Article

Remote Sensing and Kriging with External Drift to Improve Sparse Proximal Soil Sensing Data and Define Management Zones in Precision Agriculture

Hugo Rodrigues [1,*], Marcos B. Ceddia [2], Gustavo M. Vasques [3], Vera L. Mulder [4], Gerard B. M. Heuvelink [4], Ronaldo P. Oliveira [3], Ziany N. Brandão [5], João P. S. Morais [5], Matheus L. Neves [1] and Sílvio R. L. Tavares [3]

1 Postgraduate Program in Agronomy—Soil Science, Federal Rural University of Rio de Janeiro, BR-465, Km 7, Seropédica 23897-000, RJ, Brazil; matheusl.dasneves@gmail.com
2 AgroTechnologies and Sustainability Department, Federal Rural University of Rio de Janeiro, BR-465, Km 7, Seropédica 23897-000, RJ, Brazil; marcosceddia@gmail.com
3 Brazilian Agricultural Research Company (Embrapa) Soils, Rio de Janeiro 22460-000, RJ, Brazil; gustavo.vasques@embrapa.br (G.M.V.); ronaldo.oliveira@embrapa.br (R.P.O.); silvio.tavares@embrapa.br (S.R.L.T.)
4 Soil Geography and Landscape Group, Wageningen University, 6708 PB Wageningen, The Netherlands; titia.mulder@wur.nl (V.L.M.); gerard.heuvelink@wur.nl (G.B.M.H.)
5 Brazilian Agricultural Research Company (Embrapa) Cotton, Rua Osvaldo Cruz, 1143, Centenário 58428-095, CG, Brazil; ziany.brandao@embrapa.br (Z.N.B.); joao.morais@embrapa.br (J.P.S.M.)
* Correspondence: rodrigues.machado.hugo@gmail.com

Abstract: The precision agriculture scientific field employs increasingly innovative techniques to optimize inputs, maximize profitability, and reduce environmental impacts. Therefore, obtaining a high number of soil samples to make precision agriculture feasible is challenging. This data bottleneck has been overcome by identifying sub-regions based on data obtained through proximal soil sensing equipment. These data can be combined with freely available remote sensing data to create more accurate maps of soil properties. Furthermore, these maps can be optimally aggregated and interpreted for soil heterogeneity through management zones. Thus, this work aimed to create and combine soil management zones from proximal soil sensing and remote sensing data. To this end, data on electrical conductivity and magnetic susceptibility, both apparent, were measured using the EM38-MK2 proximal soil sensor and the contents of the thorium and uranium elements, both equivalent, via the Medusa MS1200 proximal soil sensor for a 72-ha grain-producing area in São Paulo, Brazil. The proximal soil sensing attributes were mapped using ordinary kriging (OK). Maps were also made using kriging with external drift (KED), and the proximal soil sensor attributes data, combined with remote sensing data, such as Landsat-8, Aster, and Sentinel-2 images, in addition to 10 terrain covariables derived from the digital elevation model Alos Palsar. As a result, three management zone maps were produced via the *k*-means clustering algorithm: using data from proximal sensors (OK), proximal sensors combined with remote sensors (KED), and remote sensors. Seventy-two samples (0–10 cm in depth) were collected and analyzed in a laboratory (1 sample per hectare) for concentrations of clay, calcium, organic carbon, and magnesium to assess the capacity of the management zone maps created using analysis of variance. All zones created using the three data groups could distinguish the different treatment areas. The three data sources used to map management zones produced similar map zones, but the zone map using a combination of proximal and remote data did not show an improvement in defining the management zones, and using only remote sensing data lowered the significance levels of differentiating each zone compared to the OK and KED maps. In summary, this study not only underscores the global applicability of proximal and remote sensing techniques in precision agriculture but also sheds light on the nuances of their integration. The study's findings affirm the efficacy of these advanced technologies in addressing the challenges posed by soil heterogeneity, paving the way for more nuanced and site-specific agricultural practices worldwide.

1. Introduction

The diverse economic and environmental pressures on crop-producing farms are growing in tandem with the demands for food, given a growing world population [1]. The increase in agricultural productivity is much needed due to the challenges of distributing food and commodities. However, it is challenged by rising debates on environmental protection and conservation, soil, fertilizer, and water use regulation, and social and economic equity [2]. A promising alternative to optimize the use of fertilizers and irrigation water, avoiding their overuse and waste while increasing crop production on farms, is applying them at different rates across a field, which has been proposed in the precision agriculture (PA) scientific field [3–8].

Many farms manage soil using fixed-rate applications, through which fixed amounts of fertilizers, amendments, and water are applied. This approach considers an area homogeneous, disregarding soil, relief, and plant variations. From the farmers' point of view, it is a practical and convenient method to manage the farm, as it simplifies the management process. On the other hand, this approach may overestimate the inputs in some areas, leading to high costs for farmers and more significant environmental impacts, and underestimate them in other areas, leading to suboptimal production and potential soil degradation [3,9–11].

Nevertheless, the delimitation of subregions with distinct soil properties in homogeneous areas has been practiced for a long time in history [12]. The interdisciplinary field of PA recognizes this approach as management zones (MZs). Doerge (1999) defines an MZ as a sub-region of a field that expresses a homogeneous combination of yield-limiting factors for which a single rate of a specific crop input is appropriate [13].

However, one of the most time-consuming and expensive activities for identifying and delineating MZs within an agricultural area for variable-rate applications in PA is obtaining many samples to characterize the variations in soil, relief, and plant properties. Traditionally, as Nawar et al. (2017) [3] describe, the agricultural conditions of the soil are assessed using a limited number of deemed representative samples (usually a few composite samples in an area), which are analyzed in the laboratory via wet chemistry [14]. However, sampling and analyzing soils via wet chemistry is costly, limiting the number of samples. As such, the explicit recognition of variations and the delineation of homogeneous areas inside a plot to support variable-rate input applications needs a better solution to overcome the high costs and time invested in collecting and analyzing many soil samples.

With new equipment sensitive to soil properties, it is possible to distinguish them in productive sub-regions for a given crop with greater efficiency than in the past [14]. Proximal soil sensors (PSSs) have been used to overcome this issue, as they obtain more data, covering a field more efficiently. PSSs constitute geophysical equipment that measures up to 2 m from the target [14]. They can be mounted to vehicles or transported manually, allowing measurements of soil properties correlated with other properties conventionally measured in the laboratory, including clay, carbon, and moisture content [15–17]. Among the properties measured via PSSs, the apparent electrical conductivity (aEC) and magnetic susceptibility (aMS), as well as the equivalent contents of thorium (eTh) and uranium (eU), have been widely studied in soil science [3,18–22].

The aEC data measured using the Geonics EM38-MK2 sensor has been used to identify and map the spatial variation of soils affected by salts due to irrigation [23–25]. This property has also been measured to predict and map soil properties, such as soil texture and organic matter [16,26,27], and to assess soil electrical conductivity variation in a grain-production farm [22]. In addition, gamma-ray proximal sensors have been used to predict and map soil properties at various spatial scales [28–31]. In agricultural areas, proximal gamma-ray sensors such as the Medusa MS1200 have shown potential for predicting

the soil water retention curve of the gravimetric moisture, nitrogen, phosphorus, and potassium content [16,32,33].

The procedure of datafusion expects to use more than one data source with different resolutions to perform a more accurate mapping procedure. This practice is widely studied in soil science and is known as data fusion [15,16,34–36]. Remote sensors mounted on satellites hundreds of kilometers away or on aerial vehicles hundreds of meters from the ground have also been used to model and map soil properties in the field [7,37–39]. Using statistical modeling, they can be combined with proximal sensors to increase the multivariate correlation with soil properties. Fusing data from remote and proximal sensors with different spatial and temporal resolutions may lead to better soil property maps and more robust management zones for PA [16,35,40,41].

De Benedetto et al. (2013) fused data from a hyperspectral remote sensor with a ground penetration radar (GPR) data amplitude, aEC data from a PSS, and vegetation indices from a remote sensor to define management zones in a 300 m × 30 m production field in southeastern Italy [20]. First, they performed individual clustering algorithms for each sensor group and merged the data. Finally, they managed to identify three major management zones. From principal component analysis, they identified that the GPR amplitude data were sufficient to identify one of the management zones. In contrast, two other clusters were identified using the aEC data and the vegetation indexes.

Horney et al. (2005) performed a cookbook-style study, presenting practical MZ implementation methods [23]. They used the aEC dataset obtained via the EM38-MK2 as an example. They studied two areas in the San Joaquin Valley, CA, USA. They defined two objectives to describe a generalizable methodology for other cultivated varieties. Moreover, a second objective was to test the first methodology in experiments with a low germination rate.

Fraisse et al. (2001) used aEC data from the EM38-MK2 and fused it with data derived from a digital elevation model (DEM) obtained via GPS in two study areas of 36 and 28 ha near Centralia, Missouri, USA [18]. Unsupervised classification and principal components analysis of the aEC data and topographic attributes derived from the DEM were used to define clayey soil management zones. They concluded that the ideal number of management zones could vary yearly, depending mainly on the climate and the crop.

Li et al. (2008) used bulk electrical conductivity (EC_b) measurements for the soil profile (0–20 cm) using a portable WET (W stands for water, E electrical conductivity, and T temperature) sensor, the cotton yield, and normalized difference vegetation index (NDVI) data measured in a 396 grid-sampling scheme to define soil management zones in a 15-ha field on saline coastal land in Zhejiang Province, China [42]. Maps for EC_b and yield produced via geostatistics and the NDVI map were merged using fuzzy c-means clustering to define the MZ modeled to the soil and yield 396 data points via analysis of variance to assess how well the designated management zones reflected the soil properties and yield level. The analysis of variance indicated significant differences between the soil's chemical properties and the cotton yield among the three delineated management zones.

Hence, it was hypothesized that PSSs combined with remote sensing (RS) data can better identify the soil heterogeneity during the MZ mapping process and improve operations in PA procedures. During fieldwork, the operator follows transects and generates considerable data points. There is a frequent question about the optimal density and spacing of the transects, which involves a trade-off between the intensity and costs of the fieldwork and the accuracy of the resulting MZ maps. The objective of this study was to compare MZ maps produced via ordinary kriging (OK) using aEC, aMS, eTh, and eU maps and kriging with external drift (KED), considering them together with RS satellite images and digital-elevation-model-derived covariates to evaluate the potential of using these covariates together with a different mapping procedure. Finally, both MZ maps' accuracy was assessed to determine whether fusing proximal and remote sensor data fusion increases the accuracy in MZ delineation by relatively improving the index.

