# CHAPTER 5

# SPATIAL VARIATION OF SOIL AND PLANT PROPERTIES AND ITS EFFECTS ON THE STATISTICAL DESIGN A OF A FIELD EXPERIMENT

A.Venter[3*], M.F. Smith[1], D.J.Beukes[2], A.S. Claassens[3] & M. Van Meirvenne[4]

[1]ARC-Biometry Unit, Private Bag X519, Silverton, 0127, Pretoria, South Africa

[2]ARC-Institute for Soil, Climate and Water, Private Bag X79, 0001, Pretoria, South Africa

[3]Department of Plant Production and Soil Science, University of Pretoria, 0002, Pretoria, South Africa

[4]Dept. Soil Management and Soil Care, Ghent University, Coupure 653, 9000 Ghent, Belgium

## 5.1 Abstract

Natural soil variability, or previous land-use practices, can significantly reduce the ability to detect experimental treatment differences. Hence, the standard procedure in field experimentation has been to lay out blocks, striving for as homogeneous conditions as possible among plots of the same block. The classical procedures of, *inter alia*, replication, blocking and randomization have assumed spatial and temporal independence among the points of determination of any particular soil or plant property. However, geostatistical concepts dictate that a spatially dependent variance structure exists for observations of a particular property, whereby nearby observations are more similar than those taken further apart. The present study was conducted on an 18-ha lucerne stand in which a 100 m X 140 m experimental area was demarcated. To determine spatial characteristics of soil and plant properties, 48 sampling points (nodes) were laid out on a 20-m square grid with an additional 75 points on a 2.5-m grid at five random node points. A randomized complete block (RCB) design trial was superimposed on the geostatistical grid design and consisted of seven pseudo (i.e. non-existent) treatments, replicated four times. Soil and plant samples were taken at all sampling points and plots in June 2001 and analyzed for various properties, including green biomass yield. Analysis of variance of the RCB design revealed statistically non-significant differences among the pseudo treatments for various soil and plant properties, including yield. The conclusion could be made that the experimental field was

homogeneous enough to lay out a standard block design experiment. However, it was found that the estimate map of soil pH($H_2O$) showed a clear structure in spatial variability. The question was posed that if the latter spatial variation had been considered, would it have had any effect on the results of this field experiment, for example, in terms of yield? Scrutiny of the latter variability revealed that the standard RCB designs did not provide homogeneous blocks with respect to soil variability. The consequent redesign of the experiment whereby all plots were randomly allocated to treatments and replications, led to dramatically different results: significant differences were obtained for plant and soil properties as a function of the pseudo treatments. From this study it is clear that spatial variability of soil and plant properties can jeopardize the results of a standard block design field experiment. Regarding soil pH($H_2O$) as a covariate (since it correlated very well with green biomass yield) and performing an analysis of covariance, no statistical difference (as expected) among treatments was observed for green biomass yield. It is, therefore, recommended that the layout of field experiments should be designed to the cognizance of the spatial variation of a soil property that correlates highly with a chosen response variate. Hence in the final statistical analysis to test for treatment differences, the particular soil property must be treated as a covariate. Consequent experimental results can then be interpreted with much greater confidence.

## 5.2 Introduction

Agricultural researchers have long understood that the effect of locality, which is often caused by natural soil variability, or previous land-use practices, can significantly reduce the ability to detect experimental treatment differences (Dulaney, Lengnick & Hart, 1994). Present-day agronomic research has reached a point where the treatment effects being tested are small and the degree of accuracy required in such studies cannot easily be obtained with conventional experimental designs (Van Es, Van Es & Cassel, 1989). It is therefore imperative to establish a high level of experimental precision.

The adverse effects stemming from soil heterogeneity can be addressed by (1) conducting the study on uniform land, or (2) controlling the effects of soil variability through experimental design and improved statistical analysis in order to better account for the effect of field variability on experimental results (Van Es *et al.*, 1989). The latter measure includes replication, blocking, randomization, row-and-column designs and methods such as nearest neighbour and trend analysis. In general, such methods improve the detection of treatment effects, although improper block layout may actually adversely affect the analysis of experiments (Van Es & Van Es, 1993). In the presence of a significant spatial correlation over small distances, the assumption of independence between plots is violated and the researcher may be faced with contradictory results. The latter can result in clear differences in crop yields between experimental plots but no significant treatment effect (Fagroud & Van Meirvenne, 2002).

