Mapping Soil Properties with Fixed Rank Kriging of Proximally Sensed Soil Data Fused with Sentinel-2 Biophysical Parameter

Nikolaos Karapetsas 1, Thomas K. Alexandridis 1,*, George Bilas 1, Muhammad Abdul Munnaf 2, Angela P. Guerrero 2, Maria Calera 3, Anna Osann 3, Anne Gobin 4,5, Tomáš Rezník 6, Dimitrios Moshou 7, and Abdul Mounem Mouazen 2

Abstract: Soil surveys with line-scanning platforms appear to have great advantages over the traditional methods used to collect soil information for the development of field-scale soil mapping and applications. These carry VNIR (visible and near infrared) spectrometers and have been used in recent years extensively for the assessment of soil fertility at the field scale, and the delineation of site-specific management zones (MZ). A challenging feature of VNIR applications in precision agriculture (PA) is the massiveness of the derived datasets that contain point predictions of soil properties, and the interpolation techniques involved in incorporating these data into site-specific management plans. In this study, fixed-rank kriging (FRK) geostatistical interpolation, which is a flexible, non-stationary spatial interpolation method especially suited to handling huge datasets, was applied to massive VNIR soil scanner data for the production of useful, smooth interpolated maps, appropriate for the delineation of site-specific MZ maps. Moreover, auxiliary Sentinel-2 data-based biophysical parameters NDVI (normalized difference vegetation index) and fAPAR (fraction of photosynthetically active radiation absorbed by the canopy) were included as covariates to improve the filtering performance of the interpolator and the ability to generate uniform patterns of spatial variation from which it is easier to receive a meaningful interpretation in PA applications. Results from the VNIR prediction dataset obtained from a pivot-irrigated field in Albacete, southeastern Spain, during 2019, have shown that FRK variants outperform ordinary kriging in terms of filtering capacity, by doubling the noise removal metrics while keeping the computation cost reasonably low. Such features, along with the capacity to handle a large volume of spatial information, nominate the method as ideal for PA applications with massive proximal and remote sensing datasets.

Keywords: geostatistical interpolation; VNIR spectrometer; NDVI; fAPAR; precision agriculture

1. Introduction

Mapping soil variability at field scale is widely recognized as a key concept in precision agriculture (PA) [1,2]. Traditional soil surveys have been using discrete soil map units in an effort to describe soil variability at landscape or regional scales, but are considered
inadequate to provide detailed and accurate information on soil variability at field scale [3]. An effective solution is offered by using proximal soil scanners to record the soil data at fine spatial resolution [4]. Different types of sensors, which can map soil moisture, micro- and macro-component contents, texture or other soil properties, use a variety of measurement techniques (electromagnetic induction, electrical resistivity, ground penetrating radar, near infrared, gamma ray, multi- and hyperspectral spectrometers) in conjunction with GPS (global positioning systems) [5]. Due to their frequent repetitive measurements, massive datasets consisting of several thousand scanned measurements with point geo-reference are typically the output of soil sensors in each field.

Due to the nature of line-scanning platforms used in proximal soil surveys that involve movement of the soil sensor along parallel transects of a specific interval, several error sources are introduced during data acquisition. Especially in the case of visible and near infrared (VNIR) spectroscopy, many challenging factors interfere in the process of data collection; among others, surface roughness, texture, moisture content, plant residues, and gravels are the most negatively affecting parameters [6]. Variation in soil-to-sensor distance, machine vibrations, and trajectory deviations are additional sources of error [7,8]. In addition, soil VNIR spectra inherently contain much redundant information, due to the strong correlation between neighboring wavelengths [4]. Although postprocessing techniques such as noise filtering are usually applied in the recorded spectra prior to the modeling and prediction effort of soil properties, inevitably the point-support predictions exhibit a strong fine-scale spatial variability from which it is difficult to receive a meaningful interpretation in PA applications.

Spatial interpolation is commonly used in PA applications to provide useful and easily assimilated, smooth continuous surfaces of soil properties from the exhaustive point-support soil spectra datasets, upon which site-specific management plans can be developed. Geostatistical interpolation or kriging methods are often preferred by soil scientists since they provide the criterion of optimality in spatial prediction; kriging interpolation methods yield the best in statistical terms of spatial prediction but the process requires the inversion of an $N \times N$ matrix, where $N$ is the size of the dataset, consisting usually of several thousand points in a typical field-scale application. The inherent constraint of kriging interpolation is the computational bottleneck of processing such large datasets [9,10]. Ad hoc methods of subsetting the data have been formalized by the local neighborhood approach, where only a limited number of neighbors is used for prediction purposes. Although this local neighborhood method has little effect on the optimality of the kriging estimator, it creates abrupt variations in the predicted maps. This can severely affect the appearance of soil variation patterns in the resulting maps, introducing local variability which is usually represented as noisy output in the interpolated maps. Several authors acknowledged kriging in local neighborhood as inadequate, providing no guarantee that long-range spatial patterns will be represented appropriately [11–13]. The local nature of kriging implies that what happens over large distances is of little consequence for the estimates, thus the notion of local stationarity applies without taking account of long-range fluctuations. The assumptions underpinning the method are not violated and it is this feature that has made kriging with local prediction the ‘workhorse’ of geostatistics [14]. Maps produced this way are regularly used in the assessment of spatial variation of soil fertility, which eventually leads to the delineation of site-specific management zones (MZ). The key activity of PA is the determination of management units within a field that have such a degree of internal homogeneity that they can be subjected to a specific form of management [15]. Filtering out the fine-scale component of spatial variation from the resulting maps is essential, and its removal is required so that the medium and large scale components of variation can become visible, paving the way for a meaningful interpretation of spatial patterns in soil properties. Research on the delineation of MZ based on soil data collected with VNIR spectroscopy sensors can be found in the literature [16,17]. Multivariate geostatistical techniques are widely used in the assessment of the physical and chemical properties of soil, involving filtering of the noisy fine-scale variation component using factorial (co)kriging [18]. This
approach is based on the assumption that the observed soil properties are the result of a linear combination of several subprocesses that generally exhibit different scales of variation with different spatial dependencies. Still, even through the local neighborhood approach, multivariate factorial kriging analysis in massive VNIR datasets cannot be applied efficiently due to the predicament of the involved computational cost.

The spatial prediction of soil properties using auxiliary variables or covariates is a well-established approach, commonly used by soil scientists in an effort to assess soil variability across landscapes and within-field scale, based on the environmental correlation between explanatory variables and soil attributes [19]. Earth observation (EO) data-based vegetation indices can provide dynamic spectral measurements of vegetation activity, and in conjunction with other canopy biophysical parameters such as the fraction of photosynthetically active radiation (fAPAR, 400–700 nm) absorbed by vegetation, are among the most widely used satellite products in monitoring ecosystems and agriculture. Non-geostatistical techniques such as multiple linear regression (MLR), as well as geostatistical methods such as ordinary kriging (OK) or cokriging (COK), and hybrid methods such as regression-kriging (RK) have proven to be robust and practical for the prediction of various soil properties, involving the use of environmental covariates, therefore allowing the exploitation of auxiliary information. Castrignanò et al. [3] used multivariate geostatistical techniques (COK) to fuse data measured with different geophysical sensors, namely, electromagnetic induction and ground-penetrating radar, for delineating the homogeneous within-field MZ for PA application. In such data-fusion approaches, the change of support problem commonly emerges, referring to the size or volume associated with each data value, support or footprint which also includes the geometrical size, shape, and spatial orientation of the region associated with the measurements. This crucial problem is usually addressed by applying block-kriging on datasets with a different footprint and proceeding with the spatial prediction on an equal support or estimation cell. Although this approach is efficiently applied in several case studies [3,20], it still is highly impractical to implement in the case of massive datasets, unless the notion of local prediction is applied.

Cressie et al. [9,10] developed a flexible, nonstationary spatial statistical model that resolves the computational bottleneck of $N \times N$ matrix inversion. Fixed rank kriging (FRK), namely, differs from existing efforts on spatial modeling and prediction for very large datasets, by avoiding stationary–isotropic covariance and variogram models and constructing instead a spatial random effects (SRE) model on a discretized spatial domain. Distinguishing and filtering out the fine-scale variation at the estimation cell level is possible, along with the inversion of an $N \times N$ variance-covariance matrix, using a fixed rank of positive-definite matrices. Spatial statistical data fusion frameworks for remote sensing (RS) applications have been established during the last decade based on the FRK framework, involving EO data from coarse to medium resolution sensors such as MODIS and Terra [21]. The usefulness of geostatistical interpolation in combining coarse resolution RS data with proximal soil spectroscopy data has also been demonstrated in individual studies, as an option that could be employed for soil property mapping in vegetated areas, where canopy coverage obstructs the extraction of soil information [22]. Two characteristic features of FRK are highlighted in other studies [23]: its ability to predict a spatial process by separating the signal from the noise—similar to factorial kriging—based on a large number of observations that contain measurement errors, and its capacity—unlike traditional kriging—to process massive EO data sets efficiently.

