*European Journal of Soil Science*, June 2007, **58,** 763–774 doi: 10.1111/j.1365-2389.2006.00866.x

# Cokriging particle size fractions of the soil

R. M. LARK & T. F. A. BISHOP
*Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK*

**Summary**

It is often necessary to predict the distribution of mineral particles in soil between size fractions, given observations at sample sites. Because the contents in each fraction necessarily sum to 100%, these values constitute a composition, which we may assume is drawn from a random compositional variate. Elements of a $D$-component composition are subject to non-stochastic constraints; they are constrained to lie on a $D - 1$ dimensional simplex. This means we cannot treat them as realizations of unbounded random variables such as the multivariate Gaussian. For this reason, there are theoretical reasons not to use ordinary cokriging (or ordinary kriging) to map particle size distributions. Despite this, the compositional constraints on data on particle size fractions are not always accounted for by soil scientists.

The additive log-ratio (alr) transform can be used to transform data from a compositional variate into a form that can be treated as a realization of an unbounded random variable. Until now, while soil scientists have made use of the alr transform for the spatial prediction of particle size, there has been concern that the simple back-transform of the optimal estimate of the alr-transformed variables does not yield the optimal estimate of the composition. A numerical approximation to the conditional expectation of the composition has been proposed, but we are not aware of examples of its application and it has not been used in soil science.

In this paper, we report two case studies in which we predicted clay, silt and sand contents of the soil at test sites by ordinary cokriging of the alr-transformed data followed by both the direct (biased) back-transform of the estimates and the unbiased back-transform. We also computed estimates by ordinary cokriging of the untransformed data (which ignores the compositional constraints on the variables) for comparison.

In one of our case studies, the benefit of using the alr transform was apparent, although there was no consistent advantage in using the unbiased back-transform. In the other case study, there was no consistent advantage in using the alr transform, although the bias of the simple back-transform was apparent. The differences between these case studies could be explained with respect to the distribution on the simplex of the particle size fractions at the two sites.

## Introduction

The distribution of mineral particles between size fractions, typically designated sand, silt and clay (although finer divisions are also used), is commonly recorded as a basic property of the soil. The particle size distribution (psd) affects many properties of the soil, including its structure, water relations, chemistry, organic carbon dynamics and mechanical properties. It is therefore a property that we may often need to predict at unsampled sites.

One complication in the analysis of data on psd is that the data are almost always available only as a *composition*. A composition is a variate whose elements necessarily sum to one (or 100%). It might be possible to express the basic information from which the psd is derived as a *basis*. The basis might be the set of dry weights of the mineral particles in each size fraction obtained from sample units of fixed support at a particular depth. The support of a sample unit is its size and shape and orientation (e.g. a vertical cylindrical core of specified length and diameter). One can obtain the composition from the basis (by dividing each element in the latter by the sum of all the elements), but the basis cannot be derived from the composition alone, because the total dry weight of mineral particles in the sample is, in effect, not constrained and varies between samples that differ in their bulk density and content of stones and organic matter. As we note above, most data on the psd are available as a composition only, and we rarely have all the information required to recover the basis.

Correspondence: R.M. Lark. E-mail: murray.lark@bbsrc.ac.uk
Received 21 April 2006; revised version accepted 27 July 2006

A composition (unlike a basis) cannot be analysed like the variates we commonly study in soil science, where the constituent variables can be treated as random numbers drawn from unbounded distributions such as the normal. This is because the elements of a composition are subject to non-stochastic constraints. A random variate composition with three elements (such as the proportions of sand, silt and clay in a soil sample) is not drawn from the real space $\mathbb{R}^3$, but from the two-dimensional simplex plane $\mathcal{S}^3$ embedded in this space. This simplex plane is familiar to soil scientists as the triangular ternary diagram in which the textural classes of soil are customarily displayed. As Aitchison (1986) points out, these constraints not only invalidate the assumption that our variables are drawn from unbounded random processes, but also induce spurious negative correlations between our variables. In the context of spatial prediction, there are further practical problems. We may estimate the elements of a composition at an unsampled site by ordinary kriging, but there is no guarantee that the separate estimates will sum to one (or 100%). This was found in practice by Odeh *et al.* (2003).

Compositional kriging, as proposed by de Gruitjer *et al.* (1997), sets out to ensure that the basic constraints on the elements of a composition are honoured in the kriged estimates. To do this, conditions, in addition to the unbiasedness condition, are imposed on the ordinary kriging system. The constrained kriging equations must be solved numerically and this is somewhat cumbersome. Chang (2002) has proposed a development of compositional cokriging in which the composition is estimated multivariately rather than by an assemblage of univariate kriging estimators. An alternative approach has been proposed by Pawlowsky *et al.* (1995) and Pawlowsky-Glahn & Olea (2004). This is additive log-ratio (alr) cokriging.

Aitchison (1986) proposes that compositional variates are transformed into log-ratios before analysis. It has been seen that a basis cannot be recovered from the composition alone, but it is clear that the ratio of two elements in the composition will be identical to the corresponding ratio of elements in the basis. For convenience, we transform the ratios to their natural logarithms. The alr transform is one of these log-ratio transforms and, as shall be seen, it has desirable properties for kriging. Odeh *et al.* (2003) applied ordinary kriging to alrs of the soil particle size fractions and found that the resulting predictions had smaller bias and root mean square errors than those obtained with compositional kriging or ordinary kriging of the untransformed compositions. However, a problem with this approach is that the simple back-transform of the estimates of the alr of the composition is biased in the sense that the estimated conditional expectation of the composition at some location is not obtained by applying the back-transformation to the estimated conditional expectation of the transformed variable. This is because the transform is non-linear and the estimate of the conditional expectation of the transformed variable has an associated error. An unbiased back-transform is not known, but Pawlowsky-Glahn & Olea (2004), following Aitchison (1986) for the general problem of the estimation of compositions, have recently proposed a numerical method to obtain it in the case of kriging. The purpose of this paper is to demonstrate and evaluate this procedure for the spatial prediction of soil psd.

## Methods

Let $\mathbf{z}$ denote a composition that we observe and assume to be a realization of a random variate $\mathbf{Z}$.

In this paper, we are concerned with regionalized compositions, $\mathbf{Z}(\mathbf{s})$, where $\mathbf{s}$ is a vector of spatial coordinates, but for conciseness in notation we insert the coordinate vector only when it is essential. The composition consists of $D$ elements,

$$\mathbf{z} = [z_1, z_2, \ldots, z_D]^{\mathrm{T}},$$

such that

$$z_i > 0 \;\forall\; i = 1, 2, \ldots, D$$

and

$$\sum_{i=1}^{D} z_i = 1.$$

The alr transform of $\mathbf{z}$ gives us the variate $\mathbf{x}$:

$$\mathbf{x} = \mathrm{alr}(\mathbf{z}) = \left(\ln\frac{z_1}{z_D}, \ln\frac{z_2}{z_D}, \ldots, \ln\frac{z_{D-1}}{z_D}\right). \quad (1)$$

This transform therefore maps from $\mathcal{S}^D$ to $\mathbb{R}^{D-1}$. We define a vector $\mathbf{w}$ where

$$\mathbf{w} = \left[\mathbf{x}^{\mathrm{T}}, 0\right]^{\mathrm{T}}.$$

This allows us to write the inverse of the alr transform, the additive generalized logistic (agl) transform, as

$$\mathbf{z} = \frac{\exp(\mathbf{w})}{\mathbf{j}^{\mathrm{T}}\exp(\mathbf{w})}, \quad (2)$$

where $\exp(\mathbf{w})$ denotes the vector $[\exp(w_1), \exp(w_2), \ldots, \exp(w_{D-1}), 1]$ and $\mathbf{j}$ is a vector of length $D$ with all elements equal to one.

Other log-ratio transforms exist and their properties when applied to regionalized compositions are discussed by Pawlowsky-Glahn & Olea (2004). A critical finding for our purposes is that the cross-covariance structure of the alr-transformed variable, $\mathbf{X}(\mathbf{s})$, contains all the information on the spatial dependence of $\mathbf{Z}(\mathbf{s})$ that is provided by other transforms, and (unlike the centred log-ratio transform) the covariance matrices are not singular. Under the intrinsic hypothesis, the alr auto- and cross-variograms [i.e the variograms of $\mathbf{X}(\mathbf{s})$] completely specify the cross-covariance structure, and they have no disadvantages other than the assumption of symmetry and the

problems of modelling the variograms that apply equally to the geostatistical analysis of any variates. We therefore use the alr transform here and estimate the variograms of the transformed variates in the usual way (see, for example, Webster & Oliver, 2001).

