Chapter 6: Improving the accuracy and spatial resolution of vineyard soil maps using regression kriging

6.1 Introduction

As illustrated in Figure 5.1, data from viticultural/horticultural soil surveys has traditionally been presented in point form at each pit site rather than as a raster map. The use of 75-100 m grids produces point maps with a lot of empty space. In recent times this has given way to raster maps with a pixel size equal to the survey grid. This provides no additional information but, while still coarse, is easier to visualize. As outlined in Chapter 5, this approach is due to the use of the linguistic variables in the survey, a general lack of knowledge in the industry about spatial interpolation techniques and the fact that maps have always been presented this way so people have become use to them. The current point maps produced have two main limitations 1) it is difficult to visual the data on the map and 2) the scale is incompatible with other data layers. The protocol presented in Chapter 5 has addressed some of these problems, however, the level of detail in the maps is still restricted to the density of the soil survey. Analysis of within vineyard production variability (Chapter 4) indicates that considerable variation in production, particularly yield, occurs at ranges <75 m. The introduction of precision viticultural technologies, such as yield monitoring, remote sensing (aerial or satellite), electromagnetic induction (EMI) surveying and GPS-based elevation surveying, is now providing soil and production information at much finer scales than the 75-100 m grid soil survey. This new ancillary data may allow for more accurate mapping of environmental parameters and improved efficacy of vineyard and irrigation design. Of particularly interest in this study is the recent adoption of soil apparent electrical conductivity (EC_a) and GPS-based elevation surveys prior to vineyard design. The potential benefits of this data in helping to describe plant environs or "digital terroirs" is illustrated previously in Figure 5.10.

While the industry has a widely-used protocol for soil surveying and interpretation there is no protocol on the collection and analysis of these new ancillary data. Currently new fine-scale ancillary data is only visually compared with the soil survey data to confirm the soil pit findings. A lack of standards in cartography and ground-truthing has created some confusion and an over-expectation in what the ancillary data can provide. This in turn has lead some surveyors to ignore or distrust ancillary data (McKenzie, 2000). With correct analysis and a defined protocol this confusion should be avoided and the value of the ancillary information maximised. Ancillary data sets, particularly elevation, may be useful for soil prediction even though they are not direct soil measurements. This is due to the influence of many factors on the development of soil profiles. This was first outlined by Jenny (1941) with the *CL*imate *O*rganism *R*elief *P*arent material and *T*ime (CLORPT) model of soil pedogenesis. Building on the CLORPT model, McBratney *et al.* (2003) have proposed a *S*oil (or soil attributes), *C*limate, *O*rganisms, *R*elief (topography), *P*arent material, *A*ge (time) and Space (*n*) (SCORPAN) model as a basis for soil digital mapping.

Multivariate geostatistical techniques, such as universal kriging, co-kriging and regression kriging, have been developed over the past 30 years that allow multiple data layers to be combined and interpolated across an area even if the data are not collected at the same scale (McBratney *et al.*, 2000). This chapter focuses on regression kriging that has been shown to be most suited to local-scale soil mapping when a large number of predictor variables are available (Odeh *et al.*, 1995, Bishop and McBratney, 2001, Knotters *et al.,* 1995). Co-kriging with more than 4-5 variables is problematic as the interpolation becomes increasing complex as more variables are added to the model (Odeh *et al.*, 1995). Regression kriging also provides more flexibility in the interpolation as a variety of different regression techniques may be used,for example linear models, generalised linear models (GLM), generalised additive models (GAM), regression trees or neural networks (McBratney *et al.*, 2000).

The aim of this chapter is to incorporate the soil data derived in Chapter 5 with ancillary information (EC_a and elevation data) using regression-kriging to provide a more accurate and finer resolution understanding of the within-vineyard environment. A range of approaches to regression analysis will be investigated to evaluate their effectiveness for vineyard soil survey data.

6.2 Methodologies

6.2.1 Survey Sites

The vineyards used in this study are the same as those described previously in section 5.2.1. Due to the constraints with data (see §6.2.3) only 77 and 212 sites were used for Pokolbin and Canowindra respectively. All sites were used for Cowra.

6.2.2 Soil Survey Data

The soil survey data was analysed as described in section 5.2.2. For this study only three soil variables were examined; 0-30 cm clay%, 0-90 cm clay% and RAW_W. As RAW_W is the current industry standard it was preferred in this study to the alternate RAW_q estimation described in Chapter 5.

6.2.3 Ancillary Data

In 1991 a registered surveyor, using a theodolite, conducted an elevation survey of the Cowra vineyard. Approximately 3000 height measurements were taken over the vineyard site. The elevation of the Pokolbin and Canowindra vineyards was mapped with an AshTech RTK-GPS mounted on a 4WD vehicle.

The three vineyards were mapped for apparent soil electrical conductivity (EC_a) using the Veris 3100Ò soil electrical conductivity cart. This provided a measurement of the topsoil (~0-30 cm), subsoil (~30-90 cm) and whole profile (~0-90 cm) EC_a. The vineyards were mapped on ~12 m swaths along the rows. Problems occurred in the Canowindra and Pokolbin vineyards with the use of the Veris 3100®. The Pokolbin vineyard was under redevelopment at the time of the Veris survey and some blocks were unable to be mapped. An attempt to map the Canowindra vineyard was made in 2002 however the soil was too dry and the machine unresponsive. This is a known drawback of direct current ECa instruments (Dabas *et al.,* 2003). In 2003 parts of the vineyard were inaccessible at the time of surveying due to late pruning.

6.2.4 Validation Sites

The fifteen validation sites used in Chapter 5 were again used as an independent validation set for this study. Validation site selection, sampling and soil analysis is described in section 5.2.4. Due to the missing ancillary data at both Pokolbin and Canowindra only 12 and 11 of the validation sites could be used respectively. All 15 sites at Cowra were used.

6.2.5 Ordinary Kriging and DTM attribute derivation

Soil survey data was interpolated using punctual ordinary kriging with a global variogram in Vesper (Minasny *et al.*, 2002), onto the 3m x 3m grids used previously. The grids for Canowindra and Pokolbin were trimmed to areas covered by the ancillary data (particularly the Veris data).

The primary ancillary datasets (elevation and Veris 3100 data) were trimmed of any outliers by visually plotting a histogram of distribution and manually removing extreme values. For the Veris data a distribution of the raw data was used while for the AshTech data the RMSE associated with the measurement was used. The contracted surveyed elevation data was not trimmed. The ancillary data was interpolated using block kriging in Vesper onto the same 3m x 3m grid used above with a local exponential variogram structure.