To overcome the difficulties of interpolating massive and inconsistent spatial resolution data obtained by a soil scanner, the nonstationary spatial statistical model of FRK could be a solution. Limited research so far has focused on the use of FRK for line-scanning soil sensor data, using a gamma radiometer and electromagnetic scanner [24]. However, no report can be found in the literature about the implementation of FRK for VNIR soil data collected with a line scanner. The use of EO-data as covariates in FRK interpolation has not been tested extensively apart from characterizing regional soil mineral composition based on scale-dependent spatial variability observed in RS data [25]. Proximal sensor and RS data are
generally massive, taken on different spatial and temporal scales, having different footprints (spatial support), and subjected to measurement error biases. A data fusion approach could take advantage of their complementary features by combining the sensor data sets in a statistically robust manner. Data fusion can be regarded as an inference challenge, concerning heterogeneous input datasets with different statistical characteristics, focusing on the optimal estimation of the variable of interest and the uncertainty associated with this inference. Thus, this work aims to assess the potential of the FRK geostatistical interpolator combined with Sentinel-2 biophysical parameters, to improve the noise-filtering ability and estimation accuracy of key soil attributes using VNIR soil scanner data. Specific objectives are (i) to assess the performance of FRK interpolation over line-scanning VNIR measured soil data, and (ii) to compare two different FRK geostatistical approaches, namely, a univariate, and one with Sentinel-2-based covariates, to OK interpolation method.

2. Materials and Methods

2.1. Study Area

The study area lies in the lower part of an irrigated commercial field under agronomic management, located in the province of Albacete, Region Castile–La Mancha, southeastern Spain (Figure 1a) (38°46′31″N, 1°50′15″W). It covers an area of 55.3 ha across a moderate slope. The area is characterized by a semi-arid Mediterranean climate, with strong rainfall variability throughout the growing season. Counterbalanced by irrigation such a variability is magnified by the low soil water-holding capacity of the developed Calcaric Cambisols, exhibiting shallow subsurface horizons at approximately 40 to 80 cm depth. The mean annual precipitation is 340 mm and the mean annual temperature is 13.6 °C [25]. An overall increase in production has been recorded during the last decade in the region, derived from a high standard of agronomic management, use of modern technology, and improvements in the field of genetics, fertilization, irrigation, and crop protection. The field is managed under irrigation and rainfed conditions, using a center pivot system, and under no-tillage and direct seeding practices. The fertilization approach included repetitions under variable-rate treatment, established in stripes (Figure 1c). Crop for the field under investigation in the year 2019 was poppy with a growing season from April to June.

Figure 1. Study area in Albacete, southeastern Spain (a), validation soil sampling distribution (b) and variable-rate treatment fertilization zones (c).

2.2. Soil—VNIR Data

Following a 2-month time period since the harvest of poppy crop in late June 2019, which had left the soil partially covered with crop residues in compliance with no-tillage management practices, soil scanning was carried out in August 2019. Measurements were collected using a VNIR soil scanner developed by Mouazen [26]. The system uses a mobile, fiber type, VNIR spectrometer (CompactSpec from Tec5 Technology, Steinbach, Germany).
with a spectral range of 305–1700 nm to record soil spectra. A differential global positioning system (Trimble AG25, Sunnyvale, CA, USA) records the position of the spectra, which are logged together with GPS readings. Field scanning was carried out along 12 m apart parallel transects across the entire field, at an average forward travelling speed of around 3.5 km/h. Prediction of various soil properties from spectral soil response was accomplished through partial least squares regression analysis by Munnaf et al. [8]. Individual processing per soil variable generated a specific dataset of predicted values for each one explicitly. The complete resulting dataset included massive point-support predictions for several soil physical-chemical properties, constituting dense spatial data of an average count of around 30 K points per soil variable per field. The predicted output dataset of spectral modelling on VNIR data by Munnaf et al. [8], regarding Ca, Mg, Na, pH and moisture content (MC), was considered as input data in the current study (Table 1).

### Table 1. Statistical summary of soil variables estimated from VNIR spectrometer data from soil scanner with nearest neighbor and attribute statistics.

| Soil Variables | Total Points | Prediction Points | NN Min | NN Max | NN Mean | Min | 1st Qu | Mean | 3rd Qu | Max | SD | CV% | Kurtosis | Skewness |
|----------------|--------------|-------------------|--------|--------|---------|------|--------|------|--------|-----|----|-----|----------|----------|
| Ca (meq/100 g) | 33,659       | 31,874            | 0.15   | 11.3   | 1.14    | 23.53| 35.39  | 38.86| 41.13  | 85.89| 4.142| 10.88| 2.5445   | -0.0937  |
| MC (％)         | 29,444       | 29,444            | 0.24   | 13.1   | 1.18    | 6.333| 11.56  | 14.06| 16.67  | 21.5 | 3.487| 24.79| 2.8964   | -0.1803  |
| Mg (meq/100 g) | 32,071       | 31,380            | 0.15   | 11.8   | 1.16    | 2.50 | 4.971  | 5.765| 6.467  | 14.96| 1.075| 18.86| 2.3110   | 0.1633   |
| Na (meq/100 g) | 33,032       | 32,668            | 0.15   | 11.3   | 1.14    | 0.004| 0.306  | 0.415| 0.520  | 1.14 | 0.154| 37.86| 2.9403   | -0.2632  |
| pH             | 28,035       | 28,035            | 0.15   | 33.7   | 1.19    | 7.93 | 8.313  | 8.439| 8.579  | 8.8  | 0.186| 2.21 | 2.7543   | -0.2855  |

1 NN distance in (m), minimum NN = 0.1 m global threshold applied. MC: moisture content, CV: coefficient of variation, SD: standard deviation.

Considering that substantial deviation from normality was not apparent in any case, distribution shape characteristics, i.e., kurtosis, with values ≤ 3, indicated extreme outlier expectation no bigger than the one from a univariate normal distribution. The distributions were filtered for outliers based on quantile distribution and IQR (interquartile range), where values lower than Q1 – (1.5 × IQR) and higher than Q3 + (1.5 × IQR) were removed as outliers. The resulting distributions are the ‘prediction points’ column in Table 1, vs. ‘total points’ which are the initial distributions. Outlier filtering is a crucial step towards spatial prediction, since extreme soil variable values screen or cover the meaningful variation that is expected in the resulting maps. Substantial differences were recorded in the variability of the input parameters with pH appearing as the least variable soil parameter and Na the most variable, based on the coefficient of variation (CV) metrics. The spatial distribution of point-support data was in coherence with the VNIR spectrometer scanning that was carried out in parallel transects 12 m apart, with a bearing of approximately 42° from north, following the variable-rate treatment scheme established in stripes across the entire field (Figure 1c).

Additionally, a total of 100 georeferenced soil samples were collected from the surface layer during the line-scanning, at individual spatial locations randomly distributed across the entire field (Figure 1b), in order to be used as validation dataset. Descriptive statistics concerning analytical results on selected soil physical-chemical properties from the sampling population are presented in Table 2.
Table 2. Statistical summary of validation soil sample dataset.

| Soil Variables | Min.  | 1st Qu | Median | Mean  | 3rd Qu | Max.  | Stand Dev | CV%   |
|----------------|-------|--------|--------|-------|--------|-------|-----------|-------|
| Ca (meq/100 g) | 26.37 | 35.3   | 38.93  | 39.1  | 42.54  | 83.15 | 7.003     | 17.91 |
| MC (%)         | 7.027 | 10.9   | 12.88  | 13.18 | 15.45  | 20.81 | 2.883     | 21.88 |
| Mg (meq/100 g) | 3.076 | 4.812  | 5.575  | 5.878 | 6.756  | 14.42 | 1.686     | 28.68 |
| Na (meq/100 g) | 0.004307 | 0.1828 | 0.3533 | 0.3839 | 0.5536 | 1.09  | 0.2513    | 65.44 |
| pH             | 7.97  | 8.394  | 8.48   | 8.499 | 8.56   | 8.765 | 0.1331    | 1.574 |

MC: moisture content, CV: coefficient of variation, n = 100 soil samples with analytical results.