2. Materials and Methods

2.1. Study Area

The study area was a crop field under center pivot irrigation of 72 ha. It was in the municipality of Itaí, in the state of São Paulo, southeastern Brazil (Figure 1A–C), with central coordinates 23.5854° south and 48.9395° west and elevation of approximately 685 m (Figure 1D). The field was cultivated with a no-till crop rotation system of wheat, bean, and oat. In the study pivot area, the soils were described explicitly as a Latossolo Vermelho Distrófico típico ("Ferralsols" in WRB (FAO, 2015)), with a clayey texture in the surface layer (515 g kg^{-1}) and a very clayey texture in the subsurface (600 g kg^{-1}) [43].
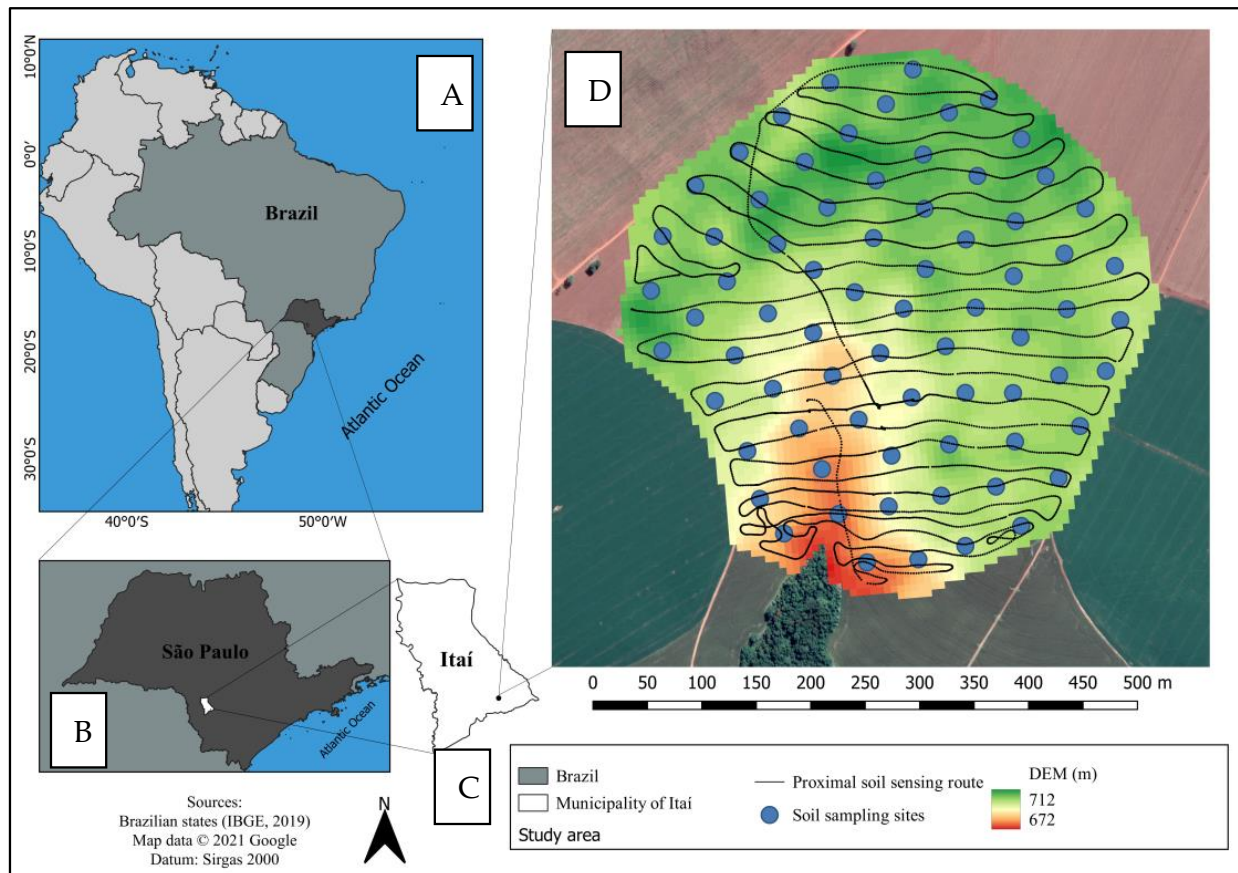
Figure 1. (**A**) Map of Brazil in Latin America; (**B**) contour of the state of São Paulo; (**C**) contour of the municipality of Itaí; (**D**) map of the study area with a digital elevation model in the background, the vehicle route equipped with proximal soil sensors (PSSs), and points collected for analysis in a laboratory.

To assess the region's climatology, two weather stations close enough to contribute to the temperature estimate at Itaí: Bauru Airport (62%, 121 km north) and Viracopos International Airport (38%, 168 km east). The historical models were built with data from 1 January 1980 to 31 December 2016. The season with the most significant precipitation occurs from October to March. However, it rains throughout the year in Itaí at a monthly average of 190 mm. Throughout the year, the temperature generally ranges from 14 °C to 30 °C and is rarely below 10 °C or above 34 °C.

2.2. Proximal Soil Sensor Data Preparation

Apparent electrical conductivity (aEC) and magnetic susceptibility (aMS) were measured using an EM38-MK2 electromagnetic induction conductivity meter (Geonics Ltd., Mississauga, ON, Canada) in Vertical Mode with 1-m coil spacing. This sensor was placed on a foam cushion inside a wooden box free of metallic parts mounted on a 1-cm-thick

rubber mat. The rubber mat was tied to the rear of a 4×4 pickup truck using a 3-m-long nylon ribbon, avoiding metallic interference from the truck (Figure 2A). In addition, equivalent thorium (eTh) and uranium (eU) contents were measured using the gamma-ray spectrometer Medusa MS1200 (Medusa, Groningen, the Netherlands) mounted on the brush grille guard of the vehicle and tied with a nylon ribbon, staying approximately 50 cm from the ground (Figure 2B).

Figure 2. (**A**) MS1200 gamma-ray spectrometer mounted to the vehicle's brush grille guard; (**B**) EM38-MK2 sensor cushioned inside a wooden box mounted on a rubber mat and tied to the vehicle's rear; (**C**) fully equipped vehicle showing the irrigation line in the background.

The field campaign to collect data from proximal sensors occurred right after harvest, with the soil covered by leftover straw. The fully equipped vehicle (Figure 2C) was driven along parallel survey lines about 40 m apart, totaling 26 lines, at a constant speed of approximately 15 km/h, totaling 90 min for the operation. Measurements were taken from both sensors every second.

The EM38-MK2 sensor measured aEC and aMS at 5790 points, while eTh and eU via the MS1200 sensor were measured at 6563 points. The difference between the total values measured using the EM38-MK2 and the Medusa MS1200 was due to the moments required to start the measurement process with each sensor. During the original EM38-MK2 data exploratory analysis, highly conductive aEC outliers and corresponding aMS samples close to the irrigation equipment were identified and removed. Also, some locations were oversampled due to vehicle stops and other maneuvers during data acquisition. The *zerodist* function from the *sp* package [44] in the software R (version 3.4.2) [45] was applied to filter, in both EM38-MK2 and MS1200 datasets, data recursively measured at the exact location.

After data filtering, the remaining PSS dataset had 4306 and 4896 samples from the EM38-MK2 and MS1200 sensors, respectively. The EM38-MK2 sensor data showed more values to be removed due to the more significant influence of the metal parts of the irrigation pivot close to the vehicle path. The Medusa MS1200 data set only had the points recursively measured at the exact location removed, allowing for a more significant number of points to be maintained.

Before the mapping procedure, the last preprocessing step was to reserve a set of data points from the PSSs to validate the maps of the attributes measured using the PSS produced by OK and the KED. The *sample* function of the *dplyr* package [46] was used

in the software R [45] to select the external validation dataset of the produced maps. Four hundred points were reserved for all datasets of the two PSSs, and the accuracy of the produced maps was evaluated using the root of the mean square error (RMSE, Equation (1)) index and the external validation dataset.

$$\text{RMSE} = \sqrt{\sum_{i=1}^{N} \frac{(y - \hat{y})^2}{N}}, \tag{1}$$

where N is the number of observations, y is the original value, and $\hat{y}$ is the predicted value.

2.3. Remote Sensor Data Preparation

Since the data from the PSSs were obtained in the first field campaign (September 2018) and the data analyzed in the laboratory were from the second campaign (October 2019), it was decided to select RS satellite data for the two dates of the campaigns to contemplate the two scenarios in which the soil of the area had been used. More about the soil sampling data for laboratory analysis is better described below. This approach consisted of maintaining a pool of orbital image data to allow the selection of possible correlations and influences between the spectrum values of the bands with the proximal sensor variables so as not to restrict the soil scenario to only that without till or cover [47].

The DEM obtained from the Alos Palsar satellite was used to complement the information about the RS covariables. Ten relief covariables were derived using the R software's *RSAGA* package [48]. The description of the satellite images and the relief variables derived from the DEM and their abbreviations can be seen in Table 1. Also, the satellite images described in Table 1 are summarized in abbreviation given that two scenarios were contemplated (2018 and 2019) during the analysis and modeling.

All the different bands' satellites and all the rasters of the covariates derived from the DEM were resampled to the final resolution of 10 m so that it was possible to stack them and extract the information from each raster to the respective coordinates of the data of PSS. For this, the *resample* function of the *raster* package [49] using the bilinear interpolation method in the R software and the extraction process were performed using the *extract* function in the *raster* package.

2.4. Selection of Remote Sensing Covariates

The many remote sensor covariates required filtering to identify those most significant to integrate the PSS data and optimize the model. These models would be used to build maps using RS and the PSSs via KED and finally define the MZ maps. The following sections better describe the fusing process to determine the MZ.

The regression method via linear models was used to model data from the PSSs as a function of RS data, and multicollinearity was minimized. The number of covariates was reduced using Pearson's correlation r values between all predictive covariates (the remote sensing satellite and DEM derivates). For this, all variables that showed Pearson's r values greater than 0.9 absolute were identified, and then these covariates were correlated again with the PSSs dataset. Those with the highest Pearson correlation absolute r values were maintained.

Finally, the remaining group of RS was applied to the *regsubsets* function present in the *leaps* package [50] in the R software [45]. This function allows for testing all possible linear combinations between the remaining satellite covariates with the proximal sensors' data. Then, the combination that presented the lowest value for Schwarz's information criterion (BIC) [51] was selected to integrate the prediction model of the PSS attributes as a function of the RS data.

2.5. Mapping via Ordinary Kriging and Kriging with External Drift

It's observed that different interpolation methods can generate more reliable thematic maps to represent specific areas, especially in locations with limited sampling units [52].

Thus, two sections describe the mapping of the original data from the PSSs and the predicted data from the PSSs according to the data from RS selected from the covariate selection procedure described above. First, the original PSS data were mapped using the ordinary kriging method (OK). Then, the data from the PSSs predicted via the selected RS covariables were mapped using kriging with external drift (KED).

Table 1. Resolution and description of the bands of the images used and covariates derived from the DEM.

Resolution	Satellite	Band	Wavelength	Abbreviation	Variables
15 m	Terra-1	VNIR	0.52–0.60 0.63–0.69 0.78–0.86	ast_B1 ast_B2 ast_B3N	Aster
10 m		Blue Green Red NIR	0.44–0.53 0.54–0.58 0.65–0.69 0.77–0.91	sent_year_B2 sent_year_B3 sent_year_B4 sent_year_B8	Sentinel 2
20 m	ESA	VRE NIR SWIR	0.69–0.71 0.73–0.75 0.77–0.80 0.85–0.88 1.34–1.41 2.07–2.31	sent_year_B5 sent_year_B6 sent_year_B7 sent_year_B8A sent_year_B11 sent_year_B12	
30 m	NASA	C/A Blue Green Red NIR SWIR TIRS	0.43–0.45 0.45–0.51 0.53–0.59 0.64–0.67 0.85–0.88 1.57–1.65 2.11–2.29 10.60–11.19 11.50–12.51	land_year_B1 land_year_B2 land_year_B3 land_year_B4 land_year_B5 land_year_B6 land_year_B7 land_year_B10 land_year_B11	Landsat 8
Resolution	Reference	Type	Abbreviation		Variables
12.5 m	Calculated from DEM	Quantitative	aspect DEM slope plan_curv prof_curv convergence twi ls_fator rsp chnd chnb		Aspect Elevation (m) Slope (%) Curvature plan Curvature depth Convergence Topographic Wetness index Length-slope factor Relative slope Position Channel network distance Channel network base level

SWIR: shortwave infrared; VRE: vegetation red edge; NIR: near-infrared; C/A: coastal/aerosol; TIRS: thermal infrared; ESA: the European Space Agency; NASA: the National Aeronautics and Space Administration; the _year_B*name* refers to 2018 and 2019 and the band spectrum length name.