The ordinary kriged elevation data for each vineyard was imported into Arc/INFO® and converted into a "grd" file. Using standard Arc/INFO® commands primary landform attributes were derived at each grid node including aspect, slope, flow direction, upslope and downslope flow, planar and profile curvature and flow accumulation. From the primary attributes the secondary landform attribute, topographic wetness index (twi) (Moore *et al.*, 1991) was derived. The landform attributes were mapped in ArcGIS® and the upslope flow, downslope flow and twi found to be nonsenscical. The elevation surveys were confined to the vineyard thus no external information on the surrounding landform was available. These three digital terrain model (DTM) attributes require information on the locality of the vineyard in the landscape to be effective. Without this information the vineyard boundary is an effective zero point and erroneous results ensue. Upslope and downslope flow and twi were removed which left eleven ancillary variables; Veris EC_a 0-30 cm, Veris EC_a 0-90 cm, Veris EC_a 30-90 cm, elevation, aspect, slope, flow direction, profile curvature, planar curvature, Eastings and Northings.

The ancillary landform data was extracted to the 3m x 3m interpolation grid using the "Pixel to ASCII" function in ERDAS IMAGINE® (Erdas LLC, 2002) and combined with the interpolated Veris data to form an "interpolation" set. Similarly the landform attributes and Veris data were extracted to the soil survey sites. This was combined with the measured and PTF manipulated soil data to form a "regression" dataset that was used to derive all the regression models. The ancillary data was also extracted to the validation sites and combined with the laboratory measured soil properties to form a "validation" set.

The derived DTM attributes and interpolated Veris data was imported into ArcGIS® and mapped for each vineyard.

6.2.6 Regression Kriging

Regression kriging (RK) is a multivariate interpolation approach that, as the name suggests, combines kriging with regression analysis. The approach was initially developed and applied in hydrosciences (Delhomme, 1974, 1978 in Knotters *et al.,* 1995) and later in soil science by Knotters *et al.,* (1995) and Odeh *et al.,* (1995). Regression kriging recognizes that the realization of an intrinsic random function, Z(x), is uncertain and this uncertainty is expressed as

$$Z(x) = z(x_i) + \varepsilon(x_i) \qquad \text{Equation 6.1}$$

where Z(x) is the intrinsic random function, $z(x_i)$ is the true value and $\varepsilon(x_i)$ is the residual error that represents the uncertainty in the system. This uncertainty may not necessarily be due entirely to the result of the regression prediction but also to other factors e.g. measurement error (Knotters *et al.,* 1995) and the broader term "kriging with uncertain data" has been suggested (Ahmed and DeMarsily, 1987). The residual errors, $\varepsilon(x_i)$, are assumed to be unsystematic, uncorrelated among themselves and uncorrelated with the variables. Initially the residual errors were incorporated into the ordinary kriging system by replacing the variances in the diagonal of the ***A*** matrix and the kriging equation was subsequently modified.

In situations where the ancillary variables are known at all the prediction points Odeh *et al.* (1995) have proposed an alternative approach which does not need modification of the *A* matrix. The regression model is established and the residuals extracted. The model is applied to the ancillary variables at each point on the raster to form an initial prediction. The residuals are kriged onto the raster and summed with the initial prediction to get a final prediction. This process is shown in Figure 6.1.

As for any regression analysis care must be taken to avoid co-linearity within the variables. This can be minimised through stepwise procedures or the application of a factor analysis to produce composite indices or Principal Components for use in the regression model (Hengl *et al.*, 2004).

6.2.6.1 Models

Regression kriging may utilise a wide variety of different regression methodologies and three different techniques have been used in this study to identify a preferred approach. These techniques are, Linear Regression, Generalized Additive Models and Neural Networks.

Multiple Linear and Stepwise linear regression (MLR/SLR): This approach utilises linear relationships to try and model the predictor variables to the dependent or target variable. Many of the variables used in the model may be strongly correlated, e.g. 0-90 cm EC_a and 30-90 cm EC_a or DTM attributes, or not correlated at all to the target variable and therefore redundant in the model. Multiple linear regression (MLR) was first run using all variables in the model. Stepwise linear regression (SLR) is a technique that allows the removal of ineffective predictor variables from the model to improve the parsimony of the model. A model is initially defined and then an iterative process of adding or subtracting predictor variables from the model and assessing the impact on the model against some assigned criteria is employed to remove ineffectual predictor variables (SAS institute Inc., 2002). SLR was performed in JMP® using a mixed model approach (i.e. both forward and backward stepping) with all variables initially entered in the model.

The data used was not transformed as the landform variables did not exhibit skewed distributions. Hengl *et al.* (2004) have proposed that logit transformation is required for landform variables however the problematic variables, upslope/downslope areas and secondary DTM attributes such as twi, are not used in this study.

Generalised Additive Model (GAM): The purpose of generalized additive models is to maximize the quality of prediction of a dependent variable from various distributions, by estimating unspecific (non-parametric) functions, often smoothing splines or loess (local regression), of the predictor variables that are "connected" to the dependent variable via a link function (SAS institute Inc., 2002). They are a melding of traditional additive models and generalized linear models. For further information on GAM readers are directed to Hastie and Tibshirani (1990).

The GAM analysis was performed in S-PLUS with three different approaches. Initially the model was run with all parameters (GAMall) using smoothing splines functions only. A stepwise GAM analysis (sGAM) in S-PLUS was then employed to improve model parsimony and avoid over-parameterisation that can be problematic in non-parametric techniques. All variables were entered and modeled either linearly, as smoothing splines or removed from the analysis. Minimisation of the Akaike Information Criteria (AIC) was used to identify the best model. An alternative approach to avoiding over-parameterisation is the use of principal component analysis (PCA). PCA was performed in JMP® on 9 of the ancillary variables (Eastings and Northings were excluded from the PCA) and 9 principal components stored. A Stepwise GAM analysis (sGAMpca), as described above, was run using the 9 principal components and Eastings and Northings as input variables.

Neural Network Analysis (NNA): NNA is an alternative method of fitting the predictor variables to the target variable non-parametrically. It is a method that seeks to simulate the human learning process using linear and S-shaped functions (SAS Institute Inc., 2002). The NNA was performed in JMP® using one layer with 3 hidden nodes. The NNA model was initially run with all parameters (NNAall) and subsequently using the output from the PCA (NNApca). Unlike the GAM approach a stepwise reduction of parameters was not available for NNA. Instead a correlation analysis was performed and the principal components which had a correlation of >0.1 or <-0.1 with a particular dependent soil property were chosen in the NNApca model for that particular soil property.