The alr-transformed variables may be estimated at unsampled sites by cokriging. As with non-compositional variates, we may either assume that the mean is known (simple cokriging) or unknown (ordinary cokriging), although it is not obvious what conditions would justify the former assumption in practice. Pawlowsky-Glahn & Olea (2004) show that alr cokriging has the attractive property of permutation invariance. That is to say, our estimates are not affected if we change the order of the elements in the composition (and so define the alr transform with a different element as the denominator in each log-ratio). Pawlowsky *et al*. (1995) and Odeh *et al*. (2003) applied alr cokriging and univariate ordinary kriging respectively to predict a composition at unsampled sites. They used the agl transform to back-transform the elements of the estimated variates. However, this back-transform is biased, and the unbiased back-transform is unknown (Pawlowsky-Glahn & Olea, 2004). This was one reason why de Gruitjer *et al*. (1997) developed compositional kriging. However, Pawlowsky-Glahn & Olea (2004) show that a back-transform can be obtained numerically (although they do not use it in their case study). If we denote by $\bar{\mu}_{\mathbf{Z}}$ the conditional expectation of $\mathbf{Z}$ at some site, then

$$\bar{\mu}_{\mathbf{Z}} = \mathrm{E}[\mathbf{Z}] = \int_{\mathcal{S}^D} \mathbf{Z} f(\mathbf{Z}) \mathrm{d}\mathbf{Z}, \tag{3}$$

which is a multivariate integration in the simplex space. If $\bar{\mu}_{\mathbf{X}}$ is our corresponding expectation for the alr-transformed variable $\mathbf{X}$, with a covariance matrix $\mathbf{C}_{\mathbf{X}}$, then we can write the probability density function (pdf) of $\mathbf{Z}$ as

$$f(\mathbf{Z}) = (2\pi)^{-\frac{D-1}{2}} |\mathbf{C}_{\mathbf{X}}|^{-\frac{1}{2}} \left( \prod_{i=1}^{D} Z_i \right)^{-1} \times \exp\left\{ -\frac{1}{2} [\mathrm{alr}(\mathbf{Z}) - \bar{\mu}_{\mathbf{X}}]^{\mathrm{T}} \mathbf{C}_{\mathbf{X}}^{-1} [\mathrm{alr}(\mathbf{Z}) - \bar{\mu}_{\mathbf{X}}] \right\}, \tag{4}$$

with the explicit assumption that $\mathbf{X}$ is a multivariate normal random variate comprising $D - 1$ variables. Equation (4) is recognizable as the multivariate normal pdf for a variate with a mean vector $\bar{\mu}_{\mathbf{X}}$ and a covariance matrix $\mathbf{C}_{\mathbf{X}}$ with an additional term $(\prod_{i=1}^{D} Z_i)$, which is the Jacobian of the agl transform, as is shown by Pawlowsky-Glahn & Olea (2004). The Jacobian and its role in the computation of the pdf of the back-transformed variable are explained in Appendix 1 of this paper.

We must evaluate the integral in Equation (3) numerically. Aitchison (1986) proposed that this is done by Gauss–Hermite (G–H) quadrature. Gauss–Hermite quadrature is a standard method for numerical integration. It is based on the result that, for some function $g(\cdot)$, of a multivariate variable $\mathbf{Y}$, the multivariate integral on the left-hand side of Equation (5) below is approximated by the multiple summation on the right-hand side,

$$\int_{\mathbb{R}^D} g(\mathbf{Y}) \exp(-\mathbf{Y}^{\mathrm{T}}\mathbf{Y}) \mathrm{d}\mathbf{Y} \approx \sum_{i_1=1}^{k} \sum_{i_2=1}^{k} \cdots \sum_{i_D=1}^{k} \lambda_{i_1} \lambda_{i_2} \ldots \lambda_{i_D} g(Y_{i_1} Y_{i_2} \ldots Y_{i_D}), \tag{5}$$

when the variable $\mathbf{Y}$ is a particular auxiliary function obtained as the $k$ zeroes of the Hermite polynomial of order $k$ (i.e. the roots of the equation obtained by equating the polynomial to zero). The $\lambda$ are associated weights derived from the Hermite polynomial. Values of $Y$ and the weights $\lambda$ are tabulated for differing $k$ (e.g. by Abramowitz & Stegun, 1964, in their table 25.10). In order to apply this method, it is necessary to find some function $g(\cdot)$ such that the expression that we want to integrate can be written down in the same form as the left-hand side of Equation (5). Pawlowsky-Glahn & Olea (2004), following Aitchison (1986), show that we can obtain an approximation to $\bar{\mu}_{\mathbf{Z}}$ by evaluating Equation (5) with

$$g(\mathbf{Y}) = \pi^{-\frac{D-1}{2}} \mathrm{agl}\left( \sqrt{2}\mathbf{R}^{\mathrm{T}}\mathbf{Y} + \bar{\mu}_{\mathbf{X}} \right), \tag{6}$$

where $\bar{\mu}_{\mathbf{X}}$ is our cokriged estimate of $\mathbf{X}$, and $\mathbf{R}$ is a factor of $\mathbf{C}_{\mathbf{X}}$, the covariance matrix of the cokriging errors. The reader who wants a fuller account of how Equation (6) is obtained is directed to Appendix 2 of this paper. A similar expression can be provided to obtain the covariance matrix of $\bar{\mu}_{\mathbf{Z}}$.

## Case studies

### *Sandford transect*

These are data from the transect near Sandford in Central England, reported by Webster & Cuanalo (1975). The soil was sampled at points on a regular transect, of spacing 10 m. There were 321 sample sites, and at each the soil was sampled from three layers, each 5 to 6 cm thick, centred at depths 8, 30 and 65 cm. In this paper, we report on the analysis of the particle size fractions (sand, silt and clay) for the second of these depth intervals. The summary statistics of these data are presented in Table 1, and Figure 1 shows them as a ternary diagram. Note that the mean values of compositional variates are a poor description of the location of the data (Pawlowsky-Glahn & Olea, 2004). We include them for completeness, but also report median values and the first and third quartiles.

Every third datum (observations at sites 3, 6, 9, ...) was removed from the data for subsequent validation, and the remaining prediction data set was used for geostatistical analysis. We computed the alr transform of the clay and silt contents, with the sand content as the denominator of the ratio. We then computed the auto- and cross-variograms of these transformed variables and fitted a linear model of coregionalization (LMCR) using the simulated annealing program presented by

**Table 1** Summary statistics of data in case studies

| | Clay | Silt | Sand | alr Clay | alr Silt |
|---|---|---|---|---|---|
| | | % | | | |
| *Sandford* | | | | | |
| Mean | 34.6 | 24.9 | 40.5 | – 0.198 | – 0.493 |
| Median | 34 | 20 | 25 | 0.134 | 0.0 |
| Quartile 1 | 15 | 10 | 15 | – 1.427 | – 1.540 |
| Quartile 3 | 50 | 35 | 70 | 1.099 | 0.693 |
| Standard deviation | 23.9 | 17.8 | 31.1 | 1.827 | 1.648 |
| Skewness | 0.37 | 0.68 | 0.66 | 0.44 | 0.65 |
| *East Creek* | | | | | |
| Mean | 49.8 | 11.3 | 38.9 | 0.250 | – 1.293 |
| Median | 50.6 | 11.75 | 38.2 | 0.277 | – 1.226 |
| Quartile 1 | 46.9 | 8.7 | 34.7 | 0.119 | – 1.563 |
| Quartile 3 | 54.1 | 13.7 | 41.8 | 0.440 | – 0.953 |
| Standard deviation | 7.02 | 3.88 | 6.66 | 0.307 | 0.480 |
| Skewness | – 0.85 | – 0.05 | 1.1 | – 1.11 | – 1.07 |
| Octile skew | – 0.1 | – 0.12 | 0.1 | – 0.08 | – 0.13 |

Lark & Papritz (2003). The estimated variograms with fitted models are shown in Figure 2(a).

We followed the same procedure with the untransformed data on clay and silt contents, and these variograms are shown in Figure 2(b).

Having fitted the LMCR, we could then compute the structural correlations between the variables. These are the correlations of the separate components of the coregionalization model, each associated either with the nugget (spatially uncorrelated) variation or a spatially dependent component with a particular variogram function (Webster & Oliver, 2001). The advantage of the structural correlations is that they allow us to measure the relationship between the spatially dependent components of variation in two or more variables, filtering out the uncorrelated variation, which often includes measurement error.