Some work has been done to evaluate the use of geostatistics in the design of agricultural field experiments (Dulaney *et al.,* 1994; Van Es *et al.,* 1989; Fagroud & Van Meirvenne, 2002). According to Dulaney *et al.* (1994), geostatistical techniques have the potential to provide better field characterization, improve plot layout, increase the power of the consequential statistical techniques and can be used to select an optimal sampling strategy for characterization of soil spatial variability at the experimental field site. This is relevant because the costs associated with conducting long-term agricultural experiments make it imperative to obtain at least some level of assurance that the data used to establish field trials are precise enough for their intended purpose.

The hypothesis of this study was that the natural variation of the soil properties would have an effect on the results of a field experiment if the spatial structure of those properties in the field were not taken into consideration when designing the trial. The purpose of this study, therefore, was to quantify the spatial variation of soil and plant properties in relation to a statistically laid out field experiment.

## 5.3 Materials and Methods

### *5.3.1 Field and analytical methods*

The study was conducted on an 18-ha lucerne stand in the Brits district of the North West Province of South Africa (27°49'E, 25°33'S). A rectangular area of 100 m X 140 m was demarcated as a study area and the soil was classified (based on a field survey) as a deep (1000 mm) Shortlands form (Pyramid family) (Soil Classification Working Group, 1991) (USDA Soil Taxonomy; Typic Rhodustults). Forty-eight sampling points (nodes) were laid out on a 20-m square grid with an additional 75 sampling points laid out on a 2.5-m square grid (sampling total = 123) at five randomly selected node points to ensure that the total spatial structure would be identified. All sampling points were georeferenced using a Ground Positioning System (GPS) and marked with flat metal discs. A randomized complete block (RCB) design trial layout was superimposed on the geostatistical grid design and consisted of seven pseudo treatments (i.e. applying no actual treatments) replicated four times. A plot size of 25 m x 20 m was decided on to fit all the plots in the available area of the original lucerne stand.

In June 2001, plant and soil samples were taken at each of the node points. Plant sampling was done by cutting the aboveground plant parts within a 0.6-m square around each of the node points and weighing to determine green biomass yield. Three soil samples were also collected within the 0.6-m square (0 – 300 mm deep) at each of the sampling points and mixed to serve as a composite sample. Sampling of the RCB design was done by cutting a 10 $m^2$ area of plants in each of the plots and weighing the samples to determine green biomass yield. Sub-samples were taken for analysis purposes. In each of the plots a composite soil sample was taken of the 0 – 300 mm layer.

Plant samples were analyzed for calcium (Ca), magnesium (Mg), phosphorus (P) and potassium (K) using standard methods (Agri Laboratory Association of Southern Africa, 1998). Soil samples were analyzed for pH($H_2O$), organic carbon (C), P (Ambic), Ammonium acetate extractable Ca, K, Sodium (Na), Mg, electrical resistance, particle size and water retention using standard methods (Non-Affiliated Soil Analysis Work Committee, 1990).

### *5.3.2 Statistical methods*

Descriptive statistical analyses were performed to obtain information on the frequency distribution, standard deviation and coefficient of variation of the plant and soil chemical properties and yield. All properties displayed acceptably normal distributions with homogeneous treatment variances and no transformations were necessary. Analyses of variance (ANOVA) were used to test for differences between pseudo treatments for all plant and soil properties using the statistical program GenStat (GenStat, 2000). Treatment means were separated using Fishers' protected t-test least significant difference (LSD) at the 5 % level of significance (Snedecor & Cochran, 1980).

Three different experimental designs were superimposed on the 28 experimental plots (Table 5.1); the first one (ANOVA 1) blocked in the NE-SW direction, with 4 treatments randomly allocated to each of the 7 blocks and the second (ANOVA 2) blocked in the NW-SE direction, with 7 treatments randomly allocated to each of the 4 blocks (Figure 5.1). Both experimental layouts were based on an RCB design. It is obvious from the spatial variability of soil pH($H_2O$) (Figure 5.1) that the standard way of blocking either in the NE-SW or NW-SE directions, would not provide homogeneous blocks with respect to soil variability. Consequently, for the third experiment, a completely random design was chosen and the 28 plots randomly allocated to 7 treatments and 4 replications (ANOVA 3). This meant a random distribution of plots over the experimental area (Figure 5.1). An analysis of covariance was also performed using pH($H_2O$) as a covariate to eliminate the linear effect of soil pH on yield.