2.5.1. Proximal Soil Sensing Mapped via Ordinary Kriging

All data from the PSSs were assessed for spatial distribution patterns to evaluate the normality distribution to interpolation using OK. Using isotropic adjustment variograms, the least-squares reduction adjustment procedure was performed via variogram analysis tools in the *gstat* package [53] in the R software. The semivariogram works to estimate a value for a region in which the semivariogram distance is known using data in the neighborhood of the estimation location [54].

The semivariograms of the aEC, aMS, and eTh data were adjusted using the spherical model (Equation (2), while the eU data were adjusted for the Gaussian model (Equation (3); then, these data were interpolated using OK (Equation (4) [55]. Finally, all data from the

proximal sensors (aEC, aMS, eTh, and eU) were interpolated using a spatial resolution of 10 m.

$$\gamma(h) = \left\{ c\left\{ \frac{3h}{2a} - \frac{1}{2}(h/a)^3 \right\} \atop c \right\}, \begin{matrix} h \leqslant a \\ h \geq a' \end{matrix} \tag{2}$$

where c is the sill variance, and a is the range. h is the separator, being the vector of the separation of the pairs of points in which $h = x_i - x_j$ and where $C(x_i, x_j) = \sum[\{Z(x_i - \mu)\}\{Z(x_j) - \mu\}]$, which is a constant for any h. This constancy of the mean, variance, and covariance depends only on the separation and not the absolute position [55].

$$\gamma(h) = c\left\{ 1 - exp\left(\frac{-h^2}{r^2} \right) \right\}, \tag{3}$$

where c is the sill variance, r is the distance parameter, and h is the vector of the separation of pairs of points [56].

$$Z_{KO}^{\circ}(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i), \tag{4}$$

where $Z_{KO}^{\circ}(x_0)$ is the value of the random variable to be estimated via ordinary kriging (OK), λ_i is the optimal weights calculated under two constraint conditions ((a) the estimator is not skewed and (b) the variance of the estimate is minimal), and $Z(x_i)$ is the values of the random variable at the n sample points [56].

2.5.2. Predicting and Mapping Proximal Soil Sensing Data via Kriging with External Drift

The kriging method with external drift (KED) consists of identifying the presence of a trend. Therefore, it is necessary to show how to separate the deterministic trend from the random component and estimate the components' contribution via restricted maximum likelihood [56]. The external drift method integrates into the kriging system supplementary universality conditions about one or several external drift variables measured exhaustively in the spatial domain [57].

If a target variable's tendency is the spatial coordinates, it indicates that this may be one of several cases of mixed linear models to be considered, for example. The deterministic effects may be modeled by others like $y_1, y_2, \ldots$. In the present study, these variables were the RS covariables, which can be linearly related to Z. In this case, Z will be the PSS attribute. Hence, measuring and calculating the target locations and the Z points is possible. Under these circumstances, it is possible to observe Equation (5).

$$Z(x) = \sum_{k=0}^{K} \beta_k y_k(x) + \varepsilon(x) = \beta_0 + \beta_1 y_1(x) + \beta_2 y_2(x) + \ldots + \beta_k y_k(x) + \varepsilon(x), \tag{5}$$

What was done was to replace the coordinates in the function with the values of one or more variables. The $y\_1_{(x)}, y\_2_{(x)}, \ldots, y\_k_{(x)}$ were known, and β_k was the unknown coefficients to be determined [54]. y_k, being $k = 1, 2, \ldots$ was the "external" variables distinct from the internal Z_x and kriged with them; therefore, it was known as "kriging with external drift". The KED incorporates the local trend within a neighborhood search window as a linear function of a secondary variable that varies smoothly, and the trend of the primary variable must be linearly related to the secondary variable [57].

Since the primary variable trend must be linearly related to the secondary variable, only the proximal sensor aEC and aMS data presented satisfactory R^2 and adjustable R^2 values of linear model adjustments (>0.7) using RS data as covariates. The eTh and eU data did not show a significant linear relationship, with no possible combination of RS data used in the present study. Although KED was not performed for eU and eTh, the best models were adjusted and would be shown, and these attributes would be mapped using OK only. Thus, only the aEC and aMS data were interpolated with KED using the *krige* function of the *gstat* package [53] associated with the pre-selected satellite covariate models. All maps were interpolated using 10 m as the spatial output resolution.

The regression modeling, followed by ordinary kriging of the residues for the aEC and aMS maps, was tested since these two attributes presented high adjusted R^2 values.

For this, the simple linear model (*lm* function) and the *predicted* functions present in the R software [45] were used to perform the regression after applying the selected regression model to the entire spatial pixel area of the RS covariates. Then, these maps were also evaluated for accuracy through the external validation RMSE index using the 400 previously separated samples.

2.6. Management Zones

Three management zone maps (MZs) were defined based on three combinations. The first MZ map was built using only the PSS maps via OK. Then, the second MZ map was constructed from the PSS associated with the best RS covariates through KED mapping. In this second MZ map, the eTh and eU data were used, but the OK map version was considered since KED did not show high linear adjustment. Finally, the third MZ map was built using all the RS covariates described in Table 1.

The *kmeans* function in the *stats* package presented natively in the R software [45] was used. The algorithm chosen within this function used the calculation of Hartigan and Wong (1979) [58]. The three datasets (PSS; PSS + RS; RS) were organized in a data matrix format and grouped using the *k*-means method. This method partitioned the dataset into *k* groups. The sum of the distance for the centers of the designated clusters was minimized. In the present study, three zones were chosen for *k* since many zones would not be operationally practical for the farmers. The *k*-means clustering algorithm aims to partition *n* observations into *k* clusters, where each observation will belong to a cluster with the average value closest to each other. The centers are in the mean of their sets of Voronoi polygons, which are the sets of points closest to each cluster's center. The value "1" was set to reproduce the same arrangement using the function *set.seed* present in the *stats* package in the R software [45].

Considering that the PSS and RS data were obtained from different sources all data used in the *kmeans* function were parameterized for the zero values of mean and variance one using the *scale* function present in the *stats* package of the R software.

2.7. Soil Laboratory Data Sampling as Field Truth

Soil data collected traditionally were used to validate the design of the three MZ maps. Seventy-two disturbed samples were collected at the soil surface (0–10 cm depth) and arranged in a regular grid (one sample per hectare; Figure 1C) using a Dutch auger (Figure 3A) in a second field campaign.
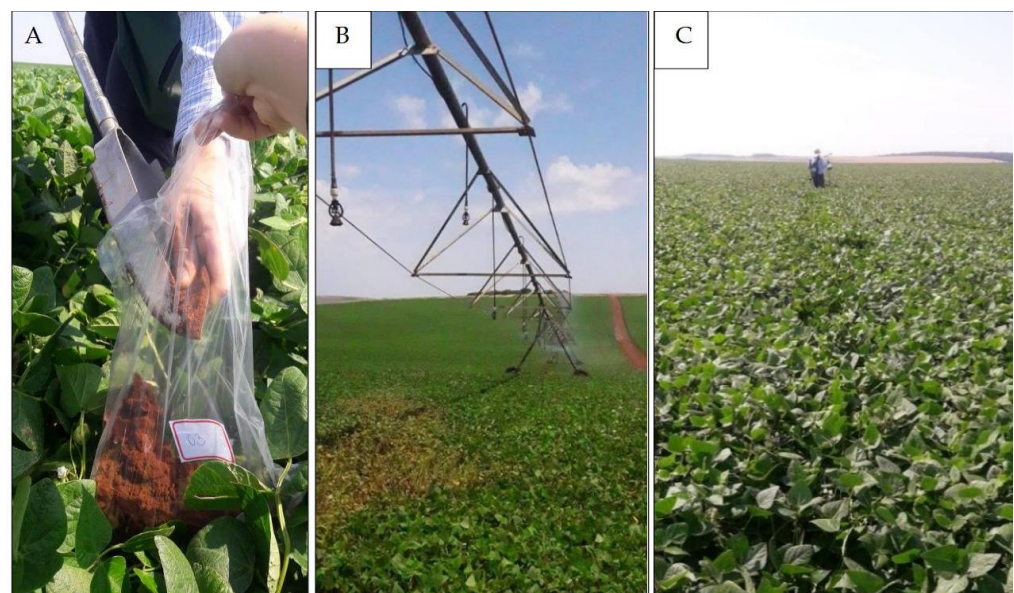
Figure 3. (**A**) Soil sampling at 0–10 cm using a Dutch auger; (**B**) study area in September 2019 with irrigated beans; (**C**) view of the study area showing the bean plantation.

This campaign was carried out on 6 October 2019, and beans covered the study area in an early stage of growth (Figure 3B,C). The stratify function in the *spcosa* package [59] was used to define and locate the points in the regular grid. This function allows to separate an area of interest in sub-areas with exact sizes, and then the spsample function present in the same package was used to locate the center of each section of the area.

The 72 disturbed samples were analyzed for soil texture and chemical contents. For texture analysis, the clay contents were measured using the sieve-and-pipette method Teixeira et al., 2017 [60].

2.8. Management Zone Validation

The 72 points analyzed in the laboratory validated the three MZ maps produced from PSSs mapped via OK, using PSS data combined with the best RS covariate models mapped via KED, and using only RS data. The zone classes were extracted from the three MZ maps for the 72 coordinates associated with the laboratory data using the *extract* function present in the *raster* package [49] in the R software [45].

Finally, the variance analysis method (ANOVA) was used to identify whether the management zones could distinguish the variance of the values of the laboratory attributes regarding their classes of management zones. The aov function of the stats package in the R software [45] was used for that.

3. Results and Discussion

The descriptive statistics of the PSS data for the training and validation dataset and the laboratory data are shown in Table 2. For all PSS datasets, the mean and median values were close. Also, the standard deviation values did not contrast much. Therefore, the asymmetry values for all PSS data can be considered moderately positive except for the aMS data. The kurtosis coefficients for the eTh and eU data, including validation data, can be regarded as leptokurtic. The coefficient corresponds to a less flat distribution than the same area's standard curve, as they are below the absolute value of 0.263 [61].

Table 2. Descriptive statistic of proximal soil sensing and laboratory data.

Kurtosis	Skewness	Std. Deviation	Variance	Median	Mean	Max.	Min.	N. Obs.	
				aEC (mS/m)					
0.22	0.68	3.39	11.49	9.3	9.58	26.25	2.62	3906	Training
0.92	0.78	3.35	11.2	9.53	9.62	25.31	3.63	400	Validation
				aMS (ppt)					
−0.80	−0.06	0.65	0.42	2.24	2.27	3.88	0.44	3906	Training
−0.58	−0.03	0.64	0.41	2.19	2.25	3.81	0.5	400	Validation
				eTh (ppm)					
−0.03	0.11	3.03	9.16	15.94	16	27.81	6.08	4496	Training
−0.04	0.21	3.24	10.47	15.78	15.97	26.17	7.02	400	Validation
				eU (ppm)					
0.06	0.09	1.21	1.46	3.12	3.14	7.26	−0.99	4496	Training
0.13	0.25	1.23	1.52	3.03	3.1	7.72	0.08	400	Validation
				Laboratory dataset					
1.42	−1.20	45.28	2050.7	420	413.33	500	280	72	Clay (g kg^{-1})
−0.34	0.26	0.88	0.77	6.25	6.31	8.5	4.6	72	Ca (g kg^{-1})
−0.50	0.28	1.66	2.74	14.8	15	19.15	10.99	72	C (g kg^{-1})
−0.33	0.24	0.28	0.08	1.9	1.93	2.6	1.4	72	Mg (g kg^{-1})

aEC: apparent electrical conductivity in mS/m (millisiemens per meter); aMS: apparent magnetic susceptibility in ppt (parts per thousand); eTh: equivalent thorium in ppm (parts per million); eU: equivalent uranium in ppm (parts per million); clay in g kg^{-1}; Ca: calcium in g kg^{-1}; C: organic carbon in g kg^{-1}; Mg: magnesium in g kg^{-1}.