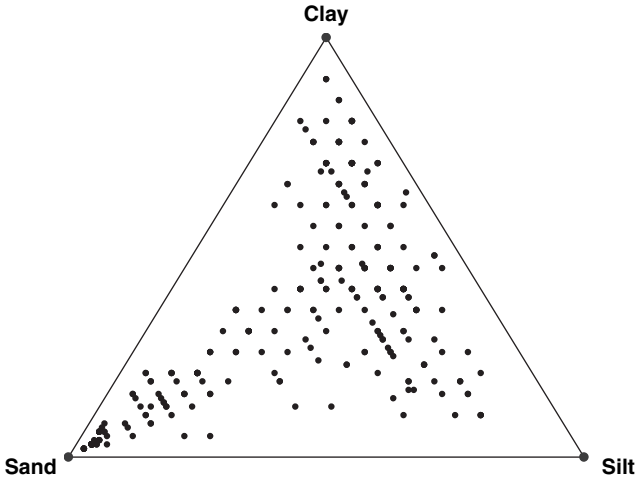


**Figure 1** Ternary diagram for the soil on the Sandford transect. Note that all ternary diagrams in this paper are labelled according to the following convention. The vertex labelled 'Clay' is the position in the simplex where the clay content of the soil is 100% and no other fractions are found. At all positions on the opposite edge of the simplex, the clay content is zero. Lines of equal clay content are parallel to this edge.

We then computed estimates of the particle size fractions of the soil at the unsampled sites using the following three procedures.

**1** The clay and silt contents were estimated by ordinary cokriging from the raw data. The estimate of the sand content was then derived by difference. The choice of sand as the variable to be obtain in this way was arbitrary.

**2** The alr-transformed values for clay and silt content were estimated by ordinary cokriging. These estimates were then back-transformed to values of clay, silt and sand by means of the agl transform in Equation (2) then re-expressed as percentages.

**3** The alr-transformed values for clay and silt content were estimated by ordinary cokriging. These estimates were then back-transformed to values of clay, silt and sand by means of the unbiased back-transform through G–H quadrature, discussed in the previous section. The values of the auxiliary function $Y$ and the weights $\lambda$ are standard numbers. We took them from table 25.10 of Abramowitz & Stegun (1964). We followed Aitchison (1986) in selecting a sufficiently large value of $k$ such that increasing it to larger values caused no change in the resulting estimates. In both case studies, $k = 7$ was satisfactory.

We then compared the estimates of the three size fractions to the measured values at the validation sites. First, we computed the mean square error for each fraction, as a measure of the precision of the predictions. Secondly, we followed Pawlowsky-Glahn & Olea (2004) and computed the standardized residual sum of squares (STRESS) as a measure of the overall similarity of the kriged estimate and the validation data. Let $\delta_{i,j}$ be some distance measure between two observed compositions $\mathbf{y}_i$ and $\mathbf{y}_j$, and let $\delta^*_{i,j}$ be the same measure for what, in mathematical terms, is a projection of the compositions onto a lower dimensional space. In practical terms, this could be a spatial smoothing of the original data. The STRESS between the observations $\mathbf{y}_i$, $i = 1, 2, \ldots$, and the smoothed values $\mathbf{y}^*_i$, $i = 1, 2, \ldots$, (on the projection) is

$$\text{STRESS} = \left\{ \frac{\sum_{i<j} \left( \delta_{i,j} - \delta^*_{i,j} \right)^2}{\sum_{i<j} \left( \delta_{i,j} \right)^2} \right\}^{\frac{1}{2}}. \qquad (7)$$

Note that STRESS is defined only if $\delta_{i,j} > 0$ for at least one combination $\{i, j\}$.

Pawlowsky-Glahn & Olea (2004) propose that the STRESS be used to compare observed values of a compostion, $\mathbf{z}$, with corresponding kriged estimates, $\mathbf{z}^*$. This is sensible because we may think of the kriging estimates as a projection of the data

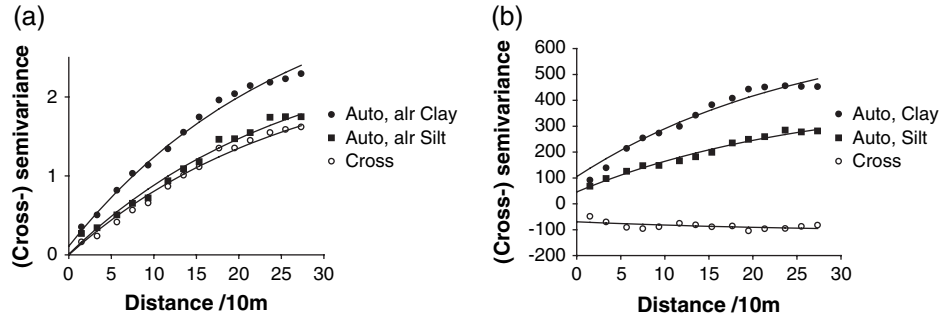


**Figure 2** Auto- and cross-variograms for clay and silt content on the Sandford transect with (a) alr transformation and (b) no transformation. Fitted models are shown by solid lines.

onto a subspace of a Hilbert space (Olea, 1999). In more intuitive terms, the kriged estimates are a more or less smoothed version of the data, and the STRESS will measure how far the estimates reproduce the variations of the observations. At one limit, **z** and **z*** correspond exactly, and STRESS is zero. At the other limit, the **z** vary spatially but the **z*** are uniform, and STRESS is one.

We used the Aitchison distance as our distance metric to compute the STRESS where the square Aitchison distance for a comparison of two compositions, **z** and **z***, is defined as

$$\delta_{\mathrm{a}}^2(\mathbf{z},\mathbf{z}^*) = \sum_{i=1}^{D}\left(\ln\frac{z_i}{\check{\mathbf{z}}} - \ln\frac{z_i^*}{\check{\mathbf{z}}^*}\right)^2, \qquad (8)$$

where $z_i$ are elements of the composition **z**, $z_i^*$ are elements of the composition **z***, and $\check{\mathbf{z}}$ denotes the geometric mean of the elements of the composition

$$\check{\mathbf{z}} = \left(\prod_{i=1}^{D} z_i\right)^{\frac{1}{D}}.$$

The Aitchison distance is the Euclidean distance between the centred log-ratio transform of the compositions. This is an alternative log-ratio transform to the alr that we are using for kriging. We use it here because while, as we note above, alr cokriging is unaffected by the order of elements in the composition, a distance metric based on the alr transform of **z** and **z*** would not be. Because the centred log-ratio transform uses all elements in the composition, the Aitchison distance as defined in Equation (8) is permutation invariant. The Aitchison distance is preferred to the Euclidean distance between the untransformed compositional variables because it reflects the constraints on their joint variation. This is discussed in detail by Aitchison (1992).

These summary statistics are all presented in Table 2.

Figure 3 shows plots of the cokriged estimates obtained with no transformation and the alr-cokriging estimates with the G–H back-transform (the alr-cokriging estimates with the simple back-transform are not shown because they were very similar to the estimates with the G–H back-transform).

The most notable effect of the alr transform on the variograms is that, while the relationship between the untransformed variables is very weak (with small negative values of the cross-variogram, and a structural correlation of -0.08 for the spatially dependent components of variation), the transformed variables are strongly (positively) correlated (a structural correlation of 0.81). There is also a difference between the predictions obtained with and without the alr transform. The mean square errors and the STRESS are smaller for predictions from the alr-transformed data then for our predictions that ignore the compositional structure of the data; note particularly the mean square error for the predictions of clay content. However, in this case, there is no evidence for an improvement in predictions when the conditional expectation is computed by G–H integration, rather than by the simple agl transform of the estimates. Figure 3 shows that the largest differences between the predictions for the methods are where the clay or sand contents are large, see for example near position 190.

Here, the alr cokriging is closer to the observations than cokriging of the untransformed data.

