0038-075X/03/16807-501–515
Soil Science

July 2003
Vol. 168, No. 7
Printed in U.S.A.

SPATIAL PREDICTION OF SOIL PARTICLE-SIZE FRACTIONS AS COMPOSITIONAL DATA

Inakwu O.A. Odeh[1], Alison J. Todd[1], and John Triantafilis[1]

Particle-size fractions (psf) of mineral soils and, hence, soil texture, are the most important attributes affecting physical and chemical processes in the soil. More often, psf data are available only at a few locations for a given area and, therefore, require some form of interpolation or spatial prediction. However, psf data are compositional and, therefore, require special treatment before spatial prediction. This includes ensuring positive definiteness and a constant sum of interpolated values at a given location, error minimization, and lack of bias. In order to meet these requirements, this study applied two methods of data transformation prior to kriging of the psf of soils in two regions of eastern Australia. The two methods are additive log-ratio transformation of the psf (ALR_{OK}) and modified log-ratio transformation ($mALR_{OK}$). The performance of the transformed values by ordinary kriging was compared with the spatial prediction of the untransformed psf data using ordinary kriging, compositional kriging (CK) (UT_{OK}), and cokriging, based on the criteria-prediction bias or mean error (ME) and precision (root mean square error (RMSE)), and validity of textural classification. ALR_{OK} and $mALR_{OK}$ outperformed UT_{OK} and CK in terms of prediction ME and RMSE. Because of the closure effect on the psf data, UT_{OK}, and, to a lesser extent, CK, did not meet all of the requirements for spatially predicting compositional data and, therefore, performed poorly. $mALR_{OK}$ outperformed all of the interpolation methods in terms of misclassification of soils into textural classes. The results show that without considering the special requirements of compositional data, spatial interpolation of psf data will necessarily produce uncertain and unreliable interpolated psf values. (Soil Science 2003;168:501–515)

Key words: Compositional soil data, particle-size fractions, log-ratio transformation, kriging, spatial prediction.

A regionalized composition is characterized by components that (i) can be modeled by *a spatial random function,* (ii) are *positive definite,* and (iii) *sum to a constant* (Pawlowsky, 1984). The study of compositional data therefore should be concerned with the relative values or ratios of the components. It is meaningless to evaluate each component in isolation. Pearson (1897) was the first to identify this problem of statistical analysis of compositional data. The problem has been referred to as a "fallacy of interpretation" (Woronow and Love, 1990) or "spurious spatial correlation" as a result of the "closure effect" (Pawlowsky, 1984). That a composition must have at least one negative correlation between a pair of its components is caused by the closure effect. A change in one component, therefore, results in a shift of all other components. However, an infinite number of different combinations or changes of the composition of the components could produce the same shifts without showing evidence of what changes have actually occurred. This invariably leads to difficulties in interpreting the corre-

[1]Australian Cotton Co-operative Research Centre, Faculty of Agriculture, Food and Natural Resources, The University of Sydney, Ross St. Building A03, NSW, 2006, Australia. Dr. Odeh is corresponding author. E-mail: i.odeh@acss.usyd.edu.au

Received Sept. 12, 2002; accepted March 24, 2003.

DOI: 10.1097/01.ss.0000080335.10341.23

lations. In geostatistical parlance, the covariance is negative, and in some instances, singular variance-covariance matrices cause the cokriging equations to be singular as well. In addition, kriging of separate components of a regionalized composition may not produce estimates that sum to a constant, a strict requirement of compositions. Therefore, the unbiased constraint fundamental to kriging may not be fulfilled.

The spurious spatial prediction of components of compositional data caused by the problems stated above can be avoided if four basic requirements are met (de Gruijter et al., 1997):

$$Z^*ij(x) \geq 0 \tag{1}$$

$$\sum_{j=1}^{c} Z^*\,ij\,(x) = \phi \qquad \phi = \text{constant} \tag{2}$$

and for:

$$Z^*_{ij}\,(x) = \sum_{i=1}^{n} \lambda_i\, Z_{ij}\,(x) \quad \sum_{i=1}^{n} \lambda_i = 1; j = 1, \ldots, k \tag{3}$$

$$E[Z^*_{ij}(x) - Z_{ij}(x)]^2 = 0 \tag{4}$$

where $Z^*_{ij}\,(x)$ is the estimate of a compositional regionalized variable, of the jth component (out of k components in the composition) at the ith location.

The first requirement, as indicated in Eq. (1), means that each of the components of a regionalized composition must be nonnegative. In the case of the second requirement (Eq. (2)), a regionalized composition must sum to a constant at every location. A third requirement (Eq. (3)) ensures that the estimate, $Z^{**}\,(x)$ are unbiased, and the fourth (Eq. (4)) is indicative of variance minimization in the kriging system of equations. Most of the spatial interpolation methods used for regionalized compositions in soil studies do not meet all four requirements.

Particle-size data are the most familiar composition in soil science. The relative proportions of the individual particle-size fraction (psf) are what constitute the soil texture. The importance of soil texture cannot be overemphasized. The soil texture, and indeed the particle-size distribution, determine, in part, water, heat, and nutrient fluxes, water and nutrient holding capacity, and soil structural form and stability. The clay fraction, in particular, as the active constituent of the composition, could be incorporated in pedotransfer functions to predict material fluxes (e.g., Arya, et al., 1999) and other soil properties (Sinowski et al., 1997). In some situations it may be necessary to predict spatially all of the components of the particle-size data at unknown locations before they are used for pedotransfer functions or modeling of other soil processes.

Either ordinary kriging or cokriging has generally been used by researchers in the spatial analysis of psf (e.g., Lookman et al., 1995; Mapa and Kumaragamage, 1996; Oberthur et al., 1999; Odeh and McBratney, 2000). These kriging techniques do not take into consideration requirements 1–4 (Eqs. (1–4)), especially requirement (2) above. Choice of an appropriate geostatistical method for spatial analysis is, therefore, critical for producing valid estimates with minimal prediction error variance. A number of statistical methods have been demonstrated to meet some of the requirements in Eqs. (1–4). The first is data transformation before kriging or cokriging, as suggested by several authors (McBratney et al., 1992; Pawlowsky et al., 1993). Another method is compositional kriging, developed by de Gruijter et al. (1997). In soil science, these methods were developed for spatial interpolation of fuzzy (continuous) soil class membership values, a form of composition. However, they have rarely been used for interpolation of particle-size data. The aim of this study is to assess the performance of these methods on particle-size data obtained for a region in New South Wales Australia. First, however, the statistical theory and the methods will be described briefly.

PREDICTION METHODS USED ON COMPOSITIONAL SOIL DATA

Log-Ratio Transformation before Kriging or Cokriging

Modeling any data requires identification of the appropriate sample space. A restricted part of real space ($\boldsymbol{R}^d$), termed the *positive simplex* ($\boldsymbol{R}^d$), is identified by Aitchison (1986, 1990a) as the appropriate sample space for compositions. A simplex is a geometric representation of attribute space, where a composition $\mathbf{Z}$ of D parts is represented by a minimum number of vertices for a space of a given number of dimensions (McBratney et al., 1992). To gain the advantage of symmetry, the d-dimensional simplex (in terms of the subvector) is embedded in D-dimensional real space (Aitchison, 1986):

$$S^d = \{(z_1, \ldots, z_D): z_1 > 0, \ldots, Z_D > 0; z_1 + \text{PPP} + z_D = 1\} \tag{5}$$

The symmetric positive simplex dovetails well with the requirements defined in Eq. (1). How-

ever, the simplex does not, cater adequately for independence and measures of dependence of the components in the absence of a satisfactory class of distributions (S^d) (Aitchison, 1982). And, although the constant-sum constraint (R^d) confines compositional vectors to a simplex, there is no guarantee that the patterns perceived in such a constrained space will necessarily have the same interpretation as in the more familiar spaces such as (R^d) (Aitchison, 1986). A solution to the problem is the transformation of the natural space (S^d) to the real space (R^d) via additive log-ratio transformation (ALR), defined as:

$$y_i = \ln \frac{z_i}{z_{d+1}} \tag{6}$$

where y_i is the log-ratio transformation of z_i. For a regionalized composition this is expressed as:

$$y_{ij}(x) = \ln \frac{z_{ij}(x)}{z_{ik}(x)} \qquad k = \mathrm{d} + 1; i = 1,\ldots, n$$

$$j = 1,\ldots, k \tag{7}$$

with inverse transformation:

$$z_{ij}(x) = \frac{\exp y_{ij}(x)}{\sum_{j=1}^{k} \exp y_k(x)} \tag{8}$$

The effect of ALR transformation is twofold: (i) the closure effect is removed (Aitchison,1982, 1990b), and (ii) through perturbation, transformed values may be closer to a normal distribution than the untransformed data. Transformation, therefore, makes the data better suited for classical statistical procedures (Aitchison, 1986). ALR has been applied to compositional geological data (e.g., Zhou et al., 1991; Pawlowsky et al., 1993; Cardenas et al., 1996) and has been used only rarely in the analysis of soil particle-size data (e.g., Lark, 1999).