2.3. Earth Observation Data

Satellite images acquired by Sentinel-2 throughout 2019 were retrieved from Copernicus Global Land Monitoring Service (land.copernicus.eu). The extracted EO biophysical parameters comprised the normalized difference vegetation index (NDVI) and fraction of photosynthetically active radiation (fAPAR) absorbed by the canopy [27]. Extensive filtering of the time-series from the open-EO platform (openEO.org) was carried out on both layer and pixel basis, using image cloud coverage percentage (pre-filtering with <25%) and pixel cloud probability (quality mask: probability < 5%), accordingly.

According to the 5-day revisit period of Sentinel-2, the time-series of RS indicators was composed of 24 images for each biophysical parameter during the cropping season (April–July), thus a total number of 48 correlated predictor variables were present (Figure 2). Principal component analysis (PCA) was carried out per biophysical parameter to avoid multi-collinearity, reduce the dimension of the dataset and decrease the number of input variables [28–30]. The first 3 individual principal components (Figure 3) for each parameter, that accounted for approximately 90% of the total variance of the original RS indicators, were initially considered as auxiliary variables in the FRK interpolation to improve the spatial prediction of soil variables based on VNIR data and assist the filtering performance.

![Figure 2.](image)

(a) Time-series plots of averages from EO biophysical parameters NDVI (a) and fAPAR (b) over study area during 2019. Red lines represent smoothened curves fitted in time-series with local regression.
2.4. Auxiliary Variables as Covariates

Sentinel-2 biophysical indicators are considered too coarse (10 m pixel) to characterize the fine-scale variation of soil properties in local farmlands [31]. The contribution of Sentinel-2 biophysical indicators in soil variability assessment lies in the large-scale component of variation, by enabling the distinction of deterministic large scale spatial trends in VNIR soil data. The strong fine-scale variability inherent in the soil-scanning data obscures the underlying soil-RS data relationship that could be explained by regression methods. Spatial aggregation applied on soil-scanning data can address this issue by reducing the observed variability of soil information, meaning that a block value is computed from all point values within the block. Model-based approaches in aggregation such as block-kriging interpolation can be applied for that purpose [32]. Although a model-based approach could potentially produce a more accurate estimate than a simple block average, the necessity to make some kind of stationarity assumption about the VNIR soil variables, makes the model-based approach questionable. The accuracy obtained from block estimation is built on the preconception that the attribute of interest satisfies the assumed spatial correlation model. Thus, scanned soil data were spatially aggregated over the extent of Sentinel-2 biophysical parameters, using simple block-average, i.e., mean value in 10 m discretized cell. Spatial aggregation is essentially a change of support in VNIR data directed upwards in terms of spatial scale, applied exclusively to assist the evaluation of the soil–RS data relationship.

2.5. Variogram Analysis

Spatial heterogeneity of soil properties is caused by a number of factors and processes acting at different spatial and temporal scales. While the CV gives a relative estimate of a property’s variability, it provides no information about how that variability is distributed spatially. Geostatistical tools such as variogram analysis allow differentiation between
spatially structured and unstructured variation [33]. Multivariate variogram analysis in the input dataset of VNIR variables included the calculation of direct semivariograms and cross-semivariograms, aiming to identify several subprocesses at different scales linked with different spatial correlation structures.

Initially, a global threshold of 0.1 m minimum distance was applied to all point distributions to eliminate erroneous, repetitive point recordings. The distribution of nearest neighbor distances for the available datasets was heavily concentrated around a mean of approximately 1 m, with a minimum of 0.15 m and a maximum of 12 m, an indication of a much higher point density along than between the transect lines. Exploratory fine-scale directional variograms, with lag distance of 0.5 m and up to 120 m, revealed strong small-scale variability, typical for all soil parameters. Presented for moisture content (MC %), the noisy fine-scale variability observed for a distance up to 10 m along the line-scanning direction $\theta$ (Figure 4a), is characteristic of the fine-scale variation and was evident in all the available soil parameters. Across the line-scanning direction or $\theta + 90^\circ$ (Figure 4b), the observed peak variation pattern matches the transition between 12 m apart scanning lines.

![Figure 4](image)

**Figure 4.** Directional fine-scale experimental variograms for visible and near infrared (VNIR) predicted moisture content (MC %) (lag = 0.5 m, cutoff = 125 m). Anisotropy is calculated along (bearing from north $\theta^\circ$) (a) and across ($\theta^\circ + 90^\circ$) line-scanning direction (b).

For multivariate analysis, the spatial intersection of the different VNIR datasets was considered, resulting in a common $\approx$26 K point dataset for each of the five VNIR soil variables. For computational efficiency, isotropic variogram analysis was carried out on the aggregated VNIR soil variables instead of the original VNIR data. Trial computations on subsets of 7–8 K points from the common dataset, revealed great similarity in the detected spatial structures between large-scale (10 m lag) variograms from original and aggregated data. As such, all 15 isotropic direct and cross variograms for the selected soil variables were computed on the aggregated common dataset of five soil variables using RGeostats Geostatistical R Package (MINES ParisTech/ARMINES 2021).

The combined effect of different sources of spatial variation that operate at 3 distinct scales, yields the nested direct and cross-semivariograms (isotropic) of Figure 5, which were modeled as the sum of a nugget effect (<1 m) and three different models of range 53 m, 164 m and 747 m accordingly. This way the observed variation can be viewed as the sum of different microscopic, short-range, medium-range and long-range spatial components. Although spatial components are mathematical constructions with no a priori physical meaning, they can be helpful to identify the main sources of spatial variation and improve our understanding of underlying physical processes in soil [34].
2.6. Ordinary Kriging

The available point-support soil scanner data of the soil’s physical-chemical properties were initially processed using the OK (geostatistical interpolation technique, which revealed the inherent limitations of such an approach. The size of the data set N is a major drawback in computing kriging spatial predictors, since its computational cost is of order N^3. Alternatively, a localized version of OK was considered using the local neighborhood approach, where only a limited number of neighbors was used for prediction purposes, while retaining the variogram modelling as global. Variance/covariance structures were modelled using mostly isotropic ‘stable’ model functions, variants of the Gaussian family of models [35]. Detailed description of variance/covariance modeling for each variable is given in Table 3, along with RMSE from cross-validation applied on the established models. ArcGIS 10.8.1 Geostatistical Analyst (ESRI) was used for variogram modelling and OK interpolation.
Table 3. Spatial structure modelling for ordinary kriging interpolation.

| Soil Variable | Model/Direction | Nugget | Partial Sill | Sill | Nugget/Sill | Range (m) | Lag Size (m) | Cut-Off (m) | Prediction Neighbors | Cross-Validation RMSE |
|---------------|-----------------|--------|--------------|------|-------------|-----------|--------------|-------------|-----------------------|----------------------|
| Ca            | stable/isotropic | 0.0171602 | 33 | 33.01716 | 0.000520 | 112 | 14 | 168 | 200 | 2.844 |
| MC            | stable/isotropic | 0.2366 | 0.808 | 1.0446 | 0.226498 | 180 | 15 | 180 | 200 | 2.397 |
| Mg            | spherical/isotropic | 0.001157 | 2.5345 | 2.535657 | 0.000456 | 112 | 14 | 168 | 200 | 0.603 |
| Na            | stable/isotropic | 0.01095 | 0.01388 | 0.02683 | 0.408125 | 244 | 18 | 360 | 200 | 0.1037 |
| pH            | stable/isotropic | 0.4913 | 0.4721 | 0.9634 | 0.509965 | 135 | 17 | 204 | 200 | 0.1421 |

MC: moisture content.

Localization of the interpolator was realized by a 4 equal-sector circular neighborhood with one axis at 42° bearing from north, following the orientation of the VNIR soil line-scanning. The constraint of 200 neighbor points for prediction, regardless of the detected range of spatial auto-correlation, was imposed by the software that has an inherent restriction of maximum neighbor points for OK interpolation. This requirement was satisfied by isotropic distances up to 15 m approximately, characteristic of the locality of the interpolator.

2.7. Fixed-Rank Kriging

Fixed rank kriging prediction is based on dimensionality reduction of large datasets, by expressing the spatial process as a linear combination of prespecified spatial basis functions, providing computational feasibility for calculating the covariance matrix. Recommendations for the spatial basis functions are that the center points vary in spatial resolution in order to capture different scales of spatial variations, while regularly covering the entire spatial domain [9]. Examples include Gaussian, exponential covariance function, bisquare, or Matérn covariance function with varying smoothness parameters. The multi-resolution bisquare functions, enabling the covariance function model to capture multiple scales of spatial variation, have the following form

\[
f_{bis}(s) := \left\{ 1 - \left( \frac{\| s - m_b \|}{d_b} \right)^2 \right\}^2, \quad \| s - m_b \| < d_b
\]

where \( m_b \) is the center point of the function at \( b \)th resolution, \( b = 1, 2, 3, \ldots \) and \( \| \cdot \| \) denotes the Euclidean distance from the center for location \( s \). The range parameter \( d_b \) controls the support of the function and is defined as 1.5 times the shortest distance between two center points for a specific scale [9].

