Using Block Kriging as a Spatial Smooth Interpolator to Address Missing Values and Reduce Variability in Maize Field Yield Data

Thomas M. Koutsos 1,*, Georgios C. Menexes 2, Ilias G. Eleftherohorinos 2 and Thomas K. Alexandridis 1

1 Department of Hydraulics, Soil Science and Agricultural Engineering, School of Agriculture, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece; thales@agro.auth.gr
2 Department of Field Crops and Ecology, School of Agriculture, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece; gmenexes@agro.auth.gr (G.C.M.); elefero@agro.auth.gr (I.G.E.)

* Correspondence: tkoutsos@agro.auth.gr

Abstract: Block kriging (a spatial interpolation method) and log¹⁰ transformation were compared for their effectiveness in reducing relative variance (coefficient of variance: CV) and estimate mean values in all harvested maize plants grown in three randomly taken field plots and for harvested plants after removing the “edge or margin” ones. The results showed that log¹⁰ transformation reduced CVs of all harvested original fresh weight (FW) plant data in the three plots from 35.6%–41.6% (original data) to 6.0%–7.5%, while the respective CVs due to block kriging were reduced to 14.5%–19.9%. The back-log¹⁰-transformed means of all harvested FW plant data were reduced by 6.8%–9.4%, while the respective reduction for plants excluding the margin ones was 1.3%–8.3%. The block kriging means for all harvested FW plant data were reduced only by 0.3%–0.4%, while the respective means of the harvested plants excluding margin ones were increased by 0.4%–4.3%. These findings strongly suggest that block kriging should be preferred over the log¹⁰ transformation method (used so far by agroscientists) as it managed to effectively reduce variability in crop data and estimate missing values that provide more precise and reliable estimates of corn yield for farmers.

Keywords: agricultural experimentation; data variability; edge effects; geoinformatics

1. Introduction

The assessment of crop productivity (plant biomass or grain yield) based on field plots includes a high degree of uncertainty due to the existence of many unmeasured and uncontrolled factors (e.g., environmental variability, genotype using environment interaction, soil fertility, fertilization, irrigation, crop protection and other cultural practices, competition, and allelopathy between plants). These factors introduce undesirable variance in the values of soil and plant characteristics, inflating the experimental error and complicating the statistical analysis and the comparison of treatment means [1,2]. By using appropriate agricultural practices, carefully selected experimental designs, plot sizes, sampling schemes, and statistical methods, agroscientists are trying to manage existing undesirable crop data variability and several issues concerning missing values and outliers in the data analysis stage to finally estimate more accurate crop metrics to assess field productivity.

The assessment of crop productivity is traditionally based on the harvest of the whole plot area because it provides the most accurate production measure, and it is regarded as the absolute standard for crop yield estimation [3–5]. However, as harvesting the whole area of the plots is impracticable due to high labor costs [4], several sampling methods were developed with the accuracy of most of them not clearly understood or noted [3,6,7].
Among these methods, the larger randomly selected plots are considered the most appropriate as they are less biased and more reliable than the smaller plots [8]. In addition, the harvest of plants grown in the central rows of the field plots is also used because the plants of these rows are considered more representative than the plants from guard rows (edge or marginal plants) that are less affected by the competition of nearby plants, and thus become more productive but less appropriate for representative sampling [7,8]. In general, the number of margin plants or the width of the crop field margin used in agronomical research experiments can vary due to the specific experiment and its objectives; in particular, in some experiments, margin plants are not used at all, while in others, a specific number of plants or a crop field margin may be used to control various factors and minimize experimental error [9,10]. It seems that there is no golden rule to this and usually only a narrow width zone is considered a margin. For example, in a four-treatment plot experiment [9], a field margin of only 0.5 m was used in the plots with dimensions of 15 × 15 m. In another study [10], the crop edge was defined as the outer only few meters of the crop, which may be part of a nature conservation strip in a conservation headland or an unsprayed crop edge, or otherwise, even a conventional crop. In the case of large areas that crop fields occupy, a field margin with a width of 3 to 10 m (median) is recommended from the farmers’ point of view [11], which can also be considered as narrow in width compared to the total crop area. However, regardless of the sampling methods or the width of the ‘crop edge’ used, issues regarding data variability and the existence of missing or extreme values affect the accuracy to provide safe means for crop production.

