A stochastic geometric model for continuous local trends in soil variation

R.M. Lark[1]

British Geological Survey, Keyworth, Nottinghamshire NG12 5GG, U.K.

Abstract

This paper develops and demonstrates a model of stochastic spatial variation. It is proposed that this model may represent soil variability according to a particular mode under which the soil varies continuously, showing short-range lateral trends induced by local effects of the factors of soil formation which vary across the region of interest in an unpredictable way. The trends in soil variation are therefore only apparent locally, and the soil variation at regional scale appears random. Such variation might be expected in a landscape where the soil varies along topographic catenas which repeat across the region in response to a drainage pattern which is not entirely regular in spacing or orientation, and is therefore unpredictable. The Continuous Local Trend (CLT) mode of soil variation may also be expected where gradients of soil properties are induced around individual plants, or plant roots.

In the stochastic model the local trend is assumed to be described by a function of distance to the nearest event in a realization of a random spatial point process. A model is developed here in which the point process shows complete spatial randomness, so it is called the Poisson Continuous Local Trend (PCLT) model. The covariance function for the PCLT with a general distance function is developed and some hypothetical examples are shown, including one in which the variogram of a soil property is inferred by using a published topofunction. The PCLT model is then fitted to the empirical variogram of some data on soil water content in a gently undulating clay landscape,

[1]Corresponding author: *E-mail address*: mlark@nerc.ac.uk (R.M. Lark).

and the multiple point statistics of the PCLT model for these data are compared with those of a corresponding multivariate normal model.

Keywords: Linear mixed model; Stochastic geometry; Voronoi tessellation; Multiple point geostatistics; Topofunction.

---

1. Introduction

Geostatisticians use mixed models to analyse and predict soil properties. In these models some of the soil variation is accounted for by fixed effects, continuous covariates or categorical factors, and the remaining variation is modelled as random effects, including a spatially correlated component (Lark *et al.,* 2006). Typically our knowledge of soil processes is put to use by selection of appropriate fixed effects for such models. The random effects account for the soil variation that we cannot explain in terms of fixed effects. Either no fixed effects can be formulated, because of the complexity of the origins of the soil variation and its dependence on contingent events in the prehistory of the landscape (Webster, 2000), or appropriate covariates are not measured at the scale of interest in the region under study.

The spatial correlation of the random effects is modelled by a covariance function typically selected from a set of authorised functions with convenient mathematical properties (Webster and Oliver, 2007). However, covariance models for the random effects would ideally be selected because they represent the processes that cause the variation. One advantage of such an approach would be that prior distributions for the covariance parameters could be specified from scientific knowledge and understanding of the underlying processes. These prior models could then be used to improve the efficiency of sampling (Marchant and Lark, 2006).

The relationship between the form of the covariance function and the underlying physical processes is well established for diffusion (Whittle, 1954; 1962) and for vari-

ables in branches of the earth sciences including hydrology (Kolvos *et al.,* 2004) and geophysics (Chilès and Delfiner, 1999), but we might reasonably observe that in most cases the factors underlying soil variation are too complex to allow a straightforward inference from process understanding to the form of the covariance function. However, we might identify a model of random variation in space that represents a general *mode* of soil variation that we can expect to encounter in particular conditions.

By a mode of soil variation is meant a simple and generalizable rule that captures how the effect of a factor of soil formation varies laterally. The mode of variation for a variable is a basis for prediction of features of its statistical distribution (e.g. Allègre and Lewin, 1995) and for decisions such as the selection of a transformation or model. For example, if soil variation is associated with microtopography in a landscape which shows pronounced and regular periodicity (e.g. ridge and furrow), then we might call the expected mode of variation *periodic*, and expect to see a variogram with a regular fluctuation. Webster and Oliver (2007) note that apparent fluctuations in the empirical variogram can be artefacts, arising, for example, from strongly clustered sampling, and advise against the routine selection of periodic variograms models just because they fit. Pedological knowledge that a variable arises from a periodic mode of variation gives us confidence both to select a variogram model with a periodic component and to interpret the wavelength of the fluctuation in the variogram as real information about the underlying mode of variation (its wavelength) and the soil-forming factors that underly it.

Lark (2009) considered another mode of soil variation where the factors of soil formation operate within *discrete domains* (different geological units, agricultural fields, catchments etc.) The Poisson Voronoi Tesselation (PVT) model was proposed for random variation of soil according to this mode, based on the partition of space into Dirichlet tiles around seed points drawn from a Poisson spatial point process. The model fitted well to the empirical variograms of soil properties measured at a range of

scales. Lark (2010) showed that the PVT model was a more plausible model of the variation in several soil data sets than was an alternative multivariate normal model. However, it is clear that a model based on discrete domains will not be universally appropriate for the random variation of the soil. It is necessary to develop a wider range of random models for other modes of soil variation.

In this paper I propose a random model for soil variation that exhibits *continuous local trends* (CLT). This mode of soil variation can be exemplified at disparate spatial scales. For example, gradients of soil properties may be induced around individual plants (Pérez, 1995) or individual rhizospheres (Youseff and Chino, 1989). Gradients of soil properties have also been reported from the centre to the margins of the polygons in patterned ground (Barrett *et al.,* 2004). Such variation is continuous (there are no step changes in the soil property), and is characterised by lateral trends. However, the trend is not global (at the scale of the whole region of interest) but rather is local induced by an underlying process such as the distribution of plants, roots or periglacial polygons whose distribution is not predictable at a global scale. The local trends therefore form a repeating pattern across the region, which cannot be regarded as a simple deterministic function (unlike a global trend across the region), and may, in the absence of an appropriate covariate (such as a remote sensor image of patterned ground) be consigned to the random effects of a mixed model.

