Biosystems Engineering, Volume 84, Issue 4, April 2003, Pages 441-453.

Soil Factors and their Influence on Within-Field Crop Variability II: Spatial Analysis and Determination of Management Zones

J.C. Taylor[1]; G.A. Wood[1]; R. Earl[1]; R.J. Godwin[1]

[1]Cranfield University at Silsoe, Bedforshire, MK45 4DT

e-mail of corresponding author: *j.c.taylor@cranfield.ac.uk*

Abstract.

Spatial variation of crop yields was examined in three trial cereal fields in England from 1994 through 1997. The fields were managed with uniform inputs but there were considerable differences in the spatial patterns and magnitudes of variation between fields and seasons. Up to 50% of the yield variation was within the tramline spacing distance (20 to 24 m) and this appeared to relate to crop management practices rather than underlying soil factors. Longer-range variation generally increased up to field-scale but was not constant between seasons. Longer-range variation was more apparent in dry years and was attributable to soil variation. Soil series differences coincided with yield differences in dry years when the soil series differences could be expected to create large differences in soil-water relationships. Soil electrical conductivity, measured by Electromagnetic Induction (EMI) was investigated as a surrogate for detailed soil coring. Field zones created by EMI also coincided with yield differences and zones were similar to those delineated by soil series with expected differences in soil-water relationships. EMI observations were found to be a useful and cost-effective surrogate for representing soil variability in fields likely to create yield variations. Sub-division of fields into management zones using multivariate K means cluster analysis of historical yield and EMI observations formed an objective basis for targeting soil samples for nutrient analysis and development of site-specific application strategies. The appropriateness of sit-specific management has to be assessed annually because magnitude and pattern of variation changes from season to season.

1 Introduction

This work examines the influence of soil and soil-nutrient spatial variability on crop performance in the context of the broader study that was aimed at the production of management guidelines for precision farming, for the farming community. The work was carried out in three trial fields in Southern and Eastern England where intensive soil and soil-nutrient observations, described in Part I of this work by Earl *et al* (2001), had been made in conventionally managed cereal crops. This was in preparation for spatially varied N management trials also reported in detail by Welsh *et al*., (2001a & b), and Wood *et al*. (2001). The focus of this paper is to explore the coincidence of spatial variability of soil nutrients (excluding N), pH and soil physical factors with spatial variation of crop yield and to consider the potential for site-specific management based on observed variations in soil and yield spatial patterns. Methodology for establishing a practical number of management zones where site-specific management is applicable is also investigated.

The fundamental objective of precision farming is to manage spatial variation in crops to give economic and/or environmental benefits. Two vital precursors to achieving this end are: accurate measurement of spatial variation; and, accurate interpretation of measurements to guide the management strategy. The ideal representations of spatial variations are maps showing the magnitude of the variable of interest at all points in the land parcel containing the crop. Measuring at all points is, however, not economically viable and so fewer, strategically located observations coupled with an appropriate model for interpolation to provide estimated values at unmeasured locations is required. The aim of sampling in this context is to enable a sufficiently accurate representation of the spatial variation in soil properties to be made at an economically viable cost.

Topographic variation and its representation in the form of contour maps is a familiar example. Production of an accurate topographic map requires an adequate number of appropriately positioned observations of land surface elevation. The positioning of sample sites is relatively straightforward as the topography is visible and rules can easily be established relating to the minimum size of features to be mapped and positioning in relation to slope breaks etc. Sample points are usually selected so that there is a linear variation in height between the points and the height anywhere between the points can be readily calculated by linear interpolation. Thus the number and density of observations for different levels of precision in representing the topographic surface is relatively easy to determine. This is not the case for most soil variations as the factors involved are not immediately visible (e.g. available water capacity) and, therefore, no visual model exists to assist assessments making sampling much more difficult. Two sampling strategies are considered in this work: grid sampling and targeted sampling in management zones determined by multivariate classification.

In grid sampling, the choice of grid dimension affects the sampling-cost and accuracy of representation of the mapped variable. Spatial variation of soil properties is generally complex and analogous to a topographic surface with many ‘hills’ and ‘valleys’ of different sizes and frequencies within the region of interest. Thus in flat or uniformly changing areas, the grid sample may result in more observations than necessary and conversely, areas with high frequency fluctuations in soil properties will be under sampled and important variations can be missed. Geo-statistical methods can be used to characterise the spatial continuity using a variogram (Isaaks and Srivastava, 1989). Numerous studies of spatial variability of soil have been carried out in recent years in the context of precision agriculture see for example Webster and McBratney (1987), Birrell *et al* (1996), Brenk *et al* (1999) and

Jaszberenyi *et al* (1999). In principle, the variogram for the soil variable being measured should be used, however, these may differ for each variable and each field. Moreover, a dense grid of observations is required to enable variograms to be constructed making this approach uneconomic for most soil variables. Crop yield variation is the ultimate issue in this work, and yield-mapping combine harvesters produce a dense grid of observations enabling the yield variogram to be estimated. A practical approach, therefore, is to use the yield variogram to guide the choice of grid size.

Targeted sampling represents a way of reducing sample size and hence cost. Various approaches including response surface analysis (Lesch *et al*, 1995), image classification (Pocknee *et al*, 1996 and Thomas *et al*, 1997) and the variance quad tree (McBratney *et al*, 1999) have been proposed, however, management zones defined by multivariate classification is also a potential method for considerably reducing the number of soil samples and hence costs, required to measure within-field soil nutrient variation. Lark *et al* (1998) created management zones by cluster analysis of yield map data and found some correspondence between soil series and yield variation. In this paper the use of soil electrical conductivity variability, measured by electromagnetic induction (EMI) as a soil surrogate to create management zones is explored and the approach by James *et al* (2000) is improved and made objective by using both EMI and historical yield data with cluster analysis.

Cereal yield-variation is potentially influenced by many factors acting simultaneously. Not all of them are related to underlying variations in soil conditions or crop management. Pest damage for example may be a regular feature in certain parts of a field because of proximity to a woodland edge.

Multivariate classification by cluster analysis enables the identification of sub-regions in the field, which internally have similar characteristics. Once the field has been sub-divided, this provides a basis for targeting sampling to investigate the reasons for yield variation using a practical number of sample samples. This approach has been used by Lark *et al.* (1998) and Cupitt and Whelan (2001) for precision farming applications and is a standard technique for the classification of multi-band remote sensing data (see for example Richards and Jia,1999). In this work the K-means clustering method has been applied. This can be viewed as the reverse of analysis of variance. Observations are the numerical values associated with each variable in the analysis (e.g. last year's yield, soil electrical conductivity etc) at a specific location. Observations are grouped and re-grouped iteratively into K classes until the within group variation is minimised and the between group variation is maximised. In the context of precision farming, once the observation points have been classified, the spatial pattern of the classes across the field can be investigated by plotting the classified points on a map. Examining the class means for each variable included in the analysis can then assess the practical importance of the classes. The K-means method also has the practical advantage of allowing the *a priori* choice of the number of classes allowed. In practice this choice of the optimum number can be based on the practical importance of differences in class means and on the spatial complexity of the resulting map. Higher numbers of classes result in spatially complex patterns with smaller and smaller differences. Practical precision farming considerations may dictate that a small number of classes are selected even though some differences may remain unresolved.

1.1 *The experimental fields and data sets*

Three sites, were selected for this study: Trent Field, Twelve Acres and Short Lane. A summary of the soil series represented in each is presented in Table 1. A summary of the soil and crop observations used in this study are given in Table 2, further details are presented in Part I.

All yield observations were made by the AGCO combine yield monitoring system (Moore, 1998). Electromagnetic induction (EMI) surveys were carried out when the fields could reasonably be assumed to be at or near field capacity moisture contents (Waine *et al*, 2000). Readings were taken 10 x 24 m grid intervals by walking along the field tramlines with an EM38 Geonics scanner. The Soil Survey and Land Research Centre (now the National Soil Resources Institute) conducted detailed soil surveys using a tractor-mounted soil-coring machine to make observations on an approximate grid spacing of 25 m in Trent Field and Twelve Acres. Two grid spacings were use in Short Lane: 25m spacing within trial areas; and, 50m spacing outside trial areas.. The soil core locations were fixed using a differential global positioning system (DGPS). The soil profiles thus obtained were classified by soil series (Avery, 1987).

Spatial variation in soil and plant observations frequently occurs at different spatial scales. Plant nutrients generally do not disperse uniformly in soil when observed at very small scales and measurements made from small samples can contain a large element of variation because of this. Interpolation between such observations, over larger distances is thus very risky because small spatial scale variations, inconsequential for practical purposes, can be greater than larger scale trends. Observations are needed that give representative values at points suitable for the desired scale of interpolation. This is achieved by increasing the ‘support’ for the observations. ‘Support’ is the area, or volume of the sample. Individual soil samples are usually very small. Sampling protocols for obtaining representative samples for whole fields entail walking in a ‘W’ pattern across a field whilst taking small soil samples at regular intervals along the way. These are bulked together and mixed thoroughly before analysis. When sampling on a grid, the aim is to collect the sample so that it is locally representative of the bulk soil properties. This can be achieved by bulking together small samples collected in the vicinity of the grid point. This problem has been thoroughly analysed by Oliver *et al* (1997). Soil Nutrients (P, K, Mg, Mn, Zn, Cu, S) Organic Matter and pH were measured at locations on an approximate 50m grid in Trent Field and Twelve Acres on two times (autumn/winter and spring). Analyses were based on a minimum of five bulked sub-samples taken within a 5m radius of the observation points. Other soil observations were based on single 50 mm diameter cores (referred to as minicores). Support for yield observations is the area is equivalent to the distance travelled by the combine times the width of cut, assuming that other errors in the yield monitoring system have been removed by signal processing (Blackmore and Moore, 1999). Support for EMI observations depends on the geometry of the coils in the instrument and is difficult to know exactly. For the EM-38 used in this work, the observations are estimated to represent a soil volume in the order of a cubic meter.