Membership grades resulting from fuzzy classification of soil types into continuous classes are the new compositional data in soil science. Continuous classes have membership values (z_{ij}) that describe the degree of membership to a class ($j, j = 1, \ldots, c$) (McBratney et al., 1992; Odeh et al., 1992). The result of fuzzy classification is a matrix of membership values, such as a compositional matrix, $\boldsymbol{Z} = c \times n$. The matrix $\boldsymbol{Z} = z_{ij}$ satisfies all the requirements of a composition. To resolve the specific case of transforming membership values of k continuous classes before kriging, McBratney et al. (1992) modified the Aitchison (1986) log-ratio transformation equation. Because of the presence of zero membership data, it is necessary to replace the zeros before transformation (Martin-Fernandez et al., 2000; Fry et al., 2000). The constant η was, therefore, introduced to cater for zero values in the data, η being one-half of the smallest membership other than zero, and Eq. (7) was modified accordingly. The modified log-ratio transformation (mALR) is expressed as:

$$y_{ij}(x) = \ln \frac{z_{ij}(x)+\eta}{\left(\prod_{j=1}^{k} (z_{ik}(x)+\eta)\right)^{1/k}} \tag{9}$$

The inverse transformation is defined as:

$$Z_{ij}(x) = \left(\frac{\exp y_{ij}(x)}{\sum_{j=1}^{k} \exp y_{ij}(x)} - \frac{\eta}{1+\sum_{j=1}^{k} \eta}\right)\left(1+\sum_{j=1}^{k} \eta\right) \tag{10}$$

For strictly positive compositional values, inclusion of the constant η is unnecessary, and Eqs. (7) and (8) are more appropriate.

Compositional Kriging

Compositional kriging (CK) was recently developed by de Gruijter et al. (1997) and used on membership classes resulting from fuzzy k means of classification of soil. They needed to produce continuous soil maps that relate soil distribution patterns to the general landscape structure. Compositional kriging is an extension of OK. It is dissimilar to cokriging, however, in that it does not assume linear correlations among the compositional components. Moreover, cross-variogram models are not required.

As with OK, CK also minimizes the error variance with respect to the unbiased constraint. In this case, the error variance can be minimized by setting its partial first derivatives, with respect to its associated Lagrange multiplier (i.e. μ_c, α_c, or β equal to zero with the following linear equations:

$$\begin{aligned}
&\sum_{j=1}^{n_c} \lambda_{jc} C_{ijc} + \mu_c + \alpha_c z_{ic} + \beta z_{ic} = C_{i0c} \;\forall i, c \\
&\sum_{i=1}^{n_c} \lambda_{ic} = 1 \qquad \forall c \\
&\sum_{i=1}^{n_c} \lambda_{ic} z_{ic} = 0 \text{ and } \alpha_c \geq 0 \qquad \forall c \\
&\sum_{c=1}^{k} \sum_{i=1}^{n_c} \lambda_{ic} z_{ic} = 1
\end{aligned} \tag{11}$$

where λ_{jc} = weight assigned to observation point j for the membership (composition) class c; C_{ijc} = covariance between observation points i and j for the membership class c; C_{i0c} = covariance between observation point i and the prediction point for the membership class c; and n_c = number of observation points used to predict the membership class c.

Therefore, the memberships at a specific prediction point are estimated by:

$$\hat{z}_c = \sum_{i=1}^{n_c} \lambda_{ic} z_{ic} \qquad \forall c \qquad (12)$$

and the error variances obtained by substituting weights via algebraic manipulation (de Gruijter et al., 1997) into a computationally more efficient expression as:

$$\sigma^2_{Rc} = \sigma^2_c - \sum_{i=1}^{n_c} \lambda_{ic} C_{i0c} - \mu_c - (\alpha_c + \beta)\hat{z}_c \quad \forall c \qquad (13)$$

where σ^2_{Rc} = variance of the prediction error in membership class c and σ^2_c = variance of the memberships to class c.

MATERIALS AND METHODS

The methods described above were applied to the spatial analysis of particle-size data obtained for the two regions, the lower Macintyre and Namoi valleys, both situated in east central Australia. First, however, we describe the two regions and also the methods of data analysis and validation.

The Study Regions

The first study region, the lower Macintyre valley, is approximately 5100 km^2 in extent and is located on the border between New South Wales and Queensland, two of the eastern states of Australia (Fig. 1). The second region is in the lower Namoi valley, which is about 200 km south of the Macintyre. Both regions are part of the Murray Darling Basin, just west of Great Dividing Range of Eastern Australia. They are of very low relief: a gentle east-west slope of approximately 1:3000. The dominant soil on the plains is deep, self-mulching cracking clays; gray clays are found on the open plains and in depressions, and brown clays are found on the slightly elevated areas (Odeh et al., 1998). Soil information for the two areas was sparse and, until recently, was sourced mainly from the "Atlas of Australian Soils" (Northcote, 1966) at a scale of 1:2 million. This study was part of a bigger project initiated to provide soil attribute information important for sustainable cotton production, a major agricultural activity in both regions.

In the lower Macintyre valley, a total of 119 sites (Fig. 1a) were visited and sampled, and more detailed sampling was conducted in the lower Namoi (Fig. 1b)). Auger samples were obtained at six prespecified depths down to 2 meters. The sampling strategy adopted for the lower Macintyre valley and part of the lower Namoi valley is described elsewhere (Odeh and McBratney, 1994). The sampling design used for the eastern part of the lower Namoi valley is described in McGarry et al., 1989. The particle-size fractions were determined by an in-house developed micropipette method. We utilized only the topsoil (0–10 cm) particle-size data for this study. Each soil sample was assigned to one of 12 texture classes as described in Soil Survey Staff (1962). Clay was differentiated into light-medium clay (40–50% clay) and heavy clay (>50% clay), making 13 overall potential classes for the purposes of the study.

Data Analysis

The following methods were examined and compared:

- kriging of each untransformed (UT_{OK}) psf directly as has been the practice;
- additive log-ratio (ALR) and modified log-ratio transformation (mALR) prior to ordinary kriging (ALR_{OK}, $mALR_{OK}$) and cokriging of the transformed value, followed by back-transformation of the kriged results;
- compositional kriging (CK) (de Gruijter et al., 1997) using all of the fractions.

Ordinary kriging (OK) and cokriging are well known within the soil science community (e.g., Wackernagel, 1995; Goovaerts, 1997) and thus will not be described here. It should be noted, however, that cokriging compositional data make sense only after log-ratio transformation as ALR removes the effect of closure (Zhou et al., 1991). Isotropic spherical variogram models were fitted to the experimental variogram of the untransformed and transformed data for all of the kriging methods.