From a statistical point of view, data variability and other issues that may exist can be managed in several ways. The most common method used to reduce the existing data variability among the experimental units (individual plants or plots) is the log transformation of raw data followed by appropriate statistical analysis, according to the experimental design and the corresponding mathematical model (general or mixed linear) [2,11,12]. However, although log transformation can reduce the variability of data and make them conform more closely to a symmetric distribution, the results of standard statistical tests performed on log-transformed data are often not relevant for the original non-transformed data (means are changed), suggesting that data transformations with this method must be applied with caution [11]. In addition, the log transformation method cannot provide estimates for missing values, a common issue in agricultural experiments, where missing values represent spots where plants did not emerge. Therefore, it can be concluded that, although the sampling methods, the sample size, and the statistical approach of applying the log transformation method seem to be effective, this procedure has several issues, and so there is still a need to find an unbiased way of reducing data variability and estimating values for missing data without altering the original mean values of data.

In recent years, the spatial approach has begun to gain interest because it exploits field heterogeneity to manage either the over- or under-application of agronomical inputs, leading to higher yields and more environmentally friendly fertilizing schemes. Spatial interpolation methods, such as kriging, a well-known geostatistical interpolation method, has been used to handle issues concerning spatial crop data [13–17]. Kriging can be used either as an exact interpolator (point kriging) or as a smooth interpolator (block kriging) [12,18]. According to researchers [19–25], ordinary kriging (as an exact interpolator) has been proven to be a very effective method for delivering yield maps [19–22,26,27] for corn grain and corn silage yield data [23], for durum wheat yield data [24], and for sugarcane data [25]. The efficacy of kriging derives from the fact that estimated values are linearly weighted of its surrounding points, while the weights are calculated by minimizing the error variance of the model of the spatial continuity for data regarding spatial distribution [24,25,28–30]. On the other hand, block kriging (as a smooth interpolator) has the potential to deliver both an estimation of values at grid blocks and smooth extreme values that may exist in agricultural data. The implementation of kriging in our previous study [31] showed that a spatial interpolation method can be used to effectively estimate missing
values and reduce existing variability in crop data, providing more valid mean values of crop parameters; however, the results were not compared to other methods used to manage data variability, such as log₁₀ transformation.

Therefore, the aim of this study is to examine the effectiveness of block kriging as a smooth interpolator versus the commonly used method of log₁₀ transformation in reducing data variability, without changing the mean values of the original data, using maize field data obtained from three randomly taken experimental plots. In addition, the effect of excluding the margin plants of plots on variability reduction without affecting mean values was studied.

More specifically, this research aims to answer the following two scientific questions:

1. Do the harvested maize plants grown in the border (margin) rows significantly affect data variability and the mean values of plants grown in the central rows of the experimental field plots?
2. Should block kriging be preferred over the most used statistical log₁₀ transformation method to reduce data variability and estimate missing values in crop data without changing the means?

2. Materials and Methods
2.1. Site Description and Sampling

A maize (hybrid AGN720, Italy) crop was established in a 3 ha field area of the Aristotle University of Thessaloniki (Greece) Farm (latitude: 40°32′1.75″ N longitude: 22°59′26.98″ E) during the 2016 growing season. The seedbed was prepared according to agricultural practices applied in the area and fertilization was made with 200 and 100 kg N and P/ha, respectively. In late April (27 April 2016), the field was sown with a 4-row pneumatic sowing machine, known as Gaspardo. Weed control was achieved with the recommended field rate of the early post-emergence-applied herbicide Modett 25/28 SE (25% terbuthylazine + 28% dimethenamid-p), whereas irrigation was performed according to the requirements of the crop plants. After crop emergence, three randomly taken plots (4 m × 4.25 m) with six rows per plot were marked. The distance between the crop rows was 80 cm and that between the plants in the same row was 17 cm. Therefore, there were 25–26 plants/row, 150 plants/plot, and 450 plants in total. The distance between the three plots was 20 m. The individual plants in the three plots (Figure 1), for the purpose of the study, were considered as the units of the target statistical population from which the samples had been taken.
At the silage stage (when the kernels began to glaze) of maize (14 weeks after sowing), all plants were harvested from each subplot and the silage yield (fresh weight—FW) of each plant was recorded. The silage stage was determined by breaking the ears of maize and visually evaluating the kernels’ stage of development.

A “crop edge” of one-plant field margin (roughly 0.7 m in width) was used for each plot (Figure 2) to study its effect on data variability. Block kriging and log transformation were also used to estimate the mean and CV values of the original, interpolated, and log10-transformed FW data for the three plots. These parameters were estimated for all harvested plants in each plot (Figure 1) and for the plants that do not include the edge (margin) ones (Figure 2). This was made to test the hypothesis that margin plants affect both variability and mean values.
2.2. Theoretical Background

2.2.1. The Use of Log Transformation in Achieving Normality and Reducing Variability in Data

Observed data from experiments and research projects are often so skewed that standard statistical analysis provides unsafe or even invalid results. Log transformation is widely used to transform skewed data to approximately conform to normality as it is easily applied and has a great effective reduction in data variability [32]. However, if the original data used do not follow a log-normal distribution, then the log transformation method does not reduce the skewness of the distribution, eliciting the risk that transformation can lead to different mean values, while making the distribution even more skewed than the original. Moreover, in contrast to popular belief, log transformation cannot always reduce the variability of data, and even in the case where this process can reduce variability, mean values must be taken into consideration.