In FRK package, the spatial distribution of basis functions is carried out automatically, based on the desired number of resolutions and the magnitude of the spatial covariance functions, controlled by the scale aperture parameter [36]. The estimation of the covariance matrix is accomplished using the SRE model, which yields maximum-likelihood estimates of the covariance matrix of basis functions and the fine-scale variation component using the expectation-maximization (EM) algorithm [37]. The EM algorithm is an iterative computational technique where dimensionality reduction is the critical feature that makes it possible to deal with a very large number of observations. Iteratively, at each filtering step of EM-estimation, the computational complexity is reduced, i.e., the required number of operations increases linearly with \( N \), ensuring in this way the computational feasibility of the method. Estimation of the measurement-error variance from the instrument is carried out using the estimated semi-variogram near the origin of distance vector. Model fitting is performed by weighted least squares linear, since the semi-variogram function may be assumed to be linear at the very small lags concerned. Kang et al. [38] showed that measurement-error can be estimated unbiasedly from the fitted line’s intercept. The establishment of SRE models was the initial FRK implementation step. Using as generic basis function the bisquare function for spatial covariance structure modeling as suggested by the previous work of Nychka et al. [39], a unique SRE model was formulated for each
soil variable at three distinct resolutions for a particular distribution of basis functions. Two different distribution schemes were constructed (Table 4) using the helper function from FRK package, with different scale aperture parameter settings:

- Scale aperture = 1, almost matched the detected spatial structures from variogram analysis, by distributing a variable number of bisquare basis functions at resolutions of 53, 174 and 696 m, approximately, according to each VNIR soil variable.
- Scale aperture = 1.25, yielded a distribution of a variable number of bisquare basis functions at resolutions of 66, 217 and 869 m, approximately, according to each VNIR soil variable. The target resolutions overlap the detected spatial structures from variogram analysis, and were adopted in order to assess the effect of the spatial covariance function magnitude on the filtering performance of the interpolator.

Table 4. Distribution of 3-resolution bisquare basis functions according to scale aperture.

| VNIR Soil Variable | Scale Aperture: 1 | Scale Aperture: 1.25 |
|--------------------|-------------------|----------------------|
|                    | Res.1  | Res.2  | Res.3  | Total count | Res.1  | Res.2  | Res.3  | Total Count |
| Ca                 | 52.68  | 173.88 | 695.52 | 480        | 66.09  | 217.34 | 869.74 | 472         |
| MC                 | 52.80  | 173.96 | 695.85 | 477        | 66.10  | 217.59 | 870.34 | 468         |
| Mg                 | 52.92  | 173.64 | 694.55 | 480        | 66.13  | 216.93 | 867.74 | 467         |
| Na                 | 52.89  | 173.44 | 693.75 | 483        | 66.23  | 217.06 | 868.25 | 471         |
| pH                 | 52.73  | 173.64 | 694.58 | 482        | 66.09  | 217.16 | 868.64 | 470         |

The different distributions of bisquare basis functions over the study area, along with their footprint or spatial support, are displayed in Figure 6.

Figure 6. Spatial distribution of bisquare basis functions in study area (red outline), at 3 resolutions for each scale aperture.

Two different variants of FRK were applied on VNIR soil variables for each distribution scheme of basis functions. In the univariate case of FRK, each dataset was processed independently, whereas in the multivariate case of FRK-BioPAR, biophysical parameters from Sentinel-2 data were summarized through their PCAs and included as predictors in the interpolation of each soil variable. In either FRK variant, each soil variable was treated as a spatially correlated random process or signal, decomposed using a linear combination of spatial basis functions, according to the scale aperture case. In univariate FRK this process corresponds to the original line-scanning soil variable, while FRK-BioPAR
uses the signal that remains after the removal of a deterministic trend, likewise regression-kriging (RK) [30]. The formulation of a unique SRE model for each variable was carried out similarly between FRK variants, using the same local basis function distribution scheme according to the scale aperture case. Processing of input datasets and FRK implementation was carried out in the R software environment (R Development Core Team, 2011).

2.8. Metrics for Evaluation

The performance of each interpolation method was quantified in terms of estimation accuracy (MAE: mean absolute error, RMSE: root mean square error), and prediction efficiency RPIQ or ratio of performance to the interquartile range. The MAE gives information on the average magnitude of prediction errors but not on the direction. It is less sensitive to outliers than the commonly used RMSE [40]. To assist the comparison between OK and both FRK interpolation methods, we computed the prediction accuracy relative improvement percentage or RI (%) of the FRK methods over the OK interpolation method, i.e., the \[
\left(\frac{\text{RMSE}_{\text{OK}} - \text{RMSE}_{\text{FRKs}}}{\text{RMSE}_{\text{OK}}}\right) \times 100
\]
[41]. The RPIQ which is defined as the interquartile range of the observed values divided by the RMSE of prediction, takes both the prediction error and the variation of observed values into account, providing a dimensionless statistic towards a more effective evaluation of model accuracy, regardless of the distribution of the input variables [42].

The denoising or filtering performance of the interpolators was assessed by comparison of all three interpolation results against the noisier spatial aggregation output of simple block-average of VNIR soil variables, i.e., mean value in 10 m discretized cell, which was considered as a benchmark reference. Upon these datasets, the following relevant metrics were computed:

MSE (noise): The actual difference between interpolation and spatial aggregation output in each case is regarded as noise and is considered to be the result of zero-mean Gaussian process. The mean-square error (MSE) for each interpolation method is the cumulative squared error between the noisier spatial aggregation output and each interpolation output, i.e.,

\[
\text{MSE}_{\text{noise}} = \frac{1}{MN} \sum_{i,j=1}^{M,N} \left( S_{ij} - \hat{S}_{ij} \right)^2,
\]

where \( S \) is the spatially aggregated output and \( \hat{S} \) the interpolated output of dimensions \( M \times N \). This metric is fundamental for the assessment of filtering performance of the interpolators, as it provides a quantitative measure of the removed fine-scale variability.

Relative differences (\( d_v \)): relative differences are often used as a quantitative indicator between interpolation methods [43]. We used \( d_v \) (%) to compare the three interpolation resulting maps against the reference (noisier) spatial aggregation output, i.e.,

\[
d_v = \left( \frac{S - \hat{S}}{S} \right) \times 100,
\]

where \( d_v \) is the relative difference (%), \( S \) is the reference spatially aggregated output, and \( \hat{S} \) is the interpolation results from the three methods involved. Individual \( d_v \) quantile differences between the three interpolation methods provide a quantitative measure of the filtering ability of each interpolation approach.

The ability of the interpolators to produce spatial patterns that exhibit the maximum possible homogeneity is essential for the evaluation of the produced maps as inputs in the delineation of site-specific MZ. High-filtering performance of an interpolation method does not necessarily imply uniformity in the generated spatial patterns of variation. In order to assess the ability of the interpolators to produce such patterns, we used the following metrics that characterize the presence and structure of patterns in categorical maps, computed upon quantile-based classified interpolation results from the three methods involved that were used to resemble homogeneous MZ [44].

Aggregation index (\( Ai \)) describes the presence of patterns and specifically if the exhibited patches (i.e., neighboring cells belonging to the same class) are rather clumped (aggregated) or tend to be isolated. \( Ai \) (%) equals 0 for maximally disaggregated and 100 for maximally aggregated classes [45].
Contiguity index (Ci) describes the structure of the exhibited patches in every class of the categorical map, by assessing the spatial connectedness (contiguity) of cells in patches for a particular class. It varies from 0 for one-pixel patches and increases to a limit of 1 for fully connected patches. The overall Ci is the average of the computed indices for each class of the categorical map [45].

3. Results
3.1. Soil-Biophysical Parameter Relationship

The Spearman’s rank correlation coefficients between block-aggregated VNIR soil scanner data and transformed RS biophysical indicators were calculated (Table 5). The individual correlations between PCAs from biophysical parameters and soil variables appeared to be very low to moderate for the majority of VNIR variables. The most promising case of correlated covariates was Mg vs PC2.NDVI ($\rho = -0.56$) and PC2.fAPAR ($\rho = -0.51$). Multi-collinearity between predictors was present. The biophysical parameter PCAs that were considered as predictors were independent by definition, but only for each particular indicator case, i.e., NDVI or fAPAR; in-between indicators NDVI and fAPAR are closely related [27]. Several studies have also shown strong linear NDVI-fAPAR relationships over a range of land cover types, insensitive to vegetation heterogeneity and scale [46]. However, the soil background influences vegetation indices encompassing soil reflectance and crop residue contributions, which, along with atmospheric and bidirectional effects can alter this relationship and affect its linearity with important consequences to scaling [47]. The strength of correlation between soil properties and RS-based biophysical parameters can also decrease with increasing heterogeneity [19]. In this context, the correlation between different biophysical parameters does not necessarily imply information redundancy and could help stabilize RS-soil models against noise [48].