The FORTRAN program, COKRIG (Carr et al., 1985) was used for generalized cokriging. Compositional kriging was performed using the program developed by de Gruijter et al. (1997). The semivariogram and cross-variogram needed for cokriging and compositional kriging were computed and modeled using the geostatistical

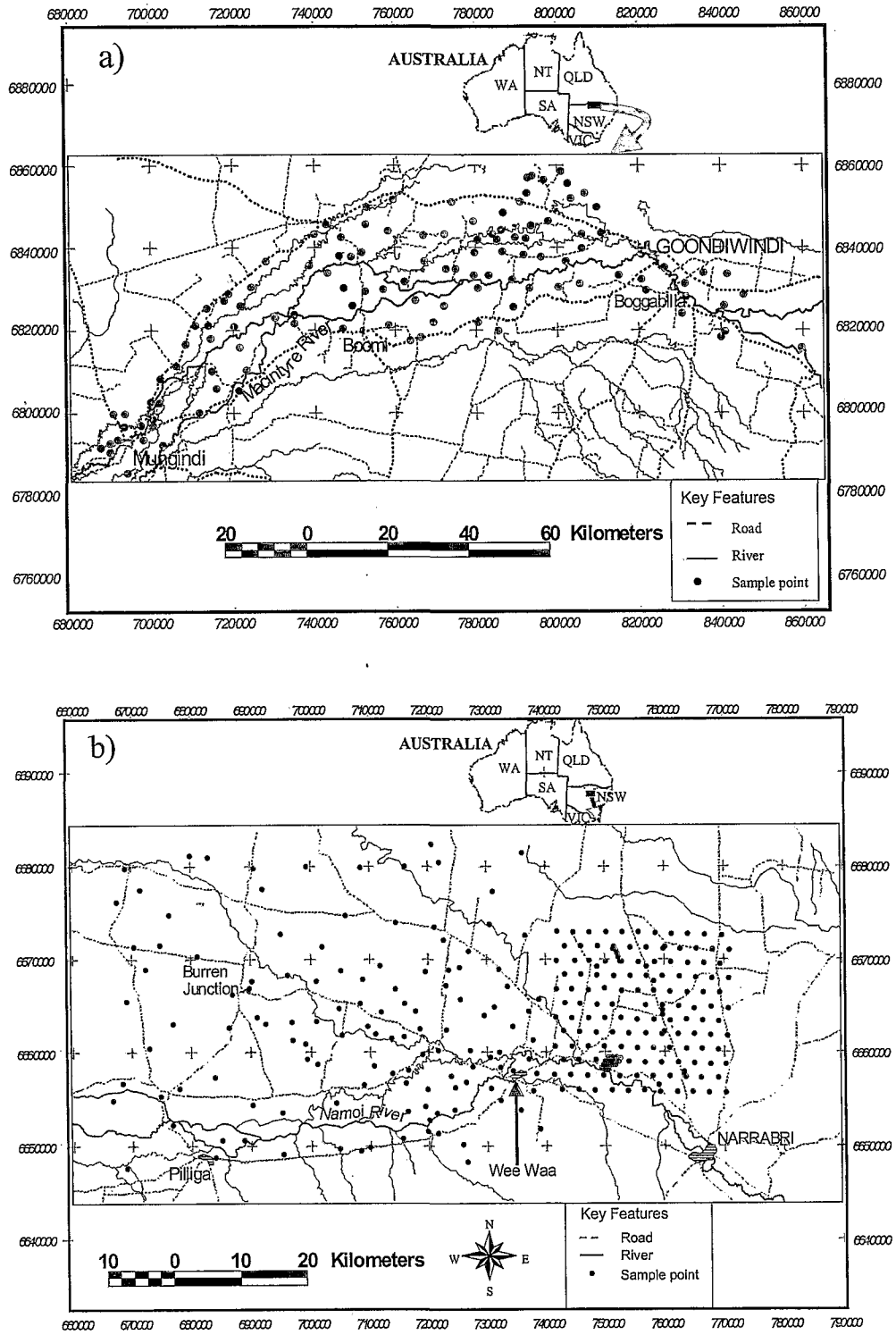

Fig. 1. Location of study area and sampling layout.

package VESPER (Minasny et al., 1999). Another program, ISATIS (Geovariances, 1997) was used for OK analysis of both the untransformed and transformed particle-size data.

Validation of the Prediction Methods

The primary validations used to test the quality of spatial prediction of soil attributes are the mean error (ME) and root mean square error (RMSE) (Voltz and Webster, 1990). Mean error can be estimated using Eq. (13):

$$ME = \frac{\sum_{i=1}^{n} z_i - \hat{z}}{n}, \quad (13)$$

RMSE is expressed as:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (z_i - \hat{z})^2}{n}} \quad (14)$$

Mean error is a measure of bias (or unbias), and RMSE is a measure of precision and bias. As RMSE is sensitive to both systematic and random errors, it could be used to estimate the accuracy of prediction (Atkinson and Foody, 2002), and it could be based on a validation sample set selected independent of the training set.

When the data set used is large, simple dichotomous sets, one for validation and the other for training, would not pose a problem. However, when the sample size is small (as is the case here: 119 for the Macintyre valley and 237 for the Namoi valley), it becomes difficult to have sufficient (and separate) sample data sets for modeling and validation. Moreover it has been suggested in the literature that a sample size of 100 is the barest minimum required for variogram estimation (Webster and Oliver, 1992). For this reason a modified jackknifing technique (Good, 1999) was used to resample the base sample data 20 times for the purpose of validation. The resampling size for validation was maintained at approximately one-sixth of the available data, i.e., 19 of 119 for the lower Macintyre valley and 58 of 328 for the lower Namoi valley. Mean error (Eq. 13), as a measure of bias, and RMSE (Eq. 14), as a measure of precision and bias, were estimated, for each of the resampled validation sets. The prediction quality of each prediction method was determined by averaging the MEs and RMSEs of the 20 jackknifed samples.

RESULTS AND DISCUSSION

The summary statistics of the two sets of data are shown in Table 1. The distributions of the sample data are typical of a composition. None of the fractions (clay, silt or sand) approximates a normal distribution. All are skewed, either positively (silt and sand) or negatively (clay). Percent clay content is by far the most variable of the fractions (SD = 13.6%), followed by sand (SD = 10.6%) and, lastly, silt (SD = 9.6%). The standard deviation (SD) is higher than one may expect for particle-size data, probably because of the large geographical extent of the study area (Fig. 1). This occurs because, although the area is relatively flat and geologically homogenous, small depressions (gilgai) are pedologically diverse from their surroundings (Hubble et al., 1983).

Illustrated here as examples, the histograms of the untransformed and transformed data for the lower Mcintyre valley are shown in Fig. 2. mALR improved normality slightly. Both transformation methods reduce skewness for sand and silt markedly, but ALR actually increased skewness for clay. Similar results were obtained for the lower Namoi valley. The correlation coefficient

TABLE 1
Summary statistics of the topsoil (0–10cm) particle size fractions data

	Lower Macintyre			Lower Namoi		
	Clay (%)	Silt	Sand	Clay	Silt	Sand
Minimum (%)	10.3	18.1	1.7	2.5	0.1	0.1
1st quartile (Q_1) (%)	44.7	26.9	11.1	39.1	16.1	17.9
Median (Q_2) (%)	51.4	31.5	16.3	51.3	19.6	25.5
Mean (%)	48.0	32.6	17.7	47.4	20.9	29.6
3rd quartile (Q_3) (%)	56.3	38.5	22.8	58.4	24.4	38.1
Maximum (%)	77.4	60.0	62.2	71.8	73.3	95.5
Range (%)	67.1	41.9	60.5	69.3	73.2	95.4
((Q3–Q1)/2) (%)	23.0	10.2	10.6	28.0	12.2	19.0
Std (%)	13.9	9.6	10.6	14.6	8.0	17.1
Skewness	−1.03	0.95	1.31	−0.76	1.47	1.02