6.2.6.2 Interpolation

Seven different models were tried for regression kriging of soil property prediction; Multiple linear regression (MLR), Stepwise linear regression (SLR), GAM with all parameters (GAMall), Stepwise GAM (sGAM), Stepwise GAM using PCA outputs (sGAMpca), NNA with all parameters (NNAall) and NNA with PCA (NNApca). In addition the results from ordinary kriging (OK) (from Chapter 5)

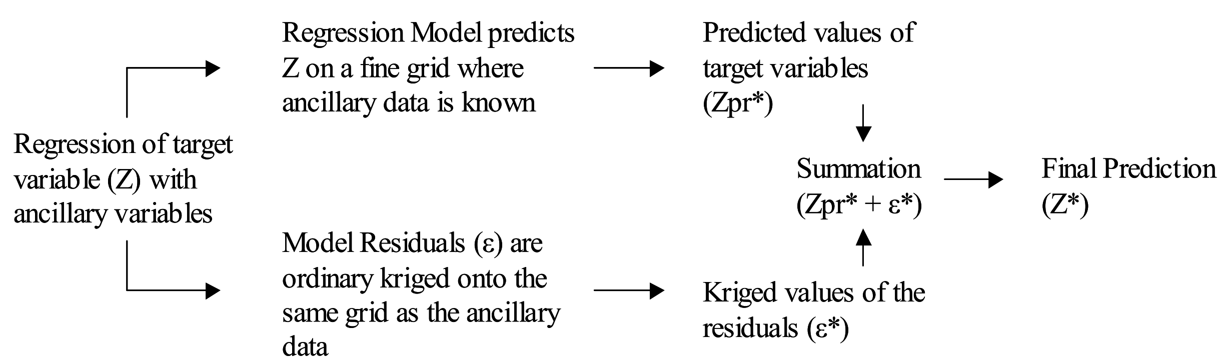

Figure 6.1: Steps involved in the regression kriging approach of Odeh *et al.* (1995) (adapted from Odeh *et al*, 1995)

were also included for discussion.

Regression equations were derived for each soil property for each approach using the "regression" dataset based on the soil survey points. The regression equations for each approach are given in Appendix 1 and the fit of the models shown in Table 6.1. The regression equations were applied to the prediction set to form an initial prediction (Zpr^*) across the vineyard. The residuals from the models were extracted and kriged onto the prediction grid using punctual ordinary kriging with a global variogram in Vesper (Minasny *et al.*, 2002) The interpolated residuals (e^*) were then added to (Zpr^*) to form the final prediction (Z^*). This process was repeated with the "validation" set to form a prediction for each soil property at the validation sites.

Results of the fit of the models and the validation data are presented as r^2 values and RMSE. The r^2 and RMSE for the regression models were ranked and an analyses of the mean rank vs standard deviation of rank performed (Laslett *et al.*, 1987)

6.3 Results and Discussion

6.3.1 Fit of Models

Table 6.1 gives details of the fits of the regression models. In general the linear regression (MLR and SLR) produced the worst fits, the neural network approaches (NNA, NNApca) the best and the GAM models (GAMall, sGAM and sGAMall) were inbetween. This is expected and reflects the complexities of the models used to fit the "regression" dataset.

6.3.2 Validation of Models

The r^2 and RMSE values for the three vineyards are given individually in Appendix 6.3. The full model outputs have been ranked from 1 to 7 in order of best fit and the summed ranking is presented in Appendix 2. Table 6.1 presents r^2 and RMSE statistics for the combined data from all three vineyards. Maps of all soil properties for each vineyard are shown in Appendix 3.

The rankings of the r^2 and RMSE values for all the models (Appendix 6.1) were very similar indicating that they are providing similar information on the fit of the data. Given this the following discussion will focus on the RMSE statistic.

The Neural Network methods produced the largest RMSE results for topsoil and subsoil clay estimation at both Cowra and Pokolbin. In contrast, at Canowindra they produced the best fit to the validation data however, when mapped, the data is noisy. In general the patterns observed in the NN maps were less spatially coherent and did not always follow the trend shown in the OK maps. The NNA and NNApca models also tended to produce over- and under-predictions in the datasets, particularly for Pokolbin. The large inconsistency in the response from the NN models (high standard deviation of ranking in Figure 6.2) would also indicate that Neural Networks are unsuited in this situation for combining soil and ancillary data in vineyards. Neural networks require an extensive training set for accurate results and there may be insufficient data in the regression datasets. Little emphasis was put on the analysis and different permutations of layers and nodes were not tried which, if attempted, may have helped to improve the final fit. Neural Networks have been used successfully previously for field-scale studies (McBratney *et al.,* 2000) however results from these three vineyards indicate that NNA is not suitable in vineyards.

The MLR and GAMall models, where all data layers were incorporated into the model, produced the highest mean ranks (Figure 6.2). In contrast the stepwise GAM and SLR approaches produced the three best mean rank results. The NNApca model was also superior to the NNA model with all parameters included. This emphasises the need for data reduction in these models to avoid over-parametising and co-linearity between variables. From Figure 6.2 the sGAM and sGAMpca produced the two lowest mean ranks with the sGAMpca marginally better. The SLR model showed the most consistent response of the models (lowest standard deviation). When the data from the three vineyards are combined the sGAMpca produced the best fit for two of the three soil variables (30 cm clay and RAW). SLR produced the best combined fit for 30-90 cm clay. On this basis the sGAMpca approach is considered to be the preferred method for regression kriging mapping of vineyard soil data.

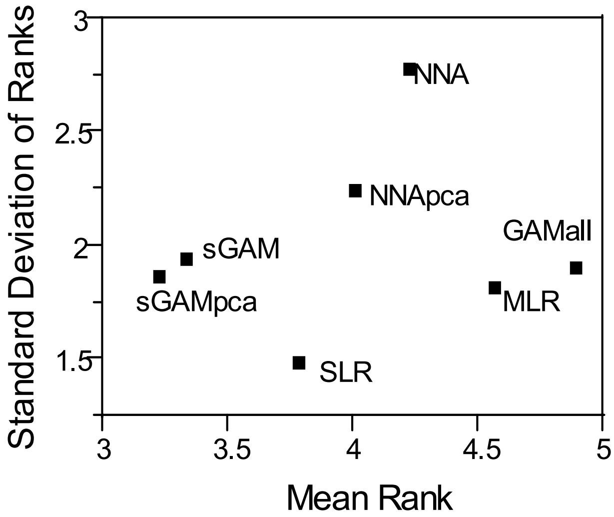

Figure 6.2 Plot of Mean Rank against the Standard deviation of Ranks for the seven regression kriging models used in this study.

