Chapter 7: Predicting Digital Terroirs to aid Vineyard Design

7.1 Introduction

As discussed in the Prologue, digital terroirs are really management zones. Precision Agriculture in general is currently operating under a management zone strategy rather than as a continuous site-specific management strategy (re Figure 1.2). Management zones are generally derived using a cluster algorithm which may utilise a hard k-means or a fuzzy k-means algorithm. The continuous nature of many environmental variables, particularly soil attributes, make these approaches preferable as they produce an estimation of class association. In the case of hard k-means this is a probability while fuzzy k-means produce partial class memberships (Boydell and McBratney, 1999). Clustering algorithms have been applied to a variety of variables to delineate areas including digital terrain model (DTM) attributes (Jones *et al.*, 1989, Sudduth *et al.*, 1997), yield data and ECa (Cupitt and Whelan, 2001), yield and soil data (Shatar and McBratney, 2001), satellite imagery (Stewart and McBratney, 2001), EC_a data (Fleming and Buchleiter, 2002), aerial imagery and farmer knowledge (Fleming and Buchleiter, 2002) and EC_a, aerial and satellite imagery and yield (Franzen and Nanna, 2002).

The previous two chapters have discussed methods of transforming data collected at a vineyard location onto a common raster. By aggregating or disaggregating data onto a common raster the information can be simultaneously analysed and some inferences about potential "digital terroirs" made. Aggregation (or disaggregation) of data presents several considerations for data analysis (see McBratney, 1998 for a more detailed discussion). One of the main considerations is what the appropriate scale of data, raw or aggregated/disaggregated, for the analysis. Modelling prior to aggregation may produce a different result to modelling after aggregation (Van Beurden and Douven, 1999). Modelling followed by aggregation appears the more sensible option as it allows analysis at the appropriate scale for the data. However environmental modelling, which draws on several baseline data sources, is usually done after aggregation/disaggregation (Van Beurden and Douven, 1999). This ensures all data are at the same scale and produces faster results. McBratney (1998) suggests a third approach whereby the modelling process also scales the output. This allows for the input of raw data and minimises the propagation of errors in the initial aggregation/disaggregation of data.

The use of clustering algorithms is fairly straightforward and has been widely adopted for management zone delineation in Precision Agriculture. The choice of inputs is not as simple and a wide range of parameters have been used (as illustrated above). Generally a crop response layer e.g. yield or canopy reflectance is used as the focus in most cropping industries is on plant production. However for a new vineyard design this data is usually unavailable. It is possible that previous cropping history at the site may be used however the value of this information is uncertain and is only now starting to be assessed in whole farm studies (Robert Bramley *pers. comm.*). In the past few years anecdotal accounts have been reported regarding the use of pasture imagery, in both New Zealand and Australia, as a surrogate indicator of soil variability (David Lamb, The University of New England, *pers. comm.*)While no crop response is available the interpolated outputs from soil and ancillary surveys provide a wealth of data that can be utilised. The major problem is which data layers to choose? It is likely that the optimal layers for vineyard design will vary from site to site and some expert knowledge may be useful in determining the cluster analysis input. Generally clustering is done with as many variables as are available. However variables are not weighted and co-linearity between attributes may bias the analysis. It is important that singular but independent data sources are not penalised by correlation between other properties.

An alternative option to clustering large numbers of variables to convert the data into an index prior to clustering. A regional site index for viticultural has been recently proposed (Tesic *et al.*, 2002) and the option for a local index to help in vineyard design is explored here.

7.1.1 Of Zones and Classes

The terms "management zone" and "management class" are frequently used in PA and often as interchangeable terms. However these terms are not identical and are defined here to clarify their meaning. A management class is defined as the area to which a particular treatment may be applied. A management zone is defined as a spatially contiguous area to which a particular treatment may be applied. Thus a management class may consist of numerous zones however a management zone can only contain one management class. A digital terroir would be defined as a management class.

The aim of this chapter is to investigate different approaches for determining the optimum number of digital terroirs at a site using a range of data inputs. The validity of the digital terroir predictions will be assessed against crop response in the vineyard.

7.2 Methodologies

7.2.1 Data Manipulation

Two vineyards, Cowra and Canowindra (described previously) were used for this study. Ordinary kriged data from Chapter 6 (Clay0-90cm$_{OK}$, Clay0-30cm$_{OK}$, Sand0-90cm$_{OK}$, Sand0-30cm$_{OK}$ and RAW$_{OK}$, Rootzone Depth$_{OK}$ and Topsoil Depth$_{OK}$), regression kriged output from Chapter 7 (Clay0-90cm$_{RK}$, Clay0-30cm$_{RK}$, RAW$_{RK}$) as well as interpolated Veris data (0-30 cm 0-90 cm and 30-90 cm), multispectral aerial imagery (B, G, R, NIR) (Cowra only) and yield and standardised yield data from Chapter 4 were used for the analyses.

The multispectral imagery was converted into the following 4 vegetative indices (after Metternicht *et al.*, 2002)

Normalised Differences Vegetative Index (NDVI)	(NIR-R/NIR+R)
Plant Cell Ratio (PCD)	(NIR/G)
Plant Pigment Ratio (PPR)	(G/B)
Photosynthetic Vigour Ratio (PVR)	(G/R)

Table 7.1: Vegetative Indices used in Precision Agriculture

Climatic data was generated using the SRAD program (Wilson and Gallant, 1997). The interpolated DEM for the vineyard was used as an input into the program. Climatic averages were generated from a 37 year average of Bureau of Meteorology averages at Cowra Airport (http://www.bom.gov.au/climate/averages/tables/cw_065091.shtml) and a 117 year average at Canowindra (http://

www.bom.gov.au/climate/averages/tables/cw_065006.shtml). The input files for the SRAD program are given in Appendix 7.1. The mean temperature for October and January were derived as was the mean daily and summed net radiation (Watts/m^2) for the period from 1/9 to 31/3.

All data, including Eastings and Northings, was collated into a single spreadsheet. The vineyards were classified separately according to blocks and varieties. Most vineyards have multiple varieties of grapes planted. Differences between varieties means that it is difficult to relate crop response across varieties. For this reason the analysis and discussion will focus on discrete varietal areas within the vineyards. For Cowra two areas were used - Chardonnay (blocks 11, 12, 21 and 25) and Shiraz (blocks 5-7. For Canowindra two areas were also used, Semillon (blocks 17,27 and 28) and Cabernet (blocks 16 ,22, 23 and 34). Maps of the varieties present in the Cowra and Canowindra vineyards are given in Figure 7.1.

The 30-90 cm silt fraction was determined by OK and RK as described in Chapters 5 and 6 respectively.

Vine phenology and architecture are important in determining the response of remote sensing (Dobrowski, *et al.*, 2002, Johnson *et al.*, 1996). The Cowra and Canowindra vineyards have been established on Australian Sprawl with machine pruning followed by hand pruning. Australian Sprawl produces large canopies, with the aim of shading the fruit from the summer sun. The large canopies extend well into the interrow which minimises the error from background noise in the remotely sensed imagery. This is a major problem in vertical-shoot positioned vineyards (Dobrowski *et al.*, 2002)

7.2.2 *A Local Site Index for Viticulture*

When determining vineyard design it may be useful to have a simple indication of the potential of a site for plant response. A regional Site Index (SI) for viticulture has been proposed by Tesic *et al.* (2002) to characterise the potential of different sites for viticulture. In this regard it is a broad index for the identification of a vineyard site rather than for vineyard design. The SI was originally designed in NZ however it has been applied successfully to the characterisation of viticultural conditions in Australian (Tesic, 2003). While the SI was originally designed for the comparison of different wine regions or subregions it has been altered here to a local site index (LSI) for application at a vineyard scale. The original index as proposed by Tesic *et al.* (2002) is

$$SI = \frac{(t_o + t_j)^2 \bullet (1 + \frac{G_p}{100})}{R_S \bullet \sqrt{1 + CS} \bullet RD} \qquad \text{Equation 7.1}$$

where t_o = mean air temperature in October (°C)
t_j = mean air temperature January (°C)
G_p = gravel percentage in topsoil
R_S = seasonal rainfall (October to April in mm)