The mean of log-transformed data (\( \log(y_i) \)), \( \mu_{LT} \) \( = \left( \frac{1}{n} \right) \sum_{i=1}^{n} \log y_i \) is used to estimate the population mean of the original data by applying the anti-log function to obtain \( \exp(\mu_{LT}) \). If \( y_i \) follows a log-normal distribution \( (\mu, \sigma^2) \), then the mean of \( y_i \) is given by \( E(y_i) = \exp(\mu + \frac{\sigma^2}{2}) \) [32]. Afterward, if we apply log transformation to \( y_i \), the transformed \( \log y_i \) will follow a normal distribution with the mean \( \mu \). Therefore, \( \hat{\mu} = \exp(\mu_{LT}) \) is an estimate of \( \exp(\mu) \). However, the mean of the original data \( y_i \) is \( \exp(\mu + \frac{\sigma^2}{2}) \), and not \( \exp(\mu) \). This means that even in the best situation, estimating the mean of the

---

**Figure 2.** Schematic presentation of the original FW plant values in each of the three plots (Plots 1-3) without the edge (margin) plants. Dark areas represent the “margin” in each one of the three randomly taken plots. Empty spots represent locations where plants did not emerge (missing values).
original \( y_i \) using the anti-log of the sample mean of the log-transformed data can lead to unsafe or even invalid and misleading estimates of the population mean of the original data. Since log transformation is frequently used in agricultural research, it is worth investigating whether the potential issues with its application extend to agricultural data derived from experiments. Thus, it is pertinent to examine or confirm if log transformation raises similar concerns in agricultural research and whether alternative methods can provide more reliable outcomes.

2.2.2. Using Kriging Interpolation to Estimate Missing Data and Reduce Outliers

The idea of ‘optimal linear prediction’ was initially attributed to Kolmogorov (1941) and finally presented by Krigge in 1951. Ordinary kriging, the most frequently used interpolation kriging type, has been used to effectively deliver visually appealing yield maps [19–22] for corn grain and silage yield data [23]. This interpolation method delivers accurate predictions in a selected area by producing new accurate grids of data (with the new estimates). Kriging differs from other interpolation methods in the fact that it uses the spatial correlation between sampled values to interpolate (estimate) all values in the spatial field. The estimation of values is based on the spatial arrangement of empirical observations, rather than on a presumed model of spatial distribution.

One of the advantages of applying interpolation in agricultural experimental data is that missing values (i.e., plants that did not emerge) can be estimated using the best linear unbiased prediction (BLUP). Thus, in the case the data set is suffering from missing values, ordinary kriging can be used to estimate values in non-sampling locations. Moreover, depending on the parameters used, interpolation can be used as an “approximate method” in such a way as to smoothly reduce the effect of existing extreme upper or lower values (outliers). Stationarity (that the mean and variance of the values is constant across the spatial field) and isotropy (uniformity in all directions) assumptions, in the case of using hybrid plant data in ordinary kriging interpolation, can be easily anticipated.

Kriging achieves the best linear unbiased prediction. We can predict the value \( Z(x_0) \) of the random function \( Z = Z(x) \) at any arbitrary location of interest \( x_0 \) based on the nearest measured observations \( z(x_i) \) of \( Z(x) \) at the sample points \( x_i \). In fact, kriging uses a weighted average value of the nearest observations based on the variogram or covariance function of \( Z(x) \). As prediction variance is minimized, the accuracy of the linear predictor increases, and finally the best linear prediction is achieved [5,8,33].

The estimation of the unsampled value \( Z \) at location \( x_0 \) is given by the following equation (Equation (1)):

\[
Z_w^*(x_0) = \sum_{i=1}^{n} w_i Z(x_i)
\]

where the weights \( w_i \) denotes kriging weights at location \( x_i \) with the sampled value \( Z(x_i) \) and they are determined such that \( Z_w^*(x_0) \) is an unbiased estimate of \( Z_w(x_0) \) and \( Z(x_0) \) denotes sampled values in the estimation region used for the calculation of values at locations \( x_0 \), which are both demonstrated in the following equation (Equation (2)):

\[
\sigma_w^2 = \text{Var}(Z_w^*(x_0) - Z(x_0))
\]

In the case of kriging used as a smooth interpolator (block kriging), the aim is to estimate the average value of a “block” centered on the grid nodes (with the size and shape of a grid cell), as defined during the interpolation process. As a result, block kriging compared to point kriging can provide estimates and smooth interpolated results using neighboring cells to estimate a value of each block [12].