The CLT mode of soil variation is exemplified at landscape scale by certain forms of catenary variation. The concept of the catena was introduced by Milne (1936) to facilitate soil survey in East Africa. Milne's catenas represent a pattern of soil variation across a valley from drainage line to interfluve. Variation along a catena may be continuous, or abrupt: for example at the transition from woodland to grassland at the margin of the dambo which occupies the bottom of the catena described by Webster (1965). In a catenary landscape soil varies predictably across a valley from one interfluve to the next, but across a region this sequence repeats, constituting a

pattern. Milne delineated map units within which a characteristic catenary pattern of variation could be discerned. One might, as Webster (2000) observed, regard the variation of a soil property at locations within such a unit as random because of the unpredictability of the drainage pattern. One such landscape is the Eldama landsystem in Western Kenya, as surveyed by Scott et al. (1971). A block diagram of this land system is shown in Figure 1a. In the mixed model context one might assign this variation to fixed effects if it can be represented by covariates, perhaps drawn from a digital elevation model, or otherwise to random effects. The CLT mode of variation would be exemplified by a repeating catenary pattern which can be represented by a continuous topofunction (Yaalon, 1975) such as those proposed for various landscapes by Walker (1966), Ruhe (1969), Walker and Ruhe (1968), Walker *et al.,* (1968) and Kleiss (1970). Continuous local trends, associated with topography are also predicted by pedogenetic models (e.g. Rosenbloom et al., 2001). In those landscapes where the drainage is strongly oriented in one direction the CLT mode of variation is essentially one-dimensional (across the drainage line), this is illustrated by the Lolimo land system in the survey of Western Kenya by Scott et al. (1971), shown in Figure 1b. A two-dimensional mode of variation could be envisaged in circumstances where the local direction of the drainage line is unpredictable for a randomly located site in the region.

In this paper I propose a stochastic model for the CLT mode of soil variation. In this model it is assumed that local trends are induced by the events in a realization of a random spatial point process (which could, for example, correspond to positions of individual plants in the example of CLT variation presented by Pérez (1995). The value of the CLT process at any location depends on the distance to the nearest event from the underlying point process. In this paper I assume complete spatial randomness of the point process, which induces a Poisson CLT (PCLT). In the remainder of this paper I derive this model in more detail and show the form of the variogram for a number of hypothetical instances. I then fit the PCLT model to the empirical variogram of some

data on the water content of soil in an undulating clay landscape in eastern England.

2. Theory

In this paper I propose a Poisson CLT (PCLT) model of random variation in which the value of the variable at some location is a function of the distance to the nearest event from a Poisson spatial point process with specified intensity. In this section I develop this model and derive the variogram function for it.

2.1 Notation

Let $\mathbf{s}$ be an arbitrary location in our region of interest which is a d-dimensional real subspace $R \subset \mathbb{R}^d$. Let $\mathbf{h} \in \mathbb{R}^d$ be a 'structuring element', i.e. a vector of unit norm and arbitrary direction (on the assumption of isotropy when $d > 1$), and let r be a lag distance.

Let Ψ be a point process in $\mathbb{R}^d$. This is a random process and a realization of it in $\mathbb{R}^d$, ψ, is a set of points with random positions, $\mathbf{x}_i \in \mathbb{R}^d,\ i \geq 0$. Denote by $\mathcal{S}$ some subspace of our d-dimensional space, $\mathcal{S} \subset \mathbb{R}^d$. By $|\mathcal{S}|$ is denoted the Lebesgue measure of S, (that is the length in one dimension, area in two dimensions, volume in three dimensions etc). Let $\psi(\mathcal{S})$ denote a random variable which is the number of events of the point process in $\mathcal{S}$. The intensity of the spatial point process is λ such that

$$\mathrm{E}\left[\psi(\mathcal{S})\right] \ = \ \lambda\,|\mathcal{S}|\,, \tag{1}$$

where E[·] denotes the expectation of the term in square brackets. The distribution of $\psi(\mathcal{S})$ is denoted by $P\left\{\psi(\mathcal{S})\right\}$, which is a Poisson distribution if the events of the process are completely spatially random and independent. I assume here that Ψ is a stationary random process (homogeneous Poisson) so that statistics such as the intensity are invariant under a translation in space.

A Poisson Voronoi tessellation of S, denoted by T, is the partition of S into non-overlapping space-filling cells which depend on ψ. The ith cell of T, $C_i(\psi)$, contains

the ith point in ψ and only the ith point because $C_i(\psi)$ is defined as the set of all points in $\mathcal{S}$ which are closer to the ith point in ψ than to any other point in ψ. The boundary of the ith cell is denoted ∂C_i, and the tessellation T is defined uniquely by its skeleton, the union of all the cell boundaries $\partial T = \mathcal{U}_{i\geq 0}\left(\partial C_i\right)$.

For any vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$, $B\left[\mathbf{u}, ||\mathbf{v}||\right] \subset \mathbb{R}^d$ denotes the closed d-ball of radius $||\mathbf{v}||$, such that $\forall \mathbf{s},\ \mathbf{s} \in B\left[\mathbf{u}, ||\mathbf{v}||\right]$ if and only if $||\mathbf{s} - \mathbf{u}|| \leq ||\mathbf{v}||$.

By $\mathcal{U}\left(B\left[\mathbf{u}, ||\mathbf{v}||\right], B\left[\mathbf{u}', ||\mathbf{v}'||\right]\right) \subset \mathbb{R}^d$ is denoted the union of the two balls that are arguments of the expression.

2.2 The Poisson continuous local trend (PCLT) model and its variogram

Consider an arbitrary point $\mathbf{s} \in \mathcal{S} \subset \mathbb{R}^d$, where $\mathcal{S}$ contains points of a realization ψ of a Poisson process with intensity λ. We denote by $K(\mathbf{s})$ a random variable

$$K(\mathbf{s}) \ = \ \min\left\{||\mathbf{s} - \mathbf{x}_i||\right\}, \quad \forall i \geq 0, \tag{2}$$

that it to say, it is the distance from $\mathbf{s}$ to its nearest neighbour in ψ. Under the PCLT model for a variable Z, it is a random function

$$Z(\mathbf{s}) \ = \ \mathcal{D}\left(K(\mathbf{s})\right), \tag{3}$$

where $\mathcal{D}\left(\cdot\right)$ is some deterministic function that we call the distance function since its argument is a distance.