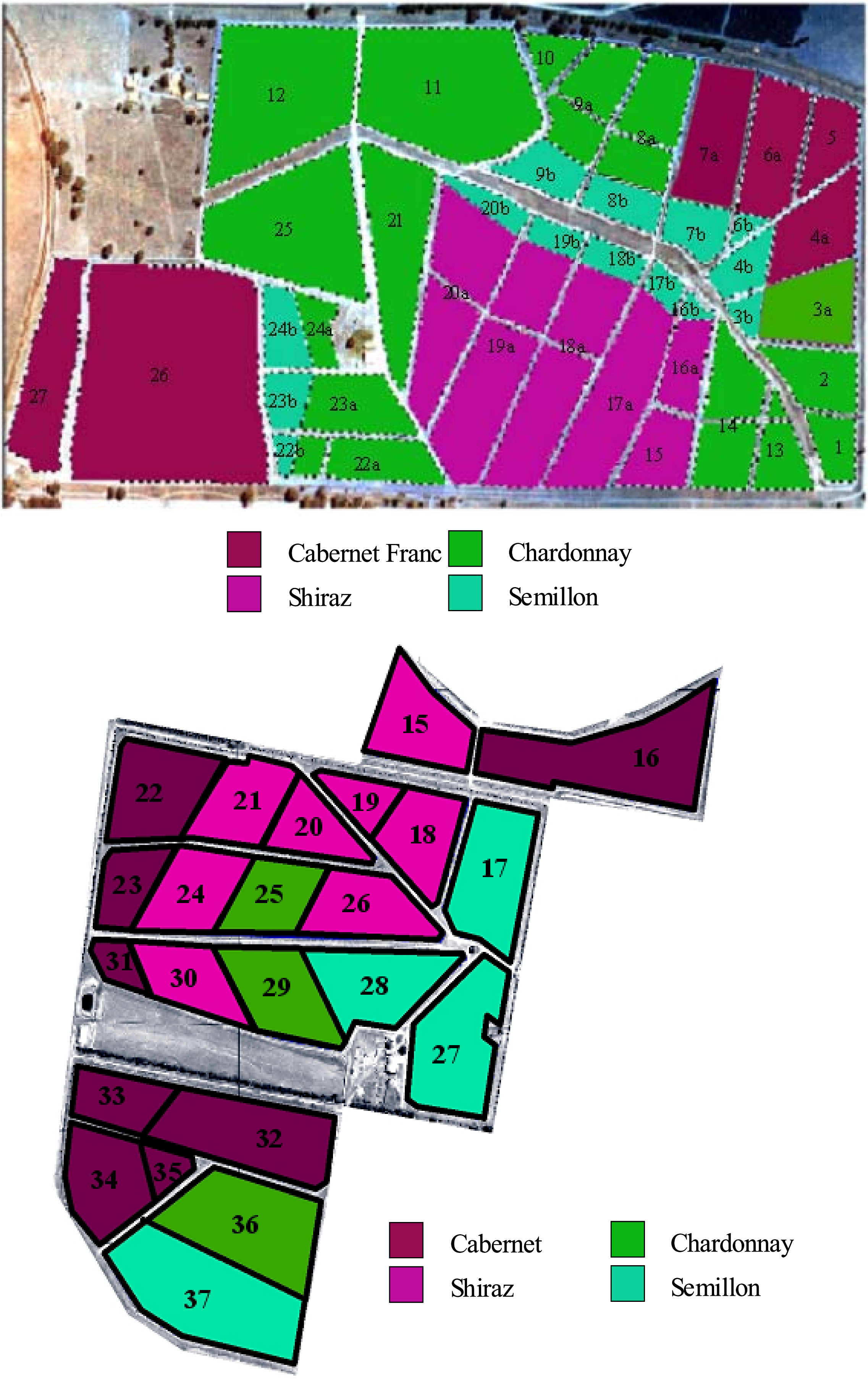


Figure 7.1: Vineyard varietal maps for Cowra (top) and Canowindra (bottom)

CS = ratio of clay percent to silt percent in the 35-70 cm zone
RD = maximum depth of rooting rone.

For this study knowledge of the gravel % in the topsoil was not recorded. The G_p factor was used by Tesic *et al.* (2002) as an indication of the thermal and moisture-holding conditions of the soil. For the Hawkes Bay region of NZ, where the SI was developed, gravel in the soil is generally considered beneficial for vine growth. This is due to its better thermal properties which permits roots, and the vine, to start metabolism earlier. A higher gravel content also improves moisture drainage in a cool and humid environment. In the majority of Australian viticultural regions gravel appears to play a smaller part in vine performance due to the warmer climate, lower rainfall and presence of irrigation. As a substitute a measure of net solar radiation at the site has been substituted into the equation. As well as incident solar radiation, this also indicates aspect and the opportunity for early morning radiation that provides faster canopy and soil warming when temperatures are lowest and most limiting (Wilson, 1998, Gladstones, 1992). The site-specific net radiation in the vineyard was obtained from the SRAD model (Wilson and Gallant, 1997). The net solar radiation is converted to a Relative Net Radiation (NR_{rel}) Value with a range of 0-1 i.e. the same scale as the G_p. The seasonal rainfall statistic, an estimate of moisture, has been replaced by the site-specific RAW_w estimate . The adjusted local SI equation (LSI) is

$$LSI = \frac{(t_o + t_j) \bullet (1 + NR_{rel})}{RAW_W \bullet \sqrt{1 + CS} \bullet RD} \qquad \text{Equation 7.2}$$

where NR_{rel} = Relative Net Radiation at site i $\left(\frac{NR_i - NR_{\min}}{NR_{\max} - NR_{\min}}\right)$

RAW_W = Readily Available Water calculated using the method of Wetherby (2000) and interpolated using OK or RK
RD = Rootzone Depth (interpolated using OK)
CS = 30-90cm Clay/30-90cm Silt and may be interpolated using OK or RK.

The LSI is based on known climatic and environmental variables that are influential in determining crop response. It has the advantage of producing a singular number that can be more easily analysed and interpreted by lay users.

A LSI was calculated using the OK data from Chapter 5 (LSI_{OK}). An alternative LSI was also calculated from the RK output from Chapter 6 (LSI_{RK}). Both of these were used in the analyses to determine if there was a benefit to using either the OK or RK data.

7.2.3 Cluster Analysis

Data was clustered in FuzMe (Minasny and McBratney, 2002) using a fuzzy k-means algorithm. This produced partial membership classes for each point and associated indices of performance (FPI and MPE). Clustering was performed on a variety of different combinations of data including the ordinary kriged data from Chapter 5, the regression kriged data from Chapter 6, ancillary data sets, LSI and combinations of all three. All the approaches were also tried with and without spatial coordinates (Eastings and Northings). A list of the different approaches and the input variables are given in Table 7.2. Only the 0-90 cm Veris measurement was used directly in this analysis (other measurements were used indirectly in the RK methodology). This was done to imitate the output from an EMI instrument which is the more common sensor used for vineyard EC_a suveys.

The spatial coordinates were included to try and introduce a usable spatial structure into the clustering process. For functionality vineyards still require a structure that is compatible with machinery use and management. The focus on spatial contiguity produces a significant difference in the output as the clusters are potentially zones not classes. When the spatial coordinates are not used management classes are not spatially constrained and may consist of multiple zones. It is hypothesised that the use of spatial coordinates will result in a higher optimal cluster number than the non-spatial approach, although the final number of zones may not necessarily be higher. For this reason the non-spatial approaches were analysed with 2-10 clusters and the spatial approaches with 2-15 clusters.

Approach	Variables
OKall	Clay0-90cm$_{OK}$, Clay0-30cm$_{OK}$, Sand0-90cm$_{OK}$, Sand0-30cm$_{OK}$ RAW$_{OK}$, Rootzone Depth$_{OK}$ and Topsoil Depth$_{OK}$
OKall_xy	Clay0-90cm$_{OK}$, Clay0-30cm$_{OK}$, Sand0-90cm$_{OK}$, Sand0-30cm$_{OK}$ and RAW$_{OK}$ Rootzone Depth$_{OK}$, Topsoil Depth$_{OK}$ and Eastings and Northings
LSI$_{OK}$	LSI$_{OK}$
LSI$_{OK}$_xy	LSI$_{OK}$, Eastings and Northings
OK1	RAW$_{OK}$ and Elevation
OK1_xy	RAW$_{OK}$, Elevation, Eastings and Northings
OK2	RAW$_{OK}$, Elevation, Clay0-90cm$_{OK}$ and Veris 0-90 cm
OK2_xy	RAW$_{OK}$, Elevation, Clay0-90cm$_{OK}$, Veris 0-90 cm, Eastings and Northings
RKall	Clay0-90cm$_{RK}$, Clay0-30cm$_{RK}$ and RAW$_{RK}$
RKall_xy	Clay0-90cm$_{RK}$, Clay0-30cm$_{RK}$, RAW$_{RK}$, Eastings and Northings
LSI$_{RK}$	LSI$_{RK}$
LSI$_{RK}$_xy	LSI$_{RK}$, Eastings and Northings
RK1	Clay0-90cm$_{RK}$, Clay0-30cm$_{RK}$, RAW$_{RK}$, Elevation, and Veris 0-90 cm
RK1_xy	Clay0-90cm$_{RK}$, Clay0-30cm$_{RK}$, RAW$_{RK}$, Elevation, Veris 0-90 cm, Eastings and Northings

Table 7.2: List of attributes used in each clustering approach

7.2.4 Determining the Optimal Number of Zones.

One of the biggest problems following cluster analysis of data is how to determine the optimal number of classes for management. The number of clusters chosen is often done subjectively however some statistical approaches have been recently proposed.

7.2.4.1 Using MPE and FPI

When a fuzzy k-means clustering algorithm is utilised, the fuzziness performance index (FPI) and modified partition entropy (MPE) (Roubens, 1982) can be calculated to assess the optimal number of classes. The FPI is a measure of the degree to which different classes share membership and is constrained to values between 0 and 1. As FPI approaches 1, membership sharing increases. As the FPI approaches 0, classes become more distinct and if FPI = 0 than classes are considered "hard". The MPE is an estimate of the amount of disorganization created by a specified number of classes and is also constrained between 0 and 1. The number of classes can be considered optimal when the FPI and MPE values are minimised (Boydell and McBratney, 1999, Fridgen *et al.*, 2000). A detailed explanation of these terms is available at on-line (http://www.usyd.edu.au/su/agric/acpa/fkme/FuzME_Theory.pdf).

