

Spatial variation of soil and plant properties and its effects on the statistical design of a field experiment

A. Venter^{1*}, M.F. Smith², D.J. Beukes¹, A.S. Claassens³ and M. Van Meirvenne⁴

¹ARC-Institute for Soil, Climate and Water, Private Bag X79, 0001, Pretoria, South Africa

²ARC-Biometry Unit, Private Bag X519, Silverton, 0127, Pretoria, South Africa

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

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

Accepted 26 August 2009

The standard procedure in field experimentation is to use a randomized complete block (RCB) design, striving for the most homogeneous conditions as possible among plots in the same block. However, geostatistical concepts dictate that a spatial dependence exists for observations of a particular property, where closely spaced are more similar than those taken at a greater distance. The present study was conducted on an 18 ha lucerne (*Medicago sativa*) stand in which a 113 m X 145 m experimental area was demarcated. To determine spatial characteristics of soil and plant properties, 48 sampling points were selected using a 20 m square grid with an additional 75 points on a 2.5 m grid at five random node points. A 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 in June 2001 at all sampling points and plots 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, as expected, for various soil and plant properties, including yield. Although large yield differences (2.1 – 5.3 t ha⁻¹) were observed, the conclusion could be made that the experimental field was homogeneous enough to lay out a RCB design experiment. However, it was found that the estimate map of soil pH_(H₂O) indicated a spatial dependence. The question was posed that if spatial variation in soil pH_(H₂O) 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 redesign of the experiment where all plots were randomly allocated to treatments and replications, led to significant differences obtained for plant and soil properties as a function of the pseudo treatments. Analysis of co-variance was then applied to eliminate variability between plots and the resultant pseudo treatments showed no significant differences. From this study it became evident that spatial variability of soil and plant properties can jeopardize the results of a standard RCB designed field experiment and should be taken cognizance of.

Keywords: Block design, covariate, geostatistics, randomization, spatial variability, yield

*To whom correspondence should be addressed. (E-mail: annari@arc.agric.za)

Introduction

Agricultural researchers have long understood that 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 *et al.*, 1994). Present-day agronomic research has reached a point where the treatment effects under investigation are small and the degree of accuracy required in such studies cannot easily be obtained with conventional experimental designs (Van Es *et al.*, 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 (Van Es *et al.*, 1989; Dulaney *et al.*, 1994; Fagroud & Van Meirvenne, 2002). According to Dulaney *et al.* (1994), geostatistical techniques have the potential to provide better field characterization, to improve plot layout, increase the power of the consequential statistical techniques and can be used to select an optimal sampling strategy to characterize 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 soil properties would have an effect on the results of a field experiment if the spatial structure of those properties in the field was 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 and its effects on

a statistically laid out field experiment.

Material and methods

Field and analytical methods

The study was conducted on an 18 ha lucerne (*Medicago sativa*) stand in the Brits district of the North West Province of South Africa (27°49'E, 25°33'S). A rectangular area of 113 m x 145 m (allowing for a 2.5 m border) 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). The lucerne stand was 2 years old when the trial commenced and had been irrigated by a sprinkler irrigation system. During the establishment of the stand, 500 kg ha⁻¹ superphosphate and 200 kg ha⁻¹ potassium chloride were applied as fertilizer. Although the coefficient of uniformity was not determined, the general value of 75% was assumed. 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 Global Positioning System (GPS) and marked with flat metal discs. Figure 1 shows an aerial photograph of the field with the sampling points depicted as small black dots.

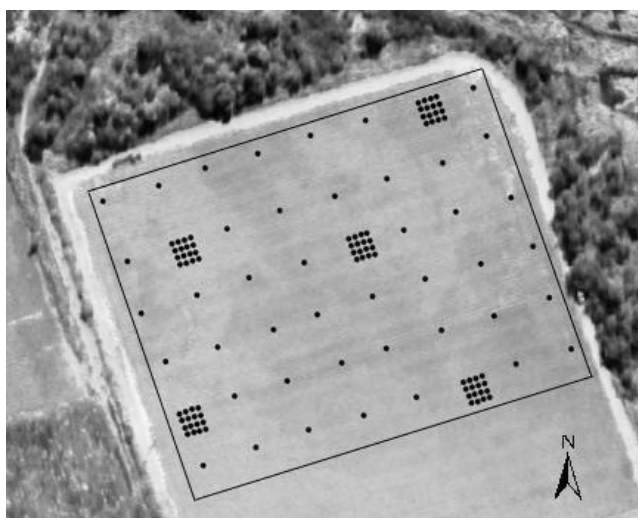


Figure 1 An aerial photograph of the field. The black dots depict the sampling points

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 chosen to fit all the plots in the available area of the original lucerne stand. Figure 2 represents a schematic presentation of the RCB design; the grid indicating the layout of the 28 experimental plots.