As defined, $Z(\mathbf{s})$ is an isotropic random function because Ψ is an isotropic process. This is not generally a realistic model for soil variation. In the following derivation isotropy is assumed, and the data used in the case study are on a transect so anisotropy is not an issue. However, the model could be extended to the anisotropic case in further work. This might be most easily done by proposing an underlying isotropic random field which is then subject to an affine transformation (Webster and Oliver, 2007).

The function $Z(\mathbf{s})$ is a random function because, although the distance function $\mathcal{D}\left(\cdot\right)$ is deterministic, its argument is a random variable. The function $Z(\mathbf{s})$ is not differentiable at the boundaries of Voronoi cells, that is when $\mathbf{s} \in \partial T$. Its differentiability

elsewhere depends on the distance function and $\mathbf{s}$. For example, if the distance function is $\mathcal{D}(k) = \alpha + \beta k^2$ then $Z(\mathbf{s})$ is differentiable for any $\mathbf{s} \notin \partial T$, whereas if $\mathcal{D}(k) = \alpha + \beta k$ then $Z(\mathbf{s})$ is not differentiable at any $\mathbf{s} \in \psi$ (that is, at seed points of the Voronoi cells).

Note that, while $Z(\mathbf{s})$ is not differentiable at $\mathbf{s} \in \partial T$, it is continuous at the boundary of Voronoi cells in the sense that, if $\mathbf{s} \in \partial C_i \cap \partial C_j$, i.e. $\mathbf{s}$ is at the boundary of the ith and jth cell, then

$$\{Z(\mathbf{s}_k) - Z(\mathbf{s}_l)\} \to 0 \text{ as } \mathbf{s}_k \to \mathbf{s} \text{ and } \mathbf{s}_l \to \mathbf{s} \quad \forall \mathbf{s}_k \in C_i(\psi) \text{ and } \mathbf{s}_l \in C_j(\psi), \tag{4}$$

see Siersma (1999). The function is similarly continuous at seed points.

This model is proposed as a random model for the CLT mode of soil variation. With appropriate choice of Ψ and $\mathcal{D}(\cdot)$ it may describe the variation of processes that show pronounced trends. For example, if the seed points in ψ represent local topographic maxima in a gently undulating landscape with more or less isotropic hillocks, and the skeleton of the Voronoi cells, ∂T, represent topographic minima, then, with an appropriate choice of $\mathcal{D}(\cdot)$ to represent a topofunction, the random function $Z(\mathbf{s})$ may represent soil properties strongly associated with topography in the landscape, such as water content.

There are some evident limitations to the model as presented here. In particular it is assumed that parameters of the function $\mathcal{D}(\cdot)$ are constant. This means, for example, that the value of $Z(\mathbf{s})$ is identical at all seedpoints, and that the lateral maximum rate of change of $Z(\mathbf{s})$ is the same for any $\mathbf{s} \notin \psi$ and $\mathbf{s} \notin \partial T$. This requirement could be relaxed so that parameters of $\mathcal{D}(\cdot)$ were random variables, but then $Z(\mathbf{s})$ would no longer be continuous at ∂T in the sense of Equation (4).

The next step is to derive the spatial covariance function, (or equivalently the variogram), for $Z(\mathbf{s})$. Under the assumption that Ψ is stationary and isotropic, the covariance $\text{Cov}\{Z(\mathbf{s}), Z(\mathbf{s} + r\mathbf{h})\}$ can be written as a function of only the scalar lag interval, $C(r)$. It is possible to obtain a function, $C(r)$, for a PCLT process. First, one derives the marginal distribution function for $K(\mathbf{s})$, $F(k)$. Next, the joint distribution

function for $\{K(\mathbf{s}), K(\mathbf{s}+r\mathbf{h})\}$ is obtained, which, under stationarity, can be denoted by $H(k, k_r)$. The joint distribution function for a PCLT process depends, as shown in the appendix, on the joint survival function of the point process, $S(k, k_r)$, which is the probability that both the distance from $\mathbf{s}$ to the nearest event in ψ is larger than k and the distance from $\mathbf{s}+r\mathbf{h}$ to the nearest event in ψ is larger than k_r. The covariance function for a random variable that depends on some distance function $\mathcal{D}(\cdot)$ can then be obtained. The detail of how this is done is presented in an appendix to this paper, Equation (20) gives the covariance function $C(r)$ for a PCLT process in terms of the distribution function, $F(k)$; the joint survival function, $S(k, k_r)$ and the distance function $\mathcal{D}(\cdot)$. One may then obtain the variogram of $Z(\mathbf{s})$ by the usual relationship to the covariance function:

$$\gamma(r) \;=\; C(0) \;-\; C(r). \tag{5}$$

2.3 Some theoretical examples

Figure 2 (continuous lines) shows variogram functions computed for four PCLT models in two dimensions. In all the intensity, λ, of the point process is 0.25×10^{-3} events per unit square of area, so, following Heinrich (1998) and Equation (15) of Lark (2009), the mean chord length of the Voronoi cells is ξ=63 units. Four distance functions are used, each is a simple function of distance or the square of distance and defined for all non-negative distances: $\mathcal{D}(k) = k$, $\mathcal{D}(k) = k^2/10$, $\mathcal{D}(k) = 10(k+1)^{-1}$ and $\mathcal{D}(k) = 10(k+1)^{-2}$. The PCLT variograms were obtained from Equation (5), the covariances were computed with Equation (20) with the joint-survival function and the stationary cdf of K obtained from Equations (17) and (13) respectively, for the two-dimensional case. The double integral in Equation (21) was evaluated numerically with the subroutine TWODQ in the IMSL library (Visual Numerics, 2006).