7.2.4.2 Class/Zone Opportunity Indices.

While the MPE/FPI approach provides a statistical indication of effectiveness of class classification it does not indicate whether or not the classes are significantly different form each other in an agronomic sense. The determination of this significant difference is problematic and not well understood. A few approaches to this problem have been proposed in recent years. Fridgen *et al.* (2000) have suggested a statistic for the comparison of within-zone variance relative to total field variance (i.e. one management zone per field). They calculate the weighted variances for each zone as

$$S_Z^2 = \frac{1}{n_z}\sum_{i=1}^{n_z}(Y_i - m)^2 \bullet \frac{n_z}{n_T} \qquad \text{Equation 7.3}$$

where

S^2_z = Weighted variance for zone Z;
Y_i = Yield measured for cell i;
m = Mean of measured yield in zone Z;
n_z = number of cells in zone Z;
n_T = Total number of cells in the map.

and used the ratio of the sum of these variances to the total variance as an indication of the performance of the zonal model. This provides an indication of the variance explained by the zonal model but not if the zones/classes are statistically different.

Cupitt and Whelan (2001) have proposed a confidence interval that utilises the kriging variance for determining if two classes are statistically different from each. They argue that the most important point for class delineation is having sufficient variability between classes to permit class-specific management. They calculate the C.I. as

$$95\%C.I. = \mu \pm \left(\sqrt{\sigma^2{}_{krig}} \bullet 1.96\right) \qquad \text{Equation 7.4}$$

and for statistically different zones

$$\left|Y_{zone_x} - Y_{zone_w}\right| \geq \left(\sqrt{\sigma^2{}_{krig}} \bullet 1.96\right) \bullet 2 \qquad \text{Equation 7.5}$$

This appears a more valid statistic approach to determining if zones are relevant as it accounts for the auto-correlation between data points. However the objective for this statistic is to determine if clustering is producing a significant crop response. For whole vineyard design, crop response is currently difficult to use due to varietal and managerial influences. The individual environmental soil properties can be analysed for significant class differences however crop response is a holistic attribute and investigating the response of multivariate attribute, such as the LSI, may be more beneficial. A potential drawback to the approach of Cupitt and Whelan (2001) is that it is currently designed only for ordinary kriged data and is not applicable in its current form to raw data or other methods of interpolations (including RK).

An alternative approach is a Management Zone Opportunity Index (O_z), based on the Opportunity Index of Pringle *et al.,* (2003). The derivation and rationale for the O_z is presented in Appendix 7.3. This approach tests the hypothesis that the spatial structure of the zones is adequate for zonal management. The spatial structure of the zones (Sz) (after Appendix 7.3) is determined by:

$$Sz = r^2\left[\frac{A}{z}\right] + J_a\left[1 - r^2\right] \qquad \text{Equation 7.6}$$

where A = Area of field
z = number of zones
r^2 = fit of the cluster model to the yield data and
Ja = the integral scale (Chapter 4)

This statistic differs from that of Fridgen *et al.* (2000) in that it considers the variability that is accounted for by the clustering technique and also the residual variability that remains that may be manageable. However, Sz is a management zone model statistic and is not applicable to management classes unless they are subdivided into discrete zones.

7.2.5 Evaluation of the Clustering Process

The MPE and FPI output was plotted for each clustering approach and the mean of the two values calculated to determine when the two statistics have been minimised. Once the optimum number of zones had been determined an ANOVA was performed on a varietal basis to determine the amount of variation in the crop response that was explained by the clusters. Parts of the vineyard were then mapped to visually compare how the clustering had accounted for variability in crop response.

7.3 Results and Discussion

7.3.1 Vineyard Site Index

The LSI for both vineyards using both OK and RK data are given in Figure 7.3 and 7.4. The LSI has been mapped using quantiles. For Cowra both the LSI_{RK} and LSI_{RK} exhibit similar patterns which is expected given the similarity in RK and OK data observed in Chapter 6. At Canowindra the LSI_{RK} and LSI_{OK} also exhibit similar patterns with more detail and sharper boundaries shown in the LSI_{RK}. Both vineyards have blocks that show large differences in the LSI within the blocks

A noticable problem in the Cowra maps is the similar LSI values observed at different locations in the landscape. The heavy clay alluvium on the river terrace (site A) produces a similar LSI value to

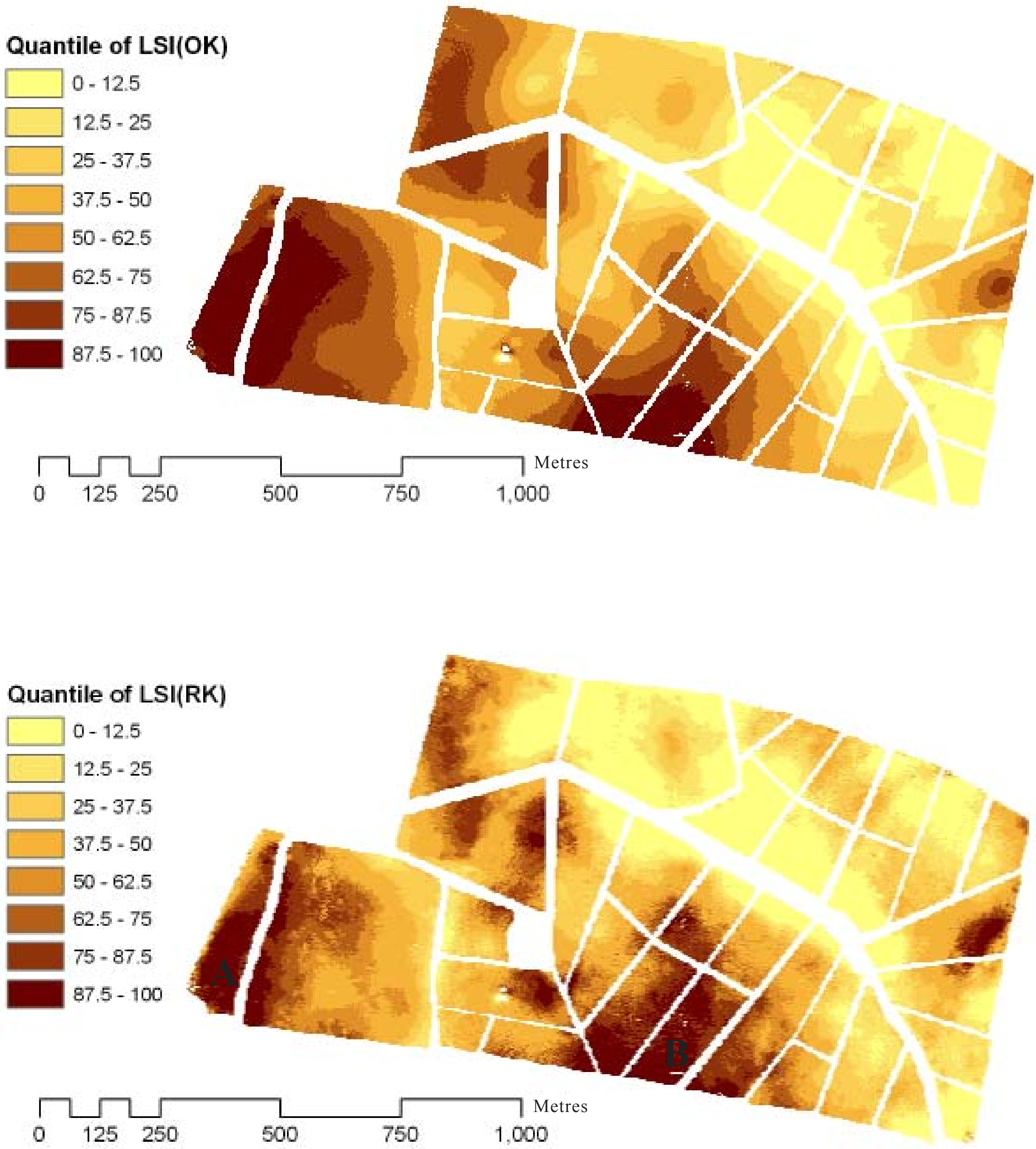


Figure 7.2: Maps of LSI_{OK} (top) and LSI_{RK} (bottom) for Cowra. (Maps are displayed as quantiles)

