRM—Serviço Geológico do Brasil. *Geologia e Recursos Minerais do Estado do Rio de Janeiro: Texto Explicativo do Mapa Geológico e de Recursos Minerais*; Serviço Geológico do Brasil: Belo Horizonte, Brazil, 2016; 182p.
31. Santos, H.G.; Jacomine, P.K.T.; Anjos, L.H.C.; Oliveira, V.A.; Lumberras, J.F.; Coelho, M.R.; Almeida, J.A.; Araujo Filho, J.C.; Oliveira, J.B.; Cunha, T.J.F. *Sistema Brasileiro de Classificação de Solos*; Embrapa: Brasília, DF, Brazil, 2018.
32. CPRM—Serviço Geológico do Brasil. *Mapa Geológico e de Recursos Minerais do Estado do Rio de Janeiro Escala 1:400.000*; Programa geologia do Brasil; Serviço Geológico do Brasil: Belo Horizonte, Brazil, 2016.

33. EMBRAPA—Empresa Brasileira de Pesquisa Agropecuária. *Sistema Brasileiro de Classificação de Solos*; Empresa Brasileira de Pesquisa Agropecuária: Brasília, DF, Brazil, 2006; 305p.
34. Beaudette, D.E.; Roudier, P.; O’Geen, A.T. Algorithms for Quantitative Pedology: A Toolkit for Soil Scientists. *Comput. Geosci.* **2013**, *52*, 258–268. [CrossRef]
35. R Core Team. *A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2023.
36. Conrad, O.; Bechtel, B.; Bock, M.; Dietrich, H.; Fischer, E.; Gerlitz, L.; Wehberg, J.; Wichmann, V.; Böhner, J. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geosci. Model Dev.* **2015**, *8*, 1991–2007. [CrossRef]
37. Köthe, R.; Lehmeier, F.; SARA—System zur Automatischen Relief-Analyse: User Manual. Tool Convergence Index available in SAGA-GIS Tool Library Documentation. 1996. Available online: <https://saga-gis.sourceforge.io/> (accessed on 1 March 2021).
38. Hjerdt, K.N.; McDonnell, J.J.; Seibert, J.; Rodhe, A. A New Topographic Index to Quantify Downslope Controls on Local Drainage: Technical note. *Water Resour. Res.* **2004**, *40*. [CrossRef]
39. Wood, J. Chapter 14 Geomorphometry in LandSerf. In *Developments in Soil Science*; Elsevier: Amsterdam, The Netherlands, 2009; Volume 33, pp. 333–349, ISBN 978-0-12-374345-9.
40. Guisan, A.; Weiss, S.B.; Weiss, A.D. GLM versus CCA Spatial Modeling of Plant Species Distribution. *Plant Ecol.* **1999**, *143*, 107–122. [CrossRef]
41. Weiss, A.D. Topographic position and landforms analysis. Poster presentation. In Proceedings of the ESRI User Conference, San Diego, CA, USA, 9–13 July 2001, Volume 200.
42. Wilson, J.P.; Gallant, J.C. Primary Topographic Attributes. In *Terrain Analysis: Principles and Applications*; Wilson, J.P., Gallant, J.C., Eds.; John Wiley & Sons: Hoboken, NJ, USA, 2000; pp. 51–85.
43. Böhner, J.; Selige, T. Spatial prediction of soil attributes using terrain analysis and climate regionalization. *Gott. Geogr. Abh.* **2006**, *115*, 13–28.
44. Riley, S.J.; De Gloria, S.D.; Elliot, R. A Terrain Ruggedness that Quantifies Topographic Heterogeneity. *Intermt. J. Sci.* **1999**, *5*, 23–27.
45. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA’s Optical High-Resolution Mission for GMES Operational Services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [CrossRef]
46. Mueller-Wilm, U.; Devignot, O.; Pessiot, L. *S2 MPC Sen2Cor Configuration and User Manual*; European Space Agency: Paris, France, 2017.