For illustrative purposes I then generated 5000 realizations of each PCLT process as follows. The intensity of the Poisson process was specified as 0.25×10^{-3} events

per unit square. The PCLT process was simulated over a 2000×2000-unit region, so the expected number of events was $\mu = 1000$. I obtained the number of events in a particular realization, n_p, as a Poisson random variables with $\mu = 1000$, using the Poisson random number generator RNPOI in the IMSL library (Visual Numerics, 2006). These n_p events were then allocated to locations within the 2000×2000-unit region independently and at random with uniform intensity. After a realization of the Poisson point process had been generated, the procedure below was followed.

1. A transect, length 1000 units and centred at the centre of the region was sampled at unit intervals.

2. At each location on the transect the distance to the nearest of the seed locations was computed, k.

3. Each of the distance functions, $\mathcal{D}(k)$, listed above was evaluated at each location on the transect.

4. For each PCLT process, corresponding to one of the distance functions, the variogram was estimated from the values on the transect, using Matheron's standard estimator as implemented for systematic sampling in one dimension by Webster and Oliver (2007).

After this had been done for each of the 5000 realizations, the average variogram over all realizations was computed for each PCLT process. These are shown as solid discs on the respective plots in Figure 2. There is good agreement between the simulations and the computed variograms.

Matérn variogram functions were fitted to these computed models by the VARIOFIT procedure in the geoR package (Ribeiro and Diggle, 2001) in R (R Development core team, 2010). The smoothness parameter ν and distance parameter ϕ are presented in Table 1 along with the effective range, at which the variogram reaches 95% of the sill variance.

All the variograms show some degree of upward-concavity at short lag distances, although this is not visually apparent for all of them in the graphs in Figure 2. This form is to be expected given that PCLT models are based on local trends. The concavity is not strong for the PCLT processes where the distance function is the reciprocal of a polynomial term; these have smaller values of the ν parameter than do the other PCLT processes. In fact the PCLT with distance function proportional to the reciprocal of distance squared has $\nu = 0.5$ which is equivalent to an exponential variogram. Note also that the PCLT models with reciprocal distance functions have effective ranges shorter than the mean chord length of the Voronoi cells, whereas the PCLT models with distance functions proportional to distance or to its square have effective ranges rather longer than the mean chord length. The interpretation of the effective range of a variogram must therefore be cautious since its relation to the length scale of an underlying process (the Voronoi cells here) depends on the form of the distance function.

Figure 3 shows marginal distribution functions (after standardization to zero mean and unit variance) for the four PCLT variables, these were generated using Equation (15) to obtain the pdf of k. Note the difference among these distributions with respect to the coefficient of skewness which is shown on the graphs. All the variables show some degree of positive skewness, but this is only strong for the variable where the distance function is proportional to the square of distance. All variables have truncated distributions, the variables with reciprocal distance functions are truncated at the maximum values (10 before standardization) and the variables with distance functions proportional to distance or the square of distance are truncated at the minimum (zero before standardization). The effect of the truncation is most apparent for the variable where the distance function is proportional to the square of distance.

One final theoretical example is considered. Walker *et al.* (1968) presented results from the analysis of the lateral variation of soil properties across a drift landscape in

central Iowa. In particular they present a topofunction that expresses the thickness of the A horizon, t_{A} (cm), defined as material with a soil organic carbon content greater than 2%), as a function of distance downslope from the local summit, d (hm). The function is a polynomial:

$$t_{\mathrm{A}} \;=\; 19.0 + 2.54d + 0.66d^2. \tag{6}$$

Walker *et al.* (1968) report a correlation coefficient from which we may infer that the coefficient of determination for the fit of this function is $R^2 = 0.79$.

I made the explicit assumption that the topofunction can be incorporated into a PCLT model in 2 dimensions, i.e. that the direction of the drainage, over a region is not uniform, and that the local summits from which d is defined in the topofunction are realizations of a Poisson point process, with some specified intensity. Given this, one can compute a covariance function for thickness of the A horizon, as predicted for the topofunction, by substituting the topofunction in Equation (6) for $\mathcal{D}(k)$ in Equation (21). I calculated this covariance function for some specified intensity, λ. I assumed that the variation in thickness of the A horizon not accounted for by the topofunction could be regarded as a pure nugget process, and that the nugget to sill ratio was approximated by $1 - R^2$ for the fitted function. Figure 4 shows the corresponding standardized variogram (i.e. standardized to a sill variance of 1.0) and with the lag distance expressed as a proportion of the mean chord length of the Voronoi cells defined by the local summits. In the absence of any further information about the variability of this property in a comparable landscape, this function summarizes the information implicit in the topofunction and the assumption that this can be incorporated into a PCLT model. Given some plausible value to use for the mean chord length, which might be inferred from the original paper to be around 200 m in this landscape, this variogram might be used to optimize a sampling scheme to produce a reliable estimate, perhaps using the procedure of Lark (2002).

3. Case Study

3.1 Data.

The case study entails the analysis of data on the gravimetric water content of soil from Central Bedfordshire. The collection of these data has been described elsewhere (Milne *et al.,* 2011). The soil was sampled with cores of diameter 44 mm from the depth interval 0–15 cm at 29.45-m intervals on a straight transect, with some additional points added at 6- and 3-m intervals from the regular locations. For purposes of this study I wanted to examine soil variation that could be expected to be dominated by topography. I therefore selected a section of the transect over a single soil association as mapped by King (1969), the Wicken Association. This association comprises Cambisols, Luvisols and Acrisols according to the World Reference Base classification (IUSS Working Group WRB, 2006). It lies over the Gault Clay, and the soils themselves were formed in the Gault Clay and overlying mixed drift. The landscape undulates gently and most of the land is under grass or arable crops. The points from the transect that were on the Wicken Association were identified from the map of King (1969). The first transect location on the Wicken Association had coordinates 508563.3, 235402.1 on the Ordnance Survey national grid of Great Britain (units are metres) and the last location was at 508897.8, 232477.0. This section of the transect comprised 111 samples. Of these 15 were collected under woodland or shrubby waste ground. The water content of these soils was much larger than the others, so the analysis was restricted to 96 samples from arable land or land under grass, of which eight were samples at shorter intervals than the basic transect. Table 1 shows summary statistics for these data. The data are weakly skewed. The empirical distribution function of the data, after they were standardized to zero mean and unit variance by the sample statistics, is shown by the open symbols in Figure 5.