1.2 *Rainfall variation*

Rainfall observations for the period 1989 through 1998 were obtained from met stations of the UK national network close to the study fields (max distance 6 km). In Trent field, the average annual seasonal rainfall (September through to following August) was 784 mm and the annual values for 1995, 1996 and 1997 were 790, 624 and 685 mm respectively. At Short Lane, the rainfall was considerably lower on average than at Trent Field. The average annual seasonal rainfall was 498 mm and the annual values for 1994, 1995 and 1996 were 572, 489 and 412 respectively. At Twelve Acres, the average annual seasonal rainfall was 731mm, slightly lower than at Trent Field. The annual values for 1995 and 1996 were 693 and 541 mm respectively. There was also considerable within-season variation of the rainfall pattern. Fig. 1 shows the monthly deviations of rainfall from the 1989-1998 averages, near the trial fields for the seasons corresponding to the yield information.

2 Within-field variability of crop yields in trial fields

Omnidirectional yield variograms for the trial fields are presented in Fig. 2. Variograms are usually depicted schematically as a continuous convex curve starting from a small intercept, rising steeply then flattening out to a maximum plateau value. The maximum height of the variogram (sill) provides an indication of the level of variation in the field. The lag at which the maximum is reached (range) is the distance between points beyond which the maximum variation is expected. The intercept at zero lag (nugget) is a measure of random variation normally attributed to measurement errors. The sill can be used to assess the importance of the within-field variation. If the sill is small, the variability is small therefore site-specific management may not be warranted. Conversely, if the sill is large then the variation may be considered important.

The actual variograms in Fig. 2 show a more complex picture. The variogram shapes for each field were different and there were considerable differences between seasons. All variograms started with a steeply rising semi-variance up to lag distances of around 20 to 24m. The proportion of the total semi-variance accounted for at these short lag distances varied considerably between fields and between seasons. Short Lane had the highest proportion with about 50% of the total variability at lags up to 20m. The ranges for the amount of variability at these short lags in the seasons presented was equivalent: 0.6 to 0.8 t/ha in Trent Field; 0.9 to 1.1 t/ha in Short Lane; and, 0.5 to 0.8 t/ha in Twelve Acres. In 1995, the total field-scale variability represented by the large lags up to 350 m in Trent Field and Twelve Acres was relatively low. In the dry year of 1996, the semi-variance increased at varying rates as the lag increased. The relative proportion of the longer-range variability in Short Lane was also greatest in the dry year. The ranges for the amount of yield variability at maximum lags for the seasons recorded were equivalent to: 0.8 to 1.4 t/ha in Trent Field; 1.2 to 1.4 t/ha in Short Lane; and, 1.1 to 2.1 t/ha in Twelve Acres. With the exception of Trent Field in 1995 and Short Lane in 1996, the yield variability was still increasing at lags approaching the maximum cross-field distances shown in Fig. 2. The variogram for 1996 in Trent field shown enlarged in Fig. 3, has a series of ‘ripples’ with local minima known as ‘holes’ (Isaaks and Srivastava, 1989) occurring at 24 m intervals. Hole effects occur when observations separated by distances corresponding to the lag where the minimum occurs are more similar than at some smaller lags. In the 1996 Trent Field variogram this is explained with reference to Fig. 4 by striping in the field caused by poor distribution of fertilizer in the lateral direction across 24m spaced tramlines reported by Taylor *et al* (1997). There is some evidence of a residual effect in the 1997 variogram although the fertilizer distribution problem had been solved. Also, the effect was not present in the 1995 variogram. There also seems to be small hole at around 48 m lag in the variograms for 1994 and 1995 in Short Lane. This might relate to some directional effect in the crop management as it corresponds to distances between every other tramline and operations would be carried out in the same direction.

If the variation is to be mapped accurately by spatial interpolation between observation points then there must be a high correlation between the observations at adjacent points. The correlation between observations can also be calculated and plotted versus lag to produce the correlogram (Isaacs and Srivastava, 1989). In correlograms, the correlation is high at low lag and decreases as the lag increases and the pattern presents itself as an inverse of the variogram. Thus the variogram pattern can be used to infer relative differences in the correlation at different lags. In the trial fields, a large proportion of the yield variability is short-range, at or below the tramline spacing. Thus, to pick up the short-range variation, the distance between adjacent sampling points has to be less than 24m. Sampling at his spacing is uneconomic also; the short-range variability is likely to include residual error from the yield monitoring system and variability induced by the crop management practices along the tramline system.

3 Influence of soil and rainfall variation on yield

Yield variation was examined in relation to soil variation as expressed by soil series determined at the minicore sites. The mean yields in a 10 m radius around each soil core were determined from the original yield observations – this eliminates most of the effects of the short-range variations discussed previously. Fig. 5 shows box (mean +/- 1 std error) and whisker (mean +/- 2 std error) plots of the yields in each soil series in a) Trent Field, b) Short Lane and c) Twelve Acres, in each season. Considerable between-series and between-season variations are evident and, therefore, the interaction of variation in rainfall with soil type was investigated as the potential cause.

3.1 *Trent Field*

Although 1996 was the driest of the three years with reference to Fig. 1a, this corresponded to the highest yields therefore the low annual rainfall did not necessarily limit yield. However, in 1996, there are significant differences between yields in the soil series. In Trent Field, there is an area of disturbed soil (Dist), which is at the foot of a long slope. In 1996 and 1997, this was the highest yielding area of the field. However, in 1995 it had the lowest yield (significantly lower then Andover, (An) but not Panholes, (pH)). This might be explained by the rainfall pattern during sowing and establishment. In 1996 and 1997, September was relatively dry whereas in 1995 it was relatively wet. Thus in 1995, wet conditions at the foot of the slope during sowing could have resulted in poor establishment. In the wetter years (1995 and 1997) the yield differences between An and pH were not significant but in the dry year, An out-yielded pH. 1997 was a low-yielding year with a generally similar rainfall pattern to 1996 and thus low yield is not explained through rainfall. Frost damage at anthesis, particularly on the more exposed slopes of the An series was a possible explanation for this low yield. Thus in years with less than average rainfall, significant yield differences are likely to occur on the different soil series – up to 0.58 ton/ha between An and pH.

3.2 *Short Lane*

1996 was only the third driest in the period of record considered (9 seasons, 1990 to 1998). In Fig. 5b, in the two wettest years 1994 and 1995, there was no difference between the mean yields in the three soil series (Ludford, (Lu), Wickham, (Wh) and Maplestead, (MM)) but there was considerable difference in the overall yield with the drier of the two years (1995) considerably out-yielding the wetter (1994). The autumn of 1994 was relatively wet (Fig. 1b) and poor establishment might have been the reason. In the dry year of 1996, the yields were significantly different and in a manner consistent with general expectation of these soils in such conditions. Wickham, a heavy clay loam over clay had the highest yield (6.42 t/ha), the Ludford, clay loam over sand was next (6.00 t/ha) and the Maplestead, sandy loam was the lowest (5.45 t/ha). Thus there was nearly 1 t/ha difference between the sandy loam and the clay loam over clay.

3.3 *Twelve Acres*

The 1996 rainfall was the lowest in the period 1989 – 1998. In Fig. 5c, there was a large seasonal difference in yield with the dry year(1996) out-yielding the other by almost a factor of 2. However, there were no significant differences between the yields on different soil types (Sherborne, (Si), Moreton, (Mor), Haselor, (Hb) and Didmarton, (dD)). Sherborne is a calcareous silty clay loam over brashy oolitic limestone, Moreton and Haselor are similar being calcareous silty clay loam over yellowish brown silty clay loam. Therefore the similarity of the upper soil profiles could account for the between series yield similarity. Didmarton comprises a deeper, silty clay topsoil derived from upslope colluvium from the other three soil types and was consistently the highest yielding, but there were only three observation sites within this soil.

3.4 *Summary*

Soil series had an important influence on within-field yield-differences when this was reflected by large differences in soil textural class. The differences were more marked in dry years. In wetter years, the differences were smaller except when soil wetness may have adversely affected crop establishment.

4 EMI as a surrogate for Soil series

Maps produced from EMI scans of the Trent Field and Short Lane at field capacity show patterns reflecting the underlying soil variations (James *et al.*, 2000, Waine *et al.*, 2000). In this work cluster analysis has been applied to the same data sets to provide an objective method of creating EMI classifications and these are compared with the soil series classifications. Mean values of EMI within a 10m radius of the minicore locations were determined from the original EMI data sets. Crop yield variation between EMI classes is investigated. Tables 3, 4 and 5 show the correspondence between EMI and soil series classifications in Trent Field, Short Lane and Twelve Acres respectively using a number of clusters equal to the number of soil series.

Inspection of Tables 3, 4 and 5 shows that the correspondence between soil series and EMI clusters is variable. For example, in there is a good agreement between An and cluster 1 in Trent Field, and between Wh, MM and clusters 3 and 1 respectively in Short Lane. However, other clusters comprise mixtures of soil series eg Cluster 2 in Trent Field and Cluster 2 in Short Lane. In Twelve Acres, the only convincing agreement is between Si and cluster 3 the remaining pattern of agreement is unclear.

This may be because the soil series in Twelve Acres are all similar in terms of textural class. Assigning the clusters to soil series, based first on obvious agreements seen in Tables 3, 4 and 5, the statistical significance of the overall agreements was assessed using the Kappa statistic (Congalton and Mead, 1983; Hudson and Ramm, 1987). The significance of the classifications was tested against the null hypothesis that the correspondence between clusters and series is random (Kappa = 0) and the results are presented in Table 6. In both Trent Field and Short lane, the agreements are statistically significant, however, in 12 Acres the agreement is not significant.

Fig. 6 shows the box (mean +/- 1 std error) and whisker (mean +/- 2 std error) plots of crop-yield for each cluster and season in the trial fields. The pattern of differences is remarkably similar to the equivalent plots for soil series. Figs 7 and 8 show plots of the soil series and clusters in Trent Field and Short Lane. The patterns are generally similar as would be expected from the cross tabulation results. In Trent, the main differences are 1) the appearance of a second region halfway along the northern boundary, similar to the disturbed soil in the northern point and, 2) the area covered by cluster 2 is more extensive that the area identified as pH series. In Short Lane, cluster 2 identifies a much larger ‘transitional’ region between the clusters related to Wh and MM series than the area covered by Lu.