6.3.3 Regression Kriging vs Ordinary Kriging

A comparison of the final maps from OK and RK (sGAMpca) are shown in Figures 6.3 - 6.5. In this study the OK approach produced lower RMSE than the sGAMpca RK method for all soil properties in the combined data set (Table 6.1). This contrasts with previous studies (Odeh *et al.*, 1995, Bishop and McBratney, 2001, Knotters *et al.*, 1995 and Goovaerts, 1999) who all reported improved predic-

	r^2			RMSE		
	Clay% (0-30cm)	**Clay% (30-90cm)**	**RAW (mm)**	**Clay 30 (0-30cm)**	**Clay 90 (30-90cm)**	**RAW (mm)**
Combined (n=38)						
MLR	0.578	0.120	0.034	7.984	13.354	9.475
SLR	0.624	0.205	0.050	7.542	12.691	9.349
NN	0.438	0.000	0.055	9.215	14.233	9.370
Nnpca	0.093	0.088	0.021	11.708	13.590	9.539
GAMall	0.556	0.024	0.051	8.191	14.062	9.391
sGAM	0.469	0.111	0.051	8.955	13.424	9.390
sGAMpca	0.652	0.060	0.190	7.251	13.797	8.677
OK	0.763	0.544	0.236	5.980	9.615	8.427

Table 6.1: Combined r^2 and RMSE of RK model responses to soil properties (topsoil clay%, subsoili clay% and RAW (mm)) for the soil validation data from Pokolbin, Canowindra and Cowra vineyards dataset from all vineyards.

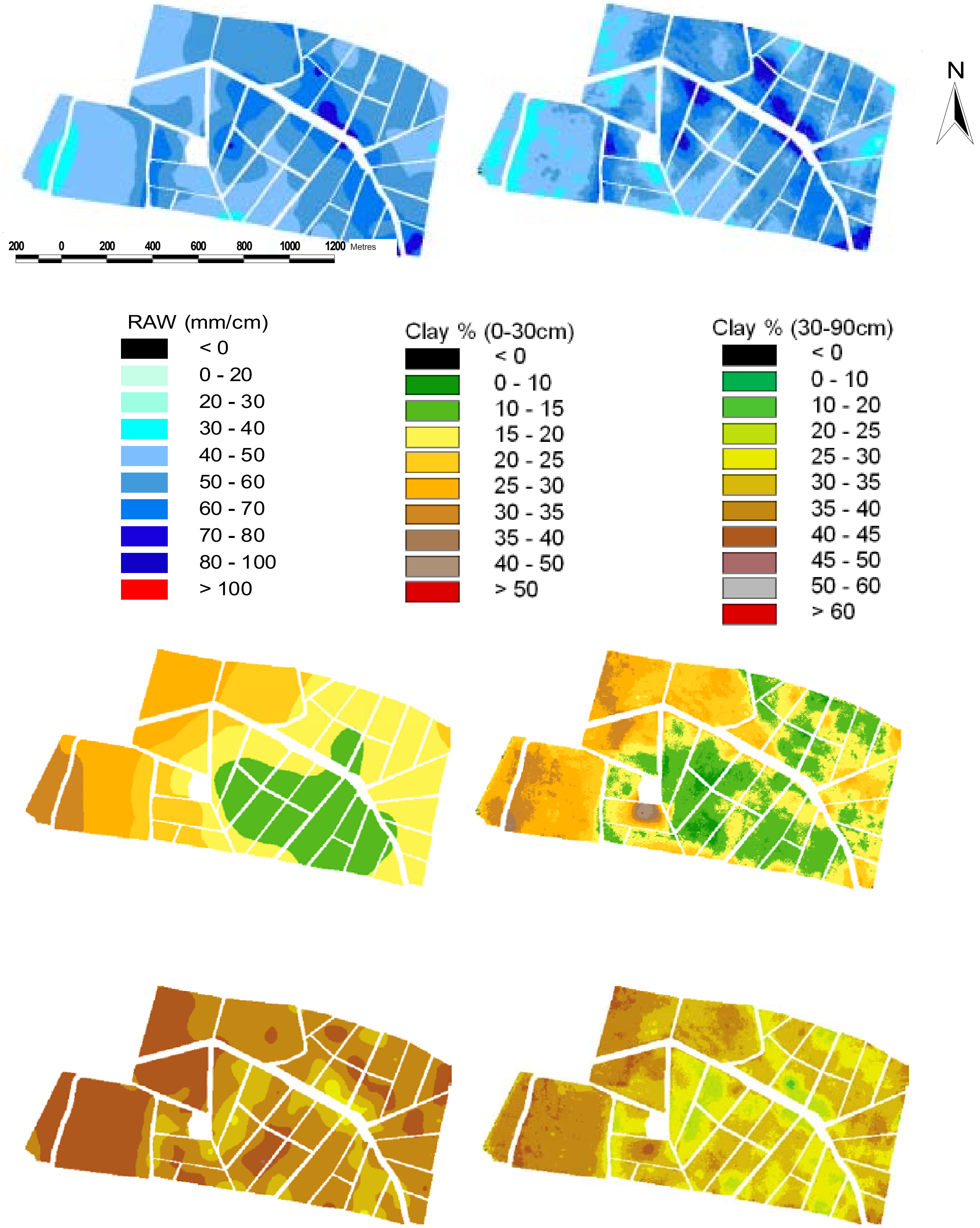

Figure 6.3: Predicted maps at Cowra for RAW (top), 0-30cm Clay% (middle) and 30-90cm Clay% (bottom) using OK (LHS) and sGAMpca RK (RHS).