Table 5. Correlation table between aggregated soil variables and PCAs from biophysical indicators.

| Aggregated Data on 10 m Grid | PC1.NDVI | PC2.NDVI | PC3.NDVI | PC1.fAPAR | PC2.fAPAR | PC3.fAPAR |
|-----------------------------|----------|----------|----------|------------|-----------|-----------|
| Ca                          | 0.21 *** | −0.52 ***| 0.07 *** | 0.14 ***  | −0.47 *** | −0.07 *** |
| MC                          | 0.07 *** | −0.17 ***| 0.04 **  | −0.08 *** | −0.19 *** | −0.02 *** |
| Mg                          | −0.24 ***| −0.56 ***| 0.06 *** | 0.15 ***  | −0.51 *** | −0.08 *** |
| Na                          | −0.26 ***| −0.14 ***| 0.08 *** | 0.22 ***  | −0.06 *** | 0.06 ***  |
| pH                          | 0.14 *** | 0.14 *** | 0.02     | −0.11 *** | 0.09 ***  | 0.04 **   |

Table presents Spearman correlation coefficients with pairwise deletion. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, MC: moisture content.

Based on the isotropic variograms of the transformed biophysical parameters over the study area (Figure 7), the first three principal components (PCs) of NDVI and fAPAR, exhibited long-range spatial structures of approximately 400–500 m range of varying magnitude (sill). PC1 and PC2 from both NDVI and fAPAR presented the highest structural spatial dependence and were included in the FRK with covariates, in order to assess their influence in the interpolation of each VNIR soil variable.

The selected covariates were imposed as a spatial trend in FRK, leaving the residual variation to be captured by the covariance decomposition process carried out subsequently. The de-trending process is carried out internally in the FRK package using ridge regression, a form of regularized regression seeking to diminish the consequences of multi-collinearity, poorly conditioned equations, and overfitting, by shrinking the coefficients of correlated predictors towards each other and reducing their variance. By separately conducting ridge regression between selected covariates and aggregated VNIR variables, using the glmnet package [49], the optimal values for the regularization parameter $\lambda$ (lambda) were extracted (Table 6), and parsed to the FRK package for regression modelling.
Figure 7. Biophysical parameter PCAs isotropic variograms (pixel-wise or 10 m lag), for NDVI-PCAs (a) and fAPAR-PCAs (b).

Table 6. Ridge regression results between selected covariates and aggregated VNIR data (n = 4965).

| Aggregated VNIR Soil Variable | RMSE  | R2    | Lambda |
|-------------------------------|-------|-------|--------|
| Ca                            | 3.063 | 0.245 | 0.169  |
| MC                            | 2.869 | 0.043 | 0.048  |
| Mg                            | 0.815 | 0.316 | 0.053  |
| Na                            | 0.126 | 0.034 | 0.002  |
| pH                            | 0.149 | 0.033 | 0.002  |

3.2. Ordinary Kriging Predictions

Application of OK on all the available soil variables, including optimization of interpolation parameters, global variogram estimation, modelling, and neighborhood selection, resulted in the maps of predicted surfaces for every soil variable in Figure 8A. The nugget-to-sill ratio indicated moderate to strong spatial autocorrelation for the available soil variables; different scales of spatial structures were detected though, with a range varying from 106 up to 244 m. Maximum nugget-to-sill ratio in pH revealed the weakest overall spatial dependence, which is mainly related to the very short extent (CV = 2.2%) of the exhibited variation in the pH VNIR dataset. In the case of extractable cations in VNIR datasets, the order of CV was Ca < Mg < Na, while the order of their spatial dependence was Ca ≈ Mg > Na. The extreme variation in the Na soil parameter and the exhibited moderate spatial dependence are largely influenced by irrigation management practices.

Lower nugget-to-sill ratios (Ca, Mg), which suggested stronger spatial dependence, were probably controlled by systematic management factors, i.e., fertilization, farming measures or cropping systems. Predictions with OK were obtained on a 10 m regular grid, fully covering the area under investigation. All prediction surfaces generated by OK geostatistical models, were block-support interpolations, i.e., each 10 m estimation cell was treated as a block estimate of a 3 × 3 grid within the cell. As expected, the OK-predicted rasters revealed their local variability and noisy appearance, evidence of the low-filtering abilities of OK (Figure 8A).
Figure 8. OK predictions (A), FRK predictions (B), FRK-BioPAR predictions (C).
3.3. Fixed Rank Kriging Predictions

The different distribution schemes of bisquare basis functions over the study area introduced two sets of scales or resolutions for FRK interpolation on VNIR soil variables. The three proposed scales at each scheme were considered adequate to capture the short to medium and long-range components of the spatial variation of soil variables, but to a different individual extent. The difference in the magnitude of the spatial covariance functions directly reflects the filtering performance of the interpolator, and was immediately visible in the FRK interpolation results for all the selected VNIR variables. Presented only for the moisture content (MC%) case in Figure 9, the effect of applying basis function scales that overlap the detected spatial structures increased the filtering performance substantially. As such, the basis function distribution scheme that corresponds to scale aperture 1.25 was adopted for the rest of the analysis and is presented in the resulting tables and maps.

![Figure 9. FRK interpolation results (moisture content %) for two different bisquare function distribution schemes.](image)

The number of estimation cells varied accordingly with the size of the VNIR datasets (Table 7), upon which the univariate FRK prediction was obtained at a spatial resolution of 10 m. FRK interpolation resulting maps are presented in Figure 8B.

In the case of FRK with covariates (FRK-BioPAR), prior to the application of the SRE models, each dataset was de-trended by means of regression using the PCAs from biophysical parameters. The large-scale spatial trends that were introduced by biophysical parameters through their PCAs are evident in the relevant isotropic variograms over the study area (Figure 7). Estimation cell size was chosen to be identical to the pixel size of Sentinel-2 images, making in this way the use of biophysical indicators possible as covariates, without the change of support problem that inevitably arises from the data-fusion interpolation process. The FRK-BioPAR interpolated maps are presented in Figure 8C. Details describing the two different FRK implementations and relevant SRE models for each soil variable are given in Table 7.
Table 7. FRK interpolations summary of spatial random effects (SRE) model’s summary for FRK and FRK-BioPAR.

| Soil Var | Method       | Covariates               | ML Estim. | Number of Iterations | Basis Functions | Estimation Cells | Mean obs. Variance at Cell | Elapsed Time (s) |
|----------|--------------|--------------------------|-----------|----------------------|-----------------|-------------------|-----------------------------|------------------|
| Ca       | FRK          | PC1,PC2,PC1,PC2,PC4      | EM        | 21                   | 476             | 5086              | 5.8437                      | 308.2            |
|          | FRK-BioPAR   | PC1,PC2,PC1,PC2,PC4      | EM        | 111                  | 472             | 5113              | 5.8437                      | 1618.1           |
| MC       | FRK          | PC1,PC2,PC1,PC2,PC4      | EM        | 14                   | 468             | 5086              | 4.1921                      | 182.7            |
|          | FRK-BioPAR   | PC1,PC2,PC1,PC2,PC4      | EM        | 26                   | 470             | 5096              | 4.1921                      | 333.9            |
| Mg       | FRK          | PC1,PC2,PC1,PC2,PC4      | EM        | 16                   | 472             | 5086              | 0.2363                      | 216.2            |
|          | FRK-BioPAR   | PC1,PC2,PC1,PC2,PC4      | EM        | 107                  | 467             | 5112              | 0.2363                      | 1452.2           |
| Na       | FRK          | PC1,PC2,PC1,PC2,PC4      | EM        | 19                   | 471             | 5086              | 0.0076                      | 250.3            |
|          | FRK-BioPAR   | PC1,PC2,PC1,PC2,PC4      | EM        | 27                   | 473             | 5119              | 0.0076                      | 361.4            |
| pH       | FRK          | PC1,PC2,PC1,PC2,PC4      | EM        | 24                   | 471             | 5086              | 0.0152                      | 276.4            |
|          | FRK-BioPAR   | PC1,PC2,PC1,PC2,PC4      | EM        | 38                   | 470             | 5086              | 0.0152                      | 444.3            |

MC: moisture content, ML: maximum-likelihood estimator, EM: expectation-maximization, BAU: basic areal unit.