### *East Creek data*

These are data from a portion of a 74-ha paddock called 'East Creek' in northern New South Wales, Australia. The data are from samples collected on two occasions. The first was in April 1996 when 110 soil cores were collected by stratified, random sampling: see Shatar & McBratney (1999) for full details. The second was in March 1999 when 109 soil cores were collected by simple random sampling: see Bishop & McBratney (2001) for

**Table 2** Results for prediction at validation sites for the Sandford data

| | Mean square error | | | |
|---|---|---|---|---|
| Kriging method | Clay | Silt | Sand | STRESS |
| Cokriging[a] | 99.3 | 61.6 | 93.5 | 0.21 |
| alr cokriging[b] | 87.1 | 56.4 | 93.0 | 0.19 |
| alr cokriging[c] | 87.1 | 57.4 | 91.9 | 0.19 |

[a]Direct cokriging of the composition.
[b]alr cokriging, with direct agl back-transformation.
[c]alr cokriging, with back-transformation by Gauss–Hermite quadrature.

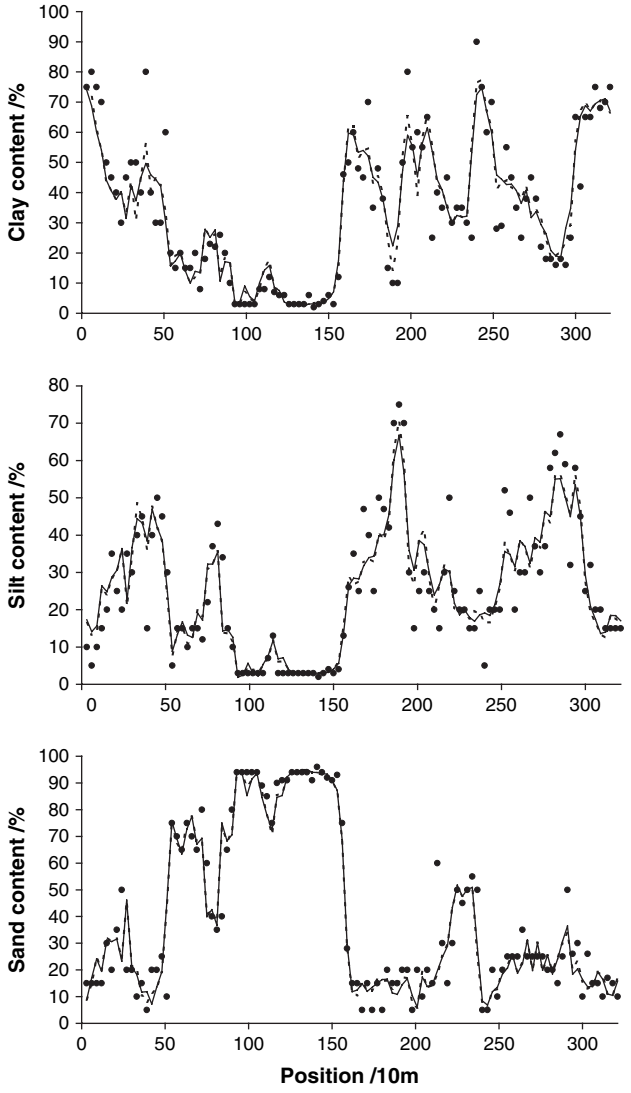


**Figure 3** Cokriged estimates of textural fractions at validation sites on the Sandford transect by ordinary cokriging with no transformation (solid line), alr transformation and G–H back-transformation (broken line). The solid points are the observed values.

full details. In this paper, we report the analysis of the particle size fractions of the 15–30 cm depth layer.

The data on clay, silt and sand contents are shown as a ternary diagram in Figure 4 and the summary statistics are in Table 1. We noted that the coefficient of skew is increased by alr transformation. However, it is clear from the ternary diagram that the sand content (which is the denominator in the transform) includes some extreme values. The histograms of the alr-transformed data are shown in Figure 5.

These could plausibly be interpreted as normal random variables with some (small) outlying values. The coefficient of skew is sensitive to outlying values because it is based on moments of the data, and it can therefore be misleading about the underlying distribution of a variable. We therefore computed the octile

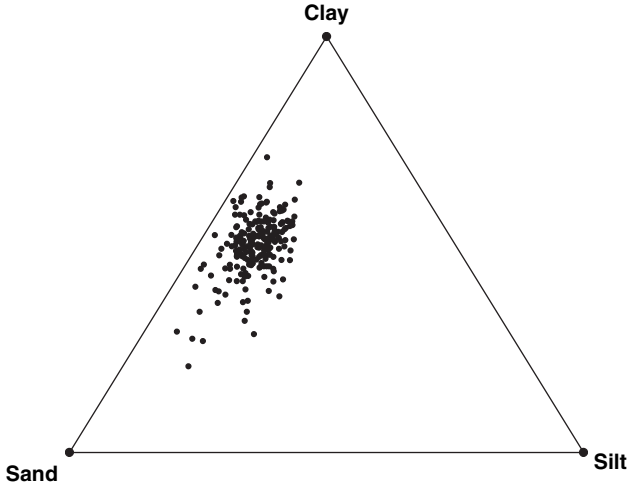


**Figure 4** Ternary diagram for the soil at East Creek. Note that all ternary diagrams in this paper are labelled according to the following convention. The vertex labelled 'Clay' is the position in the simplex where the clay content of the soil is 100% and no other fractions are found. At all positions on the opposite edge of the simplex, the clay content is zero. Lines of equal clay content are parallel to this edge.

skew of the data (Brys *et al.*, 2003). The octile skew is a measure of skew that is insensitive to outliers. When applied to data drawn from a basic distribution, with some added contaminants, the octile skew reflects the symmetry or assymetry of the basic distribution. It is defined as

$$\frac{(P_{0.875} - P_{0.5}) - (P_{0.5} - P_{0.125})}{P_{0.875} - P_{0.125}}, \qquad (9)$$

where $P_q$ is the value of the ordered datum such that the proportion $q$ of the data is smaller than $P_q$. The octile skew is zero if the first and seventh octiles are symmetric about the median. Data with an absolute conventional coefficient of skew larger than 1.0 are usually transformed (Webster & Oliver, 2001). In previous simulation studies, it has been found that random variables drawn from distributions in Tukey's $g$ family with a conventional coefficient of skew of 1.0 have an octile skew close to 0.2 (Lark *et al.*, 2006), so a rule of thumb, equivalent to that of Webster & Oliver (2001), is to consider data for transformation if the octile skew exceeds 0.2.

The octile skews are shown in Table 1. Note that all are smaller than 0.2 and that the octile skews for alr-transformed clay and silt are equal to or slightly smaller (closer to zero) than the values for the untransformed contents of these fractions. For this reason, we conclude that the alr-transformed variables can be assumed to be normally distributed but with some outliers. We do not wish to remove these outliers because there is no reason to believe that they are erroneous; rather we expect that they reflect real soil variation. The outliers are located at positions on the northern boundary of the paddock where the soils are formed in coarse-textured colluvium and sorted alluvium, much coarser than the parent material elsewhere in the field.

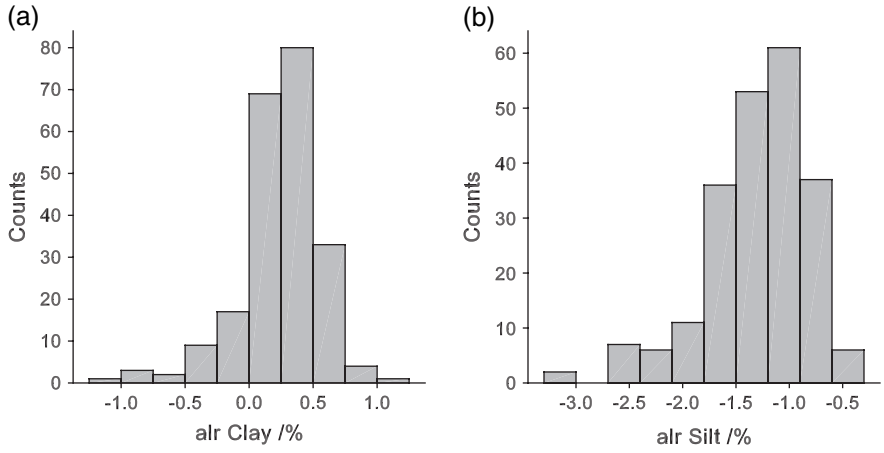


**Figure 5** Histograms of alr-transformed (a) clay content and (b) silt content at East Creek.

However, we do not want the outliers unduly to influence the fitted variograms because these will bias the kriging variances and also the covariance matrix of kriging errors, $\mathbf{C_X}$, which is required to compute the G–H back-transform. We therefore used a robust estimator of the auto- and cross-variograms, $\hat{\gamma}^{\mathrm{M}}_{u,v}(\mathbf{h})$ proposed by Lark (2003) in addition to the standard estimators (Webster & Oliver, 2001). These estimators were applied to a subset of 169 prediction data drawn at random from the full data set. The remaining 50 data were retained for validation. Figure 6 shows the estimates of the cross- and autovariograms of the alr-transformed data (robust and standard estimators) and the untransformed clay and silt contents (standard estimators). The LMCR, fitted as for the Sandford data, is also shown (solid lines).