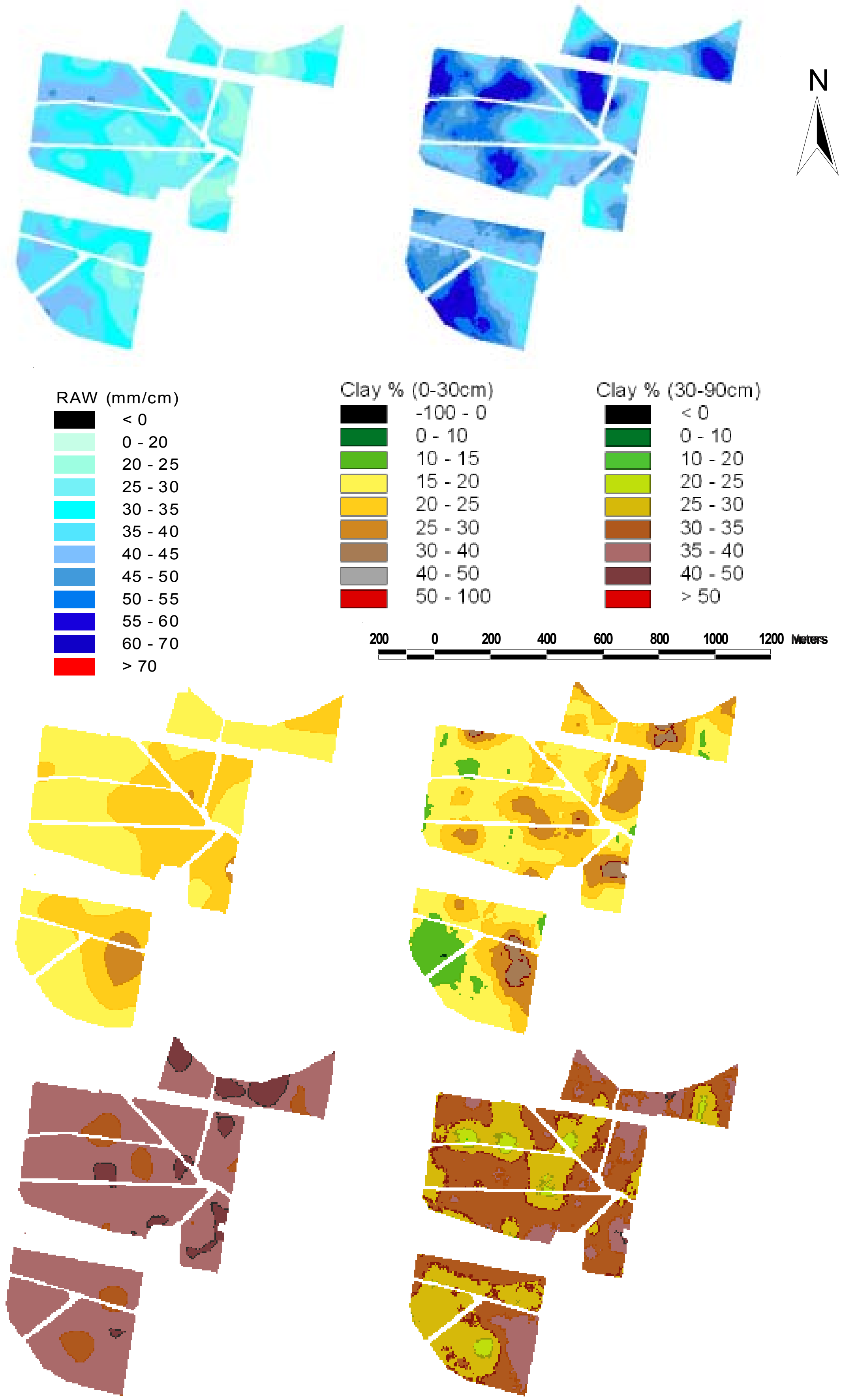

Figure 6.4: Predicted maps at Canowindra for RAW (top), 0-30cm Clay% (middle) and 30-90cm Clay% (bottom) using OK (LHS) and sGAMpca RK (RHS).

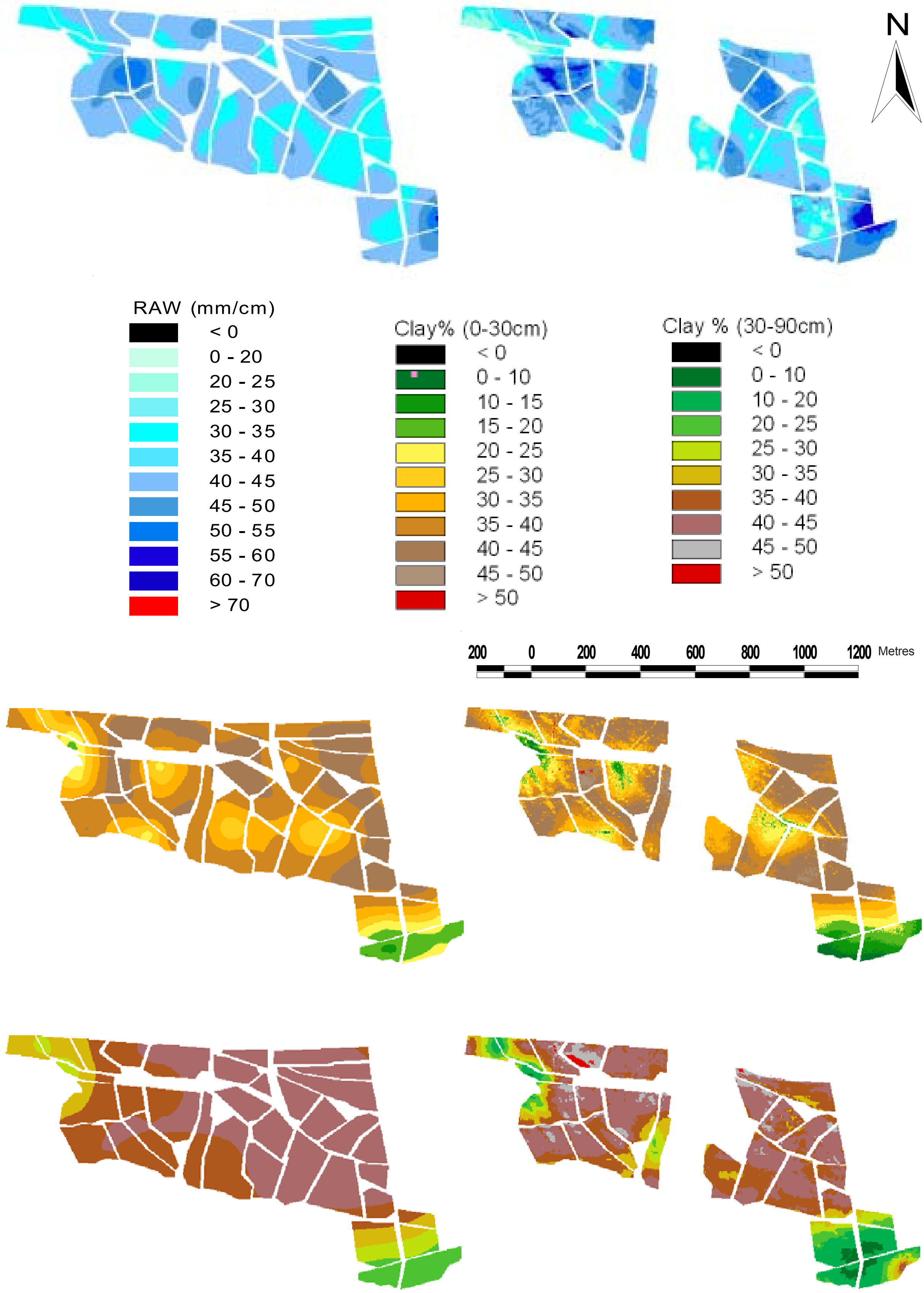

Figure 6.5: Predicted maps at Pokolbin for RAW (top), 0-30cm Clay% (middle) and 30-90cm Clay% (bottom) using OK (LHS) and sGAMpca RK (RHS).

tions using RK. There are several possible reasons for the contradictory results in this study:

i) Vineyard surveys are more intensive than the studies used previously for comparisons of RK and OK. The RMSE of RK prediction for clay was similar to that observed by McBratney *et al.* (2000) and it appears that the denser data favours OK especially when the soil survey sites are not site directed (see below)

ii) The mixture of field and laboratory measurements used for the soil survey and the validation set may also have introduced some error and bias in the analysis. The influence of the different methodologies could be eliminated by jack-knifing or boot-strapping (Bishop, 2003) the soil survey data to produce a validation set.

iii) The OK approach maintained predicted values within the range of the original soil survey data and showed the broad trends in the data. The regression predictions had a range greater than the soil survey data. A problem with the methodology used is that the soil surveys were done independently of the ancillary surveys. As a result the soil survey locations were not directed to account for the variation observed in the ancillary data. In practice, ancillary data is usually collected prior to the soil survey. If a site-directed approach is used for some of the soil sites, to account for the range of variability in the ancillary data, then the data range of the regression predictions should approach that of the soil survey and improve predictions.