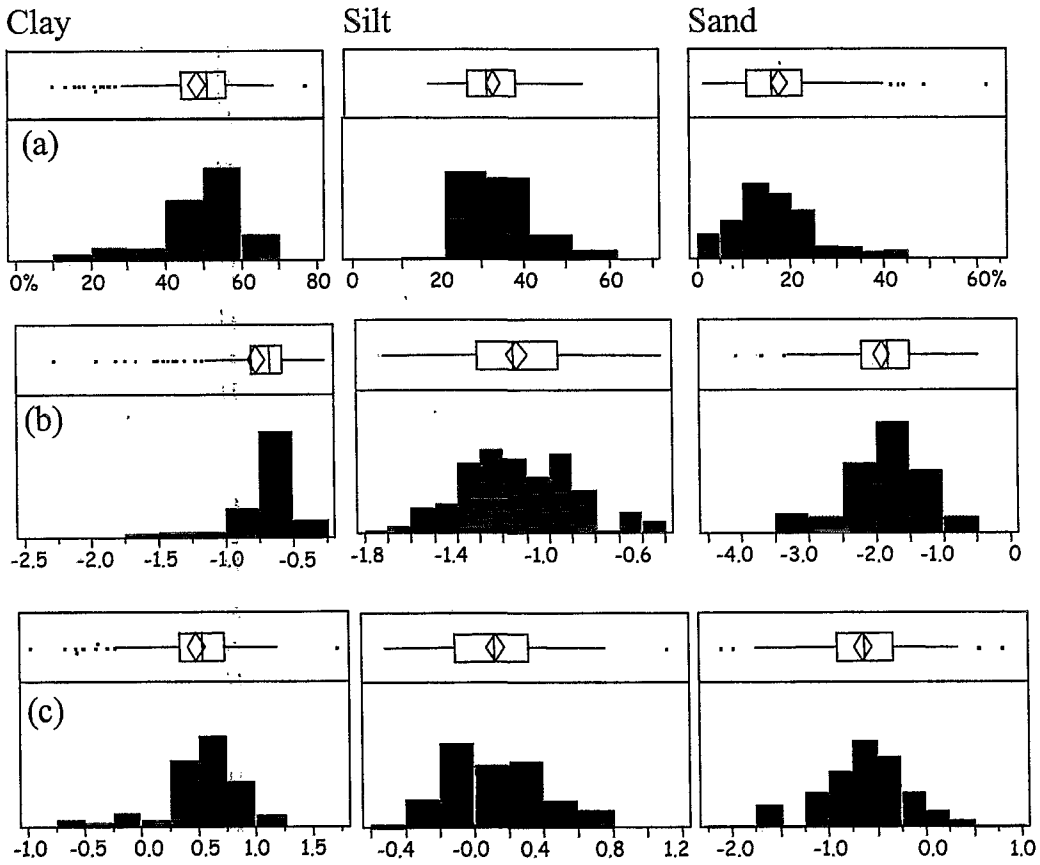

Fig. 2. Histograms and boxplots of the Macintyre particle-size data ($n = 119$), including clay (%), sand (%), and silt (%), comparing (a) untransformed (UT), (b) additive log-ratio (ALR), and (c) modified additive log-ratio (*m*ALR) transformations.

of the untransformed (UT) with transformed data are shown in Table 2. ALR decreased the correlation between clay and silt and clay and sand but improved slightly the correlation between silt and sand. *m*ALR improved considerably the correlation between sand and clay, and between sand and silt, but the correlation between silt and clay deteriorated.

A map of the spatial distribution of the sum of psf predicted by ordinary kriging of untransformed psf (UT_{OK}) for the lower Macintyre valley is shown in Fig. 3. It is evident that (UT_{OK})

TABLE 2

Correlation matrix of untransformed (UT_{OK}) and transformed sample data (ALR_{OK} and $mALR_{OK}$)

Variable	UT_{OK}			ALR_{OK}			$mALR_{OK}$		
	Clay	Silt	Sand	Clay	Silt	Sand	Clay	Silt	Sand
a) Lower Macintyre									
Clay	1.00			1.00			1.00		
Silt	−0.57	1.00		−0.49	1.00		−0.01	1.00	
Sand	−0.73	−0.15	1.00	−0.60	−0.18	1.00	−0.81	−0.57	1.00
b) Lower Namoi									
Clay	1.00			1.00			1.00		
Silt	0.12	1.00		0.23	1.00		−0.01	1.00	
Sand	0.81	−0.54	1.00	−0.33	−0.58	1.00	−0.81	−0.57	1.00

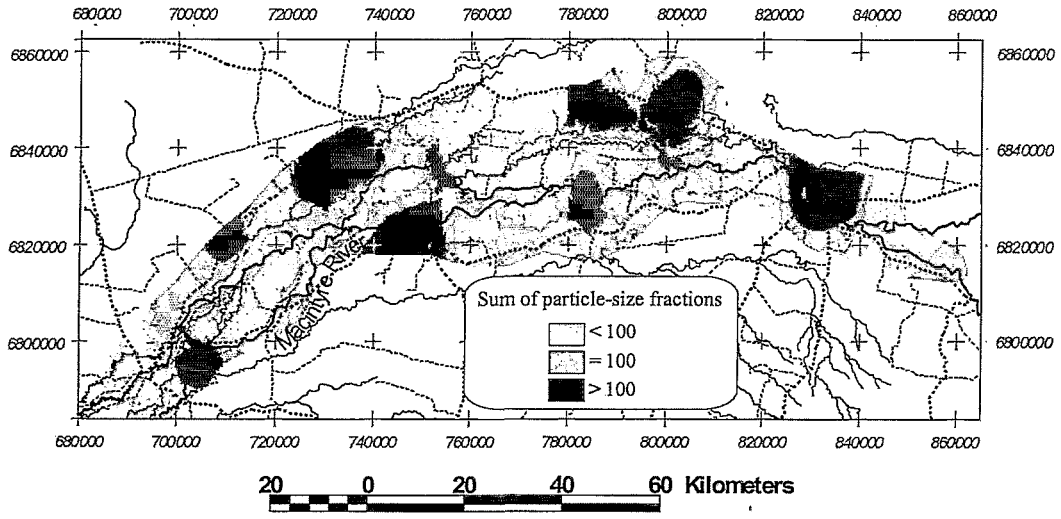

Fig. 3. Sum of the particle-size fraction produced by ordinary kriging of the untransformed particle-size data from the Macintyre valley (UT_{OK}); expected sum = 100%.

both overestimated and underestimated values of soil psf: only 35% of the predicted sites sum to 100%. It is not surprising that the constant sum requirement is not wholly met by (UT_{OK}) as the dependencies between fractions are discarded and the separate fractions are estimated independently at each point on the spatial grid. The same results were obtained for the lower Namoi valley. The results of (UT_{OK}) concur with those of other researchers who have investigated the efficacy of the approach in compositional analysis (see Pawlowsky et al., 1995; de Gruijter et al., 1997).

An operation to overcome the problem of interpolated values of untransformed psf not summing to a constant involves kriging all the fractions but one and then calculating the remaining fraction by difference. This operation, however, is not order invariant. By contrast ALR_{OK} (and $mALR_{OK}$) are order invariant (McBratney et al., 1992; de Gruijter et al., 1997) in the sense that the multivariate normality is preserved under permutation of the components of the composition (Barceló et al., 1996). These methods (and CK) produced kriged particle-size values that sum to a constant (100%). Therefore, the methods based on log-ratio transformation and CK meet the requirements in Eqs. (1)–(3). This outcome highlights the need to reassess applying OK to psf without transformation, especially psf commonly imbedded in pedotransfer functions (e.g., Sinowski et al., 1997).

Cokriging of the log-transformed two data sets produced very poor results. There are several possible reasons for this. As Table 1 and Fig. 2 show, the sample data are highly skewed, and all fractions have outliers. Although the (cross-) variogram estimator is unbiased, it is susceptible to outliers and shows nonrobust behavior toward distributional deviations (Armstrong, 1984; Dowd, 1984; Omre, 1984). In addition to outliers, the data also shows a weak trend and lack of spatial co-dependence. Surprisingly, the more intensively sampled data in the lower Namoi valley did not produce better cokriging results.