**Table 5.1** ANOVA of three different experimental designs.

| | ANOVA 1<br>RCBD | ANOVA 2<br>RCBD | ANOVA 3<br>CRD |
|---|---|---|---|
| | 4 treatments in 7 blocks | 7 treatments in 4 blocks | 7 treatments, 4 replicates |
| Source of variation | df | df | df |
| Block | 6 | 3 | - |
| Treatment | 3 | 6 | 6 |
| Error | 18 | 18 | 21 |
| TOTAL | 27 | 27 | 27 |

RCBD – Randomized Complete Block Design

CRD – Completely Random Design

Geostatistical analyses were performed on the grand total (123) number of samples to generate a map of the spatial variation of the soil pH($H_2O$) using the ArcGIS Geostatistical Analyst extension (Johnston, Ver Hoef, Krivoruchko & Lucas, 2001). Additional analyses were performed to generate a spatial map for $pH(H_2O)$ making use of sampling points on a 40-m square grid, as well as points on a 7.5-m square grid at the originally selected five node points (sample total = 37).

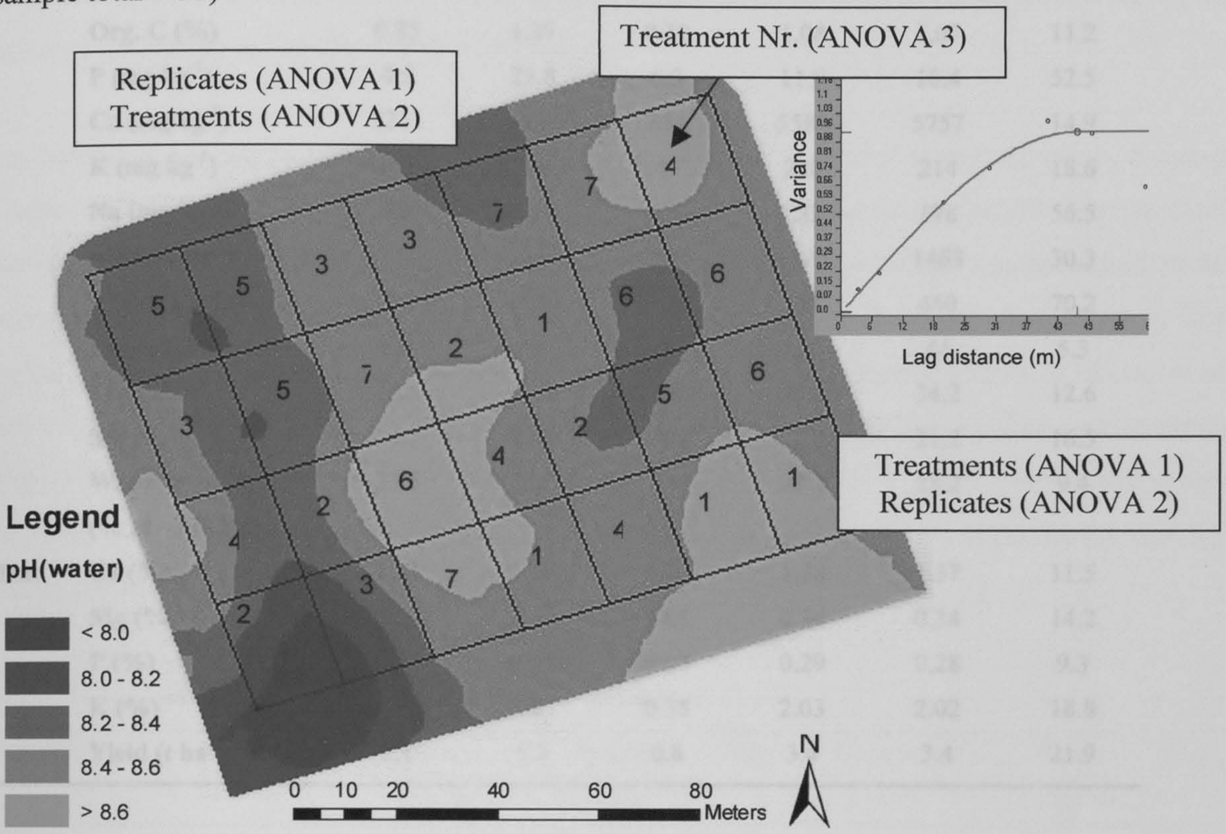

**Figure 5.1** Trial layout of the three experimental designs overlaid on the estimate map of soil $pH(H_2O)$.

## 5.4 Results and Discussion

Table 5.2 shows the mean concentrations of all the soil and plant properties for the 28 experimental plots. The observed mean nutrient concentrations in the plants are, according to Pinkerton, Smith and Lewis (1997), in the "adequate" to "high" range, although there were some spots in the field, especially for K, that showed deficiencies. The correlation matrix (Table 5.3) shows that both $pH(H_2O)$ and Na are highly negatively correlated with yield (Table 5.3; r = -0.74 and -0.68, respectively) and that there is a strong relationship between these two soil properties (Table 5.3; r = 0.76).