3.2 Fitting a PCLT model.

The statistical modelling of these data proceeds on the assumption that the water content of the soil depends largely on elevation. The land undulates gently and the local drainage has no dominant orientation; I assumed that local topographic maxima can be treated as a Poisson process in two dimensions, with intervening minima corresponding to the boundaries of the Voronoi cells around the maxima. I assumed also that soil water content increases with distance from the topographic maximum. This is a simple and speculative model, although it is also a reasonable one. I explore its implications for the spatial variability of water content and see whether the distribution of a PCLT variable arising from such a model can plausibly describe the distribution of these data. A simple choice of distance function for the PCLT process is that it is a polynomial function of distance. The data are weakly skewed, and in fact the skewness is close to the value found in the previous section for the simple distance function $\mathcal{D}(k) = k$. The distribution function for the standardized PCLT variable with this distance function strongly resembles the empirical distribution function of the standardized water data (Figure 5), so I decided to use this PCLT variable for further analysis.

I estimated the empirical variogram of water content from the data using the standard estimator due to Matheron (1962) as implemented in the VARIOG procedure of geoR (Ribeiro and Diggle, 2001), specifying 30 lag bins of width 30 m. I then fitted a model

$$\gamma(h) \;=\; c_0 \;+\; c_1\gamma_{\mathrm{PCLT}}(h|\lambda), \tag{7}$$

in which c_0 is the spatially independent nugget variance, c_1 is the variance of the spatially correlated PCLT process and $\gamma_{\mathrm{PCLT}}(h|\lambda)$ is the variogram for the PCLT process with λ the intensity of the underlying Poisson process, and the sill variance standardized to 1. This standard variogram could be obtained for any specified value of λ from Equation (5) with the required covariances computed from Equation (20) for a process in two-dimensions as done to compute the theoretical examples previously. The estimates of the three parameters, c_0, c_1 and λ were obtained by weighted least

squares with weights as recommended by Cressie (1985). I estimated λ by computing its weighted least squares profile, finding estimates of c_0 and c_1 that minimized the weighted sum of squares given some specified valued of λ, and repeating this for a range of values of λ. A profile plot of the minimized weighted sum of squares against λ is shown in Figure 6. The minimum is at an intensity of 6.59 points per km^2. The mean chord length of the corresponding Voronoi tessellation (Heinrich, 1998) computed from Equation (15) of Lark (2009) is 305.8 m. The variogram model is shown in Figure 7 along with the empirical variogram to which it was fitted.

3.4 Multiple point statistics of the PCLT model.

We have seen that the PCLT process gives rise to a marginal distribution for a variable, in this case water content, that is non-normal. This means that the multivariate distribution of the variable at a set of locations in space cannot be normal. This may have implications for the application of geostatistical methods, such as conditional simulation, which invoke an assumption of multivariate normality.

If a variable is multivariate normal at locations in space, then its joint distribution at these locations is fully accounted for by all the pairwise covariances. Many variables in the geosciences do not seem to have this property (e.g. Strebelle, 2002), which is why multiple point geostatistics has been developed. As an example in this paper I consider three locations with coordinates $\mathbf{x}_0$, $\mathbf{x}_{-1} = \mathbf{x}_0 + \{-150, 0\}$ and $\mathbf{x}_1 = \mathbf{x}_0 + \{150, 0\}$, so they lie on a straight line, 300m long. The conditional distribution of some variable $Z(\mathbf{x}_0)$ given the values at $\mathbf{x}_{-1}$ and $\mathbf{x}_1$, $F\{Z(\mathbf{x}_0)|Z(\mathbf{x}_{-1}), Z(\mathbf{x}_1)\}$ depends on the joint distribution at the three locations which can be inferred from the pairwise covariances between the locations if $Z(\mathbf{x})$ is multivariate normal. If the multivariate distribution is not normal, then it cannot be guaranteed that the joint distribution is wholly characterised by the covariance, or two-point statistics; hence the term 'multiple point statistics'. In the methods developed for multiple point statistics conditional distributions such as $F\{Z(\mathbf{x}_0)|Z(\mathbf{x}_{-1}), Z(\mathbf{x}_1)\}$ are inferred from large data sets, called

training images. All instances where the conditioning observations $Z(\mathbf{x}_{-1}), Z(\mathbf{x}_1)$ meet some criteria are found, and the empirical distribution of $Z(\mathbf{x}_0)$ over these instances is the estimate of $F\{Z(\mathbf{x}_0)|Z(\mathbf{x}_{-1}), Z(\mathbf{x}_1)\}$.