Figure 4 shows the relative proportions of the texture classes identified for the sample data set and the results of UT_{OK} CK, and the back-transformed particle-size values resulting from ALR_{OK}, $mALR_{OK}$ for the lower Macintyre valley. For the sample data set (n = 119) the proportions are sandy loam (SaLm) 2% of sites, loam (Lm) 7%, silty loam (SiLm) 2.5%, clay loam (CLm) 2.5%, silty clay loam (SiCLm) 4%, silty clay (SiC) 8%, light-medium clay (LMC) 21%, and heavy clay (HC) 53%. UT_{OK} had only half (4) the number of classes of the original sample sites. There are only five texture classes resulting from each of ALR_{OK}, $mALR_{OK}$, and CK. Figure 4 also shows the similarity of texture class distribution between ALR_{OK} and $mALR_{OK}$ and the sample data set. However, whether kriging transformed or untransformed data, all of the kriging methods have the effect of smoothing the final

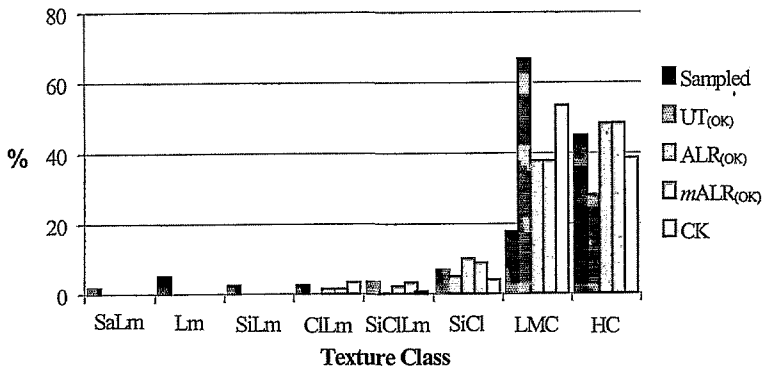

Fig. 4. Histogram of texture classes of sample particle-size data from the Macintyre valley (n = 119) and of the texture classes produced by compositional kriging (CK – n = 18722), and ordinary kriging of untransformed data (UT_{OK} – n = 18722) and kriging of transformed data (additive log-ratio transformation – ALR_{OK}; modified additive log-ratio transformation – mALR n = 18722 for both)

results. This is illustrated in maps of the soil texture classes presented in Fig. 5.

The patterns of texture class distribution produced by UT_{OK} (Fig. 5a) and CK (Fig. 5d) are similar. ALR_{OK} and $m ALR_{OK}$ have similar patterns also (Fig. 5 a and c), which is not surprising considering the similarity of the two methods. However, $m ALR_{OK}$ gives slightly better spatial continuity in the prediction. To a lesser extent, the spatial pattern of texture classes produced by CK is similar to those of ALR_{OK} and $m ALR_{OK}$ (Figs. 5b and c) except for the northeastern corner of the map of the CK results (Fig. 5d). Some patches in Fig. 5d also show some discrepancy compared with the other two methods. The histogram in Fig. 4 illustrates this, with CK clearly overestimating LMC and underestimating HC compared with ALR_{OK} and $m ALR_{OK}$. This is supported by the results of assessing the accuracy of textural classification for the two study areas as presented in Table 3.

Validation Results

Recall that we used repeated resampling of the available data to validate the prediction method. As an example, Fig. 6 shows the histograms of the RMSE of prediction for the resampled validation sets for the Macintyre valley using different prediction methods. Only the RMSE of the clay fraction exhibits distribution close to normal. This is not surprising as the clay fraction is generally characterized by a large range of values compared with silt and sand fractions (Table 1). The study regions are characterized mainly by Vertisols (which have a preponderance of clay) and a few Alfisols and Inceptisols (Soil Survey Staff, 1998). However, the histograms demonstrate that RMSE obtained by repeated resampling (multiple jackknifing) is more representative of the population than that obtained by a single jackknifing. The average values of RMSE are used to compare the quality of the prediction methods.

As shown in Table 3, ALR_{OK} is the most accurate predictor of textural classification for the lower Macintyre valley, with 72% of the validation sites classified correctly compared with 65% with $m ALR_{OK}$ and 44% for UT_{OK}. Only 55% of the validation sites were correctly classified by CK. This trend is also repeated for the lower Namoi valley. Evidently, the ME values, shown in Table 3, also indicate that all methods overestimated the silt fraction, except for UT_{OK} of the percent silt for the lower Namoi. Conversely, in most cases the methods underestimated the clay fraction. RMSE indicates that ALR_{OK} and $m ALR_{OK}$ are equally accurate for predicting percent clay, with UT as the most precise for predicting sand. Statistical testing, using the mean RMSE for ordinary kriging of the raw psf, indicates that the performance by ALR_{OK} and $m ALR_{OK}$ is significantly superior (P = 0.05) to that of UT_{OK}(Table 3). Although the difference between RMSE of ALR_{OK} $m ALR_{OK}$, and CK is minimal, what is surprising is the greater misclassification of texture classes by CK compared with the other methods. The poor performance in textural classification by CK is probably caused by the algebraic manipulation in the CK program, which was probably not as order invariant as we would like it to be.