The distribution pattern of the aEC and aMS attributes and laboratory data could be classified as platykurtic, corresponding to a flatter distribution than the normal curve of the same area, except for aEC for the leptokurtic training dataset. Despite the asymmetry

and kurtosis values of the aEC data, it was impossible to consider a normal distribution, and therefore, for mapping via OK, it was necessary to transform the data to fit the normal distribution using natural logarithmic distribution. The other attributes of the PSS (aMS, eTh, and eU) and the separate data for external validation showed a normal distribution. The distribution pattern of the values for all data sensors (the training and validation dataset) and laboratory datasets is shown in Figure 4.
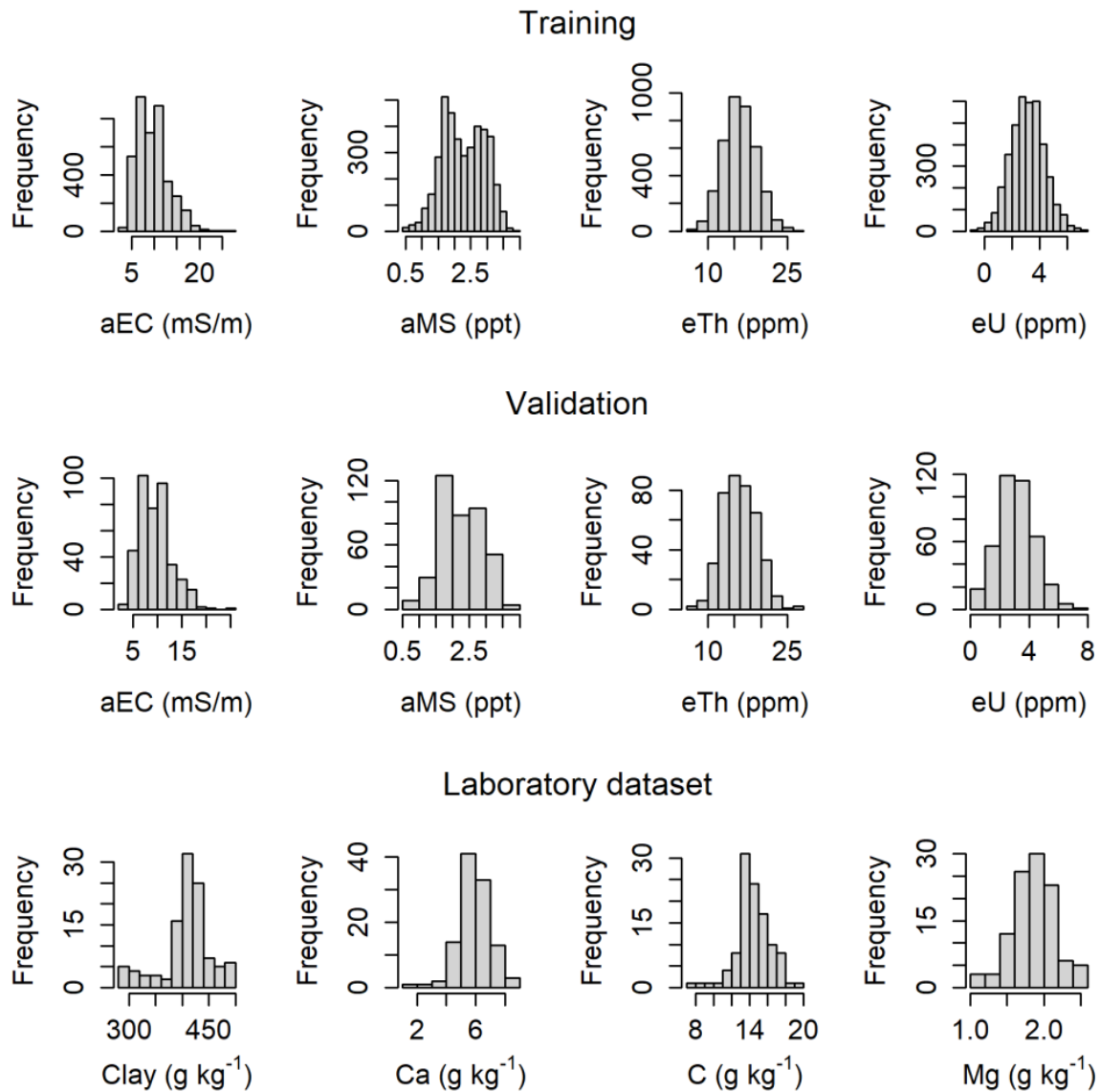
Figure 4. aEC: apparent electrical conductivity in mS/m (millisiemens per meter); aMS: apparent magnetic susceptibility in ppt (parts per thousand); eTh: equivalent thorium in ppm (parts per million); eU: equivalent uranium in ppm (parts per million); clay in g kg^{-1}; Ca: calcium in g kg^{-1}; C: organic carbon in g kg^{-1}; Mg: magnesium in g kg^{-1}. First line: histograms of proximal sensor data used for mapping via OK and KED; second line: histograms of proximal sensor data used to validate the OK and KED maps; third line: histograms of soil data collected with a Dutch auger.

The electrical conductivity and magnetic susceptibility values were similar to those presented by Castrignanò et al., 2012 [62] on an 80-ha cropping field in Corrigin, Western Australia. The thorium and uranium values were similar to those found in the study by Wong et al., 2009 [63], in which the lowest values of eTh and eU occurred as opposed to the highest values of electrical conductivity for a 200-ha experimental study area field situated 350 km north of Perth at Buntine, Western Australia.

The prediction models adjusted for the sensor attributes are shown in Table 3. The RS covariates selected from the selection procedure described in Section 2.4 via the regsubset function and lower BIC values are shown in the "Coefficient" column of Table 3 (first column). The intercept values represent the estimated response variable values when all predictors were zero. Notably, the aspect variable showed a positive association with log(aEC), indicating that an increase in aspect was linked to a slight rise in electrical conductivity (log(aEC)) by 0.01 mS/m. Conversely, the variable "chnb" exhibited negative associations with log(aEC) (−0.02 mS/m), aMS (−0.02 ppt), and eTh (−0.24 ppm), suggesting that higher values of "chnb" were associated with decreased values of these response variables. The DEM (Digital Elevation Model) variable demonstrated positive relationships with aMS (0.09 ppt), eTh (0.33 ppm), and eU (0.02 ppm), implying that an increase in DEM led to higher values of these response variables. Plan curvature ("plan_curv") negatively influenced aMS, indicating that higher plan curvature significantly decreased aMS by 15.63 ppt. Additionally, the variable "twi" showed a positive association with log(aEC) (0.01 mS/m), suggesting that increased topographic wetness index (twi) values were linked to higher electrical conductivity.

Table 3. Linear models' adjustment parameters of the proximal sensor attribute as a function of the remote sensing data using the best combination from Pearson's selection in addition to the BIC criterion and, finally, using a regsubset function.

eU (ppm)		eTh (ppm)		aMS (ppt)		log(aEC) (mS/m)		
Conf. Int (95%)	Estimates	Conf. Int (95%)	Estimates	Conf. Int (95%)	Estimates	Conf. Int (95%)	Estimates	Coefficient
−19.43−−4.22	−11.83	−48.95−−7.80	−28.37	−38.65−−33.89	−36.27	19.23−21.62	20.42	(Intercept)
						0.01−0.02	0.01	aspect
						0.01−0.00	0	ast_B1
						0.00−0.00	0	land_2018_B3
				0.00−0.00	0	0.00−0.00	0	land_2018_B5
						0.00−0.00	0	land_2018_B10
						0.00−0.00	0	land_2019_B3
						0.00−0.00	0	sent_2018_B8A
−0.00−−0.00	0	−0.01−−0.01	−0.01	−0.00−−0.00	0	0.00−0.00	0	sent_2018_B12
				−0.00−−0.00	0	0.00−0.00	0	sent_2019_B2
				−0.00−−0.00	0	0.00−0.00	0	sent_2019_B3
						0.00−0.00	0	sent_2019_B8A
		−0.34−−0.15	−0.24	−0.03−−0.01	−0.02	−0.02−−0.02	−0.02	chnb
0.01−0.03	0.02	0.23−0.44	0.33	0.08−0.11	0.09			DEM
				−21.62−−9.64	−15.63			plan_curv
				0.00−0.02	0.01			twi
−0.94−−0.35	−0.65	−8.03−−4.82	−6.42	−1.36−−1.00	−1.18			rsp
				−0.01−−0.00	0			ast_B3N
		0.00−0.00	0	0.00−0.00	0			land_2018_B7
0.00−0.00	0			−0.00−−0.00	0			land_2019_B5
				−0.00−−0.00	0			land_2019_B7
0.00−0.00	0	0.00−0.01	0.01	−0.00−−0.00	0			sent_2018_B2
				0.00−0.00	0			sent_2018_B4
		0.00−0.01	0.01	0.00−0.00	0			sent_2019_B11
		−0.00−−0.00	0					land_2019_B2
0.00−0.00	0							sent_2019_B4
4496		4496		3906		3906		Observations
0.11/0.11		0.20/0.20		0.81/0.81		0.78/0.78		R^2/adjusted R^2

aEC: apparent electrical conductivity in mS/m (millisiemens per meter); aMS: apparent magnetic susceptibility in ppt (parts per thousand); eTh: equivalent thorium in ppm (parts per million); eU: equivalent uranium in ppm (parts per million).

The variable "rsp" exhibited negative associations with log(aEC) (−1.18 mS/m), aMS (−6.42 ppt), and eTh (−0.65 ppm), indicating that higher "rsp" values led to decreased values in these response variables. Table 3 also highlights the importance of context-specific interpretation, as certain variables, such as "ast_B1", "land_2018_B3", "land_2018_B5", and others, did not show significant associations with any of the response variables within the specified confidence intervals. Considering the observed R-squared values ranging from