I used simulation to estimate the conditional distribution of the PCLT process fitted to the gravimetric water data at $\mathbf{x}_0$ conditional on the gravimetric water content's being below the first empirical quartile (36.8%) at $\mathbf{x}_{-1}$ and above the third empirical quartile (46.6%) at $\mathbf{x}_1$. In this simulation the variation represented by the nugget variance in the fitted variogram was ignored. To simulate the desired distribution I used the same procedure by which I generated realizations of PCLT processes to generate the empirical variograms shown in Figure 2. The intensity of the Poisson process in the fitted PCLT model for soil water content was 6.59 events km^2. The process was simulated within a square region with linear dimension 3 017.4 m, so the expected number of events of the underlying point process within the region was 60. As previously, the number of events in a particular realization was obtained as a Poisson random variable with mean $\mu = 60$. The value of k, the distance to the nearest seed point was then evaluated at locations $\mathbf{x}_0 = 0, 0$ and $\mathbf{x}_{-1} = \mathbf{x}_0 + \{-150, 0\}$ and $\mathbf{x}_1 = \mathbf{x}_0 + \{150, 0\}$. The value of the PCLT random function $Z(\mathbf{s}) = \mathcal{D}(k) = k$ was then evaluated at each of these three locations. The values were standardized to zero mean and unit variance, using the known parameters for the pdf of $Z(\mathbf{s})$, from Equation (15). These standardized values were then rescaled to values of soil water content by multiplying by $c_1^{0.5}$, where c_1 is the spatially dependent variance component in the PCLT variogram model fitted to the empirical variogram of water content, Equation (7). The sample mean value was then added. If the values at locations $\mathbf{x}_{-1}$ and $\mathbf{x}_1$ matched the conditions, then the value at $\mathbf{x}_0$ was retained as a sample from the desired conditional distribution. Otherwise the realization was discarded and another one drawn. This was repeated until 100 000 samples had been drawn from the conditional distribution.

A similar approach was used to sample the same conditional distribution under the assumption that it is multivariate normal. In this case a single realization at $\mathbf{x}_0 = \{0, 0\}$ and $\mathbf{x}_{-1} = \mathbf{x}_0 + \{-150, 0\}$ and $\mathbf{x}_1 = \mathbf{x}_0 + \{150, 0\}$ was drawn by the LU decomposition method (see Webster and Oliver, 2007) implemented in the RNMVN algorithm in the IMSL library (Visual Numerics, 2006). This requires a covariance matrix for the three locations. The covariances were computed from the fitted variogram function. As in the PCLT simulation, the marginal variance of the variable was equal to the variance of the spatially-dependent component, c_1 in Equation(7), and the nugget variance was ignored. Once again, a single realization was drawn, and the value at $\mathbf{x}_0$ was retained as a sample of the target conditional distribution if and only if the specified conditions were met at the other two locations. A total of 100 000 values from the conditional distribution was simulated this way.

The KERNELDENSITY procedure in GenStat (Goedhart, 2009) was used to compute empirical density functions for the two conditional distributions, and these are plotted in Figure 8. The PCLT and multivariate normal distributions had neglibly different means (41.2% and 41.9% respectively) but the former had a much smaller variance than the latter (14.6%^{2} and 42.3%^{2} respectively). Some 75% of the values of the conditional distribution for the PCLT model fell between the two empirical quartiles of the water content data used for conditioning, in comparison to 55% of the values for the multivariate normal model. These results are consistent with the two underlying models. Consider first the PCLT process. The three points are on a line close in length to the mean chord length of the underlying Voronoi tessellation. This means that, when the conditions on the distribution hold, it is likely that the locations $\mathbf{x}_{-1}$ and $\mathbf{x}_1$ correspond, respectively, to the top and bottom of a catena in our landscape with dry conditions at the top and wet at the bottom. In these circumstances we should expect the value at $\mathbf{x}_0$ to be close to the mean, and so it is not surprising that the values are tightly distributed about the mean with 76% lying between

quartile 1 and quartile 3 of the empirical distribution. In contrast, under the assumption of multivariate normality, the distribution of values at location $\mathbf{x}_o$ is much wider and only slightly more values fall between the first and third quartiles than we would expect for the marginal distribution of the random variable (50%). This shows that the two-point statistics of the process, and hence the normal model, fail to capture all features of the spatial variation of the PCLT process. This is expected, since the PCLT variable is not normally distributed. This particular example is of interest, however, because it shows how the PCLT joint distribution at three locations retains features that we would expect from a pattern of variation which comprises local trends which the commonly-assumed multivariate normal model cannot reproduce.

4. Discussion

It has been shown how the Poisson Continuous Local Trend (PCLT) model can be developed mathematically and fitted to soil data where it is plausible that the dominant sources of soil variation may be represented by the CLT mode. I have shown how the form of the variogram for a PCLT process depends on the form of the underlying distance function, and the intensity of the underlying seed process. Also, given a topo-function (such as that of Walker *et al.,* 1968), and perhaps some statistical information on its fitting, one can propose the corresponding form of the variogram, assuming that local summits can be treated as Poisson point processes. Finally, we can see that the PCLT model has multiple point statistics that cannot be accounted for from the two-point statistics, and so it is different to a multivariate normal process.

This paper has introduced the PCLT model and its properties, and shown how it can be fitted to data. The fact that such a model fits does not show that it is necessarily best, however. This is an area for further work. In general this cannot be tested by examining directly the underlying point process, or the distance function, $\mathcal{D}(k)$, since these are, usually, latent. In circumstances where they could be examined

directly then this information would normally be incorporated into a mixed model for a soil variable through the fixed effects rather than to parameterize a model for the random effects. Instead, the multiple point statistics, specifically the conditional distributions involving three or more locations, would provide a basis to compare the PCLT with other spatial models for data and to asses its practical advantages. These would require many more data than I had, and might be done using intensive sensor data, such as data from geophysical surveys of soil water content.

The PCLT is of potential practical interest for two reasons. First, in circumstances where we may expect a CLT mode of soil variation, it is a model that could be proposed *a priori*. Its likely form might also be proposed, given pedological knowledge about the likely form of a distance function such as a topofunction, and a plausible range of values for the mean chord length (i.e. the mean interval between boundaries separating the notional topographic cells, which might be obtained from a locally experienced field scientist. From such information one might generate a range of possible forms of the variogram. These might be fitted with a standard model (such as the Matérn function), to provide a range of values, and so prior distributions, for its parameters, which could then be used to optimize sampling, perhaps as proposed by Marchant and Lark (2006). The PCLT model could also be used to simulate data to plan design-based sampling to estimate global means (de Gruijter et al, 2006).