5 Identification of Management zones

The cluster analysis in the previous section is a basis for defining the management zones for precision farming. However, the relevance of both soil series-based or EMI-based management zones in relation to crop-yield variation is also variable. In the dry year of 1996, there were clear yield differences between both sets of zones in Trent Field and Short Lane although there were some differences in the field-areas defined by them. In Trent Field yields were higher in 1996 than in wetter years and it seems that soil factors were controlling spatial variation in yield. In Short Lane, the 1996 yields were lower than one of the wetter years indicating a possible overall water-limitation on the crop. In 1996, soil factors were controlling spatial variation in yield and possibly the overall yield. In Short Lane, 1996 was only the third driest year out of the 10-year period thus it is possible that the patterns observed in 1996 could be important in about 30% of years though a more extensive study of the rainfall records is necessary to confirm this.

Soil factors represented by soil series or EMI variations are not the only ones with a potential influence on yield. Soil nutrient status and pH need to be considered. Also, some sources of observed yield variation will be unrelated to soil e.g. pest damage. This points to the need to include additional information in the process of defining management zones. In practice historical yield information can be available. In the following analysis, historical yield data (from yield mapping) and EMI data are combined to divide the field into management zones using multi-variate K-means clustering. The K-means algorithm has the practical advantage of allowing some control over the total number of management zones that will be defined because the number of clusters is specified. The number of resulting management zones can be larger than the number of clusters because the same cluster may identify several similar areas within the field. Fig. 9 shows the results of cluster analysis using the 1996 yield data and EMI using 4 and 6 clusters.

In general, increasing the number of clusters results in increased spatial fragmentation of the management zones, which have smaller and smaller differences. The analysis is rapid so the best approach is to make trial analyses and to select the solution that yields a practical compromise. The spatial pattern of clusters was different in each season, reflecting the between season patterns of yield variation.

6 Variability of soil nutrients, pH and organic matter in management zones identified by cluster analysis

In this section, the variability of soil nutrients, pH and organic matter in relation to the management zones identified by clustering yield and EMI data was explored. The practical implication was to see if the management zones were a basis for soil sampling for nutrient analysis. Trent Field was used as a case study. Fig. 10 shows the relationship between soil observation points (for nutrient analysis) and the management zones identified by clustering the 1996 yield and EMI data using four clusters.

In practice, soil sampling would be done by collecting and bulking sub-samples along ‘w’ transects within each management zone, resulting in four samples (combining the two small zones identified by cluster 2 in Fig.10) instead of the 27 used in the grid sampling. To simulate this, the results of grid samples within each management zone were examined. Grid samples on the boundary between clusters were not used. The numbers of soil nutrient observations were 9, 2, 4 and 4 in clusters 1 through 4 respectively. The low number of observations in zone 2 makes the estimation of within zone variation large compared to the other zones but this would be consistent with the small area covered.

Figure 11 summarises the yields in each of the management zones for the 1996 season prior to soil nutrient sampling and the following 1997 season. In both years, the crops were managed with uniform applications of inputs and nutrient levels were not below accepted minimum levels (Chalmers *et al* (1999) and MAFF (2000)). The yield in 1997 was considerably lower than in 1996 in spite of the higher rainfall. An explanation for the lower yield is frost damage observed during anthesis in 1997. In 1996, the management zone covered by cluster 3 was the highest yielding, and that covered by cluster 1 the lowest, the yield difference being about 1.5 t/ha. In 1997, the area covered by cluster 4 was the lowest and by cluster 2 the highest, with a yield difference of about 1 t/ha. Cluster 4 is on a north-facing slope and may have suffered the more frost damage.

The pH levels in each management zone are shown in Fig. 12. The pattern of variation is the same for the two observation periods. The pH in zone 1 is significantly less than for zones 3 and 4. P and K variations between management zones are shown in Fig. 13. The patterns were similar for the two measurement periods. All zones were above limiting values but there were significant differences in P levels between zones. P levels in Zone 4 are approaching the limiting value and yields in zone 4 are among the lowest in the field. Hence, there could be a case for boosting the P fertilizer application in this zone. K levels were generally more variable but were not different between the zones. There were also significant differences in levels of Cu, Mg, Mn, and Zn detected between some of the zones. Some were approaching the recommended minimum levels.

The multivariate clustering of yield and EMI data has resulted in the identification of management zones in the field that coincide with differences in soil nutrients. Therefore, these zones would appear a good basis for targeted soil sampling and for appropriate site-specific nutrient applications. In Trent Field, the laboratory costs for targeted sampling would be 15% of those for the grid sampling approach used.

7 Conclusions

Examination of the yield semi-variograms from several seasons for the trial fields shows that a considerable proportion of the yield variability was within distances at or below tramline spacings of 20 to 24 m. Possible causes are systematic differences in crop management across tramlines caused by poor calibration of farm machinery. This short-range variation was thus not related to soil factors. Longer-range variability was seasonally variable and was coincident with differences in soil series where these reflected differences in soil texture.

EMI is a cost-effective alternative to a detailed soil survey for classifying fields into zones as not only is it cheaper, it also allows a higher density of observations. In fields containing series with large differences in soil textural classes, the EMI zoning was highly correlated to zoning by soil series and the patterns of yield variation were similar. However, direct soil observations a few locations, selected with the aid of EMI should be carried out to assist interpretation.

Coincidence of soil and yield variability was more apparent in dry seasons. In wetter seasons, the longer-range variation was smaller. On high clay content soils, drier years out-yielded wetter years indicating that excess soil moisture, particularly at establishment was perhaps the biggest influence on overall yield in the experimental fields. The wetter years in the study period were close to the long-term average rainfall. In the sandy soils of Short Lane, the yield appeared to be water-limited and this was likely to be a problem about one year in three.

The influence of transitory seasonal factors was greater than the underlying soil factors. There were considerable seasonal differences in the overall yield variability in the same field – some years large, others small enough for site-specific management to be judged as unwarranted. Therefore, it is necessary to assess the crop variation in each field every year.

Multivariate clustering of yield and EMI identified management zones that coincided with differences in soil nutrient levels and yield differences in the following season thus forming a logical basis for targeted nutrient sampling. In the case study, this reduced the number of soil samples from 27 to four and in practice, the targeting keeps the cost of soil sampling for site-specific management of crops to a minimum. The K means multivariate clustering method enables soil and crop variation to be used together to define management zones and the user can control the number of management zones identified and thus reconcile theoretical benefits with practical field management considerations.

Acknowledgements

The authors would like to thank the sponsors of this work, Home-Grown Cereals Authority, Hydro Agri and AGCO Ltd. for their support, and the contributions made by their collaborators, Arable Research Centres and Shuttleworth Farms. We would also like to thank Dr. David Pullen and Dr. Nicola Cosser for their assistance in developing the research programme and Robert Walker for implementing treatments and harvesting the experiments. Thanks must also be extended to Messrs Dines, Hart, Wilson and Welti who graciously allowed us to use their fields and gave us much support.

References

Avery B W (1987). Soil Survey Methods: A Review. Soil Survey Technical Monograph No 18, Soil Survey and Land Research Centre, Cranfield University, Silsoe, UK

Blackmore S; Moore M (1999). Remedial correction of yield map data. Precision Agriculture, 1, pp. 53-66, 1999

Birrell S J; Sudduth K A; Kitchen N R (1996) Nutrient mapping implications of short-range variability, In Precision Agriculture, Proceedings, 3rd International Conference, Eds. P. C. Robert, R. H. Rust and W. E. Larson, ASA/CSSA/SSSA pp207-216

Brenk; Pasda G; Zerulla W (1999). Nutrient mapping of soils – a suitable basis for site-specific fertilisation? In Precision Agriculture '99, (J. Stafford ed.), Society of Chemical Industry, pp 49-59

Cupitt J; Whelan B M (2001). Determining potential within-field crop management zones. In ECPA 2001, Third European Conference on Precision Agriculture, (G. Grenier and S. Blackmore eds.), Ecole Nationale Superieure Agronomique de Montpellier, Montpellier, pp7-12

Chalmers; Sinclair A H; Carver M (1999). Nutrients other than nitrogen, phosphorous and potassium (NPK) for cereals. Home-Grown Cereals Authority Research Review, No. 41, London N1 9NG, UK

Congalton R G; A R; Mead (1983). A quantitative method to test for consistency and correctness in photointerpretation. Photogrammetric Engineering and Remote Sensing, 49, 69-74

Earl R; Taylor J C; Wood G A; Bradley R I; Waine T; Welsh J P; Knight S M; Godwin R J (2001) Soil factors and their influence on within-field crop variability I: Field observation of soil variation. Submitted to Biosystems Engineering

Hudson W D; Ramm C W (1987). Correct formulation of the Kappa Koefficient of agreement, Photogrammetric Engineering and Remote Sensing, 53, 421-422

Isaaks E H; Srivastava R M (1989). Applied Geostatistics, Oxford University Press, New York, pp561

James T; Waine T W; Bradley R I; Godwin R J; Taylor J C (2000). A comparison between traditional methods and scanning for determining soil textural boundaries. Paper No. 00-PA-014, AgEng2000, Warwick, UK

Jaszberenyi; Loch J; Tamas J (1999). Evaluation of sampling patterns using geostatistical methods to develop fertilisation practice" In Precision Agriculture '99, (J. Stafford ed) Society of Chemical Industry, pp91-100

Lark E M; Bolan H C; Mayr T; Bradley R I; Burton R G O; Dampney P M R (1998). The development of cost-effective methods for analysing soil information to define crop management zones. Home-Grown Cereals Authority Project Report, No. 171, London N1 9NG, UK

Lesch S M; Strauss D J; Rhoades J D (1995) Spatial prediction of soil salinity using electromagnetic induction techniques 2. An efficient spatial sampling algorithm suitable for multiple linear regression model identification and estimation", Water Resources Research, Vol. **31**(2), pp387-398

MAFF (2000). Fertiliser recommendations for agricultural and horticultural crops (RB209), 7th ed., Stationary Office, London, UK

Moore M R (1997). An investigation into the accuracy of yield maps and their subsequent use in crop management. Unpublished PhD Thesis, Cranfield University at Silsoe

McBratney B; Whelan B M; Walvoort D J J; Minasny B (1999). A purposive sampling scheme for precision agriculture. In Precision Agriculture ’99, (J. Stafford ed.), Society of Chemical Industry, pp101-110

Oliver M A; Frogbrook Z; Webster R; Dawson C J (1997) A rational strategy for determining the number of cores for bulked sampling of soil”, In Precision Agriculture ’97, (J. Stafford ed.), Bios, pp155-162

Pocknee B; Boydell C; Green H M; Waters D J; Kvien C K (1996). Directed soil sampling”, In Precision Agriculture, Proceedings, 3rd International Conference, Eds. P. C. Robert, R. H. Rust and W. E. Larson, ASA/CSSA/SSSA pp159-168

Richards J A; Jia X (1999). Remote sensing digital analysis. 3rd edition, Springer, Berlin, pp363

Taylor J C; Thomas G; Wood G A (1997). Diagnosing sources of within-field variation with remote sensing. In Precision Agriculture ’97, (J. Stafford ed.), Bios, pp705-712.