0.20 to 0.81, the regression models explained a substantial proportion of the variance in the respective response variables. Adjusted R-squared values provide a more accurate measure of goodness of fit by considering the number of predictors, ensuring a balanced evaluation of the model's explanatory power. The R^2 and adjusted R^2 values for the aEC and aMS attributes were close to 0.8, indicating a high potential correlation with the identified satellite attributes. Finally, the graphs containing the predicted and original values for the PSS attributes according to the best models of RS covariates are shown in Figure 5.
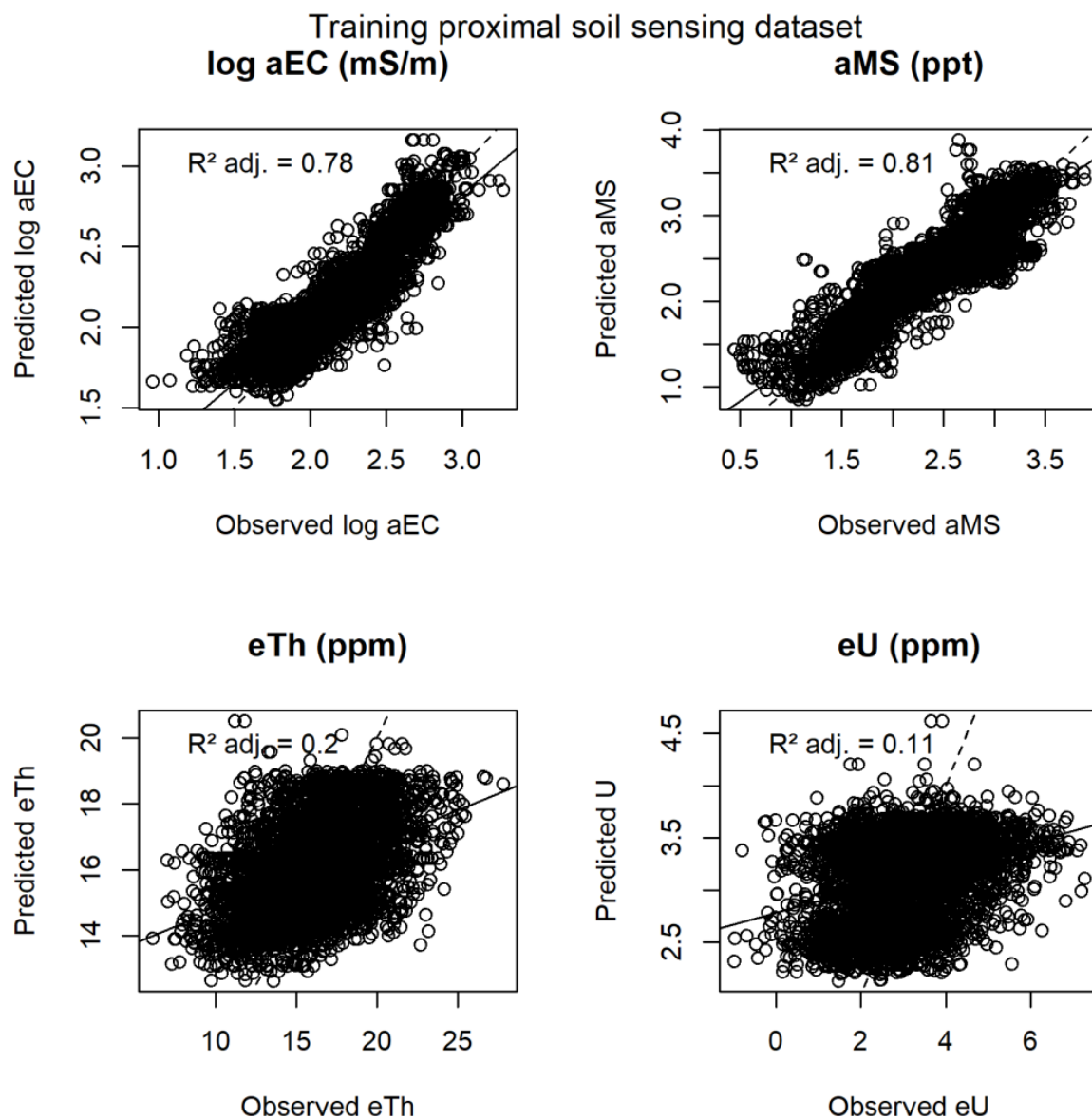
Figure 5. Predicted versus experimental plots of proximal sensor data as a function of remote sensor data using the training dataset. The continuous black lines adjust the intercept and slope for the models, while the dashed lines are the intercept and model idealized as 1 and 0, respectively. R^2 adj.: R^2 adjusted value; aEC: apparent electrical conductivity in mS/m (millisiemens per meter); aMS: apparent magnetic susceptibility in ppt (parts per thousand); eTh: equivalent thorium in ppm (parts per million); eU: equivalent uranium in ppm (parts per million).

Since the aEC and aMS data were well adjusted for the RS covariates, pixel-by-pixel regression from the created estimation model was also used. However, this method did not show significant results when using the external validation dataset through the RMSE index to evaluate the aEC and aMS maps. There was also no spatial dependence on the

predicted map residues, so the regression procedure and subsequent ordinary kriging of the residues could not be performed. Then, it was considered to remain with the KED method to compose the maps from which the MZ would be delineated.

Semivariograms were adjusted by removing the RS covariates' trend, as shown in Figure 6. The parameters used in the PSS semivariograms to map using OK and KED are shown in Table 4. All attributes of the aEC, aMS, and eTh data were adjusted for the spherical model (Figure 6), while the sensor eU was better adapted for the Gaussian model (Figure 6). The low values of the nugget parameter may have indicated a reduction in the values of variance and standard deviation of the maps.
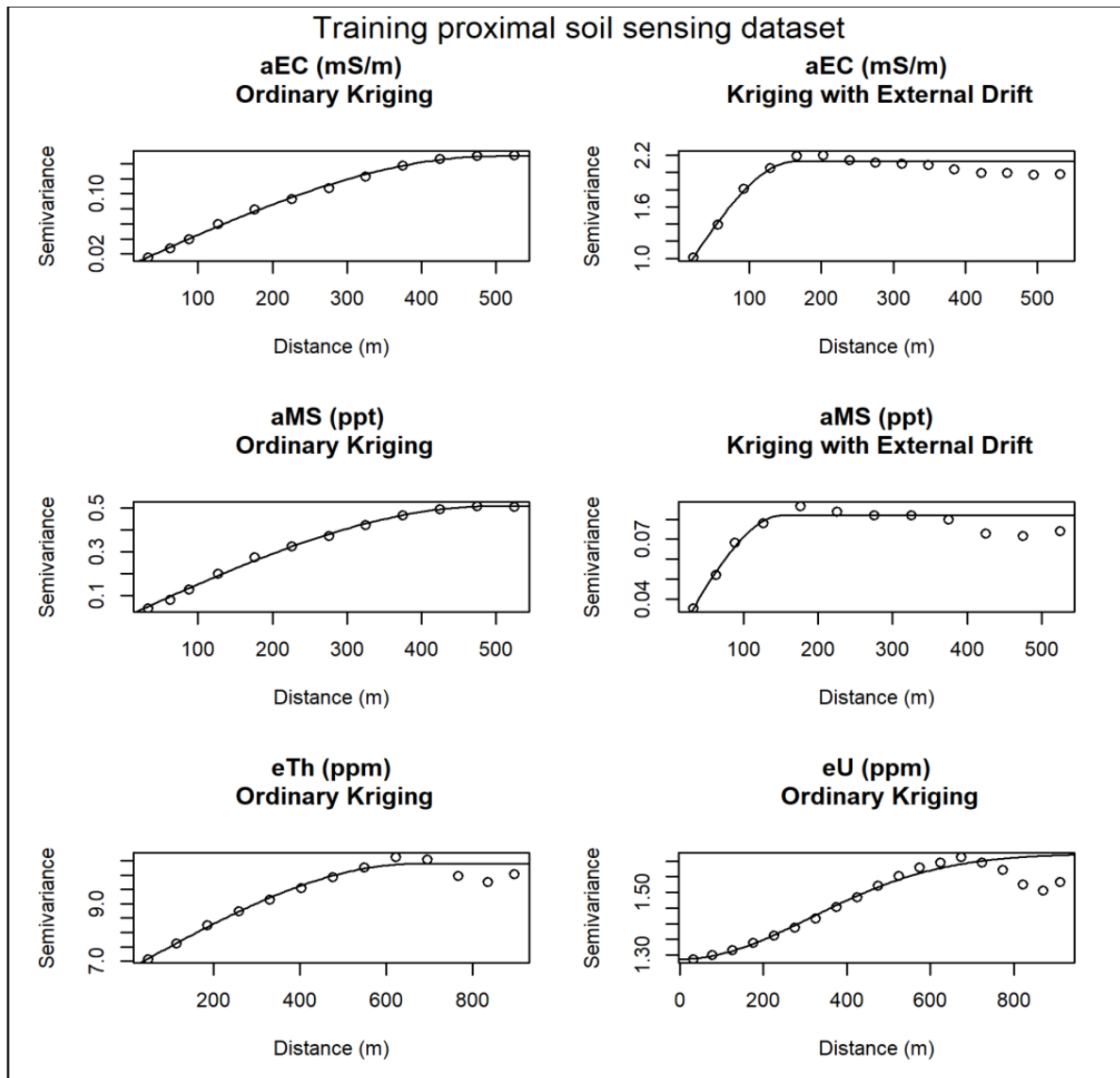
Figure 6. Empirical (circles) and adjusted (lines) semivariograms of proximal sensor variables aEC: apparent electrical conductivity in mS/m (millisiemens per meter); aMS: apparent magnetic susceptibility in ppt (parts per thousand); eTh: equivalent thorium in ppm (parts per million); eU: equivalent uranium in ppm (parts per million). First line: semivariogram of aEC using OK and KED; second line: semivariogram of aMS using OK and KED; third line: semivariogram of eTh and eU using OK. There was no adjustment of the semivariogram via KED for the last two attributes of the gamma radiometer. It did not reasonably adjust regression models in the previous steps.

The nugget/sill ratio is shown in Table 4, column 6, to present the nugget parameter's impact on the spatial structure of variation in the final representation of the semivariance.

Those ratio values were multiplied by 100 to express percentages, allowing us to better understand and discuss the errors of semivariance. The KED method for the attributes of the aEC and aMS sensors presented 34% and 17%, respectively. These values were higher than the ratio when using the OK method. The analysis of this ratio suggested that the distance of 150 m, represented by the aEC data set, for example, was the maximum adequate distance for data collection with sensors since, from the removal of the trend identified using the RS covariates, from this distance, there may have been a significant influence of randomness on the semivariance.

The interpolated maps showed similar distribution patterns for the OK method (Figure 7A) and KED (Figure 7B). The high concentrations of aEC were in the southwestern portion of the map (Figure 7A), where an ephemeral drainage channel could be observed. Although the survey with the EM38-MK2 was carried out in a dry season (July), with the soil not cultivated (therefore, without a pivot irrigation schedule) and only protected with previous harvest wheat straw, the high concentration values of aEC may have been associated with higher soil moisture related with the clay presence concentrations, as found in similar studies by Fulton et al. (2011) [64] and Islam et al. (2011) [65]. However, in the present study, the values of this sensor attribute did not have significant Pearson correlation values (−0.17) with clay (Table 5). In contrast, the lower aEC values in the northern part of the study area suggest a well-drained region with higher elevation and more sandy soils. The aEC data showed high correlation values with the chemical attributes of C and Mg. In contrast, the attributes measured via gamma radiometric sensors showed high correlation values for all measured attributes, among which C and clay presented the highest (0.57 and 0.46, respectively).

Table 4. Semivariogram parameters for proximal sensor variables for ordinary kriging (OK) and kriging with an external drift (KED).

Range (m)	Nugget/Sill (%)	Sill	Partial Sill	Nugget	Model	Method
			aEC (mS/m)			
163.33	34.03	2.13	1.4	0.72	Spherical	KED
500.31	0.96	0.15	0.15	0		OK (log format)
			aMS (ppt)			
152.77	16.78	0.08	0.07	0.01	Spherical	KED
495.37	1.93	0.52	0.51	0.01		OK
			eTh (ppm)			
668.18	64.44	10.39	3.7	6.69	Spherical	OK
			eU (ppm)			
443.37	79.27	1.62	0.34	1.29	Gaussian	OK

aEC: apparent electrical conductivity in mS/m (millisiemens per meter); aMS: apparent magnetic susceptibility in ppt (parts per thousand); eTh: equivalent thorium in ppm (parts per million); eU: equivalent uranium in ppm (parts per million).

The clay content showed a weak positive correlation with the calcium content (r = 0.19) and moderate positive correlations with carbon content (r = 0.28 *), magnesium content (r = 0.25 *), apparent magnetic susceptibility (r = 0.28 *), equivalent thorium concentration (r = 0.50 **), and equivalent uranium concentration (r = 0.46 **). Calcium content exhibited strong positive correlations with carbon content (r = 0.75 **), moderate positive correlations with magnesium content (r = 0.60 **), apparent magnetic susceptibility (r = 0.28 *), and equivalent thorium concentration (r = 0.42 **), and a slightly weaker positive correlation with equivalent uranium concentration (r = 0.43 **). Carbon content showed moderate positive correlations with magnesium content (r = 0.58 **), moderate negative correlations with apparent electrical conductivity (r = −0.42 **), and strong positive correlations with apparent magnetic susceptibility (r = 0.48 **), equivalent thorium concentration

(r = 0.65 **), and equivalent uranium concentration (r = 0.57 **). Magnesium content had a weak negative correlation with apparent electrical conductivity (r = −0.24 *), moderate positive correlations with apparent magnetic susceptibility (r = 0.32 **) and equivalent thorium concentration (r = 0.44 **), and equivalent uranium concentration (r = 0.39 **). Apparent electrical conductivity exhibited strong negative correlations with apparent magnetic susceptibility (r = −0.93 **), moderate negative correlations with equivalent thorium concentration (r = −0.68 **), and weak negative correlations with equivalent uranium concentration (r = −0.28 *). Apparent magnetic susceptibility showed strong positive correlations with equivalent uranium concentration (r = 0.79 **) and moderate positive correlations with equivalent thorium concentration (r = 0.76 **).