Second, the results in this paper suggest that standard geostatistical methods, which assume an underlying multivariate distribution, might not be well suited to soil properties that vary continuously along short-range lateral trends, reflecting a CLT mode of variation, since these may not be entirely characterized by two-point statistics. One constraint on multiple point methods, such as the algorithm of Strebelle (2002), is the availability of training data to estimate the required conditional distributions. If the PCLT model were shown to be generally plausible, then it could be used to generate unlimited training data by the simulation methods used in this paper.

Further research is required. I have already noted the need for some systematic studies with large data sets to compare the PCLT model with other plausible ones, particularly with respect to multiple point statistics. In addition, it would be useful to develop CLT models for underlying point processes other than the simple homogeneous Poisson process in which the intensity, λ, is spatially uniform. For example, in conditions where topographic variation is rather more regular than a purely random subdivision of space an underlying point process is under-dispersed relative to a homogeneous point process. Non-homogeneous point processes are available for such circumstances, as well as for circumstances where the point process is more clustered than a homogeneous Poisson process. In some cases variation on continuous local trends may be strongly anisotropic, because the drainage lines in the landscape are aligned (Figure 1b). In these conditions a one-dimensional CLT model might be fitted perpendicular to the drainage lines; and a non-homogeneous process for the topographic maxima would probably be most appropriate.

Finally, it was observed that variation at the scale of the classical catena is ideally treated not as random effects but in terms of fixed effects represented by covariates derived from digital elevation models and other sources. It would be particularly valuable, therefore, to investigate the plausibility of the PCLT model at finer spatial scales, representing microtopography, and even variation at plot to subcore scale.

5. Conclusions

The properties of the PCLT model of soil variation, a plausible model of variation that shows a strong catenary mode, have been explored. The PCLT model shows, as would be expected, effects of local drift (upward concavity of the variogram at short lags), and the shape depends on the proposed topofunction. Given a topofunction for a soil property, and assuming that local summits can be represented as Poisson point processes, it is possible to compute a proposed variogram for that property *a*

priori. The PCLT variogram model can also be fitted to the empirical variogram of soil data. The PCLT has multiple point statistics that are not reducible to its two-point statistics, so for purposes such as simulation a multivariate normal assumption would not be appropriate for such a variable. There is a need for further work to validate the PCLT model on soil properties at a range of scales.

Acknowledgements

I am grateful to Professor Richard Webster for very helpful comments on an earlier draft of this paper, and to Professor Peter Diggle for helpful discussions about the joint survival function in Equation (17). Rothamsted Research granted permission to use the data on soil water content, which I collected while in their employment. This paper is published with the permission of the Director of the British Geological Survey (NERC).

Appendix. Derivation of the covariance function for a PCLT process

Höffding (1940) showed that two random variables X and Y have covariance

$$\mathrm{Cov}\left\{X, Y\right\} \;=\; \int_{\mathbb{R}^2} \left\{H(x,y) - F(x)G(y)\right\} \mathrm{d}x\mathrm{d}y, \tag{8}$$

where $H(x,y)$, $F(x)$ and $G(y)$ are respectively the joint cumulative distribution function (cdf) of X and Y and the cdfs of X and of Y. Cuadras (2002) generalized this to

$$\mathrm{Cov}\left\{\alpha(X), \alpha(Y)\right\} \;=\; \int_{\mathbb{R}^2} \left\{H(x,y) - F(x)G(y)\right\} \mathrm{d}\alpha(x)\mathrm{d}\beta(y), \tag{9}$$

where the functions $\alpha(\cdot)$ and $\beta(\cdot)$ are defined on intervals of the real numbers and, within these intervals, both functions have bounded variation, and the expectations:

$$\mathrm{E}\left[||\alpha(X)\beta(Y)||\right],$$
$$\mathrm{E}\left[||\alpha(X)||\right],$$
$$\mathrm{E}\left[||\beta(Y)||\right],$$

are all finite.

For the moment we focus on computing the covariances of the distances, that is to say, $\mathcal{D}\left(\cdot\right)$ is the identity function. In this case

$$C_I(r) \;=\; \mathrm{Cov}\left\{K(\mathbf{s}), K(\mathbf{s}+r\mathbf{h})\right\}. \tag{10}$$

$K(\mathbf{s})$ and of $K(\mathbf{s}+r\mathbf{h})$ have the same cdf under stationarity, which is denoted $F(k)$. This is derived as follows. The probability that $K(\mathbf{s})$, the distance from an arbitrary location $\mathbf{s}$ to its nearest neighbour in ψ, is greater than some distance k is equal to the probability that no member of ψ lies within a ball of radius k centred at $\mathbf{s}$:

$$\text{Prob}\left\{K(\mathbf{s}) > k\right\} \;=\; \text{Prob}\left\{B\left[\mathbf{s},k\right] \cap \psi \;= \varnothing\right\}. \tag{11}$$

For a Poisson ψ

$$\text{Prob}\left\{B\left[\mathbf{s},k\right] \cap \psi \;= \varnothing\right\} \;=\; \exp\left\{-\lambda |B\left[\mathbf{s},k\right]|\right\}, \tag{12}$$

from the definition of the Poisson distribution. Under stationarity this is independent of $\mathbf{s}$, so we define

$$|B_o\left[k\right]| \equiv |B\left[\mathbf{0},k\right]|,$$

where $\mathbf{0}$ is the origin of $\mathbb{R}^d$, and can then write the cdf of $K(\mathbf{s})$

$$\begin{aligned} F(k) \;&=\; \text{Prob}\left\{K(\mathbf{s}) \leq k\right\} \\ &=\; 1 - \exp\left\{-\lambda |B_o\left[k\right]|\right\}. \end{aligned} \tag{13}$$