Estimates of the clay, silt and sand content at the 50 validation points were obtained by the same three approaches used with the Sandford data. In addition to this, we predicted the alr-transformed values for clay and silt content by ordinary cokriging using the variogram models fitted to the robust estimates of the auto- and cross-variogram. These predictions were then back-transformed to values of clay, silt and sand by means of the unbiased back-transform through G–H quadrature. The same validation statistics used for the Sandford data were computed at the 50 validation sites, and these results are presented in Table 3.

In addition, we predicted the particle size fractions at a fine grid across the study field by alr cokriging with the unbiased back-transform (using a robust LMCR). The results are shown in Figure 7.

The LMCRs fitted to the transformed and untransformed data differ, as we observed in the Sandford case study, although here the structural correlations of the spatially dependent components of variation are all weak: – 0.35 for the untransformed data in contrast to 0.35 (standard estimator) or 0.3 (robust estimator) for the transformed data.

There is little difference between the validation statistics for the four approaches to estimation. The STRESS values are identical. The mean square errors are slightly smaller when the back-transform is done by G–H quadrature (robust or standard variogram estimators) than when the simple agl transform is applied to the estimates. However, there is no consistent advantage of these unbiased back-transformed results over ordinary cokriging on the untransformed data.

The kriged maps of the particle size fractions are plausible given our pedological knowledge. The field includes two main soil types, a heavy-textured Grey Vertosol (Isbell, 1996) that is found along the southern boundary and in the eastern half of the field, shown as the darker colours in the map of clay content and as lighter colours in the map of sand content (Figure 7). A coarse-textured Red Chromosol (Isbell, 1996) is also found in the field, particularly in the northwestern corner of the field. In addition, as explained earlier, the northern boundary of the field has been heavily eroded and the soil here is very coarse, as shown by the darker colours in the map of sand content (Figure 7).

## Discussion and Conclusions

In the Sandford case study, there was an advantage of alr cokriging over cokriging the untransformed variables, particularly as measured by the mean square error of the predictions of clay content, but no consistent difference between the simple agl back-transform and the unbiased back-transform by G–H quadrature. In the case of East Creek, the unbiased back-transform was better than the agl, but showed no consistent improvement over the simple cokriging of untransformed variables.

A likely reason for the differences between the two study sites can be seen in Figure 8.

Here, we plot the particle size fractions for both sites on a ternary diagram, superimposed on contours that join points in the simplex where the compositional Mahalanobis distances from the mean vector of the alr-transformed data are equal. The Mahalanobis distance has been widely used in soil science for multivariate analysis (Webster & Oliver, 1990). It is a distance measure that reflects how the variables are correlated. The compositional Mahalanobis distance, $\delta_{\mathrm{m}}$, is the Mahalanobis distance between two compositions after alr transformation. On

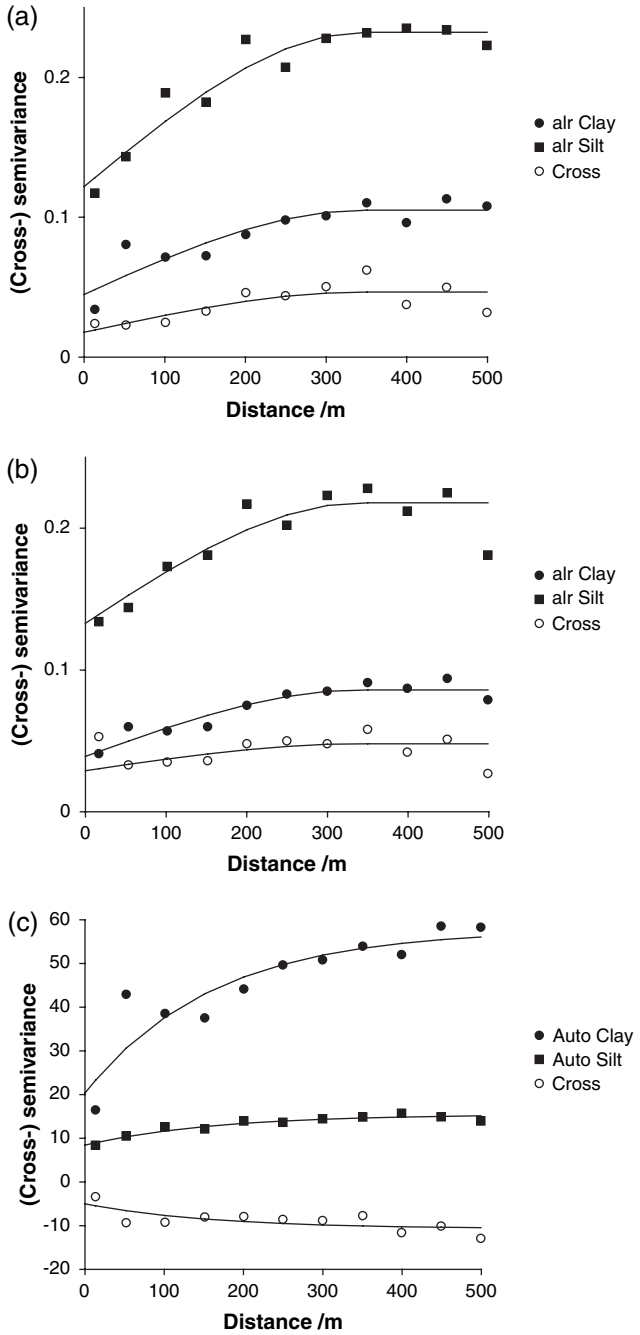


**Figure 6** Auto- and cross-variograms for clay and silt content at East Creek with: (a) alr transformation; (b) alr transformation and robust estimator; and (c) no transformation. Fitted models are shown by solid lines.

Figure 8, the contoured value at a point on the simplex that corresponds to composition $\mathbf{z}$ is

$$\delta_{\mathrm{m}}(\mathbf{z}, \bar{\mathbf{x}}) = \left\{ [\mathrm{alr}(\mathbf{z}) - \bar{\mathbf{x}}]^{\mathrm{T}} \mathbf{S}^{-1} [\mathrm{alr}(\mathbf{z}) - \bar{\mathbf{x}}] \right\}^{\frac{1}{2}}, \qquad (10)$$

where $\mathbf{S}_{\mathbf{X}}$ is the covariance matrix of the data set after the alr transform, and $\bar{\mathbf{x}}$ is the vector of mean values of the alr-transformed data.

**Table 3** Results for prediction at validation sites for the East Creek data

| | Mean square error | | | |
|---|---|---|---|---|
| Kriging method | Clay | Silt | Sand | STRESS |
| Cokriging[a] | 49.2 | 20.5 | 30.0 | 0.34 |
| alr cokriging[b] | 50.1 | 22.2 | 29.3 | 0.34 |
| alr cokriging[c] | 48.6 | 21.3 | 29.2 | 0.34 |
| alr cokriging[d] | 49.1 | 21.6 | 29.1 | 0.34 |

[a]Direct cokriging of the composition.
[b]alr cokriging, with direct agl back-transformation.
[c]alr cokriging, with back-transformation by Gauss–Hermite quadrature.
[d]alr cokriging, with back-transformation by Gauss–Hermite quadrature, robust LMCR.

If alr($\mathbf{Z}$) is a multivariate normal random variable, then two observations that correspond to the same probability density of this variable will be at the same Mahalanobis distance from the mean. When these contours are projected onto a real plane, they will take an ellipsoidal shape, reflecting the correlation between the variables on the plane. However, the projection onto the simplex shows distortion due to constraints on the distribution of the data near the edges and vertices of the simplex. This is apparent on Figure 8.

It is instructive to compare the two study sites on Figure 8. Consider first the (Figure 8b) plot for East Creek. Here, most of the data are distributed in a small ellipsoidal cluster on the simplex. Within this region, the distortion of the contours of $\delta_{\mathrm{m}}$, due to the proximity of an edge of the simplex, is rather limited. In short, for most of the observations, the assumption that the untransformed data (on the simplex) be treated as a realization of an unconstrained multivariate normal process seems reasonable. In contrast, the data from Sandford (Figure 8a) are distributed over much of the simplex and many are found near the vertices, where the distortion of the contours of $\delta_{\mathrm{m}}$, due to compositional constraints, is greatest. However, quite a few of the data are near the centre of the simplex, where the departure of the contours from an ellipsoid is much less marked.