G. Thomas, J. C. Taylor, T. Musthill and J. C. Fuller (1999). The development of remote sensing based products in support of precision farming. In Precision Agriculture ’99, (J. Stafford ed.), Society of Chemical Industry, pp191-200

Waine T W; Blackmore B S; Godwin R J (2000). Mapping available water content and estimating soil textural class using electro-magnetic induction. Paper No. 00-SW-044, AgEng2000, Warwick, UK

Webster R; McBratney A B (1987). Mapping soil fertility at Broom’s barn by simple kriging”, Journal of the Science of Food and Agriculture, Vol. 38, pp97-115

Welsh J P; Wood G A; Godwin R J; Taylor J C; Earl R; Blackmore B S (2001a). Developing strategies for spatially variable nitrogen application in cereals I: Winter Barley. Submitted to Biosystems Engineering

Welsh J P; Wood G A; Godwin R J; Taylor J C; Earl R; Blackmore B S (2001b). "Developing strategies for spatially variable nitrogen application in cereals II: Wheat", Submitted to Biosystems Engineering

Wood G A; Welsh J P; Taylor J C; Godwin R J; Knight S (2001). Real Time Measures of Canopy Size as a Basis for Spatially Varying Nitrogen at Different Seed Rates in Winter Wheat. Submitted to Biosystems Engineering

Figures for:

Soil Factors and their Influence on Within-Field Crop Variability II: Spatial Analysis and Determination of Management Zones

J.C. Taylor[1]; G.A. Wood[1]; R. Earl[1]; R.J. Godwin[1]

Table 1.
Summary of soil series in Short Lane, Trent Field and Twelve Acres

Field name and location	*Dominant soil series*	*Top soil characteristics*	*Sub soil characteristics*	*Soil water regime*
Short Lane Gamlingay Cambridgeshire, UK	Wickham.	Slowly permeable clay loam.	Stoneless clay.	Seasonally waterlogged.
	Ludford.	Deep well drained clay loam.	Sandy clay.	Well drained.
	Maplestead.	Sandy loam.	Loamy sand.	Well drained.
Trent Field Westover, Hampshire UK	Andover.	Shallow calcareous silty clay loam.	Brown silty clay loam over chalk.	Well drained, sub soil rarely wet.
	Panholes.	Brown stony silty clay loam.	Brown silty clay loam over shattered chalk.	
	Disturbed area.	Greyish brown silty clay loam.	Soil of mixed origin.	
Twelve Acres Hatherop Gloucester UK	Sherbourne.	Calcareous silty clay loam.	Brashy oolitic limestone.	Well drained.
	Moreton.	Calcareous silty clay loam.	Yellowish brown silty clay loam.	Well drained.
	Haselor	Calcareous silty clay with small platy limestone	Brown clay with abundant medium platy limestone	Well drained
	Didmarton.	Deep silty clay with small limestone	Deep stoneless silty clay.	Moderately well drained colluvium.

Table 2.

Summary of soil and crop data sets used in Short Lane, Trent Field and Twelve Acres

Field name and location	*Yield-map data and crop*	*Mean yield, t/ha*	*Standard deviation, t/ha*	*EMI*	*Soil minicore survey*	*Soil nutrient survey*
Short Lane Gamlingay Cambridgeshire, UK	1994 Winter Barley 1995 Winter Barley 1996 Winter Barley	5.7 6.8 5.9	0.65 0.63 0.76	Mar 1998	August 1998	Not available
Trent Field Westover, Hampshire UK	1995 Winter Barley 1996 Winter Barley 1997 Winter Barley	5.7 7.2 4.2	0.47 0.85 0.70	Feb 1999	August 1998?	4 Dec 1996 30 April 1997
Twelve Acres Hatherop Gloucester UK	1995 Winter Wheat 1996 Winter Wheat	4.8 8.4	0.66 0.98	Feb 1999	August 1998?	14 January 1997 25 April 1997

Table 3.
Correspondence between EMI and soil series classification in Trent Field

Series	*Cluster (assigned Series)*			
	1 (An)	*2 (pH)*	*3 (Dist)*	*Total*
Dist	0	1	6	7
An	29	20	5	54
pH	4	18	0	22
Total	33	39	11	83

Table 4.
Correspondence between EMI and soil series classification in Short Lane

Series	*Cluster (assigned Series)*			
	1 (MM)	*2 (Lu)*	*3 (Wh)*	*Total*
Lu	1	6	3	10
Wh	1	24	39	64
MM	51	19	4	74
Total	53	49	46	148

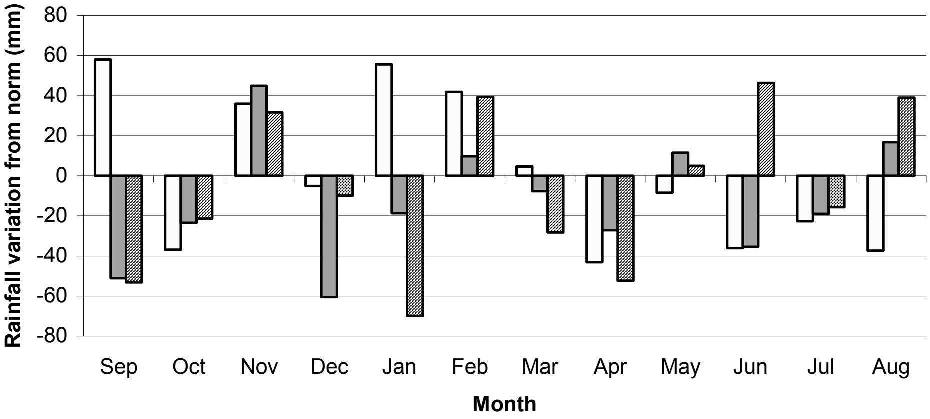
Table 5.
Correspondence between EMI and soil series classification in Twelve Acres

Series	*Cluster (assigned Series)*				
	1 (Mor)	*2 (Hb)*	*3 (Si)*	*4 (dD)*	*Total*
Si	4	9	12	3	28
Mor	8	8	2	0	18
Hb	1	2	0	0	3
dD	3	1	0	0	4
Total	16	20	14	3	53

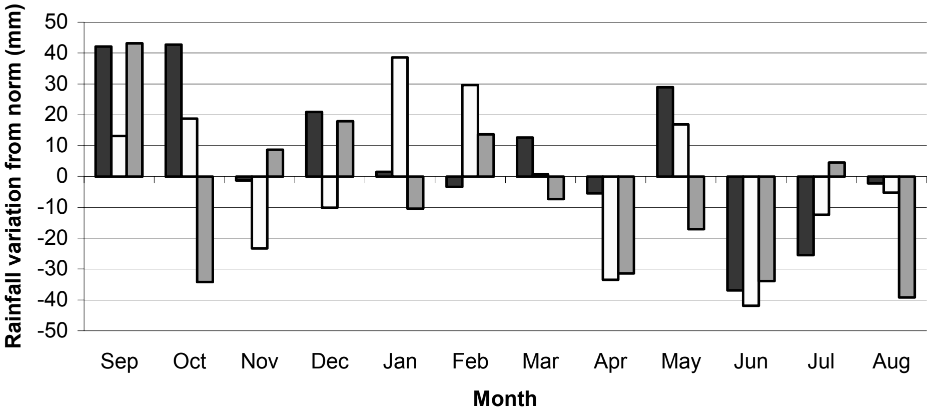
Table 6.
Agreement between EMI classification and Soil Series and statistical significance as measured using the Kappa statistic

Field	*Overall agreement %*	*Kappa %*	*Variance of Kappa*	*p*
Trent Field	64	40.32	.00690	3.44
Twelve Acres	42	20.13	.00999	1.42
Short Lane	65	47.10	.00249	6.67