While the RMSE is higher the RK methodologies did produce more detail in the final maps. However the validity of the extra detail needs to be questioned especially given the higher RMSE values. Certainly there are some artifacts in the maps from the presence of erroneous data in the original ancillary data. However this extra detail may help to better design the vineyard layout even if the accuracy of prediction within the blocks is lower. Correctly defining the areas of similarity within the natural environment (or digital terroirs) is as important, if not more so, than properly measuring the environment within these "digital terroirs". In the case of the Cowra vineyard it appears that the extra detail is unlikely to help improve vineyard design greatly. No new soil features were uncovered. Soil boundaries have possibly been more accurately defined but given the size of blocks and other constraints to vineyard design (sucha s roads and waterways) this may have little effect on the final design. The strong catenary patterns in this vineyard mean regression kriging has little advantage over ordinary kriging. However in more soil heterogeneous vineyards regression kriging may provide better information for vineyard design. Certainly in the Canowindra vineyard there is a lot more detail in the soil texture maps derived from sGAMpca analysis. The Pokolbin vineyard also demonstrates more variability in the RK maps.

The RMSE statistic does not help us identify if the OK interpolation is properly defining soil/environmental boundaries. Since the object of vineyard design is to minimise variability within blocks the best indication of whether or not a vineyard has been well designed is to investigate the vines response within the block. Unfortunately this data is not usually available when designing vineyards, unless the vineyard is being redeveloped. However vine response, in the form of canopy imagery and yield monitoring, can be used to test the validity of digital terroir predictions in existing vineyards.

Regression kriging with a Stepwise GAM model using Principal Components was found to be the best model for regression kriging. However several other models and 'hybrid' geo-statistical approaches were not tried and should be tested in future studies. Co-kriging has been found to be inferior to RK (Odeh *et al.*, 1995, McBratney *et al.*, 2000) however should be tested to validate this. Alternative regression models such as Generalised Linear Models (GLM) and Regression Trees have not been tested.

6.4 Conclusions

The aim of this study was to try and incorporate ancillary data with traditional soil survey data to improve our knowledge of soil properties prior to vineyard design. The replacement of ordinary kriging with regression kriging produced more detail in the maps however the models produced inferior fits to the ordinary kriging when compared to the validation dataset. Over-parametrisation was a problem in the models and the best result from the regression kriging results were produced from stepwise models (SLR, sGAM and sGAMpca). The soil survey was also done independently of the ancillary data and there may be discrepancies in the spatial variability between the soil and ancillary data that is contributing to the poor RK responses. In new vineyards, the soil sampling should reflect the information in the ancillary data and thus the soil and ancillary datasets should be more complimentary.

The RK methodology is a fairly complex geostatistical approach. In this regard it is not easily adopted by a lay person. In this regard the methodology is limited to those with access to data analysts with a geostatistical background. Currently these people are in short supply in the agriculture sector however with an increased demand this niche should be filled. Also, similar to OK, the RK methodology requires a certain number of data (>70 data points), thus is suited to larger vineyards.

The validity of the extra detail in the RK maps for defining "digital terroirs" is still unclear. This will be investigated in the next chapter using aerial imagery and yield maps.

6.5 References

Ahmed, S and DeMarsily, G. (1987) *Comparison of geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity.* Water Resource Research, 23, pp1717-1737

Bishop, T.F.A. (2003) Geo-data Processing Methods for site specific crop management. PhD Thesis. The University of Sydney

Bishop, T.F.A., and McBratney, A.B. (2001) *A comparison of prediction methods for the creation of field-extent soil property maps.* Geoderma 103 (1–2), pp149–160.

Dabas, M., Tabbagh, J. and Boisgontier, D. (2001) *Multi-depth continuous electrical profilihg (MuCep) for characterization of in-field variability.* In: G. Greiner and S. Blackmore (eds.). Proceedings of the Third European Conference on Precision Agriculture. June 18-20, 2001, Montpellier, France. pp361-366

Erdas (2002) Erdas IMAGINE 8.6

Goovaerts, P. (1999) *Using elevation to aid the geostatistical mapping of rainfall erosivity.* Catena 34 3–4, pp227–242

Hastie, T. and Tibshirani, R. (1990) *Generalized additive models.* Chapman and Hall, London

Hengl, T., Heuvelink, G.B.M., and Stein, A. (2004) *A generic framework for spatial prediction of soil variables based on regression-kriging*, Geoderma, In Press, Corrected Proof, Available online 19 December 2003

Jenny, H. (1941) *Factors of Soil Formation: A System of Quantitative Pedology*, McGraw-Hill, New York

281p.

Knotters, M., Brus, D.J. and Oude Voshaar, J.H. (1995) *A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations.* Geoderma, 67, pp227-246

McKenzie, D. (2000) *Soil Survey options prior to vineyard design.* The Australian Grapegrower and Winemaker Annual Technical Issue, 438a, pp144-151

Laslett, G.M., McBratney, A.B., Pahl, P.J. and Hutchinson, M.F. (1987) *Comparison of several spatial prediction methods for soil pH.* Journal of Soil Science, 38, pp325–341.

McBratney, A.B., Odeh, I.O.A., Bishop, T.F.A., Dunbar, M. and Shatar, T.M. (2000) *An overview of pedometric techniques of use in soil survey.* Geoderma 97(3–4), pp293–327.

McBratney, A.B., ,Mendonça Santos, M.L. and Minasny, B. (2003) *On digital soil mapping.* Geoderma 117, pp3–52

Minasny, B., McBratney, A.B., and Whelan, B.M. (2002) VESPER version 1.5. Australian Centre for Precision Agriculture, McMillan Building A05, The University of Sydney, NSW 2006. (http://www.usyd.edu.au/su/agric/acpa)

Moore, I.D., Grayson, R.B., Ladson, A.R. (1991) *Digital terrain modelling: a review of hydrological, geomorphological, and biological applications.* Hydrological Processes, 3, pp3-30.

Odeh, I.O.A., McBratney, A.B., and Chittleborough, D.J. (1995) *Further results on the prediction of soil properties from terrain attributes: heterotopic cokriging and regression kriging.* Geoderma, 67, pp215-226.

SAS Institute Inc. (2002) JMP® Statistics and Graphics Guide, Version 5. Cary, NC, USA

Appendix 6.1 - Regression Equations

For the GAMall approach the same equations was used for each Vineyard. The Neural Network equations are not given. Due to nodal equations these are quite convoluted and the failure of the approach makes them some what irrelevant. For the other approaches, the regression model equations are listed under the vineyard.