Table 5. r of Pearson correlation between the proximal and laboratory data variables.

| eU (ppm)	eTh (ppm)	aMS (ppt)	aEC (mS/m)	Mg (g kg^{-1})	C (g kg^{-1})	Ca (g kg^{-1})	Clay (g kg^{-1})	
0.46 **	0.50 **	0.28 *	−0.17	0.25 *	0.28 *	0.19	1	Clay (g kg^{-1})
0.43 **	0.42 **	0.28 *	−0.15	0.60 **	0.75 **	1		Ca (g kg^{-1})
0.57 **	0.65 **	0.48 **	−0.42 **	0.58 **	1			C (g kg^{-1})
0.39 **	0.44 **	0.32 **	−0.24 *	1				Mg (g kg^{-1})
−0.28 *	−0.68 **	−0.93 **	1					aEC (mS/m)
0.45 **	0.79 **	1						aMS (ppt)
0.76 **	1							eTh (ppm)
1								eU (ppm)

aEC: apparent electrical conductivity in mS/m (millisiemens per meter); aMS: apparent magnetic susceptibility in ppt (parts per thousand); eTh: equivalent thorium in ppm (parts per million); eU: equivalent uranium in ppm (parts per million); clay in g kg^{-1}; Ca: calcium in g kg^{-1}; C: organic carbon in g kg^{-1}; Mg: magnesium in g kg^{-1}; levels of statistical significance: * $p < 0.05$ and ** $p < 0.01$.
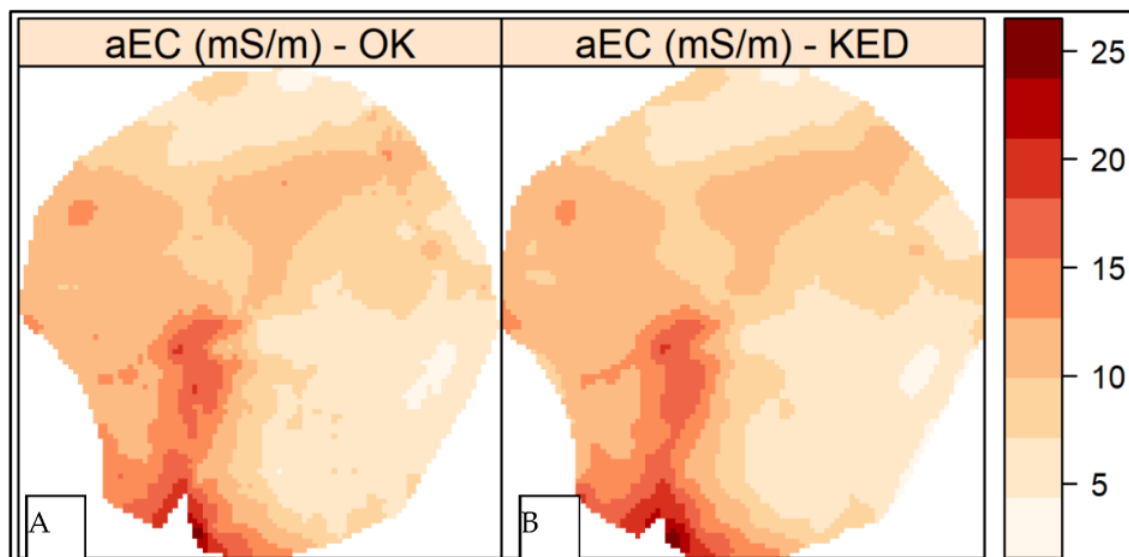
Figure 7. (**A**) shows the apparent electrical conductivity maps by ordinary kriging (KO) and (**B**) external drift kriging (KED). aEC: apparent electrical conductivity in mS/m (millisiemens per meter).

The aMS maps shown in Figure 8 presented a distribution inversely proportional to the distribution of the concentration values of the aEC maps (Figure 7). In addition, the OK mapping methods of the aMS (Figure 8A) and KED (Figure 8B) data showed similar distribution patterns for the entire study area.
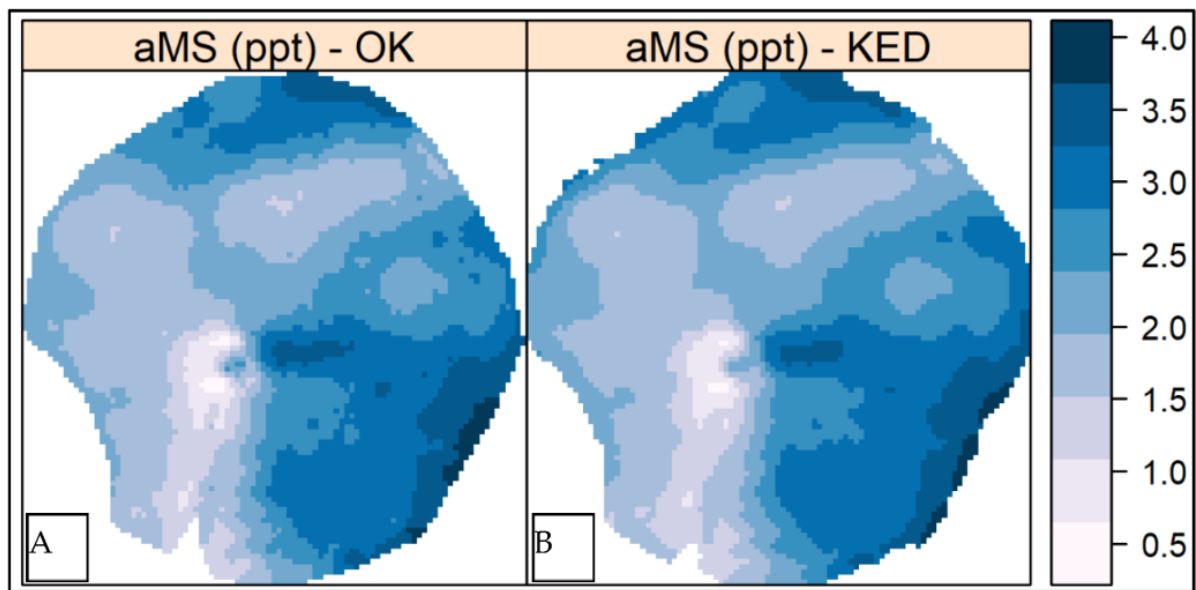
Figure 8. (**A**) Apparent magnetic susceptibility maps via ordinary kriging (KO) and (**B**) external drift kriging (KED). aMS: apparent magnetic susceptibility in ppt (parts per thousand).

The attributes of the Medusa MS1200 sensor showed concordant distributions between eTh and eU (Figure 9A,B, respectively). According to the literature, the high values of the ratio of eU over eTh may indicate a tendency to erode since the eU attribute is more mobile in the environment [66]. However, in Figure 9, it is possible to observe that the two attributes present patterns of distribution of values in similar regions, indicating that agricultural practice, even if implementing no-till, can minimally revolve the soil and homogenize the soil's attributes that control the preferential distribution of eU compared to eTh. Furthermore, from the spatial structure represented in the eU semivariograms and the map produced via OK, it was possible to notice monotonous variability in the study area, indicating growth in values towards the southeast.
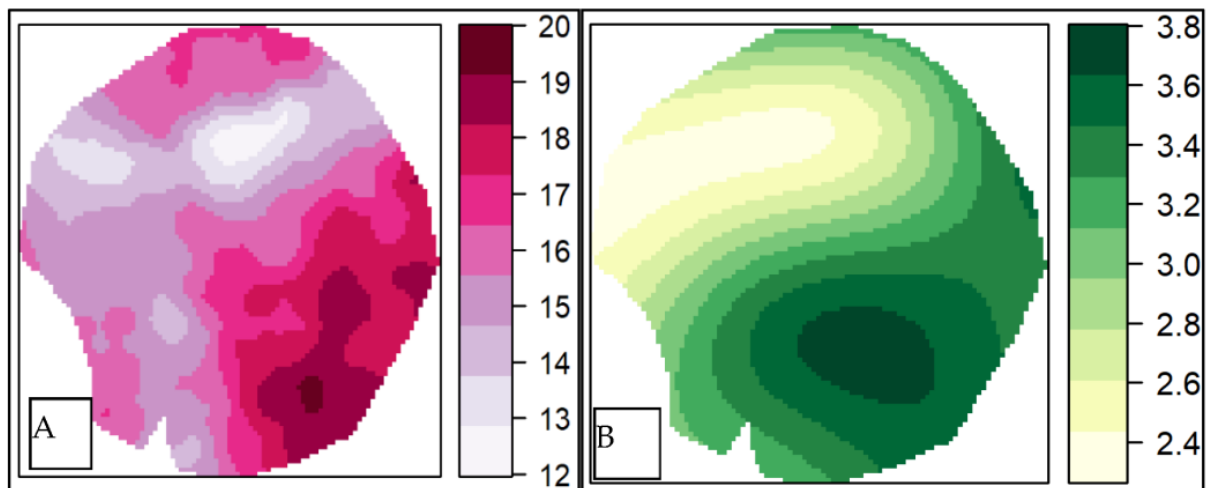
Figure 9. (**A**) Equivalent thorium and (**B**) equivalent uranium maps produced via ordinary kriging. eTh: equivalent thorium in ppm (parts per million); eU: equivalent uranium in ppm (parts per million).

The values of the uncertainties of the sensor maps are presented in Table 6. Although the hypothesis of the work was to use RS covariables to optimize the mapping of the data from proximal sensors, the values of the aEC and aMS attributes mapped using both OK

and KED presented values close to each other, indicating that there were no benefits of using satellite variables to optimize data from proximal sensors. The aEC and aMS maps produced with OK did not differ when mapped with the RS variables via KED. This was expected since the RMSE values were also close to each other.

Table 6. Root mean square error (RMSE) of the external validation of the proximal sensor data variables mapped via ordinary kriging (OK) and kriging with an external drift (KED).

| RMSE	Method	Attribute
0.56	OK	aEC (mS/m)
0.62	KED	
0.09	OK	aMS (ppt)
0.09	KED	
2.8	OK	eTh (ppm)
1.18	OK	eU (ppm)

aEC: apparent electrical conductivity in mS/m (millisiemens per meter); aMS: apparent magnetic susceptibility in ppt (parts per thousand); eTh: equivalent thorium in ppm (parts per million); eU: equivalent uranium in ppm (parts per million).

The maps of the management zones based on PSS (OK) and PSS plus RS (KED) classified via the *k*-means cluster algorithm are presented in Figure 10A and B, respectively. It is possible to observe that the zones outlined using the PSS maps via OK and KED were remarkably similar. This pattern may have resulted from implementing eTh and eU data in an OK format to compose the management zone map representative of the KED method.