a) Trent Field



b) Short Lane



c) Twelve Acres

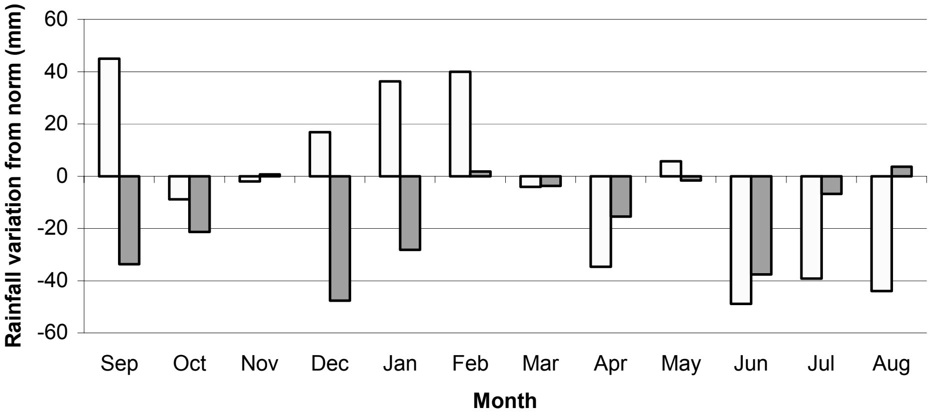
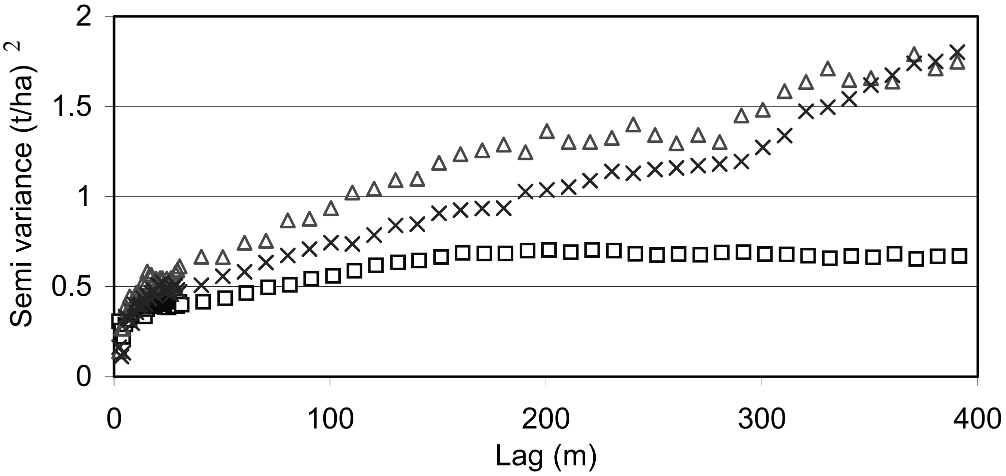
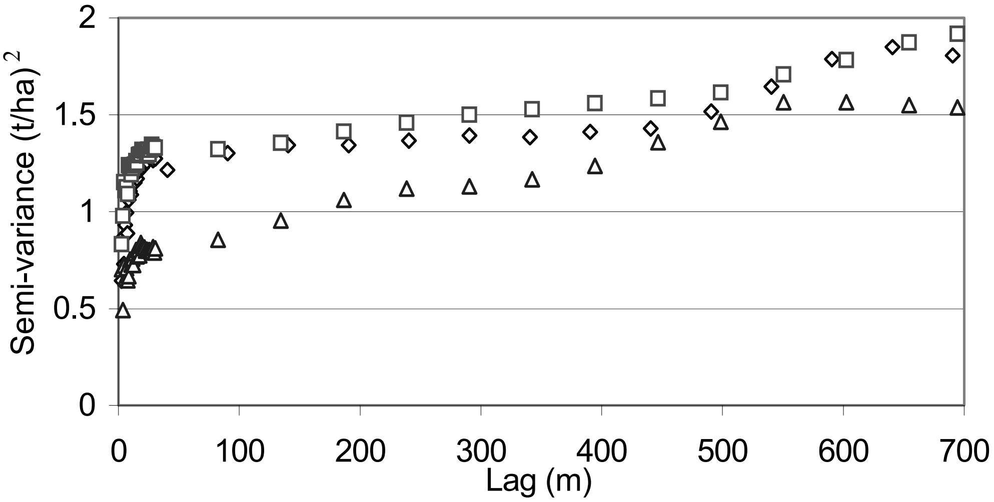


Fig. 1. Monthly deviation of rainfall from the 1989-1998 averages:
■*, 1994;*□*, 1995;* ▩*, 1996;* ▨*, 1997*

a) Trent Field



b) Short Lane



c) Twelve Acres

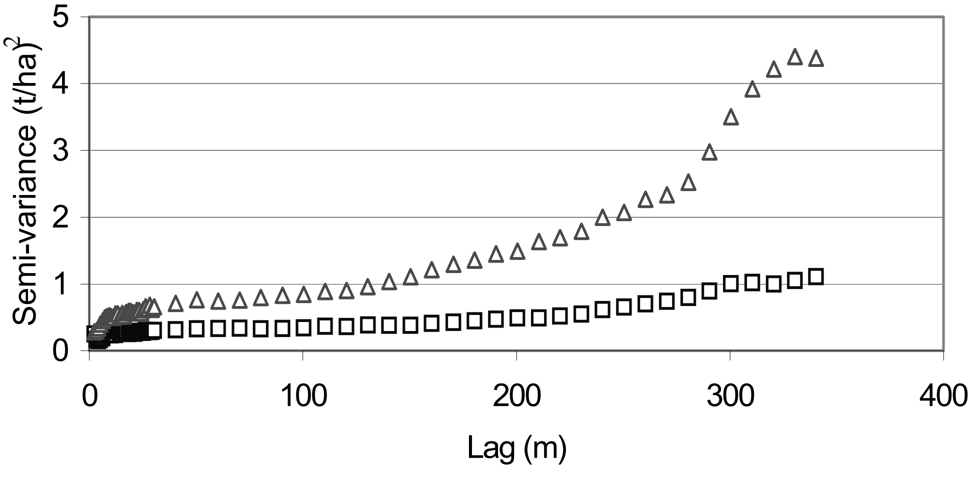


Fig. 2. Yield variograms in a) Trent Field, b) Short Lane and c) Twelve Acres: ◇, *1994*; □, *1995*; △, *1996*; ×, *1997*

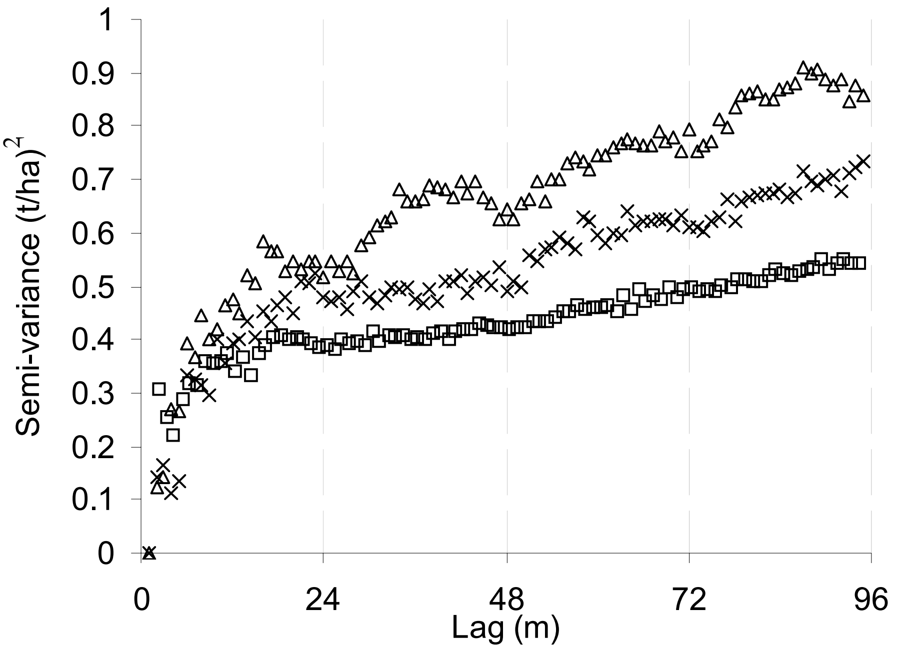


Fig. 3. Variograms of yield in Trent Field showing 'holes' in 1996 at lags equivalent to multiples of the tramline spacing of 24m: □, *1995;* △, *1996;* ×, *1997*

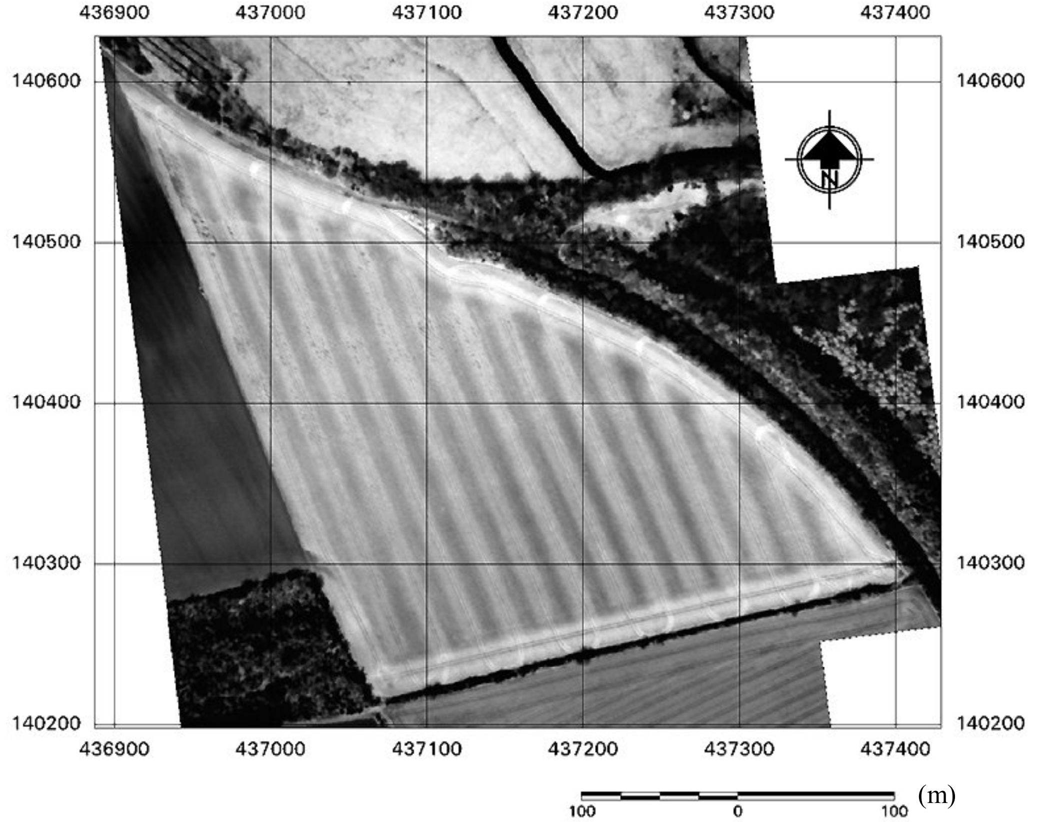


Fig. 4. Near infrared image-map of Trent Field from 1996 showing pronounced systematic variation of crop development (striping) related to the tramlines and caused by poor fertilizer distribution