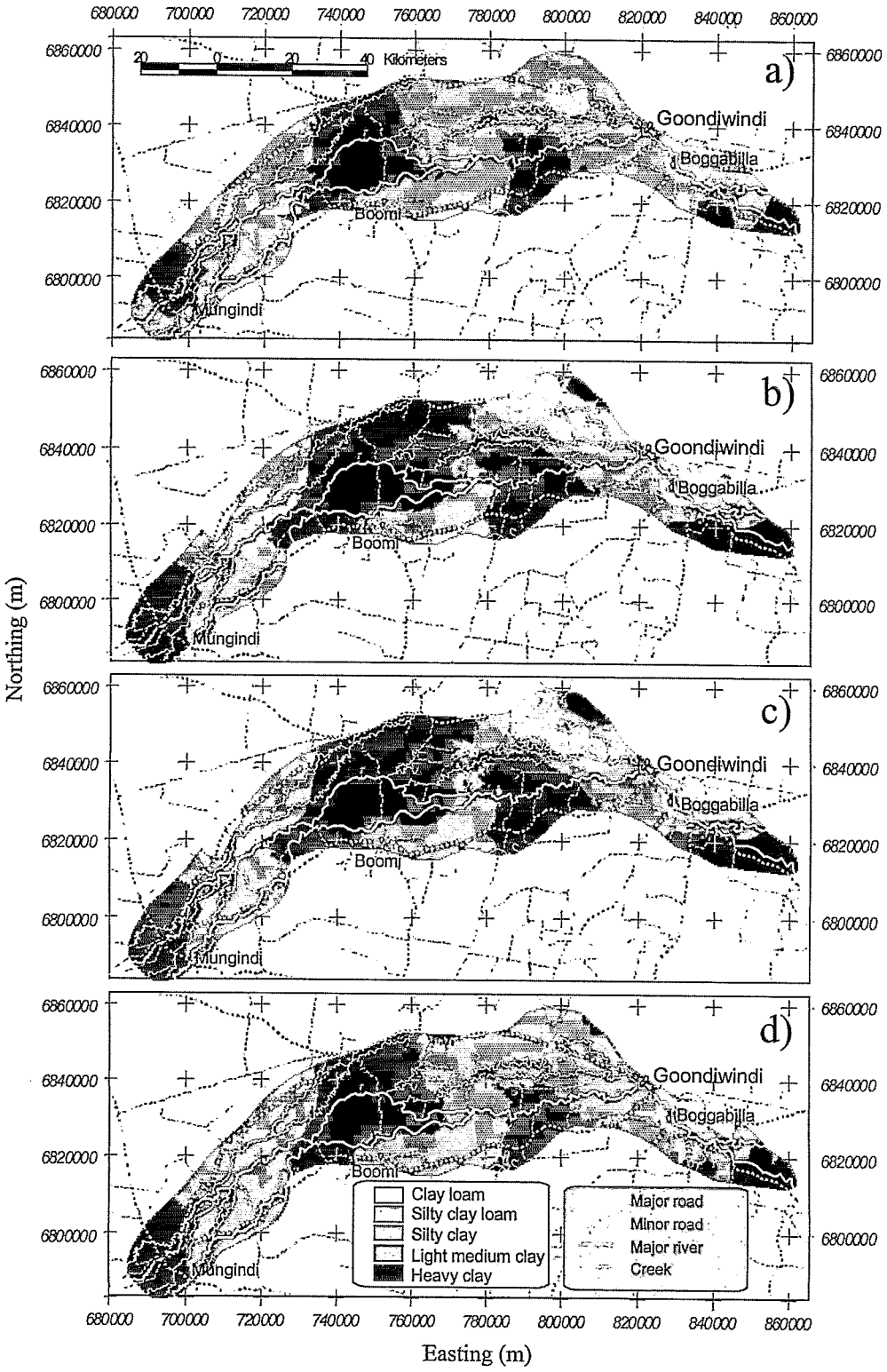

Fig. 5. Comparison of the Macintyre texture classes as predicted by a) ordinary kriging of untransformed data (UT_{OK}); b) kriging after additive log-ratio transformation ($ALRO_K$); c) kriging after modified additive log-ratio transformation ($mALR_{OK}$); and d) compositional kriging (CK).

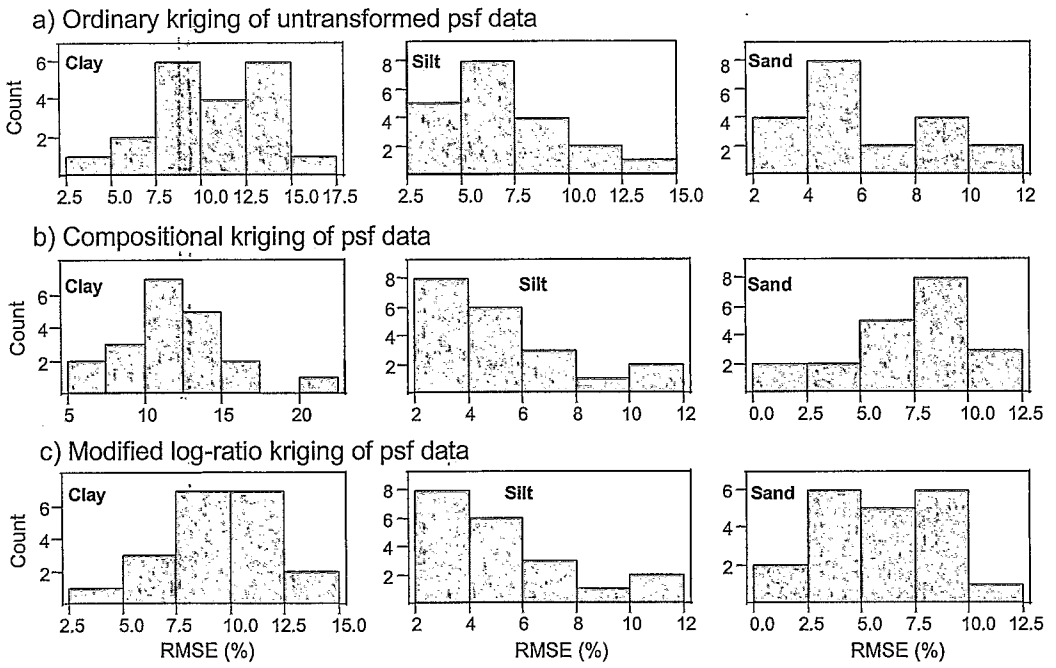

Fig. 6. Histograms of root mean square errors (RMSE) for a) Ordinary kriging of particle-size fractions (psf); b) Compositional kriging of psf; and c) Ordinary kriging of modified log-ratio transformed psf.

Figure 7 shows some of the results of $mALR_{OK}$ on the lower Namoi psf data. Figure 7a also shows the distribution pattern of the texture classes produced by $mALR_{OK}$. It is apparent that a broad band of heavier clay soil runs diagonally from the southwest corner near Pilliga to the northeast corner north of Edgeroi. Most of the irrigated cotton-growing farms are located in this area, particularly to the northeast and northwest of Wee Waa. To the south, the soil is much coarser in texture, ranging from sand to loamy sand and sandy clay loam. This is consistent with the soils derived from Pilliga Sandstone (Triantafilis et al., 2001). The area near Edgeroi is similarly characterized by silt loams and sandy clay loams, which were probably derived from washed sediments from the nearby Nandewar Range (Triantafilis et al., 2001).

The spatial distribution pattern of each of the topsoil clay fractions, as produced by $mALR_{OK}$ for the lower Namoi valley, is shown in Fig. 7b. In the areas south of Edgeroi and southeast of Wee Waa, the clay fraction (Fig. 7d) is low (<40%). In the area to the north of Wee Waa, however, the clay fraction is relatively large (>55%). This is because this area coincides with the area where the Namoi River flows into a very flat alluvial plains landscape, and, as a result, the area has seen various depositional and erosional events, and the soil tends to be predominantly fine-textured (Triantafilis et al., 2001).

In general, critics of CK may not be supportive of the embedded model in the CK program, which involves algebraic manipulation in order to solve the problem of the constraint caused by error minimization. Even though CK produced numerically valid output, inasmuch as our results are concerned, the statistical validity of the method is more or less heuristic and may not justify its being more computationally efficient. The time involved in data transformation bebfore kriging or the overparameterization and the laboriousness of constructing so many variograms or cross-variograms for kriging (de Gruijter et al., 1997) may not be reason enough to discard these techniques. However, the data transformation methods are not without problems.

CONCLUSIONS

The two study sites were large geographical areas, one of which was sparsely sampled. We identified five main texture classes in the surface layer of the soil, the dominant class of which was heavy clay. The performance of each of the pre-

TABLE 3

Validity of textural classification, and bias and accuracy of the prediction methods

Method	Texture correctly classified (%)	Bias ME (%) Clay	Silt	Sand	Accuracy RMSE (%) Clay	Silt	Sand
a) Lower Macintyre							
UT_{OK}	44	−0.83	1.40	0.86	**12.13**	**9.26**	**7.38**
ALR_{OK}	72	0.75	1.85	−1.01	9.56*	5.26**	5.89^{ns}
$mALR_{OK}$	65	−0.58	1.38	0.52	9.40**	5.17**	6.07^{ns}
CK	55	−1.43	2.54	0.85	10.58^{ns}	7.17*	6.36^{ns}
b) Lower Namoi							
UT_{OK}	49	−1.12	−0.02	0.11	**11.10**	**8.58**	**8.18**
ALR_{OK}	69	−0.25	1.84	−0.89	8.66**	6.01**	8.12^{ns}
$mALR_{OK}$	66	−1.00	0.56	−0.73	8.85**	5.81**	8.26^{ns}
CK	52	−4.58	2.15	−0.94	10.47^{ns}	7.18^{ns}	6.36*

** = Statistically different from mean for $\mathbf{UT_{OK}}$ (bold underlined) at 0.01 alpha level

* = Statistically different from mean for $\mathbf{UT_{OK}}$ (bold underlined) at 0.05 alpha level

ns = Not statistically different from mean for $\mathbf{UT_{OK}}$ (bold underlined) at 0.05 alpha level