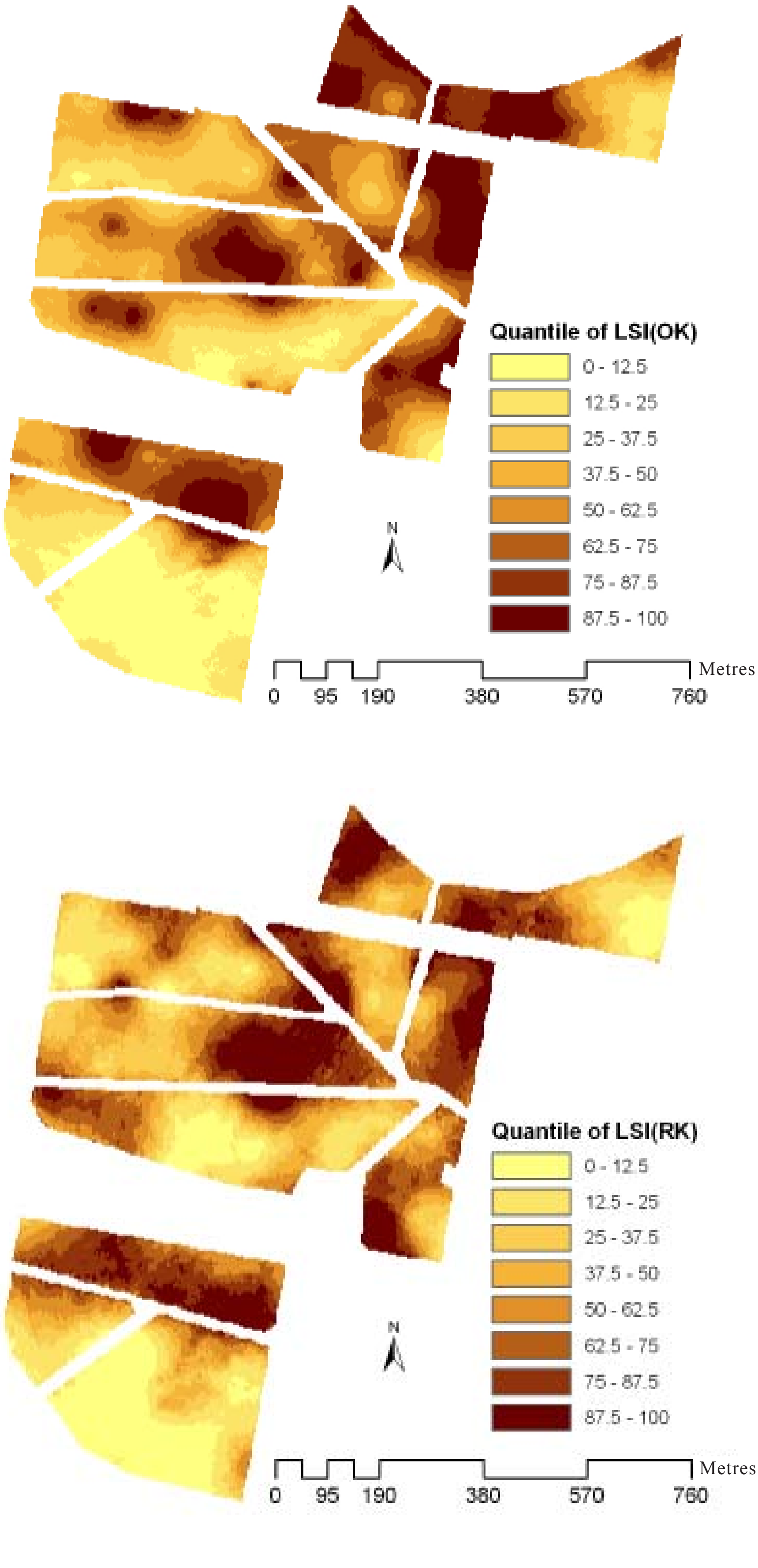


Figure 7.3: Maps of LSI_{OK} (top) and LSI_{RK} (bottom) for Canowindra. (Maps are displayed as quantiles)

the sandy hill top (site B).

7.3.2 Class optimisation

From the output of the FuzMe analysis the FPI and MPE for each class determination for each cluster approach was plotted (Appendix 7.2) and the mean of the FPI and MPE values calculated (data not shown). For Cowra, all the spatially constrained approaches had a feature at both 7 and 10 classes. As a result both classes were investigated. The spatially constrained Canowindra analysis was generally optimised at 5 classes. The optimal clusters as indicated from the FPI/MPE analysis are given in Table 7.3. The non spatially constrained clustering approaches tended to produce a lower number of clusters. This produced large areas, that were much bigger than block size, and made thus makes it difficult to accurately site blocks.

For each variety in the vineyards for which adequate crop response data was available the r^2 values of an ANOVA of optimal clusters vs crop response are given in Table 7.4. The best responses for each variable are highlighted in bold. The best fit for the data was the standardised chardonnay yield at Cowra with the LSI_{OK}_xy approach explaining 37.7% of the variation. Clustering showed no response to NDVI at Cowra with no r^2 values above 0.1. The response to the plant photosynthetic ratio (PPR) was mixed with a good response in Chardonnay but little in the Shiraz blocks. In general the inclusion of spatial coordinates improve the cluster models. As expected it produced more spatially contiguous patterns that are more useful when considering vineyard design. When the spatial coordinates are not included the clustering tends to produce gradational clusters along data contours. An example of this is shown in Figure 7.4. At Canowindra the r^2 values were better in the Cabernet blocks with the spatially constrained clustering again producing the best results.

To determine the effectiveness of the clustering parts of the vineyards were mapped with the two best cluster models, yield data, EC_a (30-90 cm) and NDVI or RVI. At Cowra the strong response in Chardonnay is shown (Figure 7.5) while at Canowindra two blocks of Semillon are mapped (Figure 7.6). At both locations the clustering appears to have failed to capture the variability that can be discerned in the yield, imagery and soil EC_a maps. This is particularly evident in the Semillon blocks at Canowindra. There is some weak correlation between the yield, RVI, EC_a and LSI_{RK} maps however the clusters do not pick up on this at all. At Cowra, the chardonnay gave the best fit of all models tried however visually much of the variation in the ancillary maps has not been captured. Although the clustering is not reflecting the variability in the yield and imagery, the LSIs, and particularly the LSI_{RK} are. The exception is the NDVI image at Cowra which shows little spatial structure to the other data.

From the information in Figures 7.5 and 7.6 the LSI_{RK} appears to be reflecting the vine response. The clustering approaches tried however appear unable to fully utilise this information. The spatial weighting may be too strong and overriding the information in the LSI layer. This appears to be a particular concern in the Cowra vineyard. When spatial coordinates are not included the data tends towards forming contours which are not suitable for vineyard design. Further studies are required to understand how clustering algorithms can help automate the prediction of digital terroirs.

Initially it was intended to test the cluster models using the statistics described in the methodologies. However this has been foregone due to the poor fit of the models. It is still uncertain if the FPI/

Method	Cowra	Canowindra
OKall	3	2
OKall_xy	7, 10	5
$\mathrm{LSI}_{\mathrm{OK}}$	8	10
$\mathrm{LSI}_{\mathrm{OK}}$_xy	7, 10	5
OK1	3	10
OK1_xy	7, 10	5
OK2	2	6
OK2_xy	7, 10	5
RKall	5	2
RKall_xy	7, 10	5
$\mathrm{LSI}_{\mathrm{RK}}$	10	10
$\mathrm{LSI}_{\mathrm{RK}}$_xy	10	5
RK1	10	2
RK1_xy	10	5

Table 7.3: Table of optimum clusters as determined by MPE/FPI

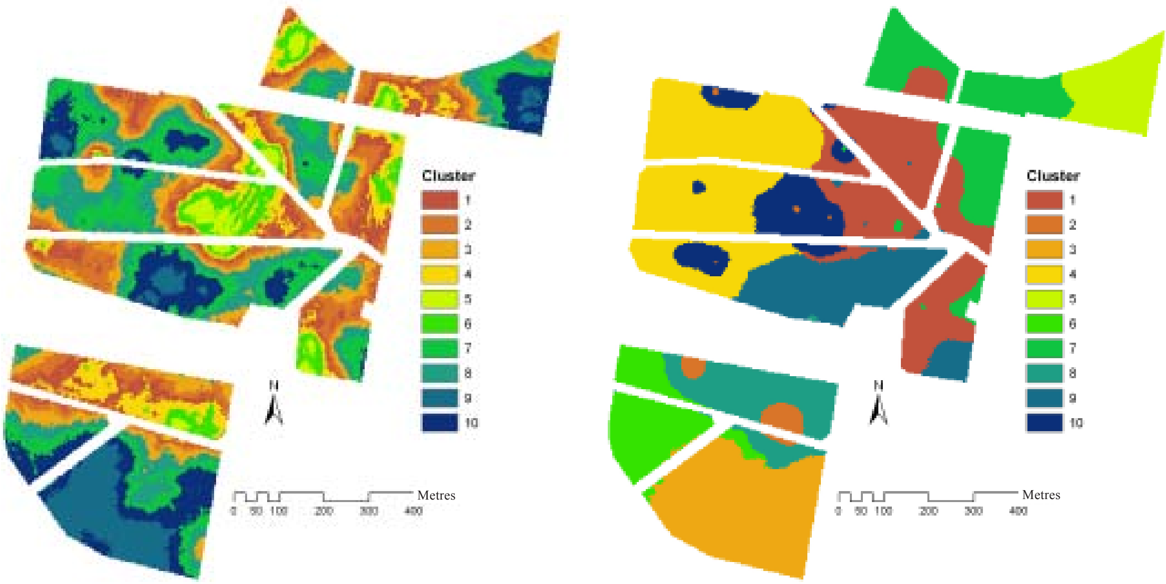


Figure 7.4: Comparison of the influence of clustering without (left) and with (right) spatial coordinates on the contiguity of managment classes.

Approach	Cowra						Approach	Canowindra	
	Chardonnay			***Shiraz***				***Semillon***	***Cabernet***
	Std. Yield	*NDVI*	*PPR*	*Std. Yield*	*NDVI*	*PPR*		*Std. Yield*	*Std. Yield*
OKall (3)	0.045	0.044	0.117	0.006	0.005	0.000	OKall (2)	0.020	0.001
OKall_xy (7)	0.264	0.061	0.184	0.016	0.005	0.013	OKall_xy (5)	0.085	0.080
OKall_xy (10)	0.169	0.051	**0.189**	0.016	0.003	0.007			
LSIOK (8)	0.220	0.053	0.110	0.044	0.045	0.018	LSIOK (10)	0.043	0.058
LSIOK_xy (7)	0.259	0.008	0.005	0.000	0.017	0.000	LSIOK_xy (5)	0.043	**0.237**
LSIOK_xy (10)	**0.377**	0.047	0.049	0.028	0.026	0.006			
OK1 (3)	0.030	0.048	0.160	0.046	0.003	0.014	OK1 (10)	**0.113**	0.162
OK1_xy (7)	0.035	0.034	0.102	0.058	0.001	0.007	OK1_xy (5)	**0.091**	0.210
OK1_xy (10)	0.189	0.047	**0.190**	**0.124**	0.003	0.016			
OK2 (2)	0.006	0.000	0.169	0.000	0.000	0.000	OK2 (6)	0.081	0.146
OK2_xy (7)	0.309	0.049	0.070	0.035	0.000	0.002	OK2_xy (5)	0.067	**0.209**
OK2_xy (10)	0.275	0.041	0.180	**0.113**	0.000	0.024			
RKall (?)	0.072	0.040	0.090	0.021	0.053	0.002	RKall (2)	0.014	0.001
RKall_xy (7)	0.268	0.039	0.079	0.003	0.022	0.011	RKall_xy (5)	0.051	0.137
RKall_xy (10)	0.153	0.033	0.126	0.015	0.047	0.015			
LSIRK (10)	0.116	0.021	0.083	0.015	0.079	0.010	LSIRK (10)	0.084	0.062
LSIRK_xy (10)	**0.332**	0.010	0.111	0.026	0.041	0.001	LSIRK_xy (5)	0.043	0.189

Table 7.4: r^2 values from ANOVA of cluster models. (The best fits are shown in bold)

MPE approach is suitable for selecting the optimal number of clusters in this case. Certainly the models identified were not particularly good but it is not known if this is due to the input data, the clustering algorithm or the MPE/FPI selection criteria. This should also be a focus for future studies.