From this, we can see that to treat the data from East Creek, the raw compositions, as approximately normally distributed in real space and to ignore the compositional constraints is not unreasonable. This is because their dispersion is relatively small and they are not centred near a vertex of the simplex. It is therefore not surprising that the STRESS of the estimates obtained by the different methods are the same and that the differences in the mean square errors are small. In contrast, many of the data at Sandford are near the vertices (particularly for large sand contents), where the effects of the compositional constraints, as shown by the contours of $\delta_{\mathrm{m}}$, are most marked. For this reason, there is an advantage in using alr cokriging, and Figure 3 shows that this is generally most apparent where the clay contents are locally very large or very small. However, many of the observations are near the centre of the simplex,

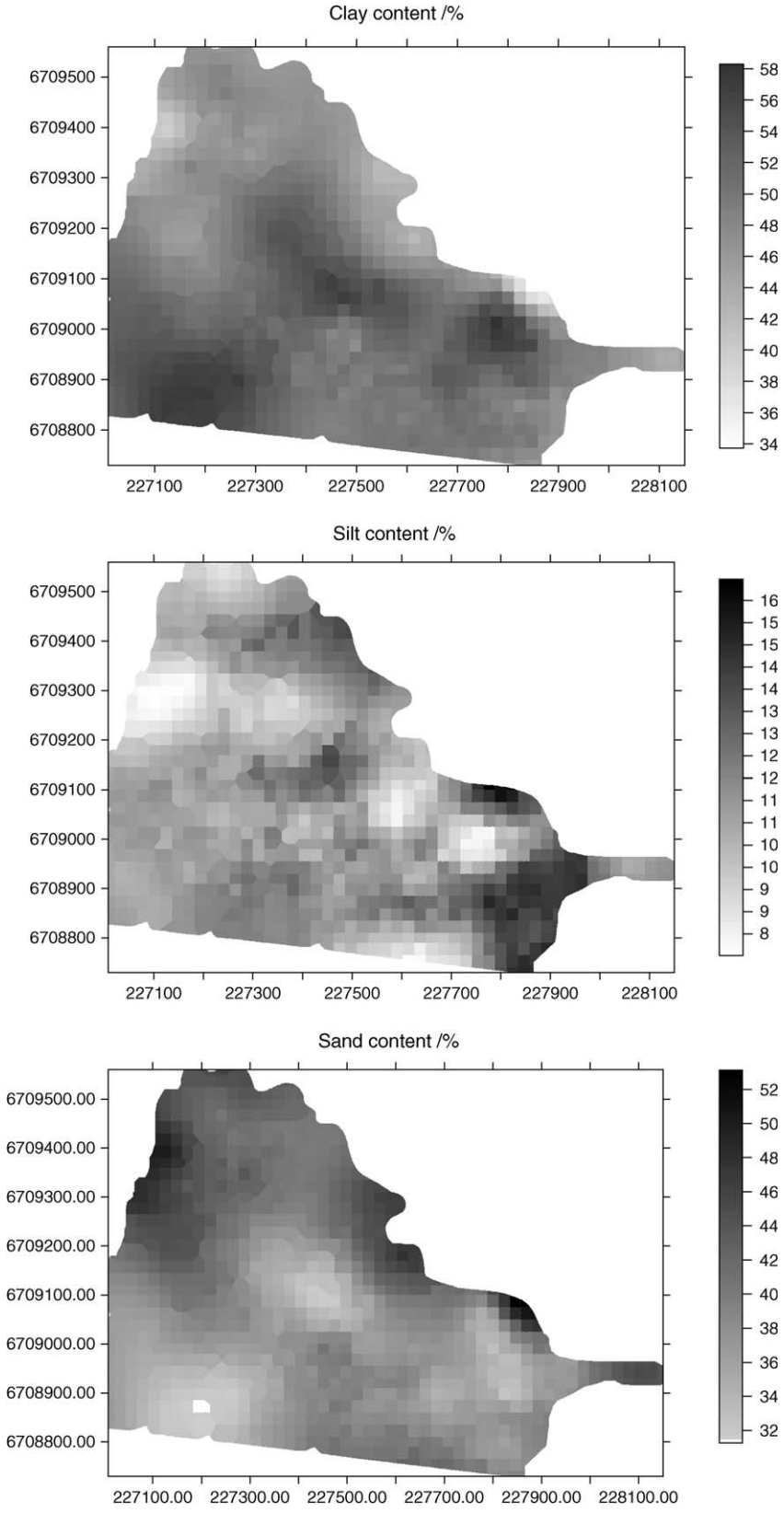


**Figure 7** Cokriged estimates of textural fractions across East Creek obtained by alr cokriging and G–H back-transformation. Coordinates are in metres according to the Map Grid of Australia.

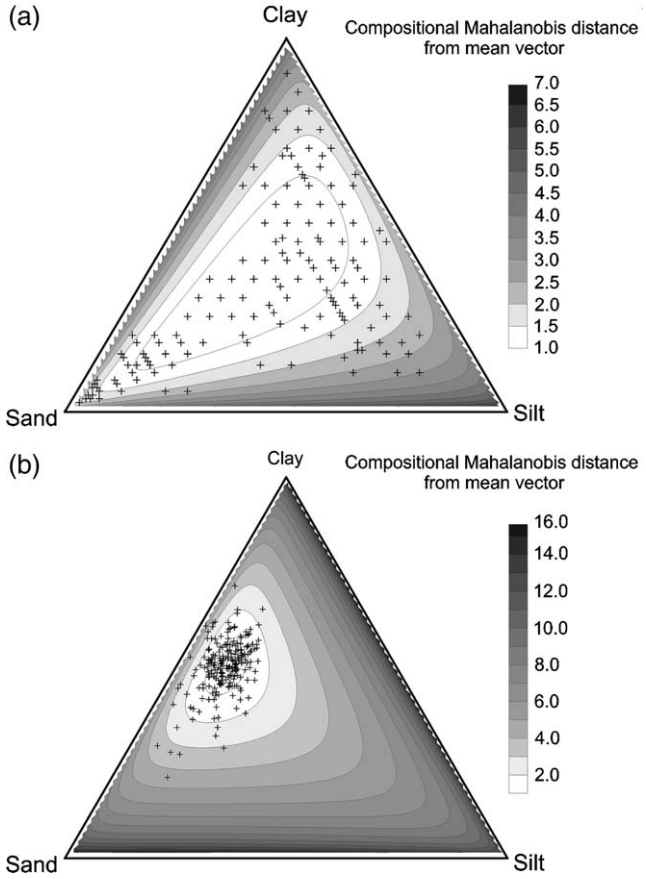


**Figure 8** Ternary diagram for (a) Sandford data and (b) East Creek. In each case, contours of equal compositional Mahalanobis distance from the mean are superimposed. Note that all ternary diagrams in this paper are labelled according to the following convention. The vertex labelled 'Clay' is the position in the simplex where the clay content of the soil is 100% and no other fractions are found. At all positions on the opposite edge of the simplex, the clay content is zero. Lines of equal clay content are parallel to this edge.

where the distortion is small, so the advantage of the alr transform is not the same everywhere.

What practical conclusions should be drawn from these case studies? An advantage of alr cokriging over cokriging of untransformed data is expected in theory, but was not seen consistently. However, the differences are explicable given the distribution of the data on the simplex (Figure 8). Statistical methods are not developed by induction from case studies, although useful rules of thumb (such as how critical certain assumptions are) can be generated this way. The fact that the alr cokriging, with the G–H back-transform, has a strong theoretical background gives us confidence in using it as a general method. We conjecture that the practical advantages of the method over ordinary cokriging of the raw compositions would be much larger than seen here in a study area where the size fractions are centred near a vertex of the simplex. A plot of the data on a ternary diagram with the contours of $\delta_m$ might be used as a diagnostic tool to decide when the more complex analysis is needed. Such practical guidelines can only emerge as soil scientists acquire experience of these methods with a range of data. In the meantime, we must note that alr cokriging has one clear advantage. If we ignore the compositional nature of our data, then we have to make some arbitrary decisions. If we cokrige, then we must decide which fraction to exclude from the estimation (and determine by difference afterwards) because the covariance matrix of the full composition is singular. If we determine each variable separately by ordinary kriging, then an arbitrary renormalization of the results (to ensure that they sum to 100%) is still needed. In alr cokriging, this is avoided altogether, as one of the fractions is used as the denominator of the transform to $\mathbb{R}^{D-1}$ space, but as we note above, the alr-cokriging estimates are not affected by this decision.