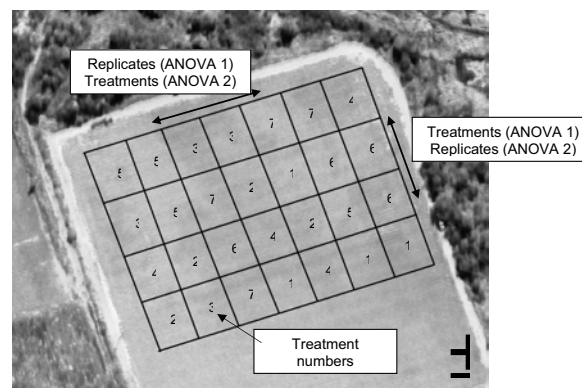


Figure 2 A schematic presentation of the RCB and CRD designs on an aerial photograph of the study area

In June 2001, plant and soil samples were collected at each of the node points. Plant sampling was done by cutting the above-ground plant parts at an early flowering stage within a 0.6 m square around each of the node points and weighing to determine green biomass yield. Three soil samples were collected within the 0.6 m square (0 – 300 mm deep, *i.e.* within the A horizon) 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 m² area of plants in each of the plots and weighing the samples to determine green biomass yield. The sample size was chosen to represent typical experimental research plots. Sub-samples (approximately 10% of the sample size) were taken for analysis purposes. In each of the 10 m² plots a composite soil sample (three samples taken randomly across the plot) 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_(H₂O), organic carbon (C), P (Ambic), ammonium acetate extractable Ca, K, sodium (Na), Mg, electrical resistance, particle size and water retention (-33 kPa), using standard methods (Non-Affiliated Soil Analysis Work Committee, 1990).

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. Analysis of variance (ANOVA) was 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 1). The first one (ANOVA 1) blocked in the NE-SW direction, with 4 treatments randomly allocated to each of the 7 blocks. The second design (ANOVA 2) was blocked in the NW-SE direction,

with 7 treatments randomly allocated to each of the 4 blocks (Figure 2). Both experimental layouts were based on a RCB design. It is obvious from the spatial variability of soil pH_(H₂O) (Figure 3) 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 (CRD) 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 2). An analysis of covariance (ANCOVA) was subsequently per-

formed using pH_(H₂O) as a covariate to eliminate the linear effect of soil pH on yield.

Geostatistical analyses were performed on the samples (123) of the geostatistical grid to generate a map showing the spatial variation of the soil pH_(H₂O) using the ArcGIS Geostatistical Analyst extension (Johnston *et al.*, 2001). Additional analyses were performed to generate a spatial map for pH_(H₂O) (Figure 4) 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).

Table 1 ANOVA of three different experimental designs

	ANOVA 1	ANOVA 2	ANOVA 3
	RCBD	RCBD	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

Results and discussion

Mean concentrations of all the soil and plant properties measured on the 28 experimental plots are shown in Table 2. The observed mean nutrient concentrations in the plants are, according to Pinkerton *et al.* (1997), in the “adequate” to “high” range, although there were some spots in the field, especially for K, which showed deficiencies. The correlation matrix (Table 3) shows that both pH_(H₂O) and Na are highly negatively correlated with yield ($r = -0.74$ and -0.68 , respectively) and that there is a strong relationship between these two soil properties ($r = 0.76$). Lanyon and Griffith (1988) quote 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 (such as the basic cations included in Table 2), the optimum pH_(H₂O) for maximum lucerne yield ranges from 6.0 to 7.5, with a negative linear relationship beyond the optimum range (Lanyon & Griffith, 1988). In this study, pH values ranged between 8.1 and 9.1 (Table 2), *i.e.* beyond the optimum range, with a negative relationship confirming the results of Lanyon and Griffith (1988).

The spatial variability of pH_(H₂O) was therefore determined with the use of a semi-variogram (Figure 3). Ordinary kriging interpolation was used to estimate the values at unsampled locations in an ArcGIS Geostatistical Analyst environment (Johnston *et al.*, 2001). This kriging method is specifically suitable for data that do not have strong trends – as was the case in the present study. For the same reason an isotropic semi-variogram model was used as no definite long and short range distances, as well as direction of spatial correlation, could be identified. The semi-variogram is calculated from the sample variance and provides a mean of quantifying the spatial variation of a range of properties by measuring the

degree of correlation between sampling points a given distance apart (Webster & Oliver, 2001). The semi-variogram had a very low nugget variance and a range of 36 m (Table 4). The low nugget value indicates that most of the variation in the soil pH_(H₂O) was accounted for with this sampling density. The estimate map (Figure 3), shows a clear trough of relatively lower values in the western part of the field, stretching across the field from the south to the north, as well as patches of relatively higher and lower values in the middle, south eastern and north eastern parts of the field.