While the clustering approach was unsuccessful, as an input into a human decision making process, the LSI_{RK} appears to be very useful in identifying where the vine:environment response is changing and delineating digital terroir. There are some concerns about the production of similar LSI values from differing environments. However these tend to be spatially separate and any decision on vine selection/management should take into account local soil information as well as the LSI values. In Figure 7.5 the LSI_{RK} shows a strong response on the right-hand side of the southern portion (Block 25). The same area exhibits poor yield and low EC_a. It is unclear if a LSI_{RK} response >0.9 is detrimental for yield growth or the high response is erroneous. This highlights the need for the methodology to be tested further. Firstly to identify what sort of numbers are optimal or suboptimal and secondly to determine how robust the model is by testing on other vineyards in other regions. The model will also continue to be tested on crop response data as it comes to hand. In particular the LSI model needs to be tested against winegrape quality parameters as well as yield.

In the previous two chapters the use of OK data has produced better fits than RK data. This was not expected and contrary to previous studies. While the OK data has outperformed the RK data statistically, when mapped the RK data appeared to contain more detail and more information on environmental boundaries. In Figures 7.5 and 7.6 this information appears beneficial and provides the LSI_{RK} with a better estimation of site-specific vine response. The LSI_{OK} maps are smoother and apparently lack the necessarily detail to accurately define boundaries.

7.4 Conclusions

A Local Site Index (LSI) for vineyards has been proposed for use in vineyard design. It is based on an existing model that has been adapted for finer-scale surveys. The LSI appears to complement crop response, particularly yield. The use of regression kriged variables in the LSI produced a more coherent response than the use of ordinary kriged variables.

Attempts to cluster the data into digital terroirs or management zones were largely unsuccessful with the clusters accounting for only a small proportion of the crop response variability. While an automated process is desirable, human (expert) input is always required to ensure that the result is plausible. Further work will continue into automating the clustering process and the derivation of indices to identify the optimal number and arrangement of classes. While clustering was not successful the derived data layers, including the LSI layers provide a lot more information to the grower to aid in vineyard design. The use of this information should promote better vineyard designs and also contribute to lower within block variability.

7.5 References

BOYDELL, B.C. & MCBRATNEY, A.B. (1999). *Identification and stability of potential within-field yield zones obtained from 11 years of cotton yield estimates derived from mid-season Landsat MS imagery.* In: J.V. Stafford (ed.). Precision Agriculture '99: Proceedings of the Second European Conference on Precision Agriculture. July 11-15, 1999, Odense Congress Centre, Denmark. pp 331-342

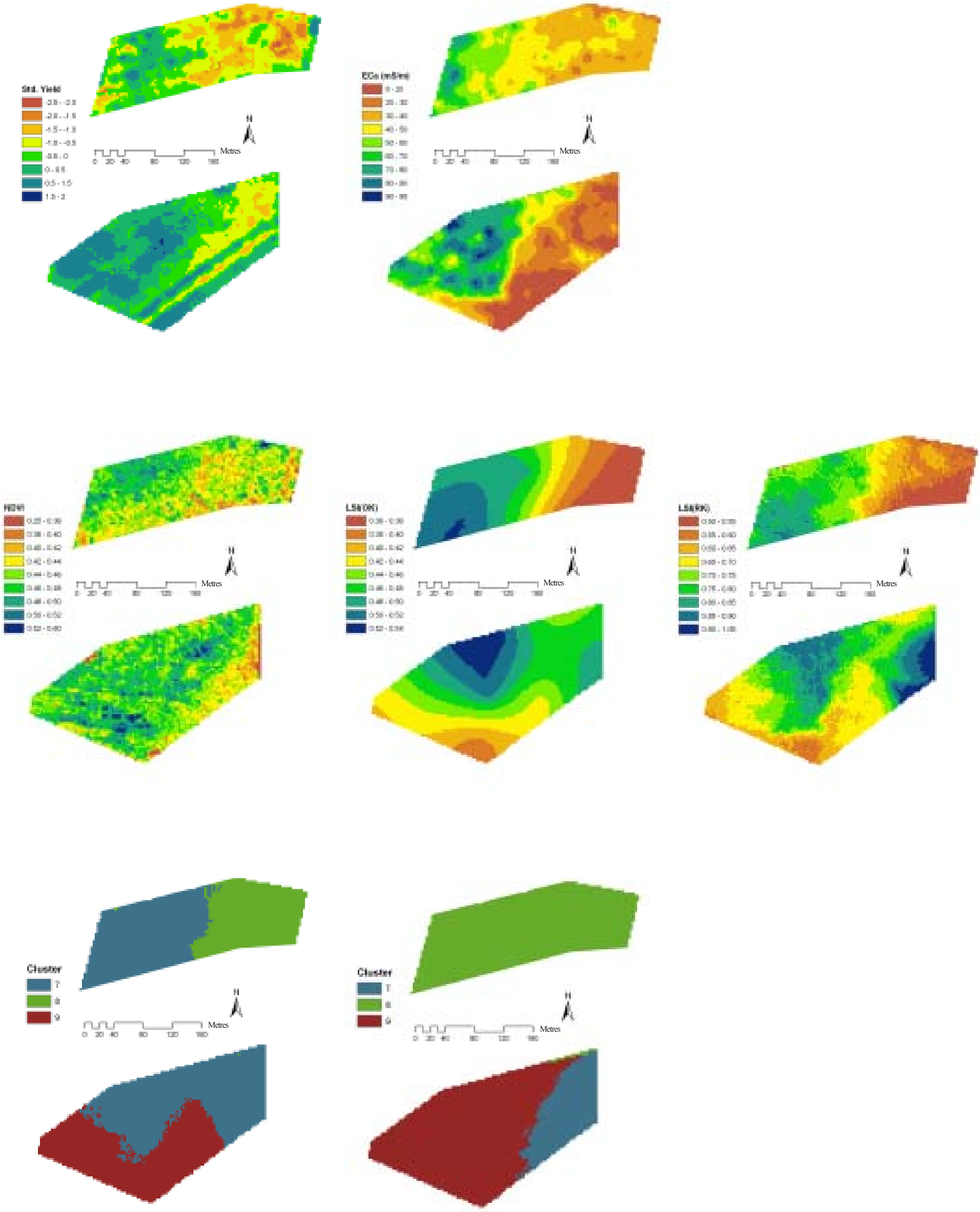


Figure 7.5: A comparison of various data layer and cluster analysis results for two chardonnay blocks at Cowra