2.3. Data Analysis

The descriptive statistics of the original, log10-transformed, and interpolated FW data were determined using IBM SPSS v.28 software and MS Excel. Spatial interpolation (point kriging and block kriging) was performed using Surfer for Windows. The main
descriptive statistics [mean, standard deviation (SD), minimum (min), maximum (max), skewness, kurtosis, and coefficient of variation (CV)] were calculated separately for each plot to compare the crop response between the three plots. Regarding skewness, this was examined to show data scattering and flatness. Descriptive statistics were calculated for all data and for those that did not include the margin plants.

The log_{10}-transformed and interpolated FW data were initially calculated and then used to estimate the CV and mean values to be compared with the respective values for two sampling scenarios: (a) all harvested plant data and (b) all excluding data from margin plants. Block kriging was used as smooth interpolator [12] to estimate missing values and to smooth existing extreme values, while point kriging was used as a reference (with results/estimates closer to the original data). It must be mentioned that block kriging was used to estimate the average value of rectangular blocks (areas with data) centered on the grid nodes (where the size and shape of each block being the same of a grid cell) of a given grid, while point kriging was used as an exact interpolator that estimates the values at the grid nodes that can provide estimated values as close as possible to the real data.

3. Results

3.1. Descriptive Statistics of the Original, Log_{10}-Transformed, and Interpolated Kriging Maize Fresh Weight Data

The removal of margin plants reduced the number of the harvested plants in the three plots from the initial 111–119 to 72–75, while the respective calculated means were reduced from 598–730 to 548–720 (Table 1). The estimated CV values for all harvested plants in the three plots ranged from 35.6% to 41.6%, while the respective CVs for the plants without the margin ones ranged from 32.2% to 41.0%. In general, the second field plot produced the lower mean value and the higher CV value as compared to the other two plots.

Table 1. Descriptive statistics of all original maize fresh weight data (g per plant) recorded in the three plots and those excluding the margin ones.

<table>
<thead>
<tr>
<th>Plot</th>
<th>Data Used</th>
<th>n</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>CV(%)</th>
<th>SD **</th>
<th>Variance</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All harvested</td>
<td>119</td>
<td>250</td>
<td>1337</td>
<td>649</td>
<td>36.9</td>
<td>239.5</td>
<td>57,351.3</td>
<td>0.5</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>No margin</td>
<td>75</td>
<td>315</td>
<td>1140</td>
<td>611</td>
<td>32.2</td>
<td>196.5</td>
<td>38,629.2</td>
<td>0.5</td>
<td>-0.2</td>
</tr>
<tr>
<td>2</td>
<td>All harvested</td>
<td>111</td>
<td>114</td>
<td>1352</td>
<td>598</td>
<td>41.6</td>
<td>248.7</td>
<td>61,830.4</td>
<td>0.5</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>No margin</td>
<td>72</td>
<td>114</td>
<td>1140</td>
<td>548</td>
<td>41.0</td>
<td>224.8</td>
<td>50,536.6</td>
<td>0.3</td>
<td>0.0</td>
</tr>
<tr>
<td>3</td>
<td>All harvested</td>
<td>116</td>
<td>133</td>
<td>1352</td>
<td>730</td>
<td>35.6</td>
<td>259.8</td>
<td>67,495.7</td>
<td>0.0</td>
<td>-0.1</td>
</tr>
<tr>
<td></td>
<td>No margin</td>
<td>75</td>
<td>183</td>
<td>1381</td>
<td>720</td>
<td>34.5</td>
<td>248.3</td>
<td>61,664.5</td>
<td>0.2</td>
<td>0.1</td>
</tr>
</tbody>
</table>

* CV: coefficient of variance, ** SD: standard deviation.

The CV values of all log_{10}-transformed harvested plant data in the three plots ranged from 6.0% to 7.4%, while the back-log_{10}-transformed means ranged from 542 to 674 (Table 2). Regarding the harvested plants without the margin ones, the CV and mean values ranged from 5.2% to 7.7% and from 496 to 673, respectively. As the original data, the second field plot produced the lower mean value and the higher CV value as compared to the other two plots.

Table 2. Descriptive statistics of all log_{10}-transformed maize fresh weight data (g per plant) recorded in the three plots and those excluding the margin ones.