Large improvements were not achieved, in the case studies, from the use of numerical quadrature to approximate the conditional expectation of the composition from the cokriged alr values. However, there is some advantage over the straight agl back-transform in the East Creek case. In general, an advantage of using the transform is that we can obtain an estimation variance for our predictions in the original units, although these have to be interpreted with caution.

Pawlowsky-Glahn & Olea (2004) point out that, under the assumption that our composition is a normal random variable under alr transformation, the simple agl back-transform of the kriged estimate of the alr variables gives an estimate of the median and mode of the conditional distribution of each element of the composition. The simple back-transform may therefore be optimal in some sense other than the least squares (it provides an estimate of the 'centre' of the conditional distribution). This may be suitable for some purposes. For example, it may be entirely satisfactory to predict the memberships in fuzzy classes created by *k*-means clustering (which constitute a composition), as done by McBratney *et al*. (1992).

Other issues require further research. In principle, it should be possible to include variables other than the alr-transformed elements of the composition in the cokriging system, and so we might improve predictions of the soil texture by including other variables such as remote sensor data or data on soil electrical conductivity. Similarly, it should be possible to compute an empirical best linear unbiased prediction (E-BLUP) from a linear mixed model for the alr-transformed data that includes a spatial trend or a regression on some external drift variable. The only complication would be to compute correctly the covariance matrix of the co-prediction errors of the alr terms, required for the back-transform.

To conclude, soil particle size fractions can be predicted from compositional data by alr cokriging, and this has advantages over ordinary cokriging without transformation. The extent of these advantages seems to depend on how far the distribution of the data in real space is actually constrained by the simplex. For many purposes, there are also advantages if the back-transformation of the alr-cokriged estimates is computed by numerical quadrature to approximate the conditional expectation.

 

## Acknowledgements

This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC) through its core grant to Rothamsted Research. T. F. A. Bishop is supported by the BBSRC through grant D20191. Two anonymous referees made helpful comments on this paper.

## References

Abramowitz, M. & Stegun, I.A. 1964. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York.

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

Aitchison, J.. 1992. On criteria for measures of compositional difference. *Mathematical Geology*, **24,** 365–379.

Bishop, T.F.A. & McBratney, A.B. 2001. A comparison of prediction methods for the creation of field-extent soil property maps. *Geoderma*, **103,** 151–162.

Brys, G., Hubert, M. & Struyf, A. 2003. A comparison of some new measures of skewness. In: *Developments in Robust Statistics* (eds R. Dutter, P. Filzmoser, U. Gather & P.J. Rousseeuw), pp. 98–113. Physica-Verlag, Heidelberg.

Chang K.-L. 2002. Optimal estimation of the granulometric composition of soils. *Soil Science*, **167,** 135–146.

de Gruitjer, J.J., Walvoort, D.J.J. & van Gaans, P.F.M. 1997. Continuous soil maps — a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models. *Geoderma*, **77,** 169–195.

Healy, M.J.R. 1986. *Matrices for Statistics*. Oxford University Press, Oxford.

Isbell, R.F. 1996. *The Australian Soil Classification*. CSIRO, Melbourne.

Krzanowski, W.J. 1988. *Principles of Multivariate Analysis*. Oxford University Press, Oxford.

Lark, R.M. 2003. Two robust estimators of the cross-variogram for multivariate geostatistical analysis of soil properties. *European Journal of Soil Science*, **54,** 187–201.

Lark, R.M. & Papritz, A. 2003. Fitting a linear model of coregionalization for soil properties using simulated annealing. *Geoderma*, **115,** 245–260.

Lark, R.M., Bellamy, P.H. & Rawlins, B.G. 2006. Spatio-temporal variability of some metal concentrations in the soil of eastern England, and implications for soil monitoring. *Geoderma*, **133,** 363–379.

McBratney, A.B., de Gruitjer, J.J. & Brus, D.J., 1992. Spacial prediction and mapping of continuous soil classes. *Geoderma*, **54,** 39–64.

Odeh, I.O.A., Todd, A.J. & Triantafilis, J. 2003. Spatial prediction of soil particle-size fractions as compositional data. *Soil Science*, **168,** 501–515.

Olea, R.A. 1999. *Geostatistics for Engineers and Earth Scientists*. Kluwer Academic Publishers, Dordrecht.

Pawlowsky, V., Olea, R. & Davis, J.C. 1995. Estimation of regionalized compositions: a comparison of three methods. *Mathematical Geology*, **27,** 105–127.

Pawlowsky-Glahn, V. & Olea, R.A. 2004. *Geostatistical Analysis of Compositional Data*. Oxford University Press, New York.

Shatar, T.M. & McBratney, A.B. 1999. Empirical modelling of relationships between sorghum yield and soil properties. *Precision Agriculture*, **1,** 249–276.

Webster, R. & Cuanalo de la C., H.E.. 1975. Soil transect correlograms of north Oxfordshire and their interpretation. *Journal of Soil Science*, **26,** 176–194.

Webster, R. & Oliver, M.A. 1990. *Statistical Methods in Soil and Land Resource Survey*. Oxford University Press, Oxford.

Webster, R. & Oliver, M.A. 2001. *Geostatistics for Environmental Scientists*. John Wiley & Sons, Chichester.

## Appendix 1: Transformation and back-transformation of a random variate

We summarize here results on the pdf of the transform of a random variate. These are needed to follow the computation of the unbiased back-transform of the estimates of the alr transform of a composition. More details can be found in textbooks of multivariate analysis, such as the one by Krzanowski (1988).

Let $\mathbf{X}$ be a random variate comprising $p$ variables; we denote by $\mathbf{x}$ an observed variate that we assume to be a particular realization of this random process. The elements of $\mathbf{x}$ are $x_1, x_2, \ldots, x_p$. We may obtain the variate $\mathbf{z}$ with elements $z_1, z_2, \ldots, z_p$ by a transformation of $\mathbf{x}$:

$$\begin{aligned} z_1 &= \phi_1(x_1, x_2, \ldots, x_p) \\ &\vdots \\ z_p &= \phi_p(x_1, x_2, \ldots, x_p). \end{aligned}$$

The Jacobian of this transform is a scalar quantity, the determinant of the matrix of partial derivatives of each element of $\mathbf{z}$ with respect to each element of $\mathbf{x}$:

$$J_\phi(\mathbf{x}) = \begin{vmatrix} \frac{\partial z_1}{\partial x_1} & \frac{\partial z_2}{\partial x_1} & \cdots & \frac{\partial z_p}{\partial x_1} \\ \frac{\partial z_1}{\partial x_2} & \frac{\partial z_2}{\partial x_2} & \cdots & \frac{\partial z_p}{\partial x_2} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial z_1}{\partial x_p} & \frac{\partial z_2}{\partial x_p} & \cdots & \frac{\partial z_p}{\partial x_p} \end{vmatrix}.$$

The Jacobian of the transform is used to determine its properties, such as the existence of a back-transform. If the back-transform exists, then we denote it by the following expression,

$$\begin{aligned} \mathbf{x}_1 &= \Phi_1(z_1, z_2, \ldots, z_p) \\ &\vdots \\ \mathbf{x}_p &= \Phi_p(z_1, z_2, \ldots, z_p) \end{aligned},$$

and its Jacobian is $J_{\Phi(\mathbf{z})} = J^{-1}_{\phi(\mathbf{x})}$.

Now, assume that we know the pdf of $\mathbf{X}$, $f(\mathbf{X})$. Our objective is to obtain the pdf $f(\mathbf{Z})$. To compute the probability density for some particular vector of values of the variables in $\mathbf{Z}$, $z_1, z_2, \ldots, z_p$, we compute the corresponding values of $x_1, x_2, \ldots, x_p$ with the transform $\Phi_i$, then we obtain the pdf for these values from $f(\mathbf{X})$and multiply the result by $J^{-1}_{\phi(\mathbf{x})}$.

Thus, in Equation (4), the pdf of our compositional variate ($\mathbf{Z}$) is obtained by substituting alr($\mathbf{Z}$)for the transformed variable ($\mathbf{X}$) in the $p$-variate Gaussian pdf and multiplying by the inverse of the Jacobian of the agl transform.