**Table 5.2** Statistical description of soil and plant properties.

| | | Min. | Max. | Std. Dev. | Mean | Median | CV (%) |
|---|---|---|---|---|---|---|---|
| **Soil** | **pH($H_2O$)** | 8.1 | 9.1 | 0.2 | 8.6 | 8.6 | 2.5 |
| | **Org. C (%)** | 0.85 | 1.37 | 0.12 | 1.08 | 1.07 | 11.2 |
| | **P (mg $kg^{-1}$)** | 4.5 | 25.8 | 6.3 | 11.9 | 10.4 | 52.5 |
| | **Ca (mg $kg^{-1}$)** | 3211 | 6852 | 834 | 5588 | 5757 | 14.9 |
| | **K (mg $kg^{-1}$)** | 162 | 308 | 41 | 220 | 214 | 18.6 |
| | **Na (mg $kg^{-1}$)** | 89 | 524 | 126 | 223 | 178 | 56.5 |
| | **Mg (mg $kg^{-1}$)** | 770 | 2058 | 400 | 1323 | 1463 | 30.3 |
| | **Resistance (ohm)** | 180 | 1600 | 408 | 581 | 440 | 70.2 |
| | **Clay (%)** | 38 | 46 | 2 | 43 | 44 | 5.3 |
| | **Sand (%)** | 26.5 | 44.7 | 4.4 | 35.2 | 34.2 | 12.6 |
| | **Silt (%)** | 14.3 | 30.3 | 3.6 | 21.8 | 21.2 | 16.5 |
| | **Water retention (% at –33 kPa)** | 23.5 | 33.7 | 2.6 | 28.1 | 28.2 | 9.4 |
| **Plant** | **Ca (%)** | 1.11 | 1.66 | 0.16 | 1.38 | 1.37 | 11.5 |
| | **Mg (%)** | 0.24 | 0.44 | 0.05 | 0.34 | 0.34 | 14.2 |
| | **P (%)** | 0.24 | 0.35 | 0.03 | 0.29 | 0.28 | 9.3 |
| | **K (%)** | 1.26 | 2.85 | 0.38 | 2.03 | 2.02 | 18.8 |
| | **Yield (t $ha^{-1}$)** | 2.1 | 5.3 | 0.8 | 3.4 | 3.4 | 21.9 |

**Table 5.3** Correlation matrix of soil properties and lucerne winter yield.

| | | Soil | | | | | | | | | Plant |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | pH | Org C | Ca | Mg | P | K | Na | Clay | Silt | Yield |
| **Soil** | pH | 1 | | | | | | | | | |
| | Org C | -0.71 | 1 | | | | | | | | |
| | Ca | 0.37 | -0.27 | 1 | | | | | | | |
| | Mg | 0.60 | -0.75 | 0.52 | 1 | | | | | | |
| | P | -0.08 | 0.23 | -0.64 | -0.41 | 1 | | | | | |
| | K | 0.20 | 0.03 | 0.16 | 0.25 | 0.26 | 1 | | | | |
| | Na | 0.76 | -0.63 | 0.21 | 0.58 | 0.09 | 0.46 | 1 | | | |
| | Clay | -0.08 | 0.13 | 0.12 | -0.04 | -0.55 | -0.28 | -0.16 | 1 | | |
| | Silt | -0.26 | 0.51 | 0.36 | -0.20 | -0.17 | 0.11 | -0.28 | 0.10 | 1 | |
| **Plant** | Yield | -0.74 | 0.57 | -0.39 | -0.58 | 0.09 | -0.23 | -0.68 | -0.04 | 0.40 | 1 |

Figure 5.1 shows the experimental layout of the geostatistical trial and the RCB design of the field experiment overlaid on the estimate map of soil pH($H_2O$), as well as the semi-variogram of the latter. The small black dots depict the 123 geostatistical sampling points (nodes) and the grid indicates the layout of the 28 experimental plots.