<table>
<thead>
<tr>
<th>Plot</th>
<th>Data Used</th>
<th>n</th>
<th>Min</th>
<th>Max</th>
<th>Mean *</th>
<th>CV (%)</th>
<th>SD **</th>
<th>Variance</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All harvested</td>
<td>119</td>
<td>2.4</td>
<td>3.1</td>
<td>605</td>
<td>6.0</td>
<td>0.2</td>
<td>0.028</td>
<td>-0.3</td>
<td>-0.6</td>
</tr>
<tr>
<td></td>
<td>No margin</td>
<td>75</td>
<td>2.5</td>
<td>3.1</td>
<td>580</td>
<td>5.2</td>
<td>0.1</td>
<td>0.020</td>
<td>-0.2</td>
<td>-0.7</td>
</tr>
<tr>
<td>2</td>
<td>All harvested</td>
<td>111</td>
<td>2.1</td>
<td>3.1</td>
<td>542</td>
<td>7.5</td>
<td>0.2</td>
<td>0.042</td>
<td>-0.8</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>No margin</td>
<td>72</td>
<td>2.1</td>
<td>3.1</td>
<td>496</td>
<td>7.7</td>
<td>0.2</td>
<td>0.043</td>
<td>-0.9</td>
<td>0.9</td>
</tr>
</tbody>
</table>
The CV values of point kriging interpolated all harvested plant data in the three plots ranged from 31.4% to 35.8%, while the calculated means ranged from 600 to 728 (Table 3). Regarding the harvested plants without the margin ones, the CV and mean values ranged from 29.5% to 35.9% and from 558 to 720, respectively. As the original and log10-transformed data, the second field plot produced a lower mean value and a higher CV value as compared to the other two plots.

Table 3. Descriptive statistics for the point kriging interpolated maize fresh weight data (g per plant) recorded in the three plots and those excluding the margin ones.

<table>
<thead>
<tr>
<th>Plot</th>
<th>Kriging Type</th>
<th>Data Used</th>
<th>n</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>CV (%)</th>
<th>SD **</th>
<th>Variance</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Point kriging</td>
<td>All harvested</td>
<td>150</td>
<td>250</td>
<td>1337</td>
<td>646</td>
<td>33.4</td>
<td>215.5</td>
<td>46,447.5</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No margin</td>
<td>92</td>
<td>315</td>
<td>1140</td>
<td>611</td>
<td>29.5</td>
<td>180.3</td>
<td>32,518.2</td>
<td>0.5</td>
<td>0.3</td>
</tr>
<tr>
<td>2</td>
<td>Point kriging</td>
<td>All harvested</td>
<td>150</td>
<td>114</td>
<td>1352</td>
<td>600</td>
<td>35.8</td>
<td>214.6</td>
<td>46,047.3</td>
<td>0.5</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No margin</td>
<td>92</td>
<td>114</td>
<td>1140</td>
<td>558</td>
<td>35.9</td>
<td>200.4</td>
<td>40,160.6</td>
<td>0.2</td>
<td>0.6</td>
</tr>
<tr>
<td>3</td>
<td>Point kriging</td>
<td>All harvested</td>
<td>150</td>
<td>133</td>
<td>1423</td>
<td>728</td>
<td>31.4</td>
<td>228.8</td>
<td>52,332.0</td>
<td>0.1</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No margin</td>
<td>92</td>
<td>183</td>
<td>1381</td>
<td>720</td>
<td>31.2</td>
<td>224.5</td>
<td>50,418.2</td>
<td>0.2</td>
<td>0.8</td>
</tr>
</tbody>
</table>

* CV: coefficient of variance, ** SD: standard deviation.

The CV values of all harvested plant block kriging interpolated data in the three plots ranged from 14.5% to 19.9%, while the calculated means ranged from 600 to 728 (Table 4). Concerning the harvested plants without the margin ones, the CV and mean values ranged from 14.3% to 18.0% and from 571 to 723, respectively. Again, the second field plot produced the lower mean values as compared to the other two plots, but its CV values were higher than those of the third plot.

Table 4. Descriptive statistics for the block kriging interpolated maize fresh weight data (g per plant) recorded in the three plots and those excluding the margin plants.