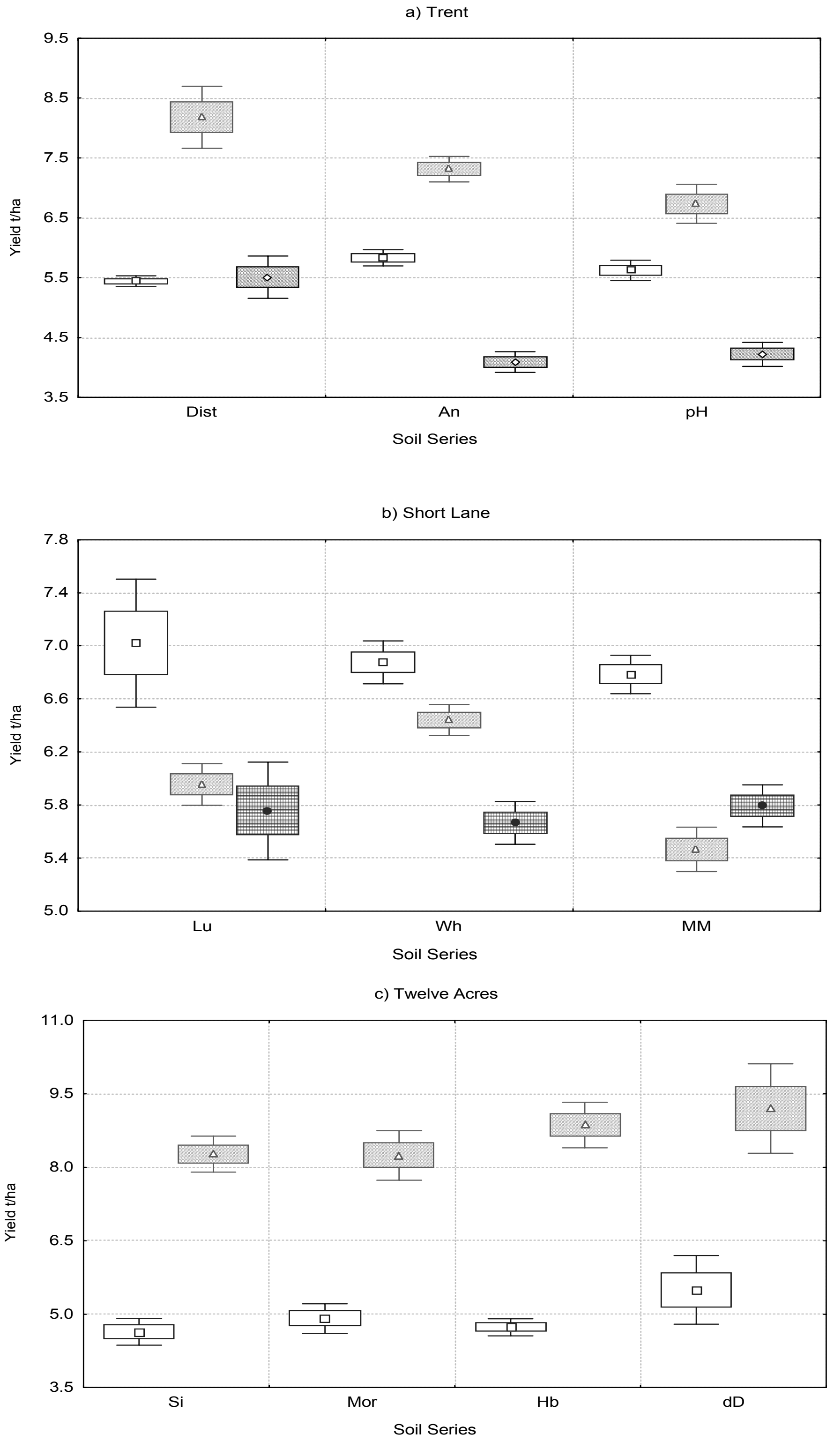
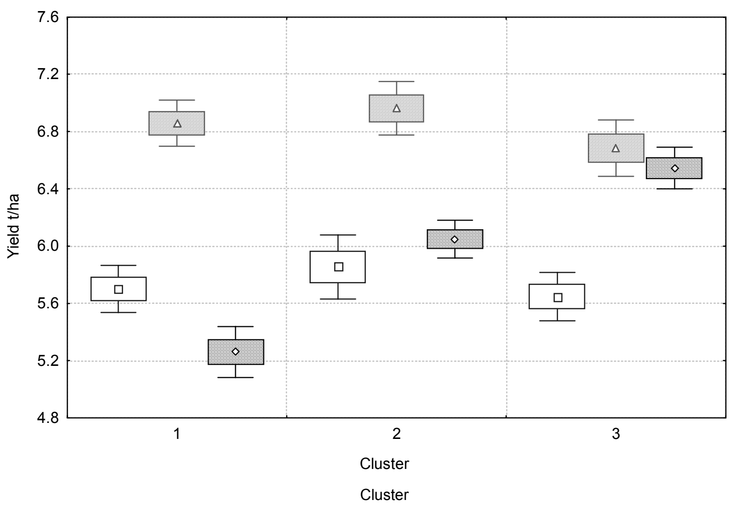


Fig. 5. Box (mean ± 1 std error) and whisker (mean ± 2 std error) plots of the yields in each soil series in a) Trent Field, b) Short Lane and c) Twelve Acres in each season: ●, 1994; □, 1995; △, 1996; ◇, 1997

b) Short Lane



c) Twelve Acres

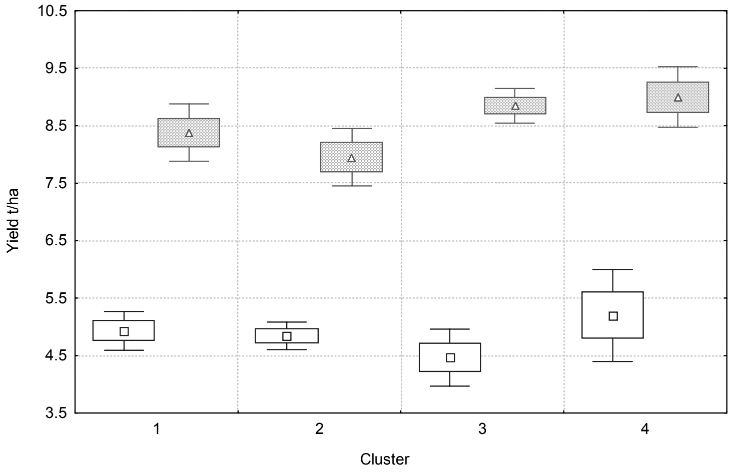


Fig. 6. Box (mean ± 1 standard. error) and whisker (mean ± 2 standard error) plots of the yields in each cluster in a) Trent Field, b) Short Lane and c) Twelve Acres in each season: ●, *1994*; □, *1995*; △, *1996*; ◇, *1997*

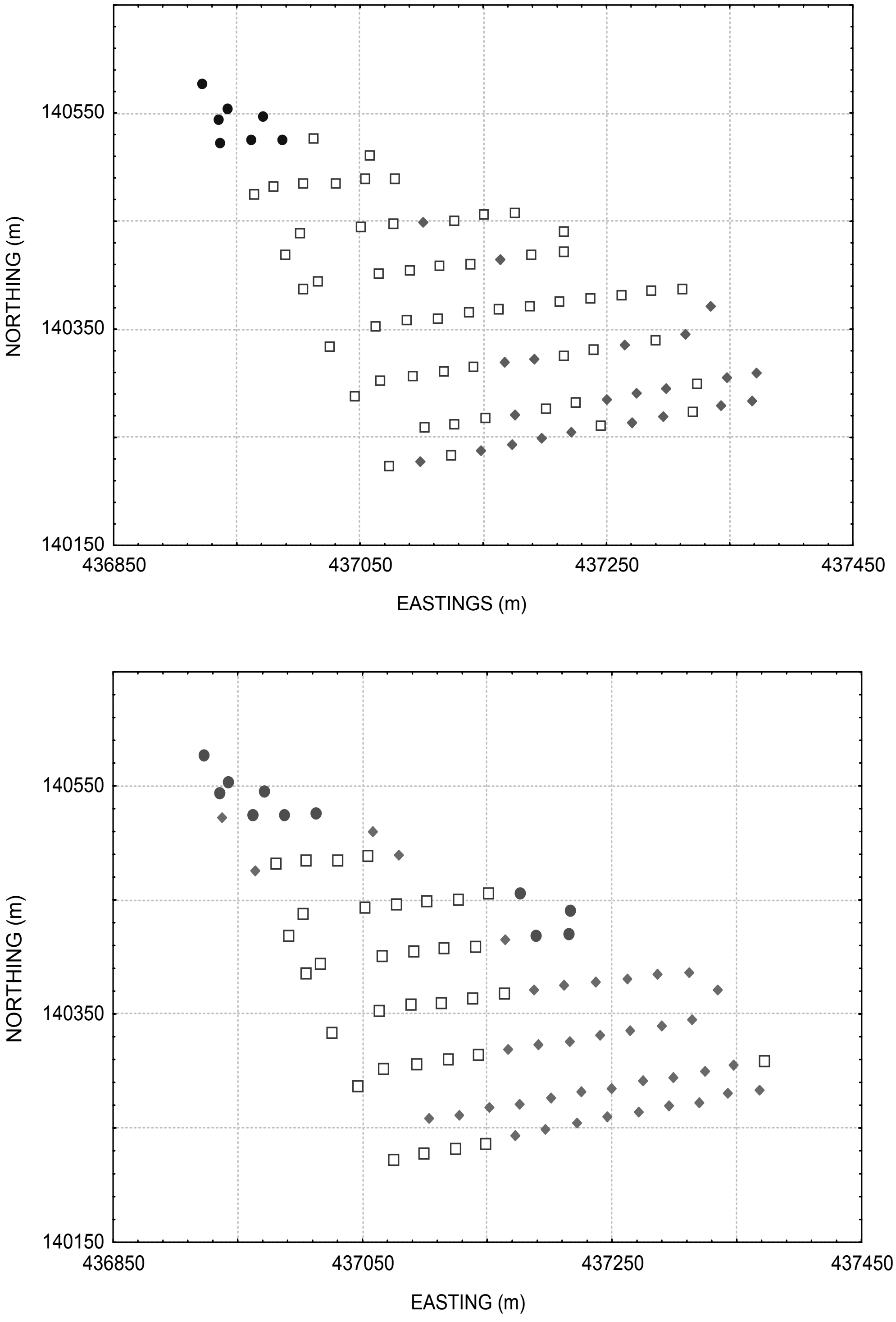
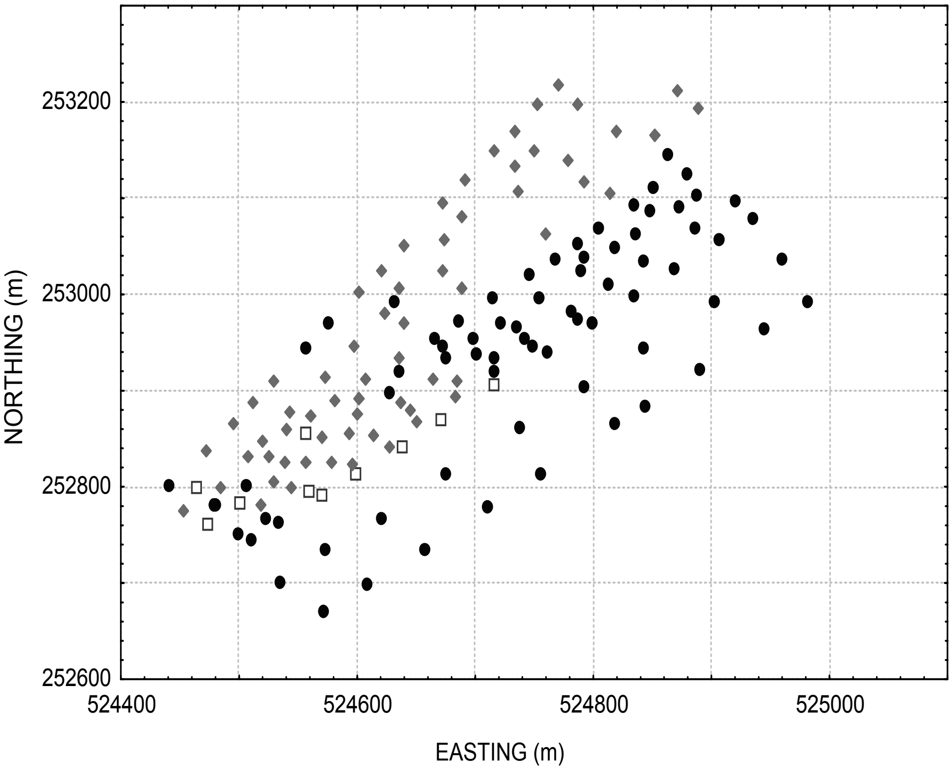