The spatial structure of pH($H_2O$) was determined with the use of a semi-variogram. Kriging interpolation was used to estimate the values at unsampled locations in an ArcGIS Geostatistical Analyst environment (Johnston *et al.*, 2001). An isotropic semi-variogram was modelled as no definite long and short range directions could be identified. The semi-variogram had a very low nugget variance and a range of 36 m (Table 5.4). The low nugget variance indicates that most of the variation in the soil pH($H_2O$) was accounted for with this sampling density. The estimate map (Figure 5.1) shows a clear trough of low values in the western part of the field, stretching across the field from the south to the north, as well as patches of high and low values in the middle, southeastern and northeastern parts of the field.

**Table 5.4** Model parameters for soil and plant properties.

| **Variables** | | **Anisotropic direction** | **Nugget** | **Sill** | **Long range (m)** | **Short range (m)** |
|---|---|---|---|---|---|---|
| **Soil** | **pH($H_2O$)** | Isotropic | 0.09 | 0.82 | 36 | - |
| | **pH($H_2O$) – 37points** | Isotropic | 0.26 | 0.98 | 35 | - |
| **Plant** | **Yield** | 170° | 0.16 | 1.01 | 42 | 26 |

The pH($H_2O$) of the soil displayed a strong negative relationship with yield (Table 5.3; r = -0.74). Lanyon and Griffith (1988) quoted several studies that found yield reductions when (1) heavy rates of lime have been applied, (2) B is potentially limiting, or (3) P is marginal and lime is applied at normal rates. Although soil pH can be influenced by soil texture, organic matter and other soil chemical properties, the optimum pH($H_2O$) for maximum lucerne yield ranges from 6.0 to 7.5 (Lanyon & Griffith, 1988). In this study, however, pH values ranged between 8.1 and 9.1 (Table 5.2). This confirms the validity of the pH range reported by Lanyon and Griffith (1988).

Table 5.5 shows that there was no significant treatment differences at the probability level $p \leq 0.05$ for yield either for ANOVA 1 or 2 (p = 0.707 and 0.489, respectively). Analyses of variance for all other properties, using the ANOVA 1 and ANOVA 2 structure were also not statistically significant (Table 5.5). The conclusion could be made that the experimental field is homogeneous enough to lay out a standard block design experiment. However, in the discussion of the spatial variation of soil pH($H_2O$), spatial heterogeneity of this property became very clear. In actual fact, the standard way of blocking in either of the two directions would not have provided homogeneous blocks with respect to soil variability. The question was then posed: If the observed spatial variation had been considered, would it have had any effect on the results of this field experiment in terms of yield, or for that matter, any of the other properties that were measured? The experimental design of ANOVA 3 (see Figure 5.1) has been an attempt to statistically take the spatial variability of soil pH($H_2O$) into consideration. ANOVA 3 (Table 5.5) exhibits significant to highly significant differences ($p \leq 0.05$ or $\leq 0.01$) for a number of properties, including yield, as a function of the "treatments", although there were actually no treatments applied. These results have serious implications for the standard method of laying out RCB field trials on what is presumed to be homogeneous land. However, if pH($H_2O$) is regarded as a covariate, as suggested by Snedecor and Cochran (1980) for the typical case where a variable (X) is linearly correlated to the final response (Y), the effects of spatial variation can be accommodated. An analysis of covariance was performed whereby treatment biomass yields were adjusted to remove the effects of pH on yield. In this way lower experimental error was obtained, as well as more precise comparisons among treatments. The results of ANCOVA 3 (Table 5.5) shows that the "treatments" had no statistically significant effect ($p = 0.191$) on biomass yield.

**Table 5.5** Summary of F probabilities (p) for the three experimental designs and covariance analysis.

| | Soil | | | | Plant | | |
|---|---|---|---|---|---|---|---|
| **Experimental designs** | **pH($H_2O$)** | **Org. C (%)** | **Ca ($mg.kg^{-1}$)** | **Silt (%)** | **Ca (%)** | **P (%)** | **Yield ($t\ ha^{-1}$)** |
| **ANOVA 1** | 0.437 NS | 0.538 NS | 0.729 NS | 0.940 NS | 0.729 NS | 0.168 NS | 0.707 NS |
| **ANOVA 2** | 0.798 NS | 0.683 NS | 0.391 NS | 0.482 NS | 0.383 NS | 0.967 NS | 0.489 NS |
| **ANOVA 3** | 0.140 NS | 0.003** | 0.426 NS | 0.036* | 0.465 NS | 0.019* | <0.009** |
| **ANCOVA 3** | n.d. | n.d. | n.d. | n.d. | n.d. | n.d. | 0.191 NS |