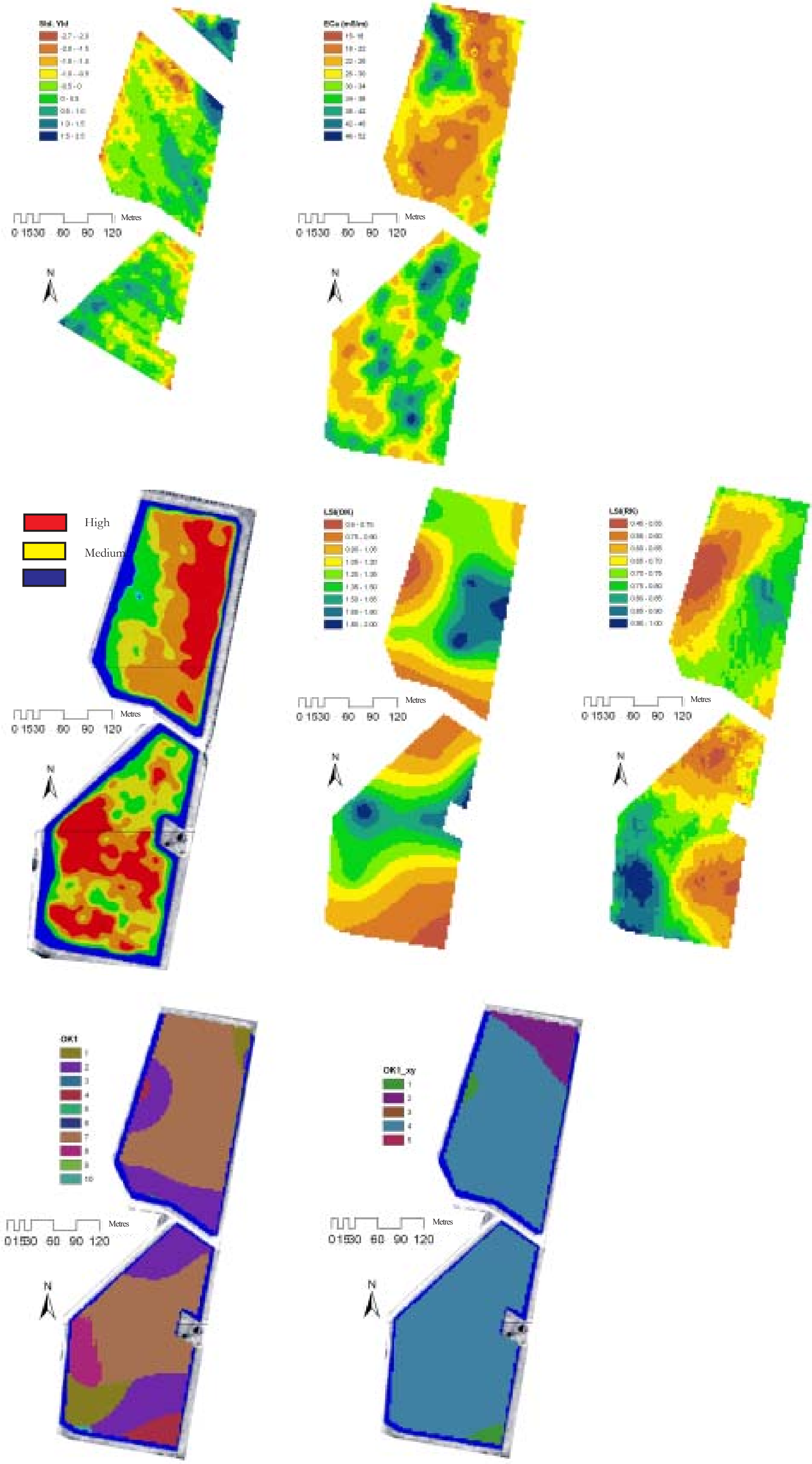


Figure 7.6: A comparison of various data layers and cluster analysis results for two semillon blocks at Canowindra

Cupitt, J. & Whelan, B.M. (2001). *Determining potential within-field crop management zones.* In: G. Greiner and S. Blackmore (eds.). Proceedings of the Third European Conference on Precision Agriculture. June 18-20, 2001, Montpellier, France. pp7-12.

Dobrowski, S.Z., Ustin, S.L., Wolpert, J.A., (2002). *Remote estimation of vine canopy density in vertically shoot positioned vineyards: Determining optimal vegetation indices.* Australian Journal of Grape and Wine Research (8) pp177-125.

Fleming, K.L. and Buchleiter G.W. (2002) *Evaluating two methods of developing management zones for precision farming.* In: P.C. Robert, R.H. Rust and W.E. Larson (eds.). Precision Agriculture and Other Resource Management: Proceedings of the Sixth International Conference on Precision Agriculture. ASA-CSSA-SSSA, Madison, WI, USA.

Franzen, D.W. and Nanna, T. (2002). *Nitrogen management zone delineation. In* 2002 Agronomy Abstracts. ASA-CSSA-SSSA. Madison, WI.

Fridgen, J.J., Kitchen, N.R. and Sudduth, K.A. (2000) *Variability of soil and landscape attributes within sub-field management zones.* In: P.C. Robert, R.H. Rust and W.E. Larson (eds.). Precision Agriculture and Other Resource Management: Proceedings of the Fifth International Conference on Precision Agriculture. ASA-CSSA-SSSA, Madison, WI, USA.

Gladstones, J.S. (1992) *Viticulture and Environment.* Winetitles, South Australia

Johnson, L., B. Lobitz, R. Armstrong, R. Baldy, E. Weber, J. DeBenedictis, and D. Bosch. (1996). *Airborne Imaging for Vineyard Canopy Evaluation.* California Agriculture, Special Issue on Phylloxera 50(4) pp14-18

Jones, A.J., Mielke, L.N., Bartles, C.A. and Miller, C.A. (1989) *Relationship of landscape position and properties tocrop production.* Journal of Soil and Water Conservation, Vol. 44, pp328-332.

McBratney, A.B. (1998) *Some considerations on methods for spatially aggregating and disaggregating soil information.* Nutrient Cycling in Agroecosytems 50, pp51-62.

Metternicht, G., Honey, F. and Beeston, G. (2002) Video-based Precision Farming. (http://www.geoplace.com/asiapac/2000/0700/0700frm.asp)

Minasny, B. and McBratney, A.B. (2000) FuzME Ver. 2.1. Australian Centre for Precision Agriculture. The University of Sydney. (www.usyd.edu.au/su/agric/acpa)

Pringle, M.J., McBratney, A.B., Whelan, B.M. & Taylor, J.A. (2003). *A preliminary approach to assessing the opportunity for site-specific crop management in a field, using a yield monitor.* Agricultural Systems 76, pp273-292.

Roubens, M. (1982) *Fuzzy clustering algorithms and their cluster validity.* European Journal of Operational Research 10, pp294-301.

Shatar, T. M. and A.B. McBratney (2001). *Subdividing a field into contiguous management zones using a K-zones algorithm.* In: G. Greiner and S. Blackmore (eds.). Proceedings of the Third European Conference on Precision Agriculture. June 18-20, 2001, Montpellier, France pp115-120

Stewart, C.M. & McBratney A.B. (2001). *Using bare soil imagery to determine management zones for the variable rate application of inputs for cotton.* In: G. Greiner and S. Blackmore (eds.). Proceedings of the Third European Conference on Precision Agriculture. June 18-20, 2001, Montpellier, France pp. 319-324

Sudduth, K.A., Drummond, S.T., Birrell, S.J. and Kitchen, N.R. (1997)*Spatial Modeling of Crop Yield Using Soil and Topographic Data.*. In: J.V. Stafford (ed.). Precision Agriculture '97: Proceedings of the First European Conference on Precision Agriculture. September 7-10, 1997, Warwick University Conference Centre, UK. pp439-447

Tesic D., Woolley, D.J., Hewett, E.W.; Martin, D.J. (2002) *Environmental effects on cv Cabernet Sauvignon (Vitis vinifera L.) grown in Hawke's Bay, New Zealand. II. Development of a site index.* Australian Journal of Grape and Wine Research 8(1), pp27-35

Tesic D. (2003). Viticultural sub-regions and their delimitation - case studies: Mornington Peninsula and Cowra. The Australian & New Zealand Grapegrower & Winemaker, 31st Annual Technical

Issue, 473a, pp135-138

Van Beurden, A. U. C. J. and Douven, W. J. A. M. (1999) *Aggregation issues of spatial information in environmental research.* Int. J. Geographical Information Science. 13(5), pp513-527

Wilson, J.E. (1998) *Terroir: The role of Geology, Climate and Culture in the making of French wines.* Mitchell Beazley. London.

Wilson, J.P., Gallant, J.C., (1997) *SRAD: a program for estimating radiation and temperature in complex terrain.* Unpublished research.

Appendix7.1 - SRAD Input Data

SRAD input for Cowra.

-33.7 -33.9
0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25
0.15 0.15 0.15 0.15 0.15 0.15 0.15 0.15 0.15 0.15 0.15 0.15
0.34 0.34 0.34 0.34 0.34 0.34 0.34 0.34 0.34 0.34 0.34 0.34
0.64 0.60 0.65 0.60 0.52 0.43 0.49 0.60 0.60 0.61 0.59 0.61
31.8 31.3 28 23.5 18.5 14.7 13.7 15.4 18.3 22.5 26.3 30.1
15.3 15.3 12.3 8.2 5.3 3 1.9 2.8 4.4 6.9 9.8 12.8
average