3.4. Performance Assessment and Comparison

Performance assessment of the interpolators in terms of estimation accuracy (MAE, RMSE) and prediction efficiency (RPIQ) was accomplished using the validation soil sample dataset depicted in Figure 1b and statistically described in Table 2. Summarized validation results are presented in Table 8. The highest RMSE relative improvement (RI) % was recorded in the MC case, where both FRK variants exhibited a slight error reduction in comparison with OK results, while remaining negligible for the rest of the soil parameters. The modest differences on RPIQ that were recorded between FRK and OK methods, favor both FRK variants over OK, and only for specific variables, i.e., pH, and soil moisture (MC) while appearing marginal for the rest of the soil parameters. The greater RPIQ translates into an improvement of the model’s predictive capacity [42].

Table 8. Validation summary of the three interpolation methods (OK, FRK, FRK-BioPAR) for different soil properties (n = 100 validation points).

| Soil Variable | Ordinary Kriging | Fixed Rank Kriging | Fixed Rank Kriging with BioPAR |
|---------------|------------------|--------------------|--------------------------------|
|               | MAE, RMSE, RPIQ  | MAE, RMSE, RPIQ    | MAE, RMSE, RPIQ                |
| Ca (meq/100 g)| 4.118, 6.242, 1.16| 4.005, 6.118, 1.176| 4.037, 6.124, 1.175, 1.89      |
| MC (%)        | 2.594, 3.149, 1.445| 2.436, 2.933, 1.527| 2.432, 2.921, 1.533, 7.24       |
| Mg (meq/100 g)| 1.022, 1.457, 1.334| 0.9833, 1.399, 1.402| 0.9912, 1.404, 1.397, 3.64      |
| Na (meq/100 g)| 0.157, 0.201, 1.846| 0.1565, 0.1995, 1.819| 0.1559, 0.1995, 1.819, 0.75      |
| pH            | 0.104, 0.144, 1.153| 0.1035, 0.139, 1.196| 0.1039, 0.1392, 1.194, 3.33      |

MC: moisture content, RI*%: identical relative improvement with scale aperture = 1.

In terms of filtering ability, or fine-scale variability removal, performance was assessed using relative differences distribution and noise statistics ($d_v$, $MSE_{noise}$). Computed noise statistics regarding the filtering results for each interpolation method and every soil variable are summarized in Table 9, along with quantile information concerning the distribution of the resulting relative differences $d_v$, and direct noise statistics accordingly, i.e., mean, SD and Shapiro–Wilk normality test (Royston, 1982). Judging from the relative differences $d_v$ distribution and $MSE_{noise}$ metrics in Table 9, the FRK-BioPAR interpolator with the integration of biophysical parameters as covariates yielded the maximum noise removal for all the available soil variables. The differences in the denoising performance between the two FRK variants is marginal. In contrast, both FRK variants exhibited substantial improvement over OK in the filtering ability of small scale variation in VNIR datasets, doubling at least the relevant noise metric of $MSE_{noise}$. Resembling a zero-mean Gaussian process, noise distribution in all variable cases exhibited only minor deviations from normality, and approximately zero-mean characteristics (Table 9) [50].
Table 9. Filtering performance of interpolation methods for different soil properties.

| Soil Variable | Method      | 2%  | 25%  | 50%  | 75%  | 98%  | MSE  | MSE*  | Noise Statistics | Mean | St. Dev | W-Test |
|---------------|-------------|-----|------|------|------|------|------|-------|-----------------|------|---------|--------|
| Ca (meq/100 g) | OK          | −9.41 | −1.518 | 0    | 1.39 | 6.47 | 1.5  | (1.5) | −4.08 × 10^{-2} | 1.22 | 0.93    |        |
|               | FRK         | −16.86 | −2.755 | 0.2539 | 2.893 | 9.61 | 4.21 | (3.95) | −1.44 × 10^{-2} | 2.05 | 0.952   |        |
|               | FRK-BioPAR  | −16.82 | −2.811 | 0.2847 | 2.928 | 9.59 | 4.27 | (4.04) | −1.91 × 10^{-2} | 2.07 | 0.955   |        |
| MC (%)        | OK          | −30.14 | −4.068 | 0    | 3.232 | 14.13 | 1.22 | (1.22) | −7.11 × 10^{-2} | 1.1  | 0.938    |        |
|               | FRK         | −52.65 | −10.182 | 0.2539 | 2.893 | 9.61 | 4.21 | (3.95) | −2.96 × 10^{-2} | 2.03 | 0.99    |        |
|               | FRK-BioPAR  | −53    | −10.062 | 0.6773 | 8.855 | 24.25 | 4.21 | (3.84) | −3.02 × 10^{-3} | 2.05 | 0.992   |        |
| Mg (meq/100 g) | OK          | −15.73 | −2.953 | 0    | 1.822 | 10.11 | 0.0881 | (0.0881) | −2.26 × 10^{-2} | 0.296 | 0.925    |        |
|               | FRK         | −30.14 | −4.91 | 0.6942 | 5.521 | 16.98 | 0.283 | (0.273) | 1.62 × 10^{-2} | 0.532 | 0.977    |        |
|               | FRK-BioPAR  | −31.71 | −5.11 | 0.6782 | 5.563 | 17.49 | 0.292 | (0.28) | 1.54 × 10^{-2} | 0.54  | 0.975    |        |
| Na (meq/100 g) | OK          | −49.8  | −4.537 | 0    | 3.853 | 21.37 | 0.0017 | (0.0017) | 1.51 × 10^{-4} | 0.0412 | 0.912    |        |
|               | FRK         | −93.27 | −9.493 | 0.0997 | 8.208 | 29.48 | 0.00424 | (0.00406) | −1.99 × 10^{-4} | 0.0651 | 0.962    |        |
|               | FRK-BioPAR  | −93.15 | −9.965 | −0.2347 | 8.062 | 29.57 | 0.00426 | (0.00407) | −1.14 × 10^{-3} | 0.0653 | 0.966    |        |
| pH            | OK          | −2.07  | −0.362 | 0    | 0.298 | 1.62 | 0.0044 | (0.0044) | −4.49 × 10^{-3} | 0.0662 | 0.925    |        |
|               | FRK         | −2.81  | −0.667 | 0.0612 | 0.75 | 2.47 | 0.0107 | (0.0102) | 9.10 × 10^{-4} | 0.103 | 0.985    |        |
|               | FRK-BioPAR  | −2.83  | −0.661 | 0.0842 | 0.791 | 2.55 | 0.011 | (0.0104) | 2.79 × 10^{-3} | 0.105 | 0.988    |        |

MC: moisture content, MSE*: identical noise metric with scale aperture = 1.

In terms of the ability to generate uniform patterns of spatial variation, performance was assessed using the relevant metrics that characterize the presence and structure of spatial patterns in quantile-based classified maps for each VNIR soil variable.

Based on the spatial pattern metrics, Ai, Ci are presented in Table 10, and in comparison with OK, a substantial improvement in the ability of both FRK interpolators to produce homogenous patterns of spatial variability was recorded in all variable cases, reaching 65% in the best case of Ci in Na variable. The difference between the two FRK variants lies mostly on the structure of the observed patterns (Ci), favoring FRK over FRK-BioPAR for all VNIR variables.

Table 10. Spatial pattern metrics of classified interpolation results: aggregation index (Ai) % and contiguity index (Ci) map average (n = 4 quantile-based classes).

| Method      | Ca | MC | Mg | Na | pH |
|-------------|----|----|----|----|----|
|             | Ai | Ai | Ai | Ai | Ai |
| OK          | 66.93 | 61.2 | 68.23 | 74.56 | 64.21 |
| FRK         | 85.23 | 80.74 | 89.43 | 87.31 | 81.7 |
| FRK-BioPAR  | 84.37 | 79.89 | 88.72 | 86.69 | 80.47 |

In terms of computational efficiency, OK has a major advantage over both FRK variants; the actual OK interpolation process on the VNIR scanner data was accomplished rapidly for every soil variable. The computational cost of either FRK method was certainly multiple times greater (200–1600 s) compared to the OK approach but reasonable enough for multiple repetitions (Table 7). The additional computational cost in the FRK-BioPAR interpolations (2–7 times more) comes as a result of the increased model complexity.
4. Discussion

FRK spatial interpolations on VNIR soil data were obtained on a discretized domain of 10 m cell size that was deemed sufficiently detailed for mapping field-scale variability without the issue of strong fine scale variability that obscures the meaningful patterns of soil variation that are commonly expected in PA applications. Despite the burden in computational efficiency, this increased processing load imposed by FRK interpolation on VNIR soil predictions, does not reflect strongly on the accuracy of the interpolator nor the efficiency of the prediction. The greatest improvement of both FRK variants over the OK approach concerns the filtering ability of the interpolators and their ability to generate homogenous patterns of spatial variation. Both FRK and FRK-BioPAR were capable of superior filtering results compared with OK, meaning that FRK has the potential to maximize the removal of fine-scale variation inherent in the VNIR datasets. The improvement of the FRK-BioPAR interpolator over the univariate FRK approach in filtering performance (noise removal) was marginally recorded for all VNIR soil variables. The ability of the FRK interpolator to generate uniform patterns of spatial variation did not seem to benefit from the inclusion of biophysical parameters as covariates in the interpolation process. This characteristic feature of the interpolator is governed by the distribution scheme of the basis functions applied, and particularly by the magnitude of the spatial covariance functions that is imposed in the interpolator across the different scales. Results have shown (Figure 10) that the homogeneity of spatial variation patterns is increased by applying basis functions distribution scales greater than the detected spatial structures that come as a result of variogram analysis and modelling. Related to the above increase in homogeneity is the fact that both RMSE relative improvement and noise removal (MSEnoise) are favored by the application of basis function distributions that correspond to scale aperture 1.25 (Tables 8 and 9).