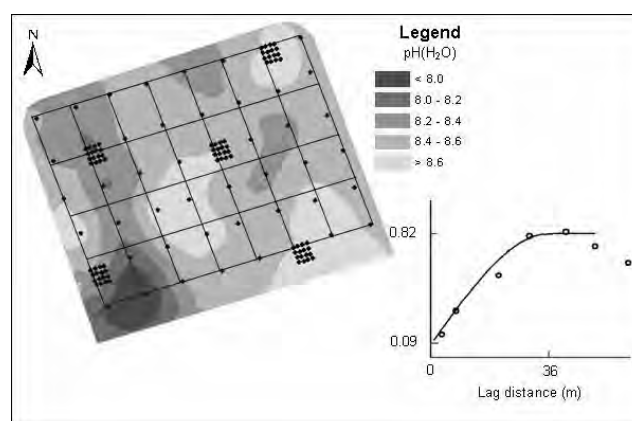


Figure 3 Semi-variogram and estimate map of soil pH_(H₂O) using 123 points

Table 2 Statistical description of soil and plant properties

		Min.	Max.	Std. Dev.	Mean	Median	CV (%)
Soil	pH _(H2O)	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	3.8	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	1.03	2.02	18.8
	Yield (t ha ⁻¹)	2.1	5.3	0.8	3.4	3.4	21.9

Table 3 Correlation matrix of soil properties and lucerne winter yield

		Soil								
		pH _(H2O)	Org C	Ca	Mg	P	K	Na	Clay	Silt
Soil	pH _(H2O)									
	Org C	-0.71								
	Ca	0.37	-0.27							
	Mg	0.60	-0.75	0.52						
	P	-0.08	0.23	-0.64	-0.41					
	K	0.20	0.03	0.16	0.25	0.26				
	Na	0.76	-0.63	0.21	0.58	0.09	0.46			
	Clay	-0.08	0.13	0.12	-0.04	-0.55	-0.28	-0.16		
	Silt	-0.26	0.51	0.36	-0.20	-0.17	0.11	-0.28	0.10	
Plant	Yield	-0.74	0.57	-0.39	-0.58	0.09	-0.23	-0.68	-0.04	0.40

Table 4 Model parameters for soil and plant properties

	Variables	Anisotropic direction	Nugget	Sill	Long range (m)	Short range (m)
Soil	pH _(H2O)	Isotropic	0.09	0.82	36	-
	pH _(H2O) 37 points	Isotropic	0.26	0.98	35	-
Plant	Yield	170°	0.16	1.01	42	26

Table 5 shows that there were no significant treatment differences at the probability level $p = 0.05$ 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). The conclusion could be made that the experimental field is homogeneous enough to lay out a standard RCB design experiment. However, as discussed above, spatial heterogeneity of soil pH_(H2O) 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 pH_(H2O) 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 third pseudo experimental design (ANOVA 3, Figure 2) was laid out as a CRD in an attempt to statistically take the spatial variability of soil pH_(H2O) into consideration. ANOVA 3 (Table 5) indicates significant treatment differences for a number of properties, including yield, although there were no actual 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 as RCB designs assume that there is a soil prop-

erty gradient in one direction only. From Figure 3 this is not true for this particular study area. 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 ANCOVA was performed whereby treatment biomass yields were adjusted to remove the effects of soil $pH_{(H_2O)}$ 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) show that the pseudo treatments had no statistically significant effect ($p = 0.191$) on biomass yield. It is recognized that there are other statistical methods that take spatial variability into account such as applying restricted maximum likelihood (Welham & Thompson, 1997).

The foregoing results are based on a sampling point total of 123 on an area of 113 x 145 m. From a sampling time and cost view, such a large number of sample points might be considered impractical. When compared to Figure 1 (123 sampling points), the estimate map (Figure 4; 37 sampling points) has lost some of the spatial variation in soil $pH_{(H_2O)}$, thereby losing some detail. However, the overall spatial trends are still discernable in Figure 4 but now time and cost can drastically be reduced.