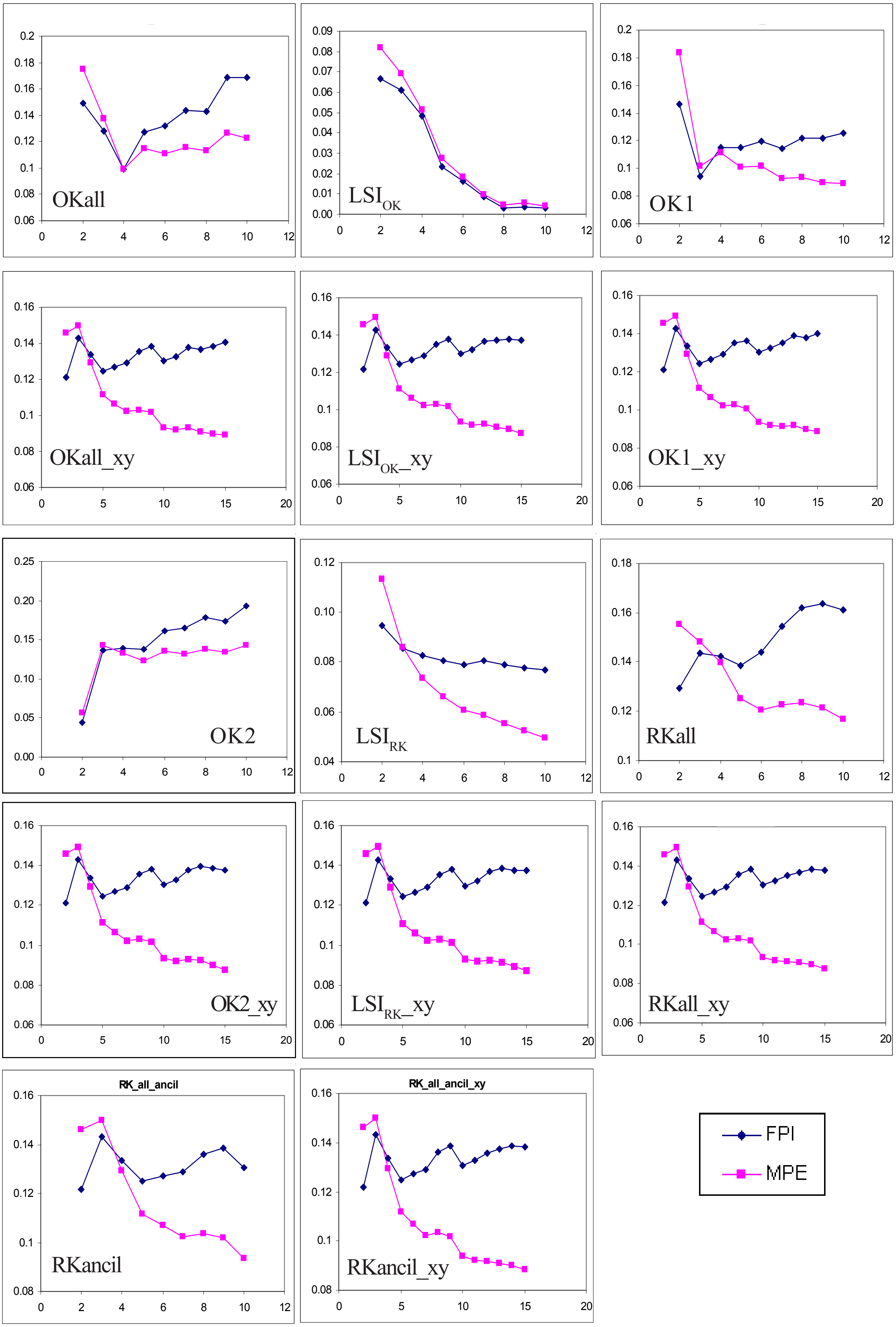
7.30 7.30 7.30 7.30 7.30 7.30 7.30 7.30 7.30 7.30 7.30 7.30
6.00 6.00 6.00 6.00 6.00 6.00 6.00 6.00 6.00 6.00 6.00 6.00
1
2.25 2.25 2.25 2.25 2.25 2.25 2.25 2.25 2.25 2.25 2.25 2.25
10.0 0.96 0.00008 25
0.65 0.66 0.67 0.68 0.69 0.70 0.71 0.70 0.69 0.68 0.66 0.65

SRAD input for Canowindra

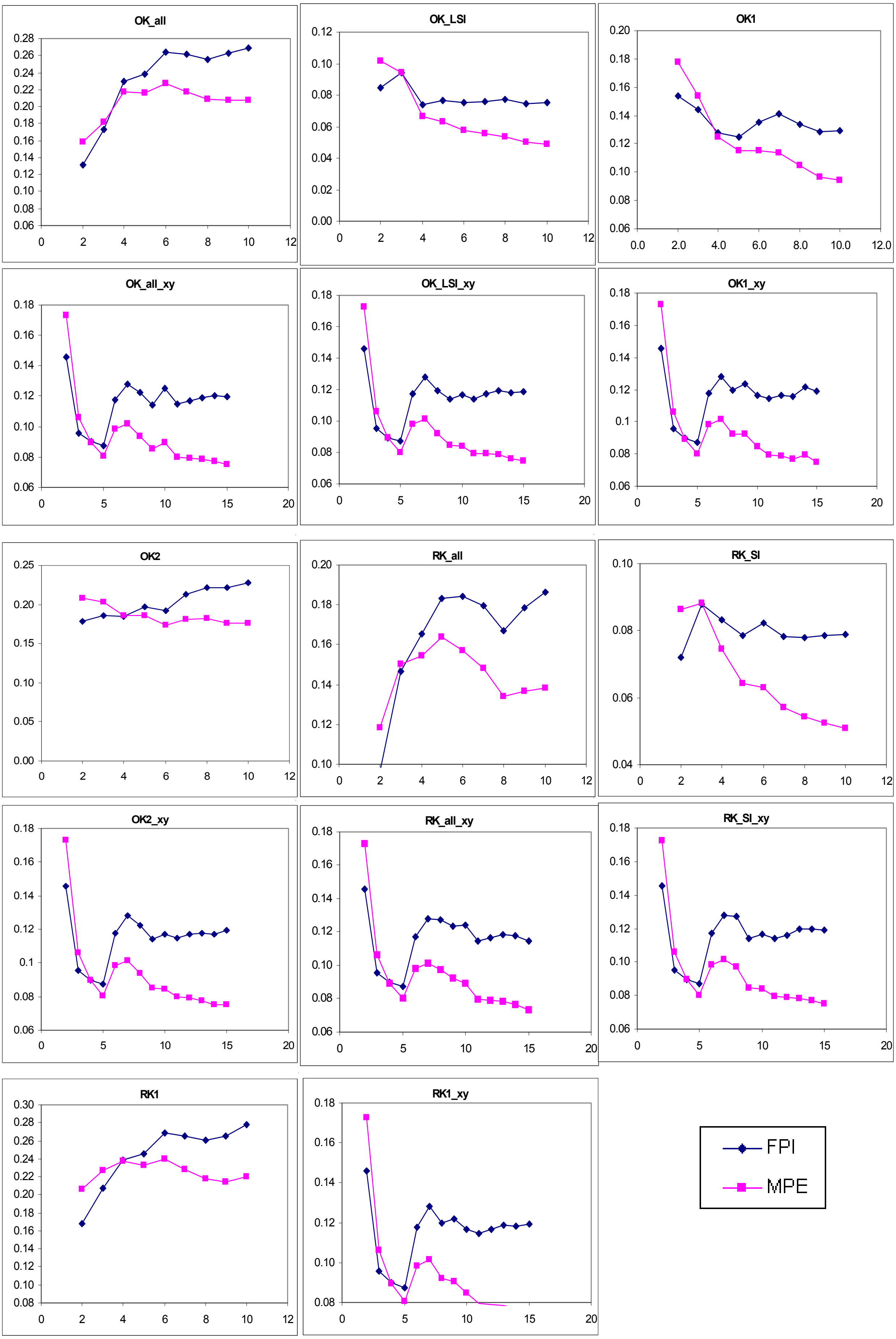
-33.4 -33.8
0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25
0.15 0.15 0.15 0.15 0.15 0.15 0.15 0.15 0.15 0.15 0.15 0.15
0.34 0.34 0.34 0.34 0.34 0.34 0.34 0.34 0.34 0.34 0.34 0.34
0.64 0.60 0.65 0.60 0.52 0.43 0.49 0.60 0.60 0.61 0.59 0.61
34 33.3 30.5 24.9 19.6 15.6 15 17 21.1 25.2 28.9 32.5
15.6 15.3 12.6 7.9 4.5 2.4 1.6 2.4 4.6 7.8 10.8 13.4
average

1
2.25
10 0.96 0.00008 25
0.65 0.66 0.67 0.68 0.69 0.7 0.71 0.7 0.69 0.68 0.67 0.65

Appendix 7.2 - MPE and FPI plots for Cowra



Appendix 7.2 - MPE and FPI plots for Canowindra



Appendix 7.3 - A Spatial Opportunity Index for Precision Agriculture

These are some thoughts on the derivation of a spatial opportunity index for Precision Agriculture arising from the development of the continuous opportunity index proposed by Pringle *et al.* 2003. The data used for this was a wheat field in Northern NSW that formed part of Pringle et al's original paper.

An index based on statistical parsimony (Lark, in press) derived from the Akaike Information Criterion could be used:

$$AIC = n \cdot \ln(RMS) + 2p$$

where: n = the number of observations
RMS = the residual mean square for the model fit
p = the number of parameters in the fitted model.

In this context we could write the opportunity index as:

$$O_{zr}(z) = 2z - n \cdot \ln(r^2)$$

where: z = the number of zones or spatial units in the field.
r^2 = the fit of the model

In this situation z is not the number of classes that might be fitted by a (fuzzy) *k*-means algorithm but rather the number of discrete contiguous spatial zones. These contiguous spatial zones can be formed by including the spatial co-ordinates as well as the yield in the numerical classification. (N.B. these zones are not the same as treatment classes). While individual zones may be discrete treatment classes, other treatment classes may be composed of two or more zones (that are spatially discrete from each other). An additional 2 zone model determined "by eye" (E_2) was also added (discussed later). For the example we have chosen West Creek 1997 as it has a high O_c and the large feature on the left hand side indicates it is suitable for zone management.

The goodness of fit of the zone models (in describing yield variation) was determined from the r^2 of a one-way analysis of variance model. The use of n is problematic however, because it is the number of independent observations. For this discussion the method of Bishop *et al.* (in press) has been adapted to raw yield data to determine n using 5000 data points. The number of zones (z), which minimises O_{zr}, can be regarded as optimal.

From Table A we can see that O_{zr} becomes asymptotic after 7 zones indicating that there is little benefit in managing additional zones. The amount of variation accounted for in the management zones can be determined by comparing the yield variogram to that of the residuals. In this case a spherical model has been fitted to the yield data and the exponential model to the residual data. There has been no penalty imposed to models with more parameters, thus as the number of zones increases the amount of variance explained by the model increases and the sill decreases.