The remote sensing covariables that collected information from the irrigation pivot area when the soil was planted (2019) did not optimize or significantly improve the distribution of the observed patterns of the PSS attributes spatially. When beans were grown, the satellite images representative of 2019 had a significative edge effect, recognizing the transition between exposed soil and planting on the pivot area's borders that can be observed in the management zone map using only RS data (Figure 11). On the other hand, DEM-derived data did not present edge problems. Therefore, the DEM-derived attributes selected to compose the KED models described in Section 2.4 and shown in Table 3 could represent the variables controlling the study area's water behavior.
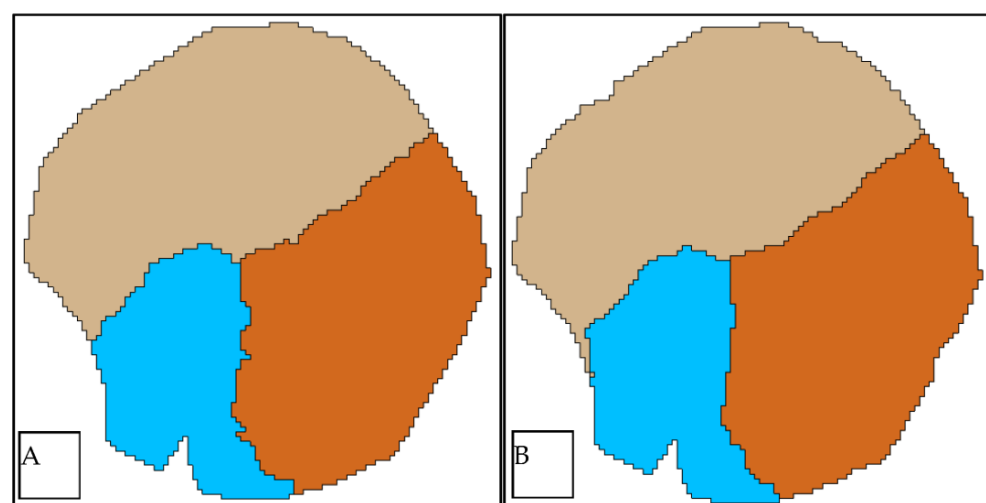
Figure 10. (**A**) Management zones made via *k*-means using proximal sensor maps with ordinary kriging (OK) and (**B**) management zones made via *k*-means using proximal sensor maps with external drift kriging (KED).
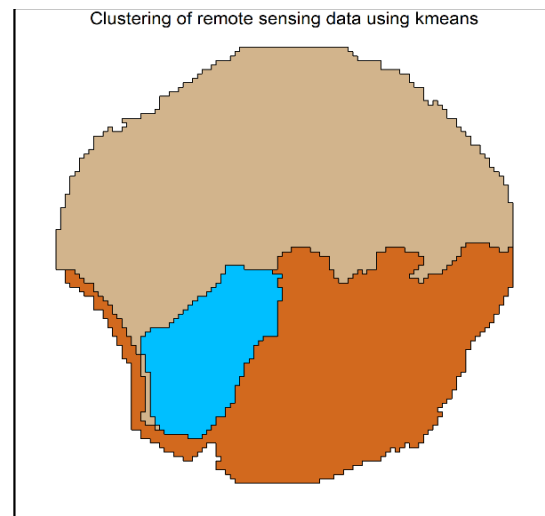
Figure 11. Management zones produced from all remote sensing data rasters using a *k*-means clustering algorithm.

The MZ created from RS data alone (Figure 11) could only separate the zone to the north of the study area, similar to the MZ maps via OK and KED (Figure 10A,B). It is possible to observe that the orange management zone in Figure 11 was like the zones in Figure 10 for both methods used (OK and KED).

The management zone in the blue area of Figure 11 needed to be better classified due to the edge effect of some satellite images in 2019 that identified the planting limits between beans and exposed soil. Although it was impossible to identify improvements for optimizing the PSS data via KED, generating an MZ map to separate soil management zones was possible using only RS data via satellite images and DEM covariates.

Table 7 shows that the classes created with the PSS data groupings via OK, KED, and RS only could distinguish the distribution regions for all the soil data analyzed in the laboratory. However, the MZ map using only remote sensor data could not be determined, with the same level of significance as OK and KED, the Ca and Mg attributes. The potential to treat different growth areas using proximal soil sensing was also presented by Van Meirvenne et al. (2013) [67] and De Benedetto et al. (2013) [68], who also tested the inclusion of remote sensor variables associated with the cluster clustering method.

Table 7. Analysis of variance (ANOVA) of the classes of the management zone map produced from the data of proximal sensors via OK (PSS); proximal sensors combined as the best satellite covariates via KED (PSS and RS); and remote sensing (RS) data raster.

F	Mean Sq	Sum Sq	DF	MZ
Clay (g kg^{-1})				
0.001 ***	13,743	27,486	2	PSS
0.001 ***	13,743	27,486	2	PSS and RS
0.001 ***	14,180	28,361	2	RS
Ca (g kg^{-1})				
0.003 **	4.12	8.24	2	PSS
0.003 **	4.12	8.24	2	PSS and RS
0.03 *	2.64	5.29	2	RS
C (g kg^{-1})				
3.9×10^{-7} ***	33.88	67.77	2	PSS
3.9×10^{-7} ***	33.88	67.77	2	PSS and RS
6.2×10^{-4} ***	18.75	37.5	2	RS

Table 7. *Cont.*

F	Mean Sq	Sum Sq	DF	MZ
Mg (g kg^{-1})				
0.001 **	0.49	0.98	2	PSS
0.001 **	0.49	0.98	2	PSS and RS
0.012 *	0.33	0.66	2	RS

Clay in g kg^{-1}; Ca: calcium in g kg^{-1}; C: organic carbon in g kg^{-1}; Mg: magnesium in g kg^{-1}; levels of statistical significance: * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$. DF: degrees of freedom; Sum Sq: sum of the squares; Mean Sq: mean square.

Although the MZ map using only RS data decreased in the degrees of significance to distinguish soil attributes, it was possible to observe that there was potential for distinguishing between heterogeneous soil zones using only RS data. This approach reveals the strong opportunity to implement remote sensing by-products into data fusion mapping to distinguish management zones in precision agriculture.

4. Conclusions

The maps of proximal sensor attributes produced with ordinary kriging showed the same external validation errors as the maps of proximal sensors built via kriging with external drift using the remote sensing data as covariables. There was no improvement in mapping proximal sensor attributes from the fusion of remote sensor data for the 72-ha analysis scale.

The ordinary-kriging and kriging-with-external-drift management zone maps produced with the *k*-means cluster algorithm were similar. However, the remote sensing management zone map delineated the management zones with less accuracy but similar spatial pattern. Furthermore, it shows promise for implementing and analyzing agricultural areas with more significant extensions.

Author Contributions: Conceptualization, H.R., G.M.V., V.L.M., G.B.M.H. and M.B.C.; methodology, H.R., G.M.V., V.L.M., G.B.M.H., M.B.C. and Z.N.B.; formal analysis, H.R., G.M.V., V.L.M., G.B.M.H., M.B.C. and Z.N.B.; investigation, H.R., G.M.V., R.P.O., S.R.L.T. and M.B.C.; resources, H.R., G.M.V., R.P.O., S.R.L.T., M.L.N., M.B.C. and J.P.S.M.; data curation, H.R., G.M.V., M.B.C., V.L.M. and G.B.M.H.; writing—original draft preparation, H.R., V.L.M. and G.B.M.H.; writing—review and editing, H.R., V.L.M., G.B.M.H., Z.N.B., M.L.N. and J.P.S.M.; supervision, G.M.V., G.B.M.H., V.L.M. and M.B.C.; project administration, G.M.V. and M.B.C.; funding acquisition, G.M.V. and M.B.C. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data is not publicly available due to intellectual property on the part of the Research Funding Company.

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

References

1. Baudron, F.; Giller, K.E. Agriculture and nature: Trouble and strife? *Biol. Conserv.* **2014**, *170*, 232–245. [CrossRef]
2. Boardman, J.; Poesen, J.; Evans, R. Socio-economic factors in soil erosion and conservation. *Environ. Sci. Policy* **2003**, *6*, 1–6. [CrossRef]

3. Nawar, S.; Corstanje, R.; Halcro, G.; Mulla, D.; Mouazen, A. *Delineation of Soil Management Zones for Variable-Rate Fertilization: A Review*, 1st ed.; Elsevier: Amsterdam, The Netherlands, 2017.

4. Peralta, N.R.; Costa, J.L.; Balzarini, M.; Franco, M.C.; Córdoba, M.; Bullock, D. Delineation of management zones to improve nitrogen management of wheat. *Comput. Electron. Agric.* **2015**, *110*, 103–113. [CrossRef]

5. Groher, T.; Heitkämper, K.; Walter, A.; Liebisch, F.; Umstätter, C. Status quo of adoption of precision agriculture enabling technologies in Swiss plant production. *Precis. Agric.* **2020**, *21*, 1327–1350. [CrossRef]

6. Hedley, C. The role of precision agriculture for improved nutrient management on farms. *J. Sci. Food Agric.* **2015**, *95*, 12–19. [CrossRef] [PubMed]

7. Demattê, J.A.M.; Alves, E.R.; Barbosa, R.N.; Morelli, J.L. Precision agriculture for sugarcane management: A strategy applied for brazilian conditions. *Acta Sci. Agron.* **2014**, *36*, 111. [CrossRef]

8. Molin, J.P.; Tavares, T.R. Sensor systems for mapping soil fertility attributes: Challenges, advances, and perspectives in Brazilian tropical soils. *Eng. Agrícola* **2019**, *39*, 126–147. [CrossRef]

9. Stafford, J.V. Implementing Precision Agriculture in the 21st Century. *J. Agric. Eng. Res.* **2000**, *76*, 267–275. [CrossRef]

10. Sparovek, G.; Schnug, E. Soil tillage and precision agriculture: A theoretical case study for soil erosion control in Brazilian sugar cane production. *Soil Tillage Res.* **2001**, *61*, 47–54. [CrossRef]

11. Ohana-Levi, N.; Bahat, I.; Peeters, A.; Shtein, A.; Netzer, Y.; Cohen, Y.; Ben-Gal, A. A weighted multivariate spatial clustering model to determine irrigation management zones. *Comput. Electron. Agric.* **2019**, *162*, 719–731. [CrossRef]

12. Fleming, K.L.; Westfall, D.G.; Wiens, D.W.; Brodahl, M.C. Evaluating Farmer Defined Management Zone Maps for Variable Rate Fertilizer Application. *Precis. Agric.* **2000**, *2*, 201–215. [CrossRef]

13. Doerge, T.A. *Site-Specific Management Guidelines*; Potash & Phosphate Institute: Norcross, GA, USA, 1999.

14. Viscarra Rossel, R.A.; Bouma, J. Soil sensing: A new paradigm for agriculture. *Agric. Syst.* **2016**, *148*, 71–74. [CrossRef]

15. 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]

16. Mahmood, H.S.; Hoogmoed, W.B.; Van Henten, E.J. Combined sensor system for mapping soil properties. In *Precision Agriculture '09*; van Henten, E.J., Goense, D., Lokhorst, C., Eds.; Wageningen Academic Publisher: Wageningen, The Netherlands, 2009; p. 992.

17. Mahmood, H.S.; Hoogmoed, W.B.; Van Henten, E.J. Estimating soil properties with a proximal gamma-ray spectrometer using windows and full-spectrum analysis methods. In *The Second Global Workshop on Proximal Soil Sensing*; McGill University: Montreal, QC, Canada, 2011; p. 4.