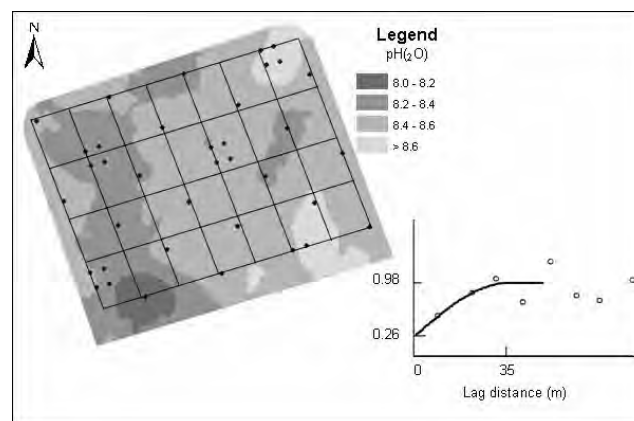


Figure 4 Semi-variogram and estimate map of soil $pH_{(H_2O)}$ using 37 points

Table 5 Summary of F probabilities (p) for the three experimental designs and covariance analysis

Experimental designs	$pH_{(H_2O)}$	Soil			Plant			Yield ($t\ ha^{-1}$)
		Org. C (%)	Ca ($mg\ kg^{-1}$)	Silt (%)	Ca (%)	P (%)		
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 ($p = 0.05$)

** - Statistically highly significant ($p = 0.01$)

NS – Not statistically significant

n.d. – not determined

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. However, a map of soil $pH_{(H_2O)}$ showed clear spatial dependence. The question was posed that if this spatial variation was considered, would it have had any effect on the results of this field experiment, for example, in terms of yield? The redesign of the experiment whereby all plots were randomly allocated to treatments and replications, led to dramatically different results: significant treatment differences were obtained for some plant and soil properties, which gave erroneous results as no real treatments were applied. From this study it is clear that spatial variability of soil properties can jeopardize the results of a standard RCB designed 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 ANCOVA indicated no statistical differences (as expected) among treatments observed for green biomass

yield. It is, therefore, recommended that field experiments should be designed to take cognizance of the spatial variation of a soil property that correlates well 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.

Acknowledgements

This research was funded by the ARC-Institute for Soil, Climate and Water, Pretoria. Thanks are due to Mrs. Esmé Lazenby and other technical staff for their help in collecting the samples.

References

- AGRI LABORATORY ASSOCIATION OF SOUTHERN AFRICA, 1998. Feed and Plant Analysis Methods. AgriLASA, Pretoria.
- DULANEY, W.P., LENGNICK, L.L. & HART, G.F., 1994. Use of Geostatistical Techniques in the Design of an Agricultural Field

- Experiment. p. 183–187. *In*: Proceedings of the Survey Research Methods Section, American Statistical Association. Contributed papers by topic, I. Remote Sensing.
- FAGROUD, M. & VAN MEIRVENNE, M., 2002. Accounting for Soil Spatial Autocorrelation in the Design of Experimental Trials. *Soil Sci. Soc. Am. J.* 66, 1134–1142.
- GENSTAT FOR WINDOWS, 2000. Release 4.2. Fifth Edition. Oxford: VSN International.
- JOHNSTON, K., VER HOEF, J.M., KRIVORUCHKO, K & LUCAS, N., 2001. Using ArcGIS Geostatistical Analyst. ESRI, 380 New York Street, Redlands, CA 92373-8100..
- LANYON, L.E. & GRIFFITH, W.K., 1988. Nutrition and Fertilizer Use. p. 333–372. *In*: A.A. Hanson (ed.). Alfalfa and Alfalfa Improvement – Agronomy Monograph 29, ASA-CSSA-SSSA, Madison.
- NON-AFFILIATED SOIL ANALYSIS WORK COMMITTEE, 1990. Handbook of Standard Soil Testing Methods for Advisory Purposes. Soil Science Society of South Africa, Pretoria.
- PINKERTON, A., SMITH, F.W. & LEWIS, D.C., 1997. Pasture Species. p. 307–311. *In*: D.J. Reuter & J.B. Robinson (eds.). Plant Analysis – an Interpretation Manual, 2nd edn, CSIRO, Collingwood.
- SNEDECOR, G.W. & COCHRAN, W.G., 1980. Statistical Methods 7th edn. Iowa State University Press, Iowa.
- SOIL CLASSIFICATION WORKING GROUP, 1991. Soil Classification – A Taxonomic System for South Africa. Soil and Irrigation Research Institute, Department of Agricultural Development, Pretoria.
- VAN ES, H.M. & VAN ES, C.L., 1993. Spatial Nature of Randomization and its Effects on the Outcome of Field Experiments. *Agron. J.* 85, 420–428.
- VAN ES, H.M., VAN ES, C.L. & CASSEL, D.K., 1989. Application of Regionalized Variable Theory to Large-Plot Field Experiments. *Soil Sci. Soc. Am. J.* 53, 1178–1183.
- WEBSTER, R. & OLIVER, M.A., 2001. Geostatistics for Environmental Scientists. John Wiley & Sons, Chichester.
- WELHAM, S.J. & THOMPSON, R., 1997. Likelihood ratio tests for fixed model terms using residual maximum likelihood. *Journal of the Royal Statistical Society, Series B* 59, 701-714.