Figure B shows the variograms for the yield and residuals from the ANOVA for 5, 6 and 7 zone models. The area between the sills of the yield and the individual residual variograms can be considered representative of the amount of variation explained by the zonal models. Taking the variance at 1000m, the 5, 6, and 7 zone models explain 51%, 63% and 67% of the variation in yield respectively. These numbers are comparable to the r^2 derived from the ANOVA of the cluster means however they also include a spatial component.

z	n	r^2	O_{zs}	O_{zi}
1	112	-	-	-
2	112	0.002	389.51	2.99
3	112	0.061	189.67	6.70
4	112	0.181	158.87	9.54
5	112	0.484	140.09	13.64
6	112	0.582	124.57	13.63
7	112	0.616	49.86	12.98
8	112	0.660	48.82	12.56
9	112	0.692	49.33	12.12
10	112	0.700	48.84	11.57
E_2	112	0.257	156.17	15.73

Table A. Estimates of Opportunity for managing different numbers of zones for West Creek 1997.

While determining which model best describes variation the calculation of O_{zs} does not take into account issues such as differences between means of zones, gross margins etc. Ideally a better opportunity index would be an economic one (O_{ze}), measured in dollars (per hectare):

$$O_{ze} = \sum_{i=1}^{z}\left[\frac{G_i}{A_i}\right]$$

where A_i = the area of zone i.

G_i = the gross margin for zone i which is calculated from;

$G_i = P_i - C_i - F_i$

where: P_i = value of production

C_i = agronomic cost of production

F_i = environmental cost of production (which is still difficult to calculate).

Here the assumption is that the zones are suitable for PA. This time, the optimal z is the number that maximises O_{ze}.

Currently, it may be difficult to obtain all the data to calculate O_{ze} but developing methods to obtain these data should be an aim of further research. In the meantime, we might think of using a compromise between the statistical and economic indices, which is really what our O_c is. If we replace S in Equation 4.5 by

$$S(z) = r^2\left[\frac{A}{z}\right] + J_a\left[1 - r^2\right]$$

where
A = Area of field
z = number of zones
r^2 = fit of the model and
J_a = the integral scale (Chapter 4)

it is possible to calculate O_{zi}, which should be maximised. Results of this are presented in Table 4. O_{zi} values indicate that the optimal number of management zones in this field is either 5 or 6 (less than that indicated by the O_{zs} analysis). It should be noted that these values cannot be compared directly to the O_c values as they are situated on a different scale. From Figure A, maps with only 3 and 4 zones are reflecting the heavy weighting of the spatial coordinates in the analysis thus have poor r^2 values when compared with the yield. When compared with the yield map, diagrams with 5, 6 and 7 zones highlight the main management zones in the field. Diagrams with 8 or more zones are starting to identify small areas in the field, which are probably not viable units with current variable-rate technologies.

In this field we would expect two management zones to have a reasonable opportunity due to the large feature on the left-hand side. However the heavy weighting with spatial coordinates in the numerical classification negates this in the 2 zone model. By applying expert knowledge we can segregate this feature into z_1 and specify the rest of the field as z_2 and analyse this model (E_2). The r^2 for the ANOVA between z_1 and z_2 is 0.257 significantly higher than the 2 and even the 3 and 4 zone model derived using yield and spatial coordinates in Table 4. Plotting the variogram of the residuals against the yield shows that this minimal segregation already accounts for 28% of the variation in yield.

The O_{zi} for this model (15.73) is the highest of any of the models due to S_z being weighted to minimise zones. As expected two large discrete contiguous zones provide a good opportunity for PA. Whether it should have a higher O_{zi} than a more complex model that better fits yield variation is a point for further research and discussion. The r^2 O_{zs} and O_{zi} values from E_2 highlight the need for a better algorithm for deriving the discrete contiguous zones.

It should be emphasised that this is only a preliminary model and presented here as an example. Considerable work still needs to be done especially on the development of a zonal algorithm. Further we would like to reiterate that while this analysis has been done on a field for one year, data from several years will be necessary before a true indication of the opportunity will be known.

References

Pringle, M.J., McBratney, A.B., Whelan, B.M. & Taylor, J.A. (2003). *A preliminary approach to assessing the opportunity for site-specific crop management in a field, using a yield monitor.* Agricultural Systems 76, pp273-292.

Bishop, T.F.A., McBratney, A.B. and Whelan, B.M. (2001). *Measuring the quality of digital soil maps using information criteria.* Geoderma, 103, pp95-111.

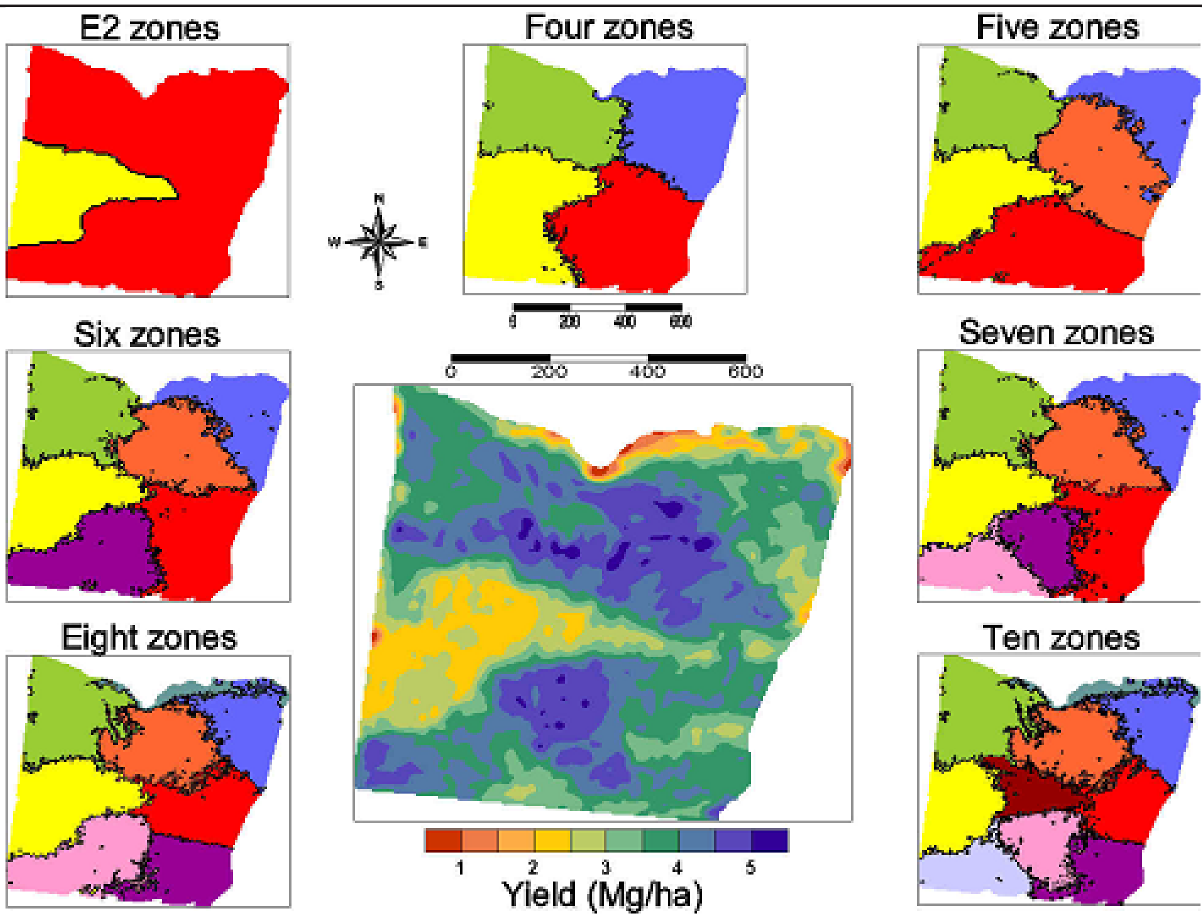


Figure A: Cluster maps of Yield, Eastings and Northings showing the spatial distribution of potential management zones. 2, 3 and 9 zone models not shown.

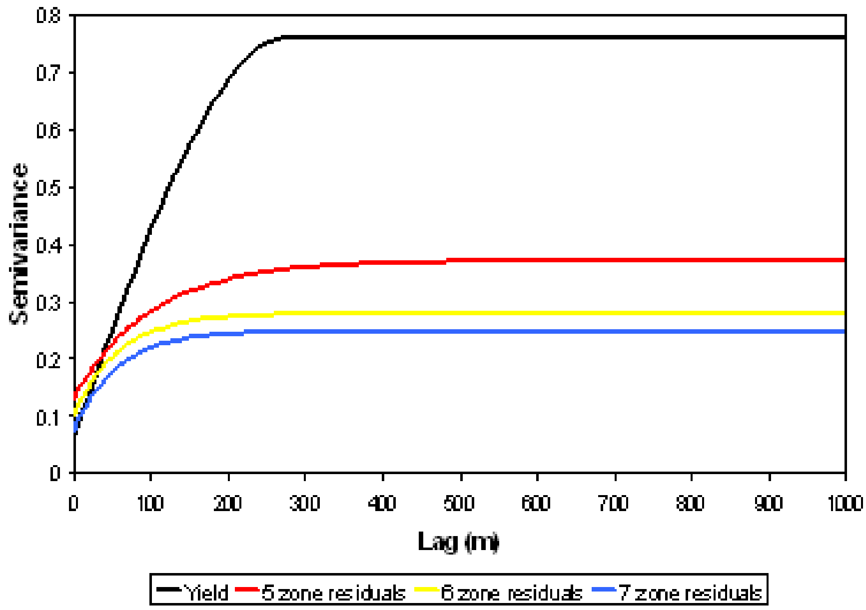


Figure B: Variograms of residuals from the 5, 6 and 7 zone models compared with the raw yield variogram