* - Statistically significant ($p \leq 0.05$)

** - Statistically highly significant ($p \leq 0.01$)
NS – Not statistically significant
n.d. – not determined

The foregoing results are based on a sampling point total of 123 on an area of 100 X 140 m. Out of a sampling time and cost view, such a large number of sample points might be considered as being impractical. When compared to Figure 5.1 (123 sampling points), the estimate map (Figure 5.2; 37 sampling points) has lost some of the spatial variation in soil pH($H_2O$). However, the overall spatial trends are still discernable in Figure 5.2.

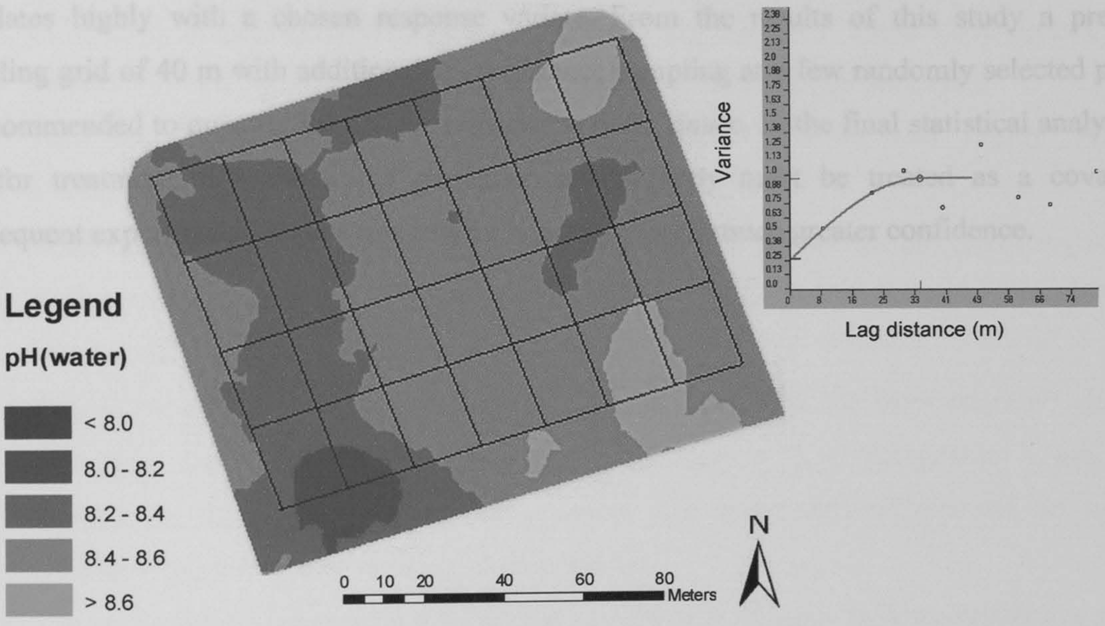

**Figure 5.2** Semi-variogram and estimate map of pH($H_2O$) using only 37 points.

## 5.5 Conclusions

Analysis of variance of a randomized complete block design that consisted of pseudo treatments with replications revealed statistically non-significant differences among treatments for various soil and plant properties, including yield. The conclusion could be made that the experimental field is homogeneous enough to lay out a standard block design experiment. A spatial map of soil pH showed a clear structure in spatial variability. The question was posed that if the latter spatial variation had been considered, would it have had any effect on the results of this field experiment, for example, in terms of yield? The consequent redesign of the experiment whereby all plots were randomly allocated to treatments and replications, led to dramatically different results: significant differences were obtained for plant and soil properties as a function of the pseudo treatments. From this study it is clear that spatial variability of soil and plant properties can jeopardize the results of a standard block design field experiment. However, it was found that soil pH($H_2O$) correlated very well with green biomass yield. Consequently, regarding soil pH($H_2O$) as a covariate an analysis of covariance indicated no statistical difference (as expected) among treatments observed for green biomass yield. It is, therefore, recommended that field experiments should be designed to the cognizance of the spatial variation of a soil property that correlates highly with a chosen response variate. From the results of this study a pre-trial sampling grid of 40 m with additional short distance sampling at a few randomly selected points is recommended to quantify the chosen response variate. Hence, in the final statistical analysis to test for treatment differences, the particular soil property must be treated as a covariate. Consequent experimental results can now be interpreted with much greater confidence.