<table>
<thead>
<tr>
<th>Plot</th>
<th>Kriging Type</th>
<th>Data Used</th>
<th>n</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>CV (%)</th>
<th>SD **</th>
<th>Variance</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Block kriging</td>
<td>All harvested</td>
<td>150</td>
<td>404</td>
<td>1033</td>
<td>646</td>
<td>19.9</td>
<td>128.3</td>
<td>16,468.4</td>
<td>0.5</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No margin</td>
<td>92</td>
<td>404</td>
<td>948</td>
<td>621</td>
<td>18.0</td>
<td>112.0</td>
<td>12,545.9</td>
<td>0.3</td>
<td>0.2</td>
</tr>
<tr>
<td>2</td>
<td>Block kriging</td>
<td>All harvested</td>
<td>150</td>
<td>323</td>
<td>950</td>
<td>600</td>
<td>18.1</td>
<td>108.6</td>
<td>11,799.2</td>
<td>0.4</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No margin</td>
<td>92</td>
<td>323</td>
<td>849</td>
<td>571</td>
<td>17.3</td>
<td>99.0</td>
<td>9804.2</td>
<td>0.2</td>
<td>0.4</td>
</tr>
<tr>
<td>3</td>
<td>Block kriging</td>
<td>All harvested</td>
<td>150</td>
<td>452</td>
<td>1036</td>
<td>728</td>
<td>14.5</td>
<td>105.3</td>
<td>11,085.3</td>
<td>0.1</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No margin</td>
<td>92</td>
<td>495</td>
<td>1015</td>
<td>723</td>
<td>14.3</td>
<td>103.3</td>
<td>10,662.48</td>
<td>0.4</td>
<td>0.3</td>
</tr>
</tbody>
</table>

* CV: coefficient of variance, ** SD: standard deviation.

The above findings (Tables 1–4) indicate that log10 transformation resulted in a greater reduction in the data variability (CV) of all harvested original FW maize plant data and those excluding margin plants in the three plots as compared with that of point and block kriging. However, the block kriging CVs were reduced more than the respective ones of point kriging. The back-log10-transformed means of all harvested FW plant data were reduced by 6.8%–9.4% as compared to the original data, while the respective reduction in the plants (excluding margin ones) was 1.3%–8.3%. The point kriging means for all harvested plants and for those excluding the margin ones were reduced by 0.1%–0.5% as compared to the original data, while the respective reduction in block kriging means for all harvested plant data was 0.3%–0.4%. By contrast, the estimated means of the harvested plants excluding margin ones were increased by 0.4%–4.3% as compared to the original data. In response to the first scientific question raised, it can be concluded that the harvested maize plants grown in the border (margin) rows did not significantly affect the
data variability and the mean values of plants grown in the central rows of the experimental field plots. Additionally, block kriging achieved a reduction in data variability, keeping the means as close as possible to the original data (only up to 0.4%) compared to log10 transformation data that altered the means (up to 10%).

3.2. Box Plots of the Original and Interpolated Kriging Maize Fresh Weight Data

The box plots, which were constructed from the minimum value, the maximum value, the first quartile, the median, and the third quartile of the data, revealed that the decreasing order of the locality, spread, and skewness groups of numerical data through their quartile was as follows: “all original harvested plants in the three field plots” > “original plants without the margin ones” > “interpolated with Point Kriging” > interpolated with lock Kriging” (Figure 3).

![Figure 3](image)

**Figure 3.** Box plots of all original and those excluding (“No margin”) the margin maize fresh weight (FW) data recorded in the three plots (Plots 1–3), along with data interpolated using point and block kriging methods.

3.3. Diagrammatic Presentation of the Original, Point, and Block Kriging Data

A visual representation of the original fresh weight (FW) data in the three field plots, along with the estimated values obtained through point and block kriging methods, reveals that both techniques successfully estimated the missing values. However, it is apparent from the diagram that the block kriging method also effectively smoothed out extreme values (higher and lower) in the original data, while point kriging delivered estimates very close to the original data. This smoothing effect of block kriging can be observed in Figure 4 for all three plots.
Figure 4. Schematic cross-sections in the three field plots (Plot 1-3) of all original measurements (points) and the corresponding estimated values using point kriging (circles) and block kriging (squares). Capital letters (A-G) represent the starting and ending points of each row (section)/plot.

Although point kriging, as an exact interpolator, can achieve more precise estimations, in this case, the application of interpolation was not aimed at getting as close as possible to the original data. Rather, the objective was to smoothen out lower and higher values while keeping the means almost constant.

As shown in Figure 5, a comparison between measured and estimated fresh weight data graphs reveals that block kriging produced a smoother outcome than point kriging and this smoothing effect mainly impacted the lower and higher values (which may be considered as not being representative from the agronomic perspective). In addition, the scatter plot for point kriging reveals that all estimated values align precisely with the reference line, presenting the expected result of an exact interpolator. On the other hand, the scatter plot for block kriging shows that lower and higher values were smoothed out, while values near the mean were only slightly affected. This representation highlights how, despite modifying several data points, the mean remains constant and almost identical to the mean of the original data.
Figure 5. Original fresh weight (FW) vs. estimated data using point and block kriging in the three field plots (Plots 1–3).