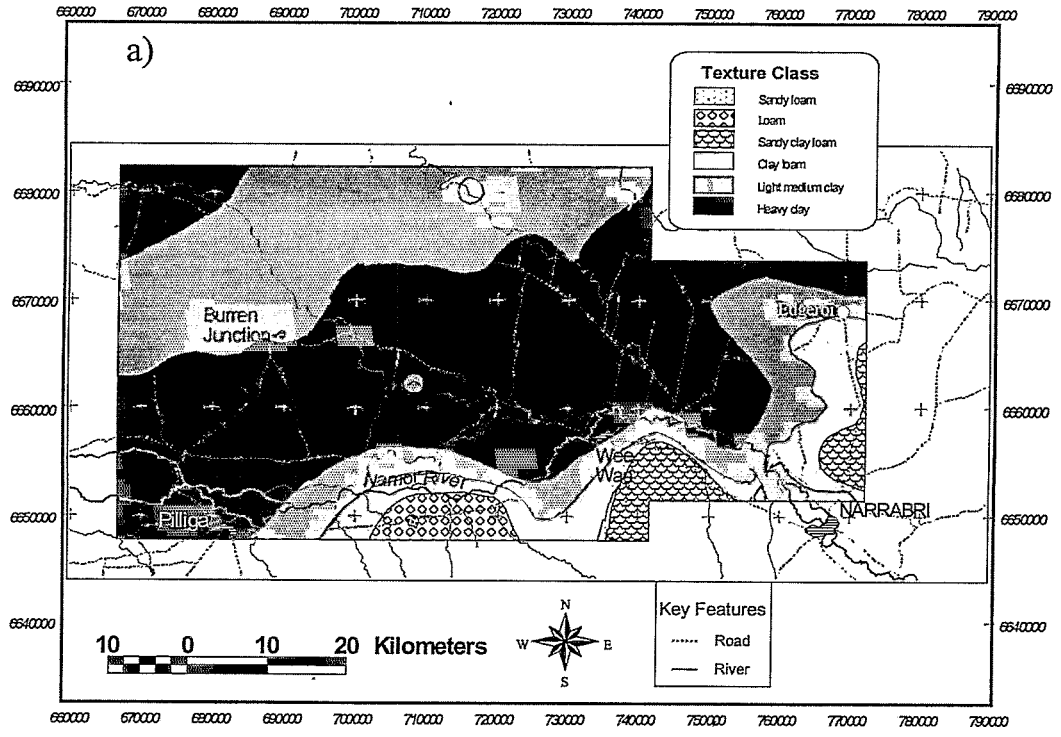

Fig. 7. Some results of $mALR_{OK}$ on particle-size fractions of soils in the lower Namoi valley (a) patterns of topsoil textural classes, and (b) spatial patterns of topsoil % clay fraction. (continued)

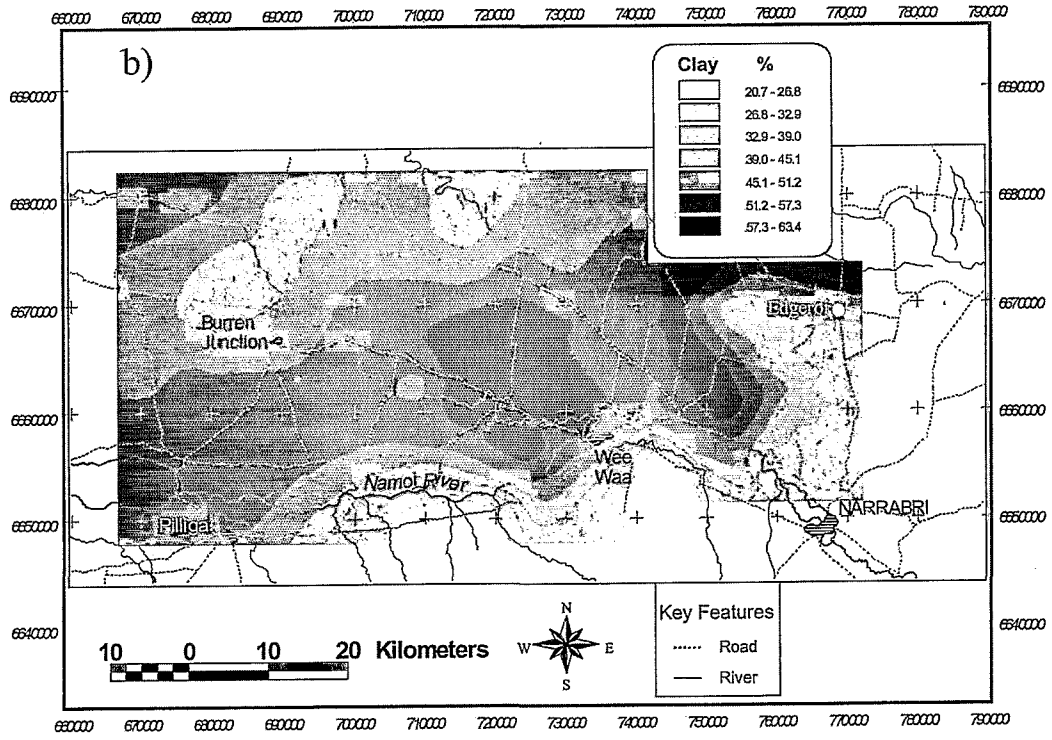

Fig. 7. (b) spatial patterns of topsoil % clay fraction.

diction methods was assessed based on four constraints: (i) predictions should be greater than or equal to zero, (ii) each prediction point should sum to a constant, (iii) the predictions should be unbiased, and (iv) the variance of the prediction errors should be minimized.

We, like other authors (Pawlowsky et al., 1995; de Gruijter et al., 1997), were unable to assess requirement (iv) using ALR_{OK} and $mALR_{OK}$, although the problem has recently been solved by Pawlowsky-Glahn and Egozcue (2002). However, all methods predicted the particle-size values which were greater than zero. Ordinary kriging on the untransformed psf led to output values not summing to a constant at many locations. All other methods (CK, ALR_{OK}, and $mALR_{OK}$) produced results that sum to a constant at every location. Nevertheless, although the accuracy of CK was similar to that of the other methods, it performed poorly in predicting texture classes. ALR and *m*ALR transformation before kriging usually outperformed UT_{OK} and CK. In this case we believe the success was the result of the transformation before kriging. There was evidence of a slight trend in the data for both study areas. Perhaps further improvement can be achieved using universal cokriging after ALR (Stein and Corsten, 1991). Nevertheless, data cannot always be modeled adequately using log-ratio transformation, and alternatives such as Box-Cox family (Barceló et al., 1996) or an adaptation of the Renner (1996) transformation may need to be assessed. The use of standardized residual sum of square (STRESS) has also recently been proposed (Martin-Fernandez et al., 2001) as an alternative criterion for testing prediction quality. We will focus on these issues in a future work.

This study also highlights the problem of relying on one method to validate a particular method of data analysis. An obvious example is the use of RMSE. RMSE of prediction for UT_{OK} and CK seems not to reflect the inaccuracy of predicting the texture classes across the study area.

The analysis and regionalization of compositional data present specific problems for soil scientists. There are new compositional soil data (e.g., fuzzy membership classes). There is also a need to incorporate components of psf in pedotransfer functions. Therefore, to ensure that we

do not make use of spurious interpolated results, it is important that we use the appropriate method (such as ALR or, perhaps, CK) rather than rely on interpolation of untransformed components of compositions, which has previously been the practice.

REFERENCES

Aitchison, J. 1986. The Statistical Analysis of Compositional Data. Chapman & Hall, London.

Aitchison, J. 1982. The statistical analysis of compositional data. J. Royal Stat. Soc. B. 44:139–177.

Aitchison, J. 1990a. Comment on "Measures of variability for geological data" by D. F. Watson and G. M. Philip. Letters to the Editor. Math. Geol. 22: 223–225.

Aitchison, J. 1990b. Relative variation diagrams for describing patterns of compositional variability. Math. Geol. 22:487–511.

Atkinson, P. M., and G. M. Foody. 2002. Uncertainty in remote sensing and GIS: Fundamentals. *In* Uncertainty in Remote Sensing and GIS. G.M. Foody & P. M. Atkinson (eds.). Wiley & Sons, pp. 1–18.