Concerning the appearance of patterns in the resulting maps, line-scanning proximal sensors collect VNIR spectral data at much higher resolution along the travel direction than across. Due to this difference in resolution, interpolation often suffers from irregular spatial distribution, where measurement lines are clearly visible on the resulting maps [51]. The spatial distribution of VNIR spectrometer data in the current study was governed by the applied scanning scheme of parallel transects 12 m apart, with a bearing of approximately 42° from north, following the variable-rate treatment scheme that was established in stripes across the entire field (Figure 1c). Present in every VNIR dataset, this characteristic distribution significantly affected the appearance of common linear patterns in all the interpolated maps (Figure 8). FRK interpolators tend to minimize this effect of uneven scanning resolution by aggregating adjacent linear patterns together, presenting a more uniform spatial variation pattern in the resulting maps.
FRK is an advanced, nonstationary spatial interpolation method that can be implemented on very large spatial datasets, providing the necessary geostatistical framework for data-fusion applications of massive remote and proximal sensing datasets. The reduction in the dimensionality of the computations with VNIR datasets that was accomplished with FRK implementation was from an average size of dataset \( N = 30,917 \) to \( r = 480 \), i.e., the average total number of basis functions applied. The filtering abilities of the FRK interpolator coupled with the generation of homogenous spatial patterns, along with the capacity to handle a large volume of spatial information, nominate the method as ideal for PA applications with massive proximal and remote sensing datasets. As has been pointed out, modeling the covariance function through the SRE model, gives a unique change-of-support properties to the FRK interpolator [37]. The multiresolution nature of the basis functions—such as the bisquare type—in SRE models, enables the covariance function model to capture multiple scales of variation; a large spatial scale that is missed by the detrending procedure can potentially be recovered by some of the spatial random-effects components of the model.

The greatest constraints regarding the implementation of FRK, arise from its own formulation and establishment of the involved SRE model. The construction and distribution of the adopted basis functions greatly affects the computational cost of FRK methods. The different scales that are imposed by the formulation of basis functions should reflect the pragmatic scales of spatial variation for all the available soil parameters. Such a requirement is difficult to meet with an overall local basis function distribution scheme. Exploring the similarities on spatial structure between soil variables can be helpful for partitioning the datasets prior to interpolation and connecting them with different basis-function schemes, on the basis of resolution and domain coverage extent. An overdetailed modelling approach, with abundance in the distribution of basis functions across multiple scales of variation would disproportionately extend the computational cost. Future work may focus on alternative RS-based indicators or types of covariates (terrain-based) and specific aspects of FRK interpolation such as the application of different types and schemes of multiresolution basis functions in the modeling process.

5. Conclusions

An FRK non-stationary geostatistical interpolation method was used to overcome the computational bottleneck of processing massive line-scanning soil datasets and to realize the production of smooth, interpolated maps of soil properties. The results from FRK application in univariate and RS data covariate cases have shown an improved performance over the popular kriging variant of OK, both in terms of filtering performance and the ability to generate uniform patterns of spatial variation. The latter is of great value for the delineation of site-specific MZ, which constitutes a major objective in PA applications. A data fusion approach was adopted using the FRK geostatistical framework, in order to take advantage of the complementary features provided by Sentinel-2 biophysical crop parameters. The incorporation of auxiliary RS data-based indicators as covariates in the interpolation, although feasible, did not justify the computational cost imposed by the overall optimization process. Overall, it facilitated the FRK interpolator to maximize the filtering results, with no major impact on the accuracy and the efficiency of the predictions.

**Author Contributions:** Conceptualization, N.K., T.K.A. and A.M.M.; formal analysis and validation, N.K.; field surveys, M.A.M., A.P.G., M.C. and A.O.; data curation, N.K., M.A.M. and A.P.G.; writing—original draft preparation, N.K., T.K.A., M.A.M. and A.M.M.; writing—review and editing, N.K., T.K.A., G.B., A.G., T.R., D.M. and A.M.M.; funding acquisition, D.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This paper is part of the SIEUSOIL project ‘Sino-EU Soil Observatory for Intelligent Land Use Management’, which received funding from the European Union’s Horizon 2020 Research and Innovation Framework Programme under grant agreement No. 818346.