3.4. Contour Maps of the Point and Block Interpolation Grids

In all plots, the same hybrid was utilized, and consistent agronomical practices were implemented across the three plots situated in the same region. Consequently, the observed variability in yield data (fresh weight) can likely be attributed to factors such as environmental conditions and soil characteristics, rather than the hybrid itself. Hence, it is inferable that interpolation can be effectively applied to the original point data within each plot, treating the yield as a continuous variable. In this context, contour maps were constructed with “Surfer for Windows” to provide a better visual view of the point and block interpolation grids (Figure 6). The purpose of displaying the contour maps is to illustrate the spatial discrepancies between the results obtained from point and block kriging, with a specific focus on the smoothing effect provided by block kriging. Moreover, the contour maps provide a visual representation of the estimated values, allowing for a more detailed examination of the spatial distribution of data. In particular, the contour maps of block kriging (a2, b2, and c2) had less brown/dark blue color areas corresponding to low/high (extreme) values than the respective areas of point kriging (a1, b1, and c1) because of the smoothing effect on data. By comparing the contour maps generated from both interpolation methods, it becomes apparent that block kriging smooths out the extreme values, resulting in a more gradual transition between areas with high and low values. This smoothing effect is especially evident in areas where there were missing values or extreme values in the original data. Therefore, in this case, the contour maps serve as a valuable tool in understanding the differences between point and block kriging, and the advantages of using block kriging to address missing values and reduce variability in agricultural data.
3.5. Fitted Variogram Variables for Point and Block Kriging

The most suitable fitted model for both point and block kriging interpolated maize data in the three field plots was the exponential (Table 5) and it was selected after comparing the performance of other models based on their \( R^2 \) values and root mean square errors (RMSEs). Table 5 presents the results of point kriging and block kriging for the three different plots, providing information about the variogram model parameters; goodness-of-fit measures; and the following accuracy measures: (a) residual sum of squares (RSS), which measures the sum of squared differences between the observed values and the predicted values from the kriging model (RSS), and (b) root mean square error (RMSE), which is the standard deviation of residuals (differences between observed and predicted values). Overall, both point kriging and block kriging have high \( R^2 \) values (\( R^2 \) represents the coefficient of determination, which measures the goodness-of-fit of the variogram model to the data) across all plots, indicating a very good fit of variogram models to the fresh weight data.

Figure 6. Contour maps for interpolated fresh weight (FW) data in the three plots, where (a1–c1) correspond to point kriging and (a2–c2) to block kriging.
Table 5. Geostatistical variables of the best-fitted variogram models for both point and block kriging data recorded in the three field plots.

<table>
<thead>
<tr>
<th>Kriging Type</th>
<th>Plot</th>
<th>( n )</th>
<th>Variogram</th>
<th>Nugget</th>
<th>Sill</th>
<th>Range</th>
<th>( R^2 )</th>
<th>RSS</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Point kriging</td>
<td>1</td>
<td>150</td>
<td>Exponential</td>
<td>0</td>
<td>53470</td>
<td>0.96</td>
<td>0.99</td>
<td>9.5 \times 10^{-12}</td>
<td>2.5 \times 10^{-7}</td>
</tr>
<tr>
<td>Block kriging</td>
<td>1</td>
<td>150</td>
<td>Exponential</td>
<td>0</td>
<td>53510</td>
<td>0.97</td>
<td><strong>0.93</strong></td>
<td>14.5 \times 10^{5}</td>
<td>98.5</td>
</tr>
<tr>
<td>Point kriging</td>
<td>2</td>
<td>150</td>
<td>Exponential</td>
<td>0</td>
<td>60600</td>
<td>0.70</td>
<td>0.99</td>
<td>6.3 \times 10^{-12}</td>
<td>2.04 \times 10^{-7}</td>
</tr>
<tr>
<td>Block kriging</td>
<td>2</td>
<td>150</td>
<td>Exponential</td>
<td>0</td>
<td>60270</td>
<td>0.72</td>
<td><strong>0.94</strong></td>
<td>1.9 \times 10^{5}</td>
<td>113.5</td>
</tr>
<tr>
<td>Point kriging</td>
<td>3</td>
<td>150</td>
<td>Exponential</td>
<td>0</td>
<td>68000</td>
<td>0.53</td>
<td>0.99</td>
<td>8.2 \times 10^{-12}</td>
<td>2.34 \times 10^{-7}</td>
</tr>
<tr>
<td>Block kriging</td>
<td>3</td>
<td>150</td>
<td>Exponential</td>
<td>0</td>
<td>68000</td>
<td>0.62</td>
<td><strong>0.96</strong></td>
<td>24.7 \times 10^{-5}</td>
<td>128.4</td>
</tr>
</tbody>
</table>

where \( R^2 \) = coefficient of determination; \( RSS \) = residual sum of squares, \( RMSE \) = root mean square error. The bold print shows the highest values per plot.