18. Fraisse, C.W.; Sudduth, K.A.; Kitchen, N.R. Delineation of Site-Specific Management Zones by Unsupervised Classification of Topographic Attributes and Soil Electrical Conductivity. *Trans. ASAE* **2001**, *44*, 155–166. [CrossRef]

19. McBratney, A.; Minasny, B. On measuring pedodiversity. *Geoderma* **2007**, *141*, 149–154. [CrossRef]

20. De Benedetto, D.; Castrignanò, A.; Rinaldi, M.; Ruggieri, S.; Santoro, F.; Figorito, B.; Gualano, S.; Diacono, M.; Tamborrino, R. An approach for delineating homogeneous zones by using multi-sensor data. *Geoderma* **2013**, *199*, 117–127. [CrossRef]

21. Shaddad, S.M.; Madrau, S.; Castrignanò, A.; Mouazen, A.M. Data fusion techniques for delineation of site-specific management zones in a field in UK. *Precis. Agric.* **2016**, *17*, 200–217. [CrossRef]

22. Rodrigues, H.M.; Vasques, G.M.; Oliveira, R.P.; Tavares, S.R.; Ceddia, M.B.; Hernani, L.C. Finding suitable transect spacing and sampling designs for accurate soil ECa mapping from EM38-MK2. *Soil Syst.* **2020**, *4*, 19. [CrossRef]

23. Horney, R.D.; Taylor, B.; Munk, D.S.; Roberts, B.A.; Lesch, S.M.; Plant, R.E. Development of practical site-specific management methods for reclaiming salt-affected soil. *Comput. Electron. Agric.* **2005**, *46*, 379–397. [CrossRef]

24. Nouri, H.; Borujeni, S.C.; Alaghmand, S.; Anderson, S.J.; Sutton, P.C.; Parvazian, S.; Beecham, S. Soil Salinity Mapping of Urban Greenery Using Remote Sensing and Proximal Sensing Techniques; The Case of Veale Gardens within the Adelaide Parklands. *Sustainability* **2018**, *10*, 2826. [CrossRef]

25. Huang, J.; Subasinghe, R.; Malik, R.; Triantafilis, J. Salinity hazard and risk mapping of point source salinisation using proximally sensed electromagnetic instruments. *Comput. Electron. Agric.* **2015**, *113*, 213–224. [CrossRef]

26. Triantafilis, J.; Lesch, S. Mapping clay content variation using electromagnetic induction techniques. *Comput. Electron. Agric.* **2005**, *46*, 203–237. [CrossRef]

27. Sudduth, K.; Kitchen, N.; Wiebold, W.; Batchelor, W.; Bollero, G.; Bullock, D.; Clay, D.; Palm, H.; Pierce, F.; Schuler, R.; et al. Relating apparent electrical conductivity to soil properties across the north-central USA. *Comput. Electron. Agric.* **2005**, *46*, 263–283. [CrossRef]

28. Becegato, V.A.; Ferreira, F.J.F. Gamaespectrometria, resistividade elétrica e susceptibilidade magnética de solos agrícolas no noroeste do estado do Paraná. *Rev. Bras. Geofis.* **2005**, *23*, 371–405. [CrossRef]

29. Loonstra, E.; van Egmond, F. On-the-go measurement of soil gamma radiation. In *Precision Agriculture, Proceedings of the 7th European Conference on Precision Agriculture, Wageningen, The Netherlands, 6–8 July 2009*; van Henten, E.J., Goense, D., Lokhorst, C., Eds.; Wageningen Academic: Wageningen, The Netherlands, 2009.

30. Taylor, M.J.; Smettem, K.; Pracilio, G.; Verboom, W. Relationships between Soil Properties and High-Resolution Radiometrics, Central Eastern Wheatbelt, Western Australia. *Explor. Geophys.* **2002**, *33*, 95–102. [CrossRef]

31. Holland, J.; Biswas, A.; Huang, J.; Triantafilis, J. Scoping for scale-dependent relationships between proximal gamma radiometrics and soil properties. *Catena* **2017**, *154*, 40–49. [CrossRef]

32. Ji, W.; Adamchuk, V.I.; Chen, S.; Su, A.S.M.; Ismail, A.; Gan, Q.; Shi, Z.; Biswas, A. Simultaneous measurement of multiple soil properties through proximal sensor data fusion: A case study. *Geoderma* **2019**, *341*, 111–128. [CrossRef]

33. Pelegrino, M.H.P.; Weindorf, D.C.; Silva, S.H.G.; de Menezes, M.D.; Poggere, G.C.; Guilherme, L.R.G.; Curi, N. Synthesis of proximal sensing, terrain analysis, and parent material information for available micronutrient prediction in tropical soils. *Precis. Agric.* **2018**, *20*, 746–766. [CrossRef]

34. Higginbottom, T.P.; Symeonakis, E.; Meyer, H.; van der Linden, S. Mapping fractional woody cover in semi-arid savannahs using multi-seasonal composites from Landsat data. *ISPRS J. Photogramm. Remote Sens.* **2018**, *139*, 88–102. [CrossRef]

35. Landrum, C.; Castrignanò, A.; Mueller, T.; Zourarakis, D.; Zhu, J.; De Benedetto, D. An approach for delineating homogeneous within-field zones using proximal sensing and multivariate geostatistics. *Agric. Water Manag.* **2015**, *147*, 144–153. [CrossRef]

36. Rossel, R.V.; Adamchuk, V.I.; Sudduth, K.A.; McKenzie, N.J.; Lobsey, C. Proximal soil sensing. In *An Effective Approach for Soil Measurements in Space and Time*; Elsevier: Amsterdam, The Netherlands, 2011.

37. de Queiroz, D.M.; de Freitas Coelho, A.L.; Valente, D.S.M.; Schueller, J.K. Sensors Applied to Digital Agriculture: A Review. *Rev. Ciência Agronômica* **2020**, *51*, 1–15. [CrossRef]

38. Dalmolin, R.S.D.; Gonçalves, C.N.; Klamt, E.; Dick, D.P. Relação entre os constituintes do solo e seu comportamento espectral. *Ciência Rural* **2005**, *35*, 481–489. [CrossRef]

39. Vasques, G.M.; Demattê, J.A.M.; Rossel, R.A.V.; López, L.R.; Terra, F.S.; Rizzo, R.; Filho, C.R.D.S. Integrating geospatial and multi-depth laboratory spectral data for mapping soil classes in a geologically complex area in southeastern Brazil. *Eur. J. Soil Sci.* **2015**, *66*, 767–779. [CrossRef]

40. Rodrigues, F.; Bramley, R.; Gobbett, D. Proximal soil sensing for Precision Agriculture: Simultaneous use of electromagnetic induction and gamma radiometrics in contrasting soils. *Geoderma* **2015**, *243–244*, 183–195. [CrossRef]

41. Pantazi, X.E.; Moshou, D.; Mouazen, A.M.; Alexandridis, T.; Kuang, B. Data Fusion of Proximal Soil Sensing and Remote Crop Sensing for the Delineation of Management Zones in Arable Crop Precision Farming. In Proceedings of the 7th International Conference on Information and Communication Technologies in Agriculture, Food and Environment, Kavala, Greece, 17–20 September 2015; pp. 765–776.

42. Li, Y.; Shi, Z.; Wu, C.-F.; Li, H.-Y.; Li, F. Determination of potential management zones from soil electrical conductivity, yield and crop data. *J. Zhejiang Univ. B* **2008**, *9*, 68–76. [CrossRef]

43. FAO. World reference base for soil resources 2014. In *International Soil Classification System for Naming Soils and Creating Legends for Soil Maps*; FAO: Rome, Italy, 2014.

44. Bivand, R.S.; Pebesma, E.; Gómez-Rubio, V. *Applied Spatial Data Analysis with R*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2013.

45. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2023.

46. Wickham, H.; François, R.; Henry, L.; Müller, K.; Vaughan, D. dplyr: A Grammar of Data Manipulation. R Package Version 1.1.4. Available online: https://dplyr.tidyverse.org (accessed on 19 April 2023).

47. 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]

48. Brenning, A.; Bangs, D.; Becker, M. RSAGA: SAGA Geoprocessing and Terrain Analysis. R Package Version 1.4.0. 2022. Available online: https://CRAN.R-project.org/package=RSAGA (accessed on 22 April 2023).

49. Hijmans, R. raster: Geographic Data Analysis and Modeling. R Package Version 3.6-26. 2023. Available online: https://CRAN.R-project.org/package=raster (accessed on 22 April 2023).

50. Lumley, T. leaps: Regression Subset Selection. R Package Version 3.1. 2020. Available online: https://CRAN.R-project.org/package=leaps (accessed on 13 May 2023).

51. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1986**, *6*, 461–464. [CrossRef]

52. da Silva, N.C.; Santos, R.C.; Zucca, R.; Geisenhoff, L.O.; Cesca, R.S.; Lovatto, J. Enthalpy Thematic Map Interpolated with Spline Method for Management of Broiler Chicken Production. *Rev. Bras. Eng. Agric. Ambient.* **2020**, *24*, 431–436. [CrossRef]

53. Gräler, B.; Pebesma, E.; Heuvelink, G. Spatio-Temporal Interpolation using gstat. *R J.* **2016**, *8*, 204–218. [CrossRef]

54. Wackernagel, H. *Multivariate Geostatistics: An Introduction with Applications*, 3rd ed.; Springer: Berlin/Heidelberg, Germany, 2003; ISBN 9783662052945.

55. Webster, R.; Oliver, M.A. *Geostatistics for Environmental Scientists*, 2nd ed.; John Wiley & Sons, Ltd.: Hoboken, NJ, USA; The Atrium: Chichester, UK, 2007.

56. Diggle, P.J.; Ribeiro, P.J., Jr. *Model-Based Geostatistics*; Springer Science+Business Media, LLC: New York, NY, USA, 2007; ISBN 9780387329079.

57. Goovaerts, P. *Geostatistics for Natural Resources Evaluation*; Oxford University Press: New York, NY, USA, 1997.

58. Hartigan, J.A.; Wong, M.A. Algorithm AS 136: A *k*-means Clustering Algorithm. *J. R. Stat. Soc. Ser. Appl. Stat.* **1979**, *28*, 100–108. [CrossRef]

59. Walvoort, D.J.J.; Brus, D.J.; de Gruijter, J.J. spcosa: Spatial Coverage Sampling and Random Sampling from Compact Geographical Strata. Volume 28. 2023. Available online: https://CRAN.R-project.org/package=spcosa (accessed on 13 May 2023).

60. Teixeira, P.C.; Donagemma, G.K.; Fontana, A.; Teixeira, W.G. *Manual de Métodos de Análise de Solo*, 3rd ed.; Embrapa: Brasília, Brazil, 2017.

61. DeCarlo, L.T. On the meaning and use kurtosis. *Psychol. Methods* **1997**, *2*, 292–307. [CrossRef]
62. Castrignanò, A.; Wong, M.; Stelluti, M.; De Benedetto, D.; Sollitto, D. Use of EMI, gamma-ray emission and GPS height as multi-sensor data for soil characterisation. *Geoderma* **2012**, *175–176*, 78–89. [CrossRef]
63. Wong, M.T.F.; Oliver, Y.; Robertson, M.J. Gamma-Radiometric Assessment of Soil Depth across a Landscape Not Measurable Using Electromagnetic Surveys. *Soil Sci. Soc. Am. J.* **2009**, *73*, 1261–1267. [CrossRef]
64. Fulton, A.; Schwankl, L.; Lynn, K.; Lampinen, B.; Edstrom, J.; Prichard, T. Using EM and VERIS technology to assess land suitability for orchard and vineyard development. *Irrig. Sci.* **2011**, *29*, 497–512. [CrossRef]
65. Islam, M.M.; Saey, T.; Meerschman, E.; De Smedt, P.; Meeuws, F.; Van De Vijver, E.; Van Meirvenne, M. Delineating water management zones in a paddy rice field using a Floating Soil Sensing System. *Agric. Water Manag.* **2011**, *102*, 8–12. [CrossRef]
66. 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]
67. Van Meirvenne, M.; Islam, M.M.; De Smedt, P.; Meerschman, E.; Van De Vijver, E.; Saey, T. Key variables for the identification of soil management classes in the aeolian landscapes of north–west Europe. *Geoderma* **2013**, *199*, 99–105. [CrossRef]
68. De Benedetto, D.; Castrignano, A.; Diacono, M.; Rinaldi, M.; Ruggieri, S.; Tamborrino, R. Field partition by proximal and remote sensing data fusion. *Biosyst. Eng.* **2013**, *114*, 372–383. [CrossRef]