**Institutional Review Board Statement:** Not applicable.
References

1. Auernhammer, H. Precision farming—the environmental challenge. *Comput. Electron. Agric.* 2001, 30, 31–43. [CrossRef]
2. Reznik, T.; Lukas, V.; Charvát, K.; Charvát, K.; Jr; Horáková, Š.; Krivánek, Z.; Hermana, L. Monitoring of in-field variability for site specific crop management through open geospatial information. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* 2016, 41, 1023–1028. [CrossRef]
3. Castrignano, A.; Buttafuco, G.; Quarto, R.; Parisi, D.; Rossel, R.V.; Terribile, F.; Langella, G.; Venezia, A. A geostatistical sensor data fusion approach for delineating homogeneous management zones in Precision Agriculture. *Catena* 2018, 167, 293–304. [CrossRef]
4. Viscarra Rossel, R.; Chappell, A.; De Caritat, P.; McKenzie, N. On the soil information content of visible–narrow infrared reflectance spectra. *Eur. J. Soil Sci.* 2011, 62, 442–453. [CrossRef]
5. Kuan, B.; Mahmood, H.S.; Quraishi, M.Z.; Hoogmoed, W.B.; Mouazen, A.M.; Van Henten, E.J. Sensing soil properties in the laboratory, in situ, and on-line: A review. *Adv. Agron.* 2012, 114, 155–223.
6. Mouazen, A.M.; Alexandridis, T.; Buddenbaum, H.; Cohen, Y.; Moshou, D.; Mulla, D.; Nawar, S.; Sudduth, K.A. Chapter 2—Monitoring. In *Agricultural Internet of Things and Decision Support for Precision Smart Farming*; Castrignano, A., Buttafuco, G., Khosla, R., Mouazen, A.M., Moshou, D., Naud, O., Eds.; Academic Press: Cambridge, MA, USA, 2020; pp. 35–138.
7. Mouazen, A.M.; Maleki, M.; Cockx, L.; Van Meirvenne, M.; Van Holm, L.; Merckx, R.; De Baerdemaecker, J.; Ramon, H. Optimum three-point linkage setup for improving the quality of soil spectra and the accuracy of soil phosphorus measured using an on-line visible and near infrared sensor. *Soil Tillage Res.* 2009, 103, 144–152. [CrossRef]
8. Munna, M.A.; Nawar, S.; Mouazen, A.M. Estimation of secondary soil properties by fusion of laboratory and on-line measured Vis–NIR spectra. *Remote Sens.* 2019, 11, 2819. [CrossRef]
9. Cressie, N.; Johannesson, G. Fixed rank kriging for very large spatial data sets. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* 2008, 70, 209–226. [CrossRef]
10. Cressie, N.; Shi, T.; Kang, E.L. Fixed rank filtering for spatio-temporal data. *J. Comput. Graph. Stat.* 2010, 19, 724–745. [CrossRef]
11. Sun, W.; Minasny, B.; McBratney, A. Analysis and prediction of soil properties using local regression-kriging. *Geoderma* 2012, 171, 16–23. [CrossRef]
12. Van der Meer, F. Remote-sensing image analysis and geostatistics. *Int. J. Remote Sens.* 2012, 33, 5644–5676. [CrossRef]
13. Walter, C.; McBratney, A.B.; Douaoui, A.; Minasny, B. Spatial prediction of topsoil salinity in the Chelif Valley, Algeria, using local ordinary kriging with local variograms versus whole-area variogram. *Soil Res.* 2001, 39, 259–272. [CrossRef]
14. Khosla, R.; Westfall, D.; Reich, R.; Malal, J.; Gangloff, W. Spatial variation and site-specific management zones. In *Geostatistical Applications for Precision Agriculture*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 195–219.
15. Lal, R.; Stewart, B.A. *Soil-Specific Farming: Precision Agriculture*; CRC Press: Boca Raton, FL, USA, 2015; Volume 22.
16. Nawar, S.; Corstanje, R.; Halcro, G.; Mulla, D.; Mouazen, A.M. Delineation of soil management zones for variable-rate fertilization: A review. *Adv. Agron.* 2017, 143, 175–245.
17. Shaddad, S.; Madrau, S.; Castrignano, A.; Mouazen, A. Data fusion techniques for delineation of site-specific management zones in a field in UK. *Precis. Agric.* 2016, 17, 200–217. [CrossRef]
18. Bevington, J.; Scudiero, E.; Teatini, P.; Vellidis, G.; Morari, F. Factorial kriging analysis leverages soil physical properties and exhaustive data to predict distinguished zones of hydraulic properties. *Comput. Electron. Agric.* 2019, 156, 426–438. [CrossRef]
19. Keskin, H.; Grunwald, S. Regression kriging as a workhorse in the digital soil mapper’s toolbox. *Geoderma* 2018, 326, 22–41. [CrossRef]
20. Casa, R.; Castaldi, F.; Pascucci, S.; Basso, B.; Pignatti, S. Geophysical and hyperspectral data fusion techniques for in-field estimation of soil properties. * Vadose Zone J.* 2013, 12, 1–10. [CrossRef]
21. Nguyen, H.; Cressie, N.; Braverman, A. Spatial statistical data fusion for remote sensing applications. *J. Am. Stat. Assoc.* 2012, 107, 1004–1018. [CrossRef]
22. Wulf, H.; Mulder, T.; Schaepman, M.E.; Keller, A.; Jörg, P.C. *Remote Sensing of Soils*; University of Zurich: Zurich, Switzerland, 2015; pp. 71.
23. Mulder, V.; De Bruin, S.; Weyermann, J.; Kokaly, R.F.; Schaepman, M.E. Characterizing regional soil mineral composition using spectroscopy and geostatistics. *Remote Sens. Environ.* 2013, 139, 415–429. [CrossRef]
24. Cressie, N.; Kang, E.L. High-resolution digital soil mapping: Kriging for very large datasets. In *Proximal Soil Sensing*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 49–63.
25. Campos, J.; Gonzalez-Gomez, L.; Villodre, J.; Calera, M.; Campoy, J.; Jiménez, N.; Plaza, C.; Sánchez-Prieto, S.; Calera, A. Mapping within-field variability in wheat yield and biomass using remote sensing vegetation indices. *Precis. Agric.* 2019, 20, 214–236. [CrossRef]
26. Mouazen, A. *Soil Survey Device*. *International Publication Published under the Patent Cooperation Treaty (PCT)*; World Intellectual Property Organization, International Bureau: Geneva, Switzerland, 2006.
27. Weiss, M.; Baret, F. ATBD S2ToolBox Level 2 Products: LAI, FAPAR, FCOVER (Version 1.1). 2016. Available online: http://step.esa.int/docs/extra/ATBD_S2ToolBox_L2B_V1.1.pdf (accessed on 27 February 2022).

28. Goovaerts, P. Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties. *Soil Sci. Soc. Am. J.* **2000**, *64*, 1423–1433. [CrossRef]

29. Goovaerts, P.; Deckers, J.; Feyen, J. Soil-landscape modelling to quantify spatial variability of soil texture. *Phys. Chem. Earth Part B: Hydrology, Oceans, Atmos.* **2001**, *26*, 41–45. [CrossRef]

30. Hengl, T.; Toomanian, N.; Reuter, H.; Malakouti, M.J. Methods to interpolate soil categorical variables from profile observations: Lessons from Iran. *Geoderma* **2007**, *140*, 417–427. [CrossRef]

31. Xu, Y.; Smith, S.E.; Grunwald, S.; Abd-Elrahman, A.; Wani, S.P. Incorporation of satellite remote sensing pan-sharpened imagery into digital soil prediction and mapping models to characterize soil property variability in small agricultural fields. *ISPRS J. Photogramm. Remote Sens.* **2017**, *123*, 1–19. [CrossRef]

32. Heuvelink, G.B.; Pebesma, E.J. Spatial aggregation and soil process modelling. *Geoderma* **1999**, *89*, 47–65. [CrossRef]

33. Duffera, M.; White, J.G.; Weisz, R. Spatial variability of Southeast US Coastal Plain soil physical properties: Implications for site-specific management. *Geoderma* **2007**, *137*, 327–339. [CrossRef]

34. Goovaerts, P. Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties. *Biol. Fertil. Soils* **1998**, *27*, 315–334. [CrossRef]

35. Wackernagel, H. *Multivariate Geostatistics: An Introduction with Applications*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2003.

36. Zammit-Mangion, A.; Cressie, N. FRK: An R package for spatial and spatio-temporal prediction with large datasets. *J. Stat. Softw.* **2021**, *98*, 1–48. [CrossRef]

37. Katzfuss, M.; Cressie, N. Spatio-temporal smoothing and EM estimation for massive remote-sensing data sets. *J. Time Ser. Anal.* **2011**, *32*, 430–446. [CrossRef]

38. Kang, E.L.; Liu, D.; Cressie, N. Statistical analysis of small-area data based on independence, spatial, non-hierarchical, and hierarchical models. *Comput. Stat. Data Anal.* **2009**, *53*, 3016–3032. [CrossRef]

39. Nychka, D.; Wikle, C.; Royle, J.A. Multiresolution models for nonstationary spatial covariance functions. *Stat. Model.* **2002**, *2*, 315–331. [CrossRef]

40. Janssen, P.; Heuberger, P. Calibration of process-oriented models. *Ecol. Model.* **1995**, *83*, 55–66. [CrossRef]

41. Wang, B.; Waters, C.; Orgill, S.; Gray, J.; Cowie, A.; Clark, A.; Li Liu, D. High resolution mapping of soil organic carbon stocks using remote sensing variables in the semi-arid rangelands of eastern Australia. *Sci. Total Environ.* **2018**, *630*, 367–378. [CrossRef] [PubMed]

42. Bellon-Maurel, V.; Fernandez-Ahumada, E.; Palagos, B.; Roger, J.-M.; McBratney, A. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *TrAC Trends Anal. Chem.* **2010**, *29*, 1073–1081. [CrossRef]

43. Řezník, T.; Herman, L.; Trojanová, K.; Pavelka, T.; Leitgeb, Š. Interpolation of data measured by field harvesters: Deployment, comparison and verification. In Proceedings of the International Symposium on Environmental Software Systems, Wageningen, The Netherlands, 5–7 February 2020; pp. 258–270.

44. McGarigal, K. *FRAGSTATS: Spatial Pattern Analysis Program for Quantifying Landscape Structure*; US Department of Agriculture, Forest Service, Pacific Northwest Research Station: Portland, Oregon, USA, 1995; Volume 351.

45. Hesselbarth, M.H.; Sciaini, M.; With, K.A.; Wiegand, K.; Nowosad, J. landscapemetrics: An open-source R tool to calculate landscape metrics. *Ecography* **2019**, *42*, 1648–1657. [CrossRef]

46. Huete, A.; Miura, T.; Yoshioka, H.; Ratana, P.; Broich, M. Indices of vegetation activity. In *Biophysical Applications of Satellite Remote Sensing*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 1–41.

47. Jiang, Z.; Huete, A.R.; Li, J.; Chen, Y. An analysis of angle-based with ratio-based vegetation indices. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2506–2513. [CrossRef]

48. Miller, B.A.; Koszinski, S.; Wehrhan, M.; Sommer, M. Impact of multi-scale predictor selection for modeling soil properties. *Geoderma* **2015**, *239*, 97–106. [CrossRef]

49. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **2010**, *33*, 1. [CrossRef]

50. Royston, J.P. An extension of Shapiro and Wilk’s W test for normality to large samples. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **1982**, *31*, 115–124. [CrossRef]

51. Mouazen, A.; Maleki, M.; De Baerdemaeker, J.; Ramon, H. On-line measurement of some selected soil properties using a VIS–NIR sensor. *Soil Tillage Res.* **2007**, *93*, 13–27. [CrossRef]