Fig. 7. Classification of soil minicore locations by soil series (above) and by cluster (below) in Trent field: ●, *Dist*; □, *An*; ◆, *pH;*
□, *Cluster 1* ◆, *Cluster 2;* ●, *Cluster 3*



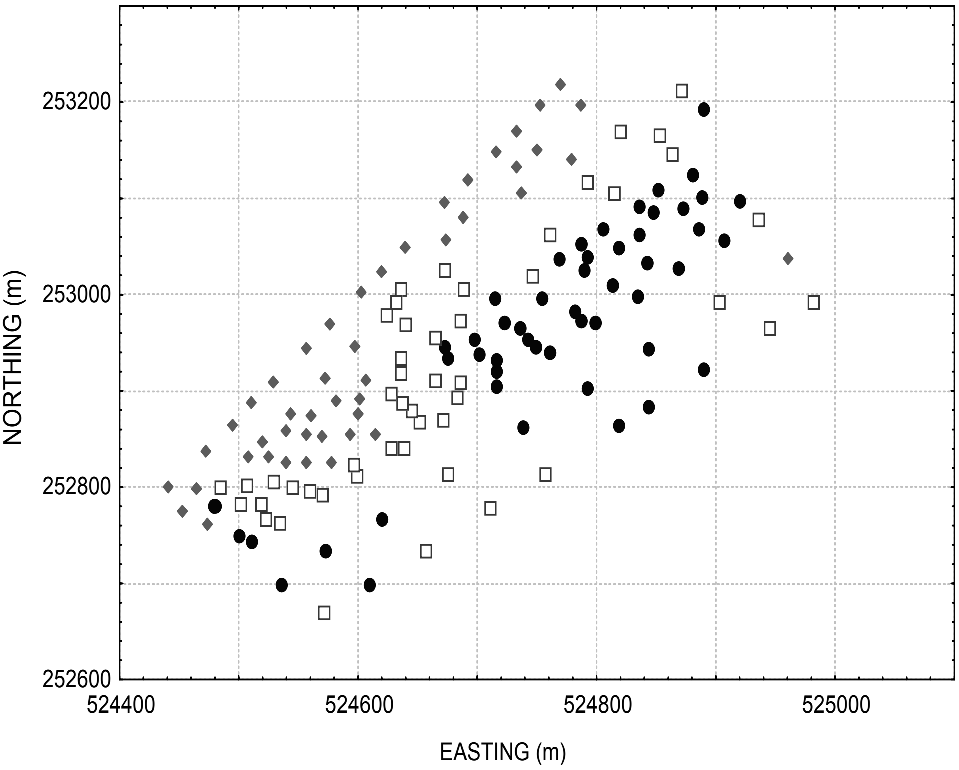
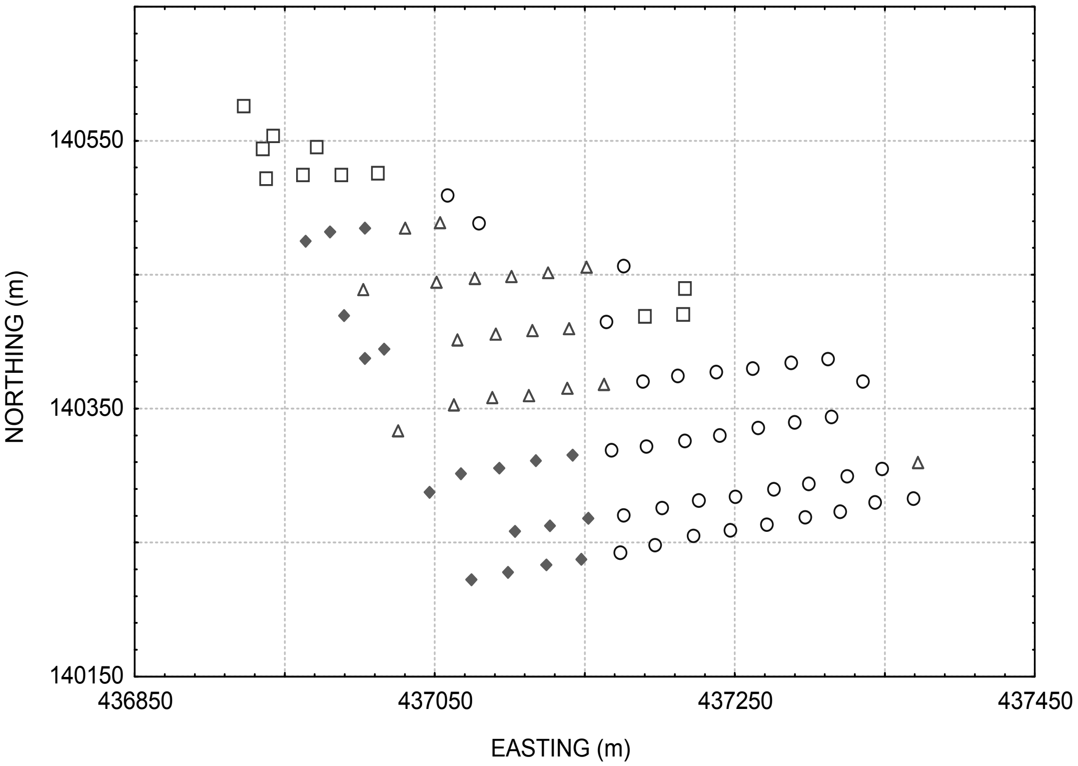


Fig. 8. Classification of soil minicore locations by soil series (above) and by cluster (below) in Short Lane:
□, Lu; ◆, Wh; ●, MM; ●, Cluster 1; □, Cluster 2; ◆, Cluster 3



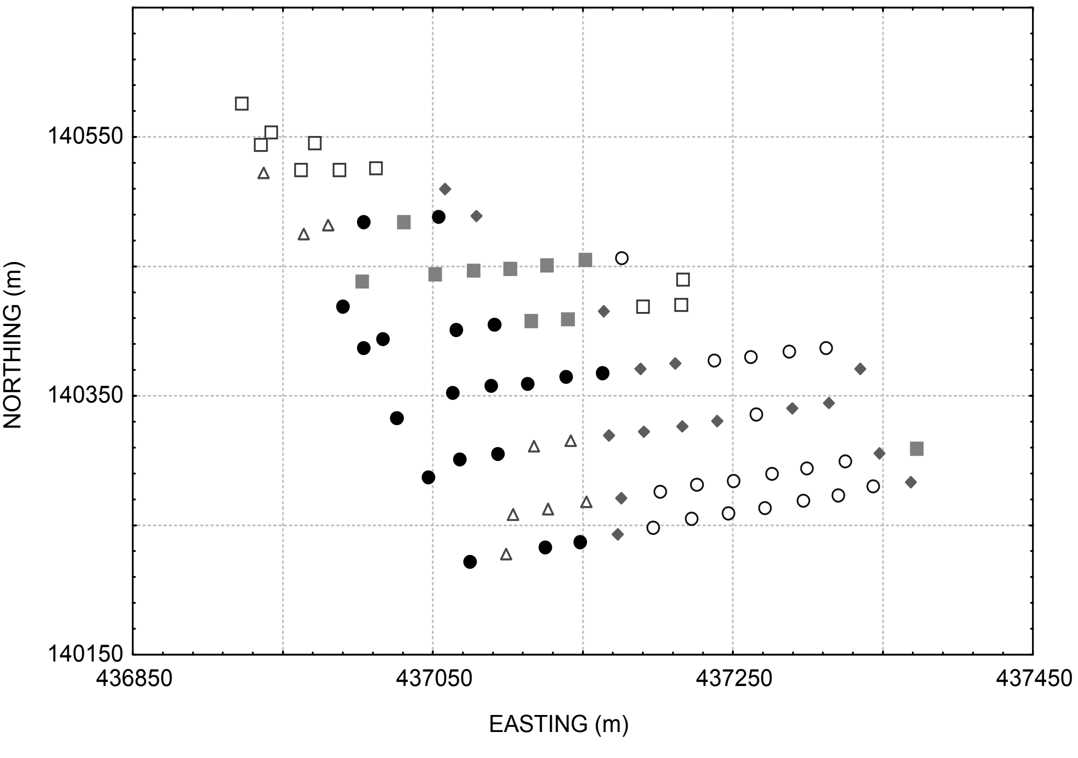


Fig. 9. Maps of Trent Field showing increased fragmentation of management zones with cluster number:
○*, Cluster 1;* □*, Cluster 2;* ◆*, Cluster 3;* △*, Cluster 4;* ●*, Cluster 5;* ■*, Cluster 6*