4. Discussion

This study was conducted to investigate and assess the performance of \( \log_{10} \) transformation and block kriging interpolation on reducing maize fresh weight data variability. The \( \log_{10} \) transformation method was tested because it is the most widely used method in agricultural, biomedical, and psychosocial research to stabilize and reduce variability of data and ensure more close conformity to the normal distribution or, in general, to asymmetric distribution [2,11,12]. On the other hand, the kriging interpolation method was selected and compared with the \( \log_{10} \) transformation method because several previous studies have indicated its importance in reducing data variability and adequately estimating missing and mean values [13,20,22,31].

The significant reduction in variability (CV) due to \( \log_{10} \) transformation, as compared to the respective CVs of all original harvested FW plant data and those excluding margin plants in all plots, justifies its popular use in agricultural research for reducing data variability. However, the fact that back-transformed estimated means after \( \log_{10} \) transformation were reduced as compared with the original ones makes the interpretation of results meaningless, since the adjusted crop yield means of the expected real values were underestimated on the original scale [2]. Several studies [2,11] have indicated that \( \log_{10} \) transformation has been misused, even by statisticians, leading to incorrect interpretations of experimental results [2,11]. Therefore, it is suggested that, if \( \log_{10} \) transformation must be used, the researcher must be mindful about its limitations, particularly when interpreting the relevance of transformed data analysis for the hypothesis of interest regarding original data.

The greater reduction in block kriging CVs as compared to those of point kriging suggests its possible use in agricultural research. In addition, the lower change in the estimated block kriging mean values as compared to the back-\( \log_{10} \)-transformed values strongly suggests that block kriging, in addition to the adequate estimation of missing values at non-sampled locations (places where plants did not emerge), can effectively replace the most commonly used statistical method of \( \log_{10} \) transformation since it can offer both a reduction in data variability and a more precise crop yield estimate. It is worth mentioning that the above findings were confirmed by descriptive statistics of the original; \( \log_{10} \)-transformed and interpolated kriging maize fresh weight data; box plots of the original and interpolated kriging maize fresh weight data; a diagrammatic presentation of the original, point, and block kriging data; the contour maps of point and block interpolation grids; and fitted variogram variables for point and block kriging data.

Achieving better estimates for crop production via block kriging is crucial; for example, the 10% less estimated maize crop yield (using \( \log_{10} \) transformation so far), as compared to the original means, reduces the total maize yield by up to 2 tn/ha (with an average maize yield of 12–18 tn/ha) and estimates non-existent economic loss for maize growers. Therefore, it can be deduced that block kriging must be favored over \( \log_{10} \) transformation to address missing values and reduce data variability, and as a result to provide more safe and accurate crop metrics.
5. Conclusions

The findings of this study allow the following conclusions to be drawn:

1. Log_{10} transformation was found to appropriately analyze the stage of maize crop yield data as it provides a notable reduction in data variability (CV values), but it failed to estimate means leading to non-existent economic loss for the producers.

2. The block kriging interpolation method was found to adequately replace the commonly used statistical method of log_{10} transformation so far as it managed to reduce data variability without altering the means, leading to more precise estimates of crop yield. A summary of the highlighted advantages of block kriging interpolation versus log_{10} transformation showed that this method can successfully (a) estimate and fill in missing values, (b) smooth unrepresentative or extreme values (usually present in agricultural data), (c) adjust the estimated values to account for the spatial correlation of experimental units with respect to the measured characteristics, (d) reduce the data variability without altering the estimated mean values of the measured characteristics, and (e) improve the overall quality of the data.

Therefore, block kriging can efficiently replace the log_{10} transformation method, which provides a significantly reduction in CV values, but also leads to significant erroneous estimates for larger areas in most cases. Block kriging can effectively address the issue of missing values and reduce variability in maize field yield data, while keeping the means almost constant, providing more accurate crop metrics to assess field productivity.


Funding: This research received no external funding.

Data Availability Statement: All data used are presented in Figures 1 and 2.

Acknowledgments: We would like to thank alumni A. Pesios and A. Chatzopoulos for providing the data and the staff of the A.U.Th. Farm for their assistance throughout the experiment.

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

References


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