Universal equations

GAMall
Clay30 ~ s(V30) + s(V90) + s(V3090) + s(elevation) + s(aspect) + s(slope) + s(X) + s(planc) + s(profc) + s(Y)
Clay90 ~ s(V30) + s(V90) + s(V3090) + s(elevation) + s(aspect) + s(slope) + s(X) + s(planc) + s(profc) + s(Y)
RZRAW ~ s(V30) + s(V90) + s(V3090) + s(elevation) + s(aspect) + s(slope) + s(X) + s(planc) + s(profc) + s(Y)

Cowra

MLR
0-30cm Clay% = (-7288.21173902268) + -0.00335647046281367 * :X + 0.00151318743250676 * :Y + 0.152715645393666 * :elevation + 0.0123004318388898 * :aspect + -0.687268425821615 * :slope + -0.00295744168475389 * :flowd + 3.4344074069128 * :profc + -0.576852157592946 * :planc + 0.0739547080504255 * :V30 + 0.823906731797491 * :V90 + -0.454776999195015 * :V3090

30-90cm Clay% = (-11103.298932487) + -0.000872280922562966 * :X + 0.00187087640990793 * :Y + 0.0260071955311905 * :elevation + 0.00819630576449054 * :aspect + 0.482284261946008 * :slope + 0.0007915141694789 * :flowd + -5.22867842007033 * :profc + -1.65449779399077 * :planc + -0.0544470167858797 * :V30 + -0.0243104875124756 * :V90 + 0.175485320121808 * :V3090

RAW = 21323.6852143531 + 0.00922587000827356 * :X + -0.00433534348959897 * :Y + -0.698340743126899 * :elevation + -0.00558852683677011 * :aspect + -1.91231669463404 * :slope + 0.0197635828710486 * :flowd + 10.7642182065199 * :profc + 0.773474118764151 * :planc + 0.313249731404086 * :V30 + -1.19599758007894 * :V90 + 0.535722524126046 * :V3090

SLR
0-30cm Clay% = 2034.41603437609 + -0.00313353905520471 * :X + 0.114008339008906 * :elevation + 0.013873940342518 * :aspect + -0.689856545937654 * :slope + 0.966289600004594 * :V90 + -0.548027045187141 * :V3090

30-90cm Clay% = (-10449.2062139997) + 0.00167589775609224 * :Y + 0.00761936717569967 * :aspect + 0.548318734289852 * :slope + -4.01100409069783 * :profc + 0.153496324582221 * :V3090

RAW = 22306.9824166719 + 0.00912925271696523 * :X + -0.00448285315242564 * :Y + -0.69596510860798 * :elevation + -1.94338264072288 * :slope + 0.0302756476250088 * :flowd + 10.0671127031632 * :profc + -0.386601649691214 * :V90

sGAM
Clay30 ~ aspect + s(slope) + s(planc) + V90 + s(V3090) + X
Clay90 ~ s(elevation) + aspect + s(slope) + s(V90)
RZRAW ~ elevation + slope + s(planc) + profc + s(V30) + V90 + X + Y

sGAMpca
Clay30 ~ s(Prin1) + s(Prin2) + s(Prin4)
Clay90 ~ s(Prin1) + Prin2 + s(Prin3) + Prin4 + Prin7 + Prin8
RZRAW ~ s(Prin1) + Prin2 + s(Prin3) + Prin4 + Prin7 + X

Canowindra

MLR
0-30cm Clay% = (-10272.7579084567) + 0.00344526346602669 * :X + 0.00128018858051144 * :Y + 0.0456885246798395 * :elevation + -0.00407035576973728 * :aspect + 0.353758750961685 * :slope + -0.00247288646293319 * :flowd + 1.88226350877652 * :profc + 2.20089571127647 * :planc + -0.335666290073337 * :V30 + 2.64885554007646 * :V90 + -1.9124663528251 * :V3090

30-90cm Clay% = (-4424.44051912363) + 0.00314235767955931 * :X + 0.000385805796456219 * :Y + -0.018390210142314 * :elevation + -0.00424699308811194 * :aspect + -0.19557610488601 * :slope + -0.00269425721509297 * :flowd + -0.856099563239785 * :profc + 3.40449305949756 * :planc + 0.166036454899804 * :V30 + -0.202725808092715 * :V90 + 0.200156091684391 * :V3090

RAW = (-6626.91104092006) + -0.000381199443699272 * :X + 0.00111895778665659 * :Y + -0.347522905825029 * :elevation + 0.00190555387691101 * :aspect + 0.2063004701625 * :slope + 0.0206647427700119 * :flowd + 0.551091276980998 * :profc + 2.86200280727824 * :planc + 0.425934687960023 * :V30 + -1.93728288568341 * :V90 + 1.51244439765821 * :V3090

SLR
0-30cm Clay% = (-3332.23302190107) + 0.00517113181553377 * :X + 1.82575709097714 * :V90 + -1.38589957640295 * :V3090

30-90cm Clay% = (-2529.46003528479) + 0.00394733114358254 * :X + 4.07204921933393 * :planc + 0.0775612533257957 * :V3090

RAW = 166.275161420507 + -0.372178315672615 * :elevation + 0.175245576546392 * :V3090

sGAM
Clay30 ~ elevation + V90 + V3090 + s(Y)
Clay90 ~ elevation + aspect + s(planc) + s(Y)
RZRAW ~ elevation + s(profc) + V3090 + s(Y)

sGAMpca
Clay30 ~ Prin1 + Prin2 + Prin4 + Prin5 + s(Y)
Clay90 ~ s(Prin3) + s(Prin4) + X
RZRAW ~ Prin1 + Prin2 + Prin5 + s(Prin7)

Pokolbin

MLR
0-30cm Clay% = (-237437.316896067) + -0.509737853442759 * :elevation + -0.00061561162058188 * :aspect + 0.906139318717722 * :slope + -0.0263140901010238 * :flowd + 1.82946317059815 * :profc + 0.593829501483071 * :planc + -0.196627750759402 * :V30 + 0.184114928880525 * :V90 + -0.0717659658431589 * :V3090 + -0.00283016498747716 * :X + 0.0373984643270482 * :Y

30-90cm Clay% = (-207177.941213001) + -0.468498655502209 * :elevation + 0.0066126445486503 * :aspect + 1.79588670469082 * :slope + 0.0338505358179851 * :flowd + 1.18204397840833 * :profc + 4.64501999714331 * :planc + -0.0652692704278525 * :V30 + 0.316600477201543 * :V90 + -0.233055499717982 * :V3090 + -0.00318629482737473 * :X + 0.0326704762373793 * :Y