47. QGIS Development Team QGIS Geographic Information System. Open Source Geospatial Foundation Project. 2023. Available online: <http://qgis.osgeo.org> (accessed on 1 January 2023).
48. Perera, Y.Y.; Zapata, C.E.; Houston, W.N.; Houston, S.L. Prediction of the Soil-Water Characteristic Curve Based on Grain-Size-Distribution and Index Properties. In *Advances in Pavement Engineering*; American Society of Civil Engineers: Austin, TX, USA, 2005; pp. 1–12.
49. Xiao, J.; Shen, Y.; Tateishi, R.; Bayaer, W. Development of Topsoil Grain Size Index for Monitoring Desertification in Arid Land Using Remote Sensing. *Int. J. Remote Sens.* **2006**, *27*, 2411–2422. [CrossRef]
50. Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring Vegetation Systems in the Great Plains with ERTS. In Proceedings of the Third ERTS Symposium, Washington, DC, USA, 10–14 December 1973; NASA Special Publication. Volume 351, p. 309.
51. 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 Sens. Environ.* **2014**, *148*, 124–133. [CrossRef]
52. Briggs, I.C. Machine contouring using minimum curvature. *Geophysics* **1974**, *39*, 39–48. [CrossRef]
53. Geosoft. Tutorials Oasis Montaj: Bi-Directional Gridding. Geosoft Inc. 2023. Available online: <https://www.seequent.com> (accessed on 1 January 2023).
54. Baranov, V. A new method for interpretation of aeromagnetic maps: Pseudo-gravimetric anomalies. *Geophysics* **1957**, *22*, 359–382. [CrossRef]
55. Li, X. Understanding 3D Analytic Signal Amplitude. *Geophysics* **2006**, *71*, L13–L16. [CrossRef]
56. Cavazzi, S.; Corstanje, R.; Mayr, T.; Hannam, J.; Fealy, R. Are Fine Resolution Digital Elevation Models Always the Best Choice in Digital Soil Mapping? *Geoderma* **2013**, *195–196*, 111–121. [CrossRef]
57. IAEA—International Atomic Energy Agency. *Guidelines for Radioelement Mapping Using Gamma Ray Spectrometry Data*; IAEA—International Atomic Energy Agency: Vienna, Austria, 2003; ISBN 978-92-0-108303-6.
58. Gnojek, I.; Prichystal, A. A new zinc mineralization detected by airborne gamma-ray spectrometry in Northern Moravia (Czechoslovakia). *Geoexploration* **1985**, *23*, 491–502. [CrossRef]
59. Saunders, D.F.; Burson, K.R.; Branch, J.F.; Thompson, C.K. Relation of Thorium-normalized Surface and Aerial Radiometric Data to Subsurface Petroleum Accumulations. *Geophysics* **1993**, *58*, 1417–1427. [CrossRef]