## Appendix 2: The derivation of Equation (6) for the Gauss-Hermite back-transform

We need to factorize our integral in Equation (3) so that it can be presented in the form of Equation (5). This is done by setting

$$\mathbf{Y}^{\mathrm{T}}\mathbf{Y} = \left\{\frac{1}{2}[\mathrm{alr}(\mathbf{Z}) - \bar{\mu}_{\mathbf{X}}]^{\mathrm{T}}\mathbf{C}_{\mathbf{X}}^{-1}[\mathrm{alr}(\mathbf{Z}) - \bar{\mu}_{\mathbf{X}}]\right\}. \tag{A1}$$

The covariance matrix of cokriging errors, $\mathbf{C}_{\mathbf{X}}$, is a positive definite matrix (Webster & Oliver, 1990) because its elements are computed from cross- and auto-variograms that constitute an authorised LMCR (Webster & Oliver, 2001). It follows from this (see, for example, Healy, 1986) that we can compute a factorization of $\mathbf{C}_{\mathbf{X}}$ into the product of an upper and lower triangular matrix (i.e. matrices with all zeros either above or below the main diagonal), one of which is the transpose of the other:

$$\mathbf{C}_{\mathbf{X}} = \mathbf{R}^{\mathrm{T}}\mathbf{R}.$$

This is the Cholesky decomposition or lower–upper (LU) factorization of $\mathbf{C}_{\mathbf{X}}$. Other properties of the LU factorization include

$$\mathbf{C}_{\mathbf{X}}^{-1} = \left(\mathbf{R}^{\mathrm{T}}\mathbf{R}\right)^{-1} = \mathbf{R}^{-1}\left(\mathbf{R}^{\mathrm{T}}\right)^{-1} = \mathbf{R}^{-1}\left(\mathbf{R}^{-1}\right)^{\mathrm{T}},$$

so we can write Equation (A1) as

$$\mathbf{Y}^{\mathrm{T}}\mathbf{Y} = \left\{\frac{1}{\sqrt{2}}[\mathrm{alr}(\mathbf{Z}) - \bar{\mu}_{\mathbf{X}}]^{\mathrm{T}}\mathbf{R}^{-1}\frac{1}{\sqrt{2}}\left(\mathbf{R}^{-1}\right)^{\mathrm{T}}[\mathrm{alr}(\mathbf{Z}) - \bar{\mu}_{\mathbf{X}}]\right\}, \tag{A2}$$

so

$$\mathbf{Y} = \frac{1}{\sqrt{2}}\left(\mathbf{R}^{-1}\right)^{\mathrm{T}}[\mathrm{alr}(\mathbf{Z}) - \bar{\mu}_{\mathbf{X}}], \tag{A3}$$

therefore

$$\sqrt{2}\mathbf{Y} = \left(\mathbf{R}^{-1}\right)^{\mathrm{T}}[\mathrm{alr}(\mathbf{Z}) - \bar{\mu}_{\mathbf{X}}],$$

therefore

$$\sqrt{2}\mathbf{R}^{\mathrm{T}}\mathbf{Y} = [\mathrm{alr}(\mathbf{Z}) - \bar{\mu}_{\mathbf{X}}],$$

therefore

$$\sqrt{2}\mathbf{R}^{\mathrm{T}}\mathbf{Y} + \bar{\mu}_{\mathbf{X}} = \mathrm{alr}(\mathbf{Z}),$$

therefore

$$\mathbf{Z} = \mathrm{agl}\left(\sqrt{2}\mathbf{R}^{\mathrm{T}}\mathbf{Y} + \bar{\mu}_{\mathbf{X}}\right). \tag{A4}$$

We now require the Jacobian of the transform

$$\mathbf{X} = \sqrt{2}\mathbf{R}^{\mathrm{T}}\mathbf{Y} + \bar{\mu}_{\mathbf{X}},$$

where

$$\mathbf{Z} = \mathrm{agl}(\mathbf{X}).$$

Now

$$\left|\frac{\partial\mathbf{X}}{\partial\mathbf{Z}}\right| = \left(\prod_{i=1}^{D} Z_i\right)^{-1}, \tag{A5}$$

i.e. the inverse of $J_{\mathrm{alr}(\mathbf{Z})}$.

$$\begin{aligned}\left|\frac{\partial\mathbf{X}}{\partial\mathbf{Y}}\right| &= \left|\frac{\partial\sqrt{2}\mathbf{R}^{\mathrm{T}}\mathbf{Y} + \bar{\mu}_{\mathbf{X}}}{\partial\mathbf{Y}}\right| \\ &= \left|\frac{\partial\sqrt{2}\mathbf{R}^{\mathrm{T}}\mathbf{Y}}{\partial\mathbf{Y}}\right|,\end{aligned}$$

because $\bar{\mu}_{\mathbf{X}}$ is constant with respect to $\mathbf{Y}$, and so

$$\left|\frac{\partial\mathbf{X}}{\partial\mathbf{Y}}\right| = \sqrt{2}^{D-1}\left|\mathbf{R}^{\mathrm{T}}\right| = 2^{\frac{D-1}{2}}\left|\mathbf{R}^{\mathrm{T}}\right|, \tag{A6}$$

because the determinant of a triangular matrix is the product of its diagonal elements.

From Equation (A5), we may write

$$\partial\mathbf{Z} = \left(\prod_{i=1}^{D} Z_i\right)\partial\mathbf{X}.$$

Substituting in from Equation (A6) gives us

$$\begin{aligned}\partial\mathbf{Z} &= \left(\prod_{i=1}^{D} Z_i\right)2^{\frac{D-1}{2}}\left|\mathbf{R}^{\mathrm{T}}\right|\partial\mathbf{Y} \\ &= \left(\prod_{i=1}^{D} Z_i\right)2^{\frac{D-1}{2}}\left|\mathbf{C}_{\mathbf{X}}\right|^{\frac{1}{2}}\partial\mathbf{Y},\end{aligned} \tag{A7}$$

because $\left|\mathbf{R}^{\mathrm{T}}\right| = \left|\mathbf{R}\right| = \left|\mathbf{C}_{\mathbf{X}}\right|^{\frac{1}{2}}$.

From this, we can write

$$\partial\mathbf{Y} = 2^{-\frac{D-1}{2}}\left(\prod_{i=1}^{D} Z_i\right)^{-1}\left|\mathbf{C}_{\mathbf{X}}\right|^{-\frac{1}{2}}\partial\mathbf{Z}. \tag{A8}$$

We can now write out an expression for the term under the integration sign in Equation (3) by substituting in from Equation (A4) for $\mathbf{Z}$, from Equation (4) for $f(\mathbf{Z})$ and from Equation (A7) for $\partial\mathbf{Z} = \ldots\partial\mathbf{Y}$. This gives us

$$\begin{aligned}&\mathrm{agl}\left(\sqrt{2}\mathbf{R}^{\mathrm{T}}\mathbf{Y} + \bar{\mu}_{\mathbf{X}}\right)(2\pi)^{-\frac{D-1}{2}}\left|\mathbf{C}_{\mathbf{X}}\right|^{-\frac{1}{2}}\left(\prod_{i=1}^{D} Z_i\right)^{-1} \\ &\times \exp\left\{-\mathbf{Y}^{\mathrm{T}}\mathbf{Y}\right\}\left(\prod_{i=1}^{D} Z_i\right)2^{\frac{D-1}{2}}\left|\mathbf{C}_{\mathbf{X}}\right|^{\frac{1}{2}}\partial\mathbf{Y}.\end{aligned}$$

We can therefore write

$$\bar{\mu}_{\mathbf{Z}} = \int_{\mathbb{R}^{D-1}} \pi^{-\frac{D-1}{2}}\mathrm{agl}\left(\sqrt{2}\mathbf{R}^{\mathrm{T}}\mathbf{Y} + \bar{\mu}_{\mathbf{X}}\right)\exp\left\{-\mathbf{Y}^{\mathrm{T}}\mathbf{Y}\right\}\mathrm{d}\mathbf{Y}. \tag{A9}$$

Inspecting Equation (5), we see that the above integral can be evaluated by setting

$$g(\mathbf{Y}) = \pi^{-\frac{D-1}{2}}\mathrm{agl}(\sqrt{2}\mathbf{R}^{\mathrm{T}}\mathbf{Y} + \bar{\mu}_{\mathbf{X}}). \tag{A10}$$

 