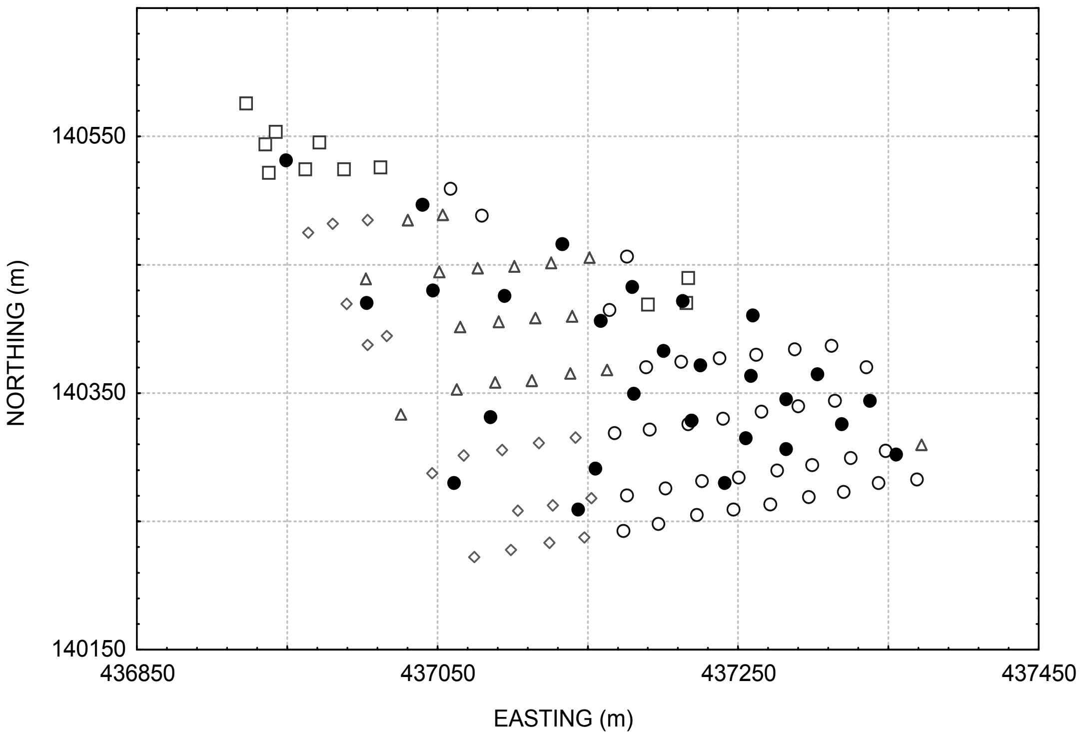


Fig. 10. Location of soil nutrient sampling points with respect to management zones determined by clustering:
○, *Cluster 1;* □, *Cluster 2;* ◆, *Cluster 3;* △, *Cluster 4;* ●, *Nutrient samples*

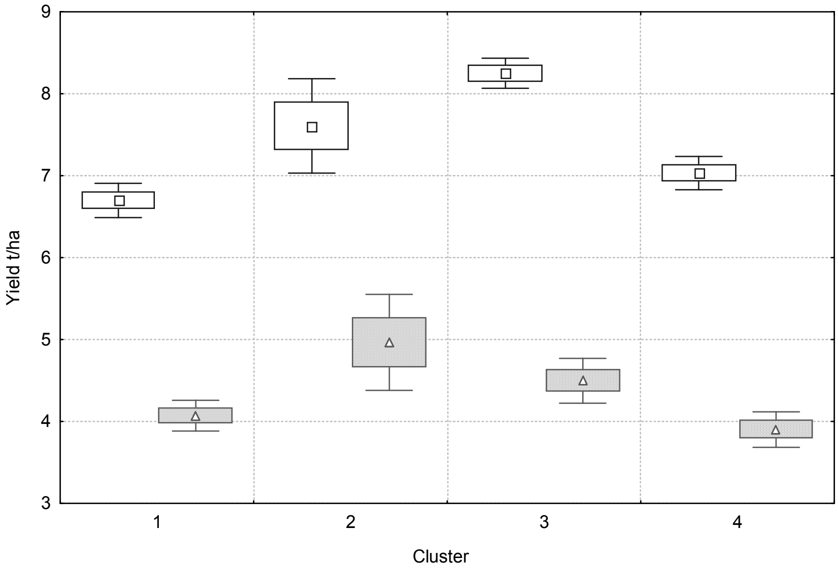


Fig. 11. Yield variation in management zones in 1996 and 1997: □, 1996; △, 1997

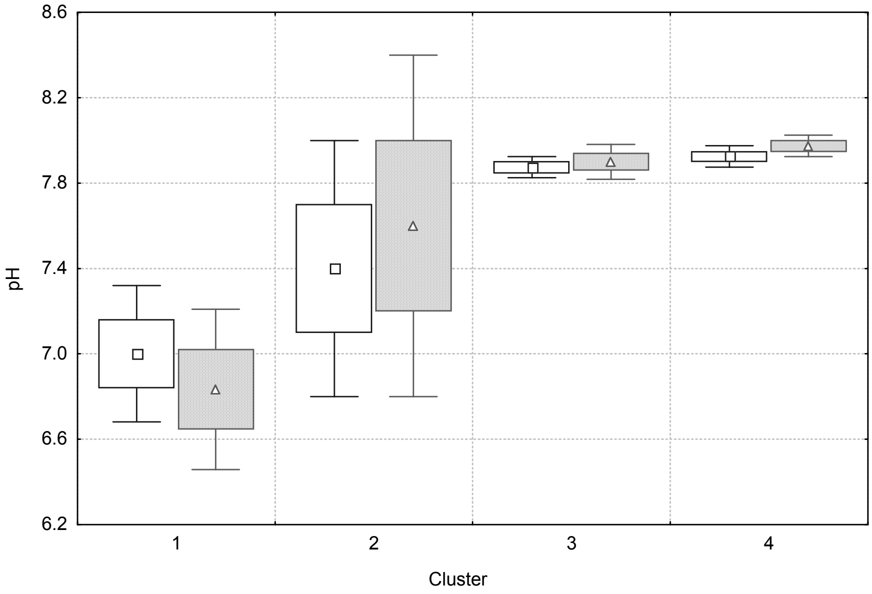
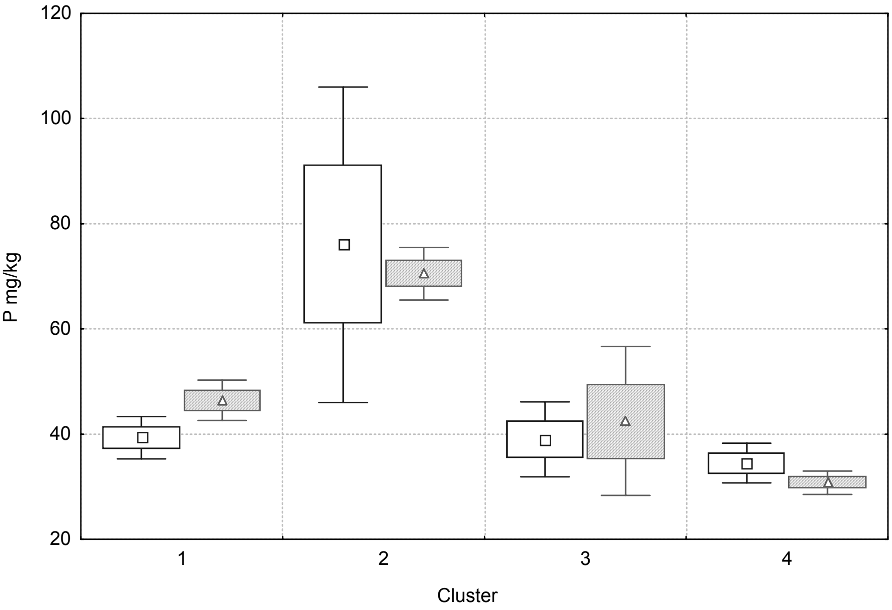


Fig. 12. Variation of pH between Trent management zones:
□, Autumn/Winter; △, Spring



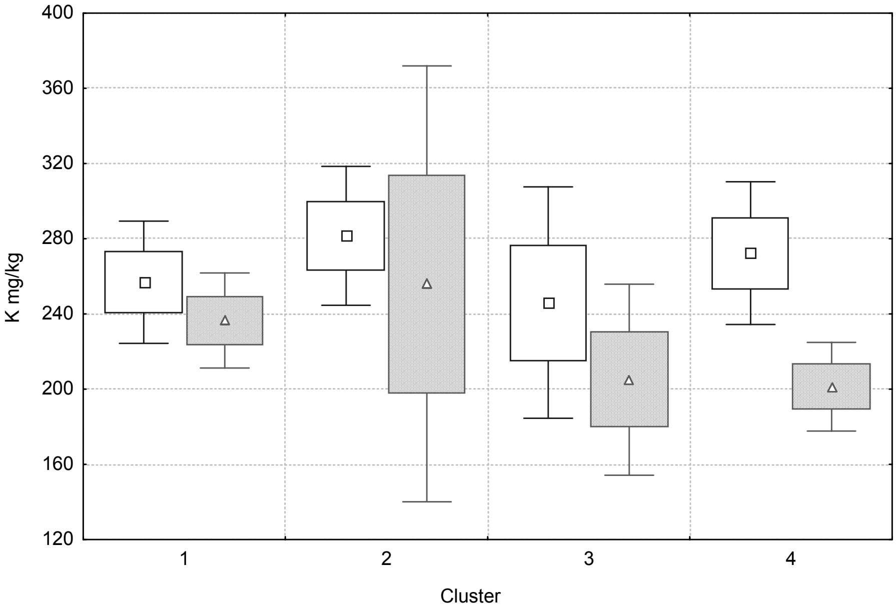


Fig. 13. Variations in P (above) and K (below) in Trent Field management zones:
□, *Autumn/Winter;* △, *Spring*