60. Kuhn, M. Building Predictive Models in R Using the caret Package. *J. Stat. Softw.* **2008**, *28*, 1–26. [CrossRef]
61. Kuhn, M.; Johnson, K. *Applied Predictive Modeling*; Springer: New York, NY, USA, 2013; ISBN 978-1-4614-6848-6.
62. Siqueira, R.G.; Moquedace, C.M.; Francelino, M.R.; Schaefer, C.E.G.R.; Fernandes-Filho, E.I. Machine Learning Applied for Antarctic Soil Mapping: Spatial Prediction of Soil Texture for Maritime Antarctica and Northern Antarctic Peninsula. *Geoderma* **2023**, *432*, 116405. [CrossRef]



63. Kaya, F.; Mishra, G.; Francaviglia, R.; Keshavarzi, A. Combining Digital Covariates and Machine Learning Models to Predict the Spatial Variation of Soil Cation Exchange Capacity. *Land* **2023**, *12*, 819. [[CrossRef](#)]
64. Meyer, H.; Reudenbach, C.; Wöllauer, S.; Nauss, T. Importance of Spatial Predictor Variable Selection in Machine Learning Applications—Moving from Data Reproduction to Spatial Prediction. *Ecol. Model.* **2019**, *411*, 108815. [[CrossRef](#)]
65. Gomes, L.C.; Faria, R.M.; De Souza, E.; Veloso, G.V.; Schaefer, C.E.G.R.; Filho, E.I.F. Modelling and Mapping Soil Organic Carbon Stocks in Brazil. *Geoderma* **2019**, *340*, 337–350. [[CrossRef](#)]
66. Vasques, G.M.; Rodrigues, H.M.; Coelho, M.R.; Baca, J.F.M.; Dart, R.O.; Oliveira, R.P.; Teixeira, W.G.; Ceddia, M.B. Field Proximal Soil Sensor Fusion for Improving High-Resolution Soil Property Maps. *Soil Syst.* **2020**, *4*, 52. [[CrossRef](#)]
67. Carvalho Junior, W.D.; Chagas, C.D.S.; Lagacherie, P.; Calderano Filho, B.; Bhering, S.B. Evaluation of Statistical and Geostatistical Models of Digital Soil Properties Mapping in Tropical Mountain Regions. *Rev. Bras. Ciênc. Solo* **2014**, *38*, 706–717. [[CrossRef](#)]
68. Ker, J.C.; Curi, N.; Schaefer, C.E.G.R.; Vidal-Torrado, P. *Pedologia: Fundamentos*; Sociedade Brasileira de Ciência do Solo: Viçosa, MG, Brazil, 2015; ISBN 978-85-86504-09-9.
69. Hung, L.Q.; Batelaan, O.; De Smedt, F. 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, Proceedings of the SPIE Remote Sensing, Bruges, Belgium, 19–22 September 2005*; SPIE: Bellingham, WA, USA; p. 59830T.
70. Falebita, D.E.; Ayua, K.J. Appraisal of Lineaments for Groundwater Prognosis in the Middle Benue Trough, Nigeria: A Case Study. *Sustain. Water Resour. Manag.* **2023**, *9*, 12. [[CrossRef](#)]
71. Brubaker, S.C.; Jones, A.J.; Lewis, D.T.; Frank, K. Soil Properties Associated with Landscape Position. *Soil Sci. Soc. Am. J.* **1993**, *57*, 235–239. [[CrossRef](#)]
72. Rawlins, B.G.; Marchant, B.P.; Smyth, D.; Scheib, C.; Lark, R.M.; Jordan, C. Airborne Radiometric Survey Data and a DTM as Covariates for Regional Scale Mapping of Soil Organic Carbon across Northern Ireland. *Eur. J. Soil Sci.* **2009**, *60*, 44–54. [[CrossRef](#)]
73. Powers, J.S.; Corre, M.D.; Twine, T.E.; Veldkamp, E. Geographic Bias of Field Observations of Soil Carbon Stocks with Tropical Land-Use Changes Precludes Spatial Extrapolation. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 6318–6322. [[CrossRef](#)]
74. Lagacherie, P.; Arrouays, D.; Bourennane, H.; Gomez, C.; Nkuba-Kasanda, L. 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* **2020**, *375*, 114503. [[CrossRef](#)]
75. Ng, W.; Minasny, B.; McBratney, A.; De Caritat, P.; Wilford, J. Digital Soil Mapping of Lithium in Australia. *Earth Syst. Sci. Data* **2023**, *15*, 2465–2482. [[CrossRef](#)]
76. Chen, S.; Richer-de-Forges, A.C.; Leatitia Mulder, V.; Martelet, G.; Loiseau, T.; Lehmann, S.; Arrouays, D. Digital Mapping of the Soil Thickness of Loess Deposits over a Calcareous Bedrock in Central France. *Catena* **2021**, *198*, 105062. [[CrossRef](#)]
77. Adler, K.; Persson, K.; Söderström, M.; Eriksson, J.; Pettersson, C.-G. Digital Soil Mapping of Cadmium: Identifying Arable Land for Producing Winter Wheat with Low Concentrations of Cadmium. *Agronomy* **2023**, *13*, 317. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.