RAW = 45181.9454477285 + -0.0300962996746195 * :elevation + -0.021177327557501 * :aspect + 1.21222475146004 * :slope + -0.0163392372706334 * :flowd + 0.906161372964967 * :profc + -6.86554039457406 * :planc + -0.135592471619576 * :V30 + -0.309813186779561 * :V90 + 0.289963475734334 * :V3090 + 0.00582232787526414 * :X + -0.00738456473124274 * :Y

SLR
0-30cm Clay% = (-239464.85917179) + 0.0375662270691156 * :Y + -0.449103292792099 * :elevation + 0.981316757887239 * :slope + -0.0316010884841454 * :flowd + -0.158593044132458 * :V30 + 0.0740260391802708 * :V90

30-90cm Clay% = (-211726.822582739) + 0.0332149070877793 * :Y + -0.403892177869599 * :elevation + 1.89797086645907 * :slope + 3.6295155735105 * :planc

RAW = (-2970.12034532094) + 0.00896915563813847 * :X + 0.97139490512123 * :slope + -7.01889914856742 * :planc + -0.247379092350317 * :V30 + 0.0829533893122906 * :V3090

sGAM
Clay30 ~ elevation + s(Y)
Clay90 ~ s(elevation) + X
RZRAW ~ s(elevation) + s(slope) + V90 + V3090

sGAMpca
Clay30 ~ s(Prin4) + s(Y)
Clay90 ~ s(Prin3) + s(Prin4) + X
RZRAW ~ Prin1 + Prin2 + Prin5 + s(Prin7)

Appendix 6.2 - Fits (r^2 and RMSE) of Regression Models to the 'Regression' dataset

	Clay% 0-30cm	Clay% 30-90cm	RAW (mm/cm)	Clay 0-30cm	Clay 30-90cm	RAW (mm/cm)
Cowra (n=15)						
MLR	0.552	0.466	0.463	4.960	3.792	8.148
SLR	0.548	*0.463*	*0.460*	4.985	*3.805*	*8.171*
NNA	**0.798**	**0.721**	**0.675**	**3.334**	**2.742**	**6.336**
NNAPCA	0.667	0.612	0.608	4.281	3.234	6.960
GAMall	0.679	0.617	0.585	4.205	3.218	7.172
SGAM	*0.502*	0.580	0.519	*13.111*	3.367	7.715
SGAMPCA	0.521	0.617	0.514	13.074	3.218	7.753
Canowindra (n=11)						
MLR	0.101	0.078	0.114	5.35	4.67	9.13
SLR	*0.90*	*0.066*	*0.110*	*5.38*	*4.70*	*9.15*
NNA	**0.457**	**0.404**	**0.384**	**4.16**	**3.76**	**5.86**
NNAPCA	0.132	0.168	0.270	5.25	4.44	8.29
GAMall	0.265	0.257	0.302	4.86	4.25	8.15
SGAM	0.124	0.102	*0.110*	5.28	4.62	*9.15*
SGAMPCA	0.115	0.100	0.130	5.31	4.62	9.06
Pokolbin (n=12)						
MLR	*0.320*	0.444	0.147	9.25	7.94	13.56
SLR	0.324	*0.420*	*0.126*	*9.29*	*8.11*	*13.73*
NNA	**0.972**	**0.983**	**0.867**	**1.88**	**1.40**	**5.36**
NNAPCA	0.467	0.883	0.711	8.25	3.64	7.89
GAMall	0.704	0.710	0.711	6.26	5.77	8.24
SGAM	0.462	0.569	0.419	8.30	7.02	11.31
SGAMPCA	0.554	0.549	0.458	7.62	7.15	10.98

Appendix 6.3 - Regression Kriging Model Fits (R2 and RMSE) for Individual Vineyards to the "Validation" dataset

	r^2			RMSE		
	Clay% (0-30cm)	Clay% (30-90cm)	RAW (mm)	Clay 30 (0-30cm)	Clay 90 (30-90cm)	RAW (mm)
Cowra						
MLR	0.528^{5}	0.469^{1}	0.178^{6}	5.421^{5}	7.131^{1}	8.501^{6}
SLR	0.591^{4}	0.460^{2}	0.213^{4}	5.048^{4}	7.192^{2}	8.315^{4}
NNA	0.217^{7}	0.400^{7}	0.334^{1}	6.986^{7}	7.580^{7}	7.652^{1}
NNApca	0.324^{6}	0.427^{5}	0.218^{3}	6.488^{6}	7.406^{6}	8.290^{3}
GAMall	0.601^{3}	0.430^{6}	0.126^{7}	4.984^{3}	7.392^{5}	8.765^{7}
sGAM	0.735^{1}	0.431^{4}	0.204^{5}	4.064^{1}	7.383^{4}	8.364^{5}
sGAMpca 0.701^{2}	0.440^{3}	0.232^{2}	4.319^{2}	7.324^{3}	8.215^{2}	
OK	0.731	0.426	0.265	4.097	7.418	8.035
Canowindra						
MLR	0.085^{4}	0.111^{7}	0.015^{4}	5.700^{4}	15.572^{7}	9.353^{4}
SLR	0.069^{6}	0.117^{5}	0.019^{3}	5.750^{6}	15.525^{6}	9.334^{3}
NNA	0.323^{2}	0.356^{2}	0.135^{1}	4.903^{2}	13.256^{2}	8.765^{1}
NNApca	0.343^{1}	0.438^{1}	0.031^{2}	4.829^{1}	12.385^{1}	9.273^{2}
GAMall	0.075^{5}	0.117^{5}	0.002^{6}	5.730^{5}	15.518^{5}	9.414^{6}
sGAM	0.110^{3}	0.165^{3}	0.001^{7}	5.621^{3}	15.094^{3}	9.416^{7}
sGAMpca 0.045^{7}	0.152^{4}	0.011^{5}	5.822^{7}	15.212^{4}	9.368^{5}	
OK	0.751	0.265	0.179	2.973	14.165	8.537
Pokolbin						
MLR	0.111^{5}	0.357^{3}	0.008^{6}	12.107^{5}	11.764^{3}	9.824^{6}
SLR	0.229^{4}	0.735^{2}	0.047^{3}	11.279^{4}	7.557^{2}	9.632^{3}
NNA	0.076^{6}	0.024^{7}	0.017^{4}	12.345^{7}	14.492^{7}	9.782^{4}
NNApca	0.083^{7}	0.027^{6}	0.013^{5}	12.301^{6}	14.477^{6}	9.801^{5}
GAMall	0.599^{1}	0.243^{5}	0.003^{7}	8.136^{1}	12.765^{5}	9.853^{7}
sGAM	0.438^{2}	0.347^{4}	0.124^{1}	9.627^{2}	11.852^{4}	9.234^{1}
sGAMpca 0.399^{3}	0.852^{1}	0.048^{2}	9.960^{3}	5.643^{1}	9.627^{2}	
OK	0.537	0.814	0.032	8.741	6.331	9.704