Armstrong, M., 1984. Improving the estimation and modelling of the variogram. *In* Geostatistics for Natural Resources Characterization, Part 1. G. Verly, M. David, A. G. Journel, and A. Maréchal (eds.). Reidel, Dordrecht, Germany, pp. 1–19.

Arya, L. M., F. J Leij., P. J Shouse, and M. Th. van Genuchten. 1999. Relationship between the hydraulic conductivity function and particle-size distribution. Soil Sci. Soc. Am. J. 63:1063–1070.

Barceló, C., V. Pawlowsky, and E. Grunsky. 1996. Some aspects of transformation of compositional data and the identification of outliers. Math. Geol. 28:501–518.

Cardenas, A. A., G. H. Girty, A. D. Hanson, M. M. Lahren, C. Knaack, and D. Johnson. 1996. Assessing differences in composition between low metamorphic grade mudstones and high-grade schists using log-ratio techniques. J. Geol. 104:279–293.

Carr, J. R., D. E. Myers, and C. E. Glass. 1985. Cokriging–A computer program. Comp. Geosci. 11:111–127.

de Gruijter, J. J., D. J. J. Walvoort, and P. F. M. van Gaans. 1997. Continuous soil maps - A fuzzy set approach to bridge the gap between aggregation levels of process and distribution models. Geoderma 77: 169–195.

Dowd, P. A. 1984. The variogram and kriging: Robust and resistant estimators. *In* Geostatistics for Natural Resources Characterization, Part 1. G. Verly, M. David, A. G. Journel, and A. Maréchal (eds.). Reidel, Dordrecht, Germany, pp. 91–106.

Fry, J. M., T. R. L. Fry, and K. R. McLaren. 2000. Compositional data analysis and zeros in micro data. Appl. Econ. 32:953–959.

Geovariances. 1997. ISTATIS version 3.1. Avon, France.

Good, P. I. 1999. Resampling Methods: A Practical Guide to Data Analysis. Birkhauser, Boston.

Goovaerts, P. 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York.

Hubble, G. D., R. F. Isbell, and K. H. Northcote. 1983. Features of Australian soils. *In* Soils: An Australian Point of View. Division of Soils, CSIRO. Academic Press, London, pp. 17–47.

Lark, R. M. 1999. Soil-landform relationships at within-field scales: An investigation using continuous classification. Geoderma 92:141–165.

Lookman, R., N. Vandeweert, R. Merckx, and K. Vlassak. 1995. Geostatistical assessment of the regional distribution of phosphate sorption capacity parameters (Fe_{ox} and Al_{ox}) in northern Belgium. Geoderma 66:285–296.

Mapa, R. B., and D. Kumaragamage. 1996. Variability of soil properties in a tropical Alfisol used for shifting cultivation. Soil Technol. 9:187–197.

Martin-Fernandez, J. A., C. Barcelo-Vidal, and V. Pawlowsky-Glahn. 2000. Zero replacement in compositional data. *In* Advances in Data Science and Classification. H.A.L Kiers, J.-P. Rasson, P.J.F. Groenen, and M. Schader (eds.). Springer-Verlag, Berlin, pp. 155–160.

Martin-Fernandez, J. A., R. A. Olea-Meneses, and V. Pawlowsky-Glahn. 2001. Criteria to compare estimation methods of regionalised composition. Math. Geol. 33:889–909.

McBratney, A. B., J. J. De Gruijter, and D. J. Brus. 1992. Spacial prediction and mapping of continuous soil classes. Geoderma 54:39–64.

McGarry, D., W. T. Ward, and A. B. McBratney. 1989. Soil Studies in the lower Namoi: methods and Data. 1. The Edgeroi Data Set. CSIRO Division of Soils, Brisbane, Australia.

Minasny, B., A. B. McBratney, and B. M. Whelan. 1999. VESPER version 1.0. Australian Centre for Precision Agriculture, The University of Sydney, NSW.

Northcote, K. H. 1966 Atlas of Australian Soils. Explanatory Data for Sheet 3, Sydney-Canberra-Bourke-Armidale Area. CSIRO, Melbourne University Press, Australia.

Oberthur, T., P. Goovaerts, and A. Dobermann. 1999. Mapping soil texture classes using field texturing, particle-size distribution and local knowledge by both conventional and geostatistical methods. Euro. J. Soil Sci. 50:457–479.

Odeh, I. O. A., and A. B. McBratney. 1994. Sampling design for quantitative inventory of the irrigated cotton soil. Proceedings of the 7th Australian Cotton Conference, BroadBeach, Gold Coast, QLD, Australia, pp. 399–404.

Odeh, I. O. A., and A. B. McBratney. 2000. Using AVHRR images for spatial prediction of clay content in the lower Namoi valley of eastern Australia. Geoderma 97:237–254.

Odeh, I. O. A., A. B. McBratney, and D. J. Chittleborough. 1992. Fuzzy-c-means and kriging for map-

ping soil as a continuous system. Soil Sci. Soc. Am. J. 56:1848–1854.

Odeh, I. O. A., J. Todd, J. Triantafilis, and A. B. McBratney. 1998. Status and trends of soil salinity at different scales: The case for the irrigated cotton-growing region of eastern Australia. Nutr. Cycl. Agroecosyst. 50:99–107.

Omre, H. 1984. The variogram and its estimation. *In* Geostatistics for Natural Resources Characterization, Part 1. G. Verly, M. David, A. G. Journel, and A. Maréchal (eds.). Reidel, Dordrecht, Germany, pp. 107–125.

Pawlowsky, V. 1984. On spurious spatial covariance between variables of constant sum. Sci. de la Terre, Inf. Geol. 21:107–113.

Pawlowsky, V., R. A. Olea, and J. C. Davis. 1993. Additive log-ratio estimation of regionalized compositional data: An application to calculation of oil reserves. *In* Geostatistics for the Next Century. R. Dimitrakopoulos (ed.). Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 371–382.

Pawlowsky, V., R. A. Olea, and J. C. Davis. 1995. Estimation of regionalized compositions: A comparison of three methods. Math. Geol. 27:105–127.

Pawlowsky-Glahn, V., and J. J. Egoscue. 2002. BLU estimators and compositional data. Math. Geol. 34: 259–274.

Pearson, K., 1897. Mathematical contributions to the theory of evolution. On a form of spurious correlation which may arise when indices are used in the measurement of organs. Proc. Royal Soc. 60:489–498.

Renner, R. M. 1996. An algorithm for constructing extreme compositions. Comp. Geosci. 22:15–25.

Sinowski, W., A. C. Scheinost, and K. Auerswald. 1997. Regionalization of soil water retention curves in a highly variable soilscape, II. Comparison of regionalization procedures using a pedotransfer function. Geoderma 78:145–159.

Soil Survey Staff. 1962. Soil Survey Manual. USDA, Soil Conservation Service, Washington, DC.

Soil Survey Staff. 1998. Keys to Soil Taxonomy. USDA, Natural Resources Conservation Service, Washington, DC.

Stein, A., and C. A. Corsten. 1991. Universal kriging and cokriging as a regression procedure. Biometrics 47:575–587.

Triantafilis, J., W. T. Ward, I. O. A. Odeh, and A. B. McBratney. 2001. Creation and interpolation of continuous soil layer classes in the lower Namoi valley. Soil Sci. Soc. Am. J. 65:403–413.

Voltz, M., and R. Webster. 1990. A comparison of kriging, cubic splines and classification for predicting soil properties from sample information. J. Soil Sci. 41:473–490.

Wackernagel, H. 1995. Multivariate Geostatistics: An Introduction with Applications. Springer-Verlag, Berlin.

Webster, R., and M. A. Oliver. 1992. Sampling adequately to estimate variograms of soil properties. J. Soil Sci. 43:177–192.

Woronow, A., and K. Love. 1990. Quantifying and testing differences among means of compositional data suites. Math. Geol. 22:837–852.

Zhou, D., H. Chen, and Y. Lou. 1991. The log-ratio approach to the classification of modern sediments and sedimentary environments in northern South China Sea. Math. Geol. 23:157–165.