For $d = 2$ this is

$$F(k) \;=\; 1 - \exp\left\{-\lambda\pi k^2\right\}. \tag{14}$$

We can obtain the pdf of k, $f(k)$ by

$$f(k) \;=\; \frac{\mathrm{d}}{\mathrm{d}k}F(k) \;=\; 2\lambda\pi k \exp\left\{-\lambda\pi k^2\right\}. \tag{15}$$

Under stationarity assumptions we can simplify the notation for the random variables, writing K and K_r respectively in place of $K(\mathbf{s})$ and $K(\mathbf{s}+r\mathbf{h})$. Next, we require an expression for the joint cdf of K and K_r , $H(k,k_r)$. First, we define the joint survival function

$$S(k,k_r) \;=\; \text{Prob}\left\{K > k, K_r > k_r\right\}. \tag{16}$$

For a Poisson and stationary Ψ this is given by

$$\begin{aligned} S(k,k_r) &= \text{Prob}\left\{\mathcal{U}\left(B\left[\mathbf{0},k\right],B\left[\mathbf{0}+r\mathbf{h},k_r\right]\right)\cap\psi\ =\varnothing\right\} \\ &= \exp\left\{-\lambda|\mathcal{U}\left(B\left[\mathbf{0},k\right],B\left[\mathbf{0}+r\mathbf{h},k_r\right]\right)|\right\}. \end{aligned} \tag{17}$$

Figure 9 represents the variate $[K,K_r]^{\mathrm{T}}$. Each variable has a lower bound at zero. The values k and k_r define two overlapping regions for which $K>k$ and $K_r>k_r$ respectively. The integral of the joint density of $[K,K_r]^{\mathrm{T}}$ over these unbounded regions is $1-F(k)$ and $1-F(k_r)$ respectively. The integral of the joint density over the overlap between these two regions is the joint survival function $S(k,k_r)$. The integral over the bounded region $[0\leq K\leq k, 0\leq K_r\leq k_r]$ is the function that we require, the joint cdf $H(k,k_r)$, so it can be seen that

$$\begin{aligned} H(k,k_r) &= 1-\left[(1-F(k))+(1-F(k_r))-S(k,k_r)\right] \\ &= S(k,k_r)+F(k)+F(k_r)-1. \end{aligned} \tag{18}$$

From Equation (8), and substituting Equation (18) for $H(k,k_r)$,

$$\begin{aligned} C_I(r) &= \int_{\mathbb{R}^2}\left\{H(k,k_r)-F(k)G(k_r)\right\}\mathrm{d}k\ \mathrm{d}k_r, \\ &= \int_{\mathbb{R}^2}\left\{S(k,k_r)+F(k)+F(k_r)-F(k)F(k_r)-1\right\}\mathrm{d}k\ \mathrm{d}k_r. \end{aligned} \tag{19}$$

For the random function with $\mathcal{D}(\cdot)$ not the identity function, it can be seen from Equation (9) that

$$\begin{aligned} C(r) &= \int_{\mathbb{R}^2}\left\{S(k,k_r)+F(k)+F(k_r)-F(k)F(k_r)-1\right\}\mathrm{d}\mathcal{D}(k)\ \mathrm{d}\mathcal{D}(k_r) \\ &= \int_{\mathbb{R}^2}\left\{S(k,k_r)+F(k)+F(k_r)-F(k)F(k_r)-1\right\}\mathcal{D}'(k)\mathrm{d}k\ \mathcal{D}'(k_r)\mathrm{d}k_r, \end{aligned} \tag{20}$$

where $\mathcal{D}(k)$ is the distance function in Equation (3) and $\mathcal{D}'(k)$ is its first derivative with respect to k. Thus, for example, if

$$\mathcal{D}(k)\ =\ \alpha\ +\beta k^2,$$

then

$$C(r) = \int_{\mathbb{R}^2}\left\{S(k,k_r)+F(k)+F(k_r)-F(k)F(k_r)-1\right\}2\beta\,\mathrm{d}k\ 2\beta\,\mathrm{d}k_r. \tag{21}$$

References

Allègre, C.J., Lewin, E. 1995. Scaling laws and geochemical distributions. Earth and Planetary Science Letters, 132, 1–13.

Barrett, J.E., Virginia, R.A., Wall, D.H., Parsons, A.N., Powers, L.E., Burkins, M.B. 2004. Variation in biogeochemistry and soil biodiversity across spatial scales in a polar desert ecosystem. Ecology, 85, 3105–3118.

Chilès, J.-P., Delfiner, P. 1999. Geostatistics, Modeling Spatial Uncertainty. John Wiley & Sons, New York.

Cuadras, C. M. 2002. On the covariance between functions. Journal of Multivariate Analysis, 81, 19–27.

Cressie, N. 1985. Fitting variogram models by weighted least squares. Mathematical Geology, 17, 563–586.

de Gruijter, J., Brus, D., Bierkens, M.F.P., Knotters, M. 2006. Sampling for Natural Resource Monitoring. Springer, Heidelberg.

Goedhart, P.W. 2009. KERNELDENSITY procedure in (ed.) R.W. Payne, GenStat Release 12 Reference Manual, Part 3 Procedure Library PL20. VSN International, Hemel Hempstead.

Heinrich, L. 1998. Contact and chord length distribution of a stationary Voronoi tesselation. Advances in Applied Probability, 30, 603–618.

Höffding, W. 1940. Masstabinvariante Korrelations-theorie. Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin, 5, 179–233.

IUSS Working Group WRB. 2006. World reference base for soil resources. 2nd edition. World Soil Resources Reports No. 103. FAO, Rome.

King, D.W. 1969. Soils of the Luton and Bedford District. Special Survey No 1. Soil Survey of England and Wales. Lawes Agricultural Trust, Harpenden.

Kleiss, H.J. 1970. Hillslope sedimentation and soil formation in northeastern Iowa. Soil Science Society of America Proceedings, 34, 287–290.

Kolvos, A., Christakos, G., Hristopulos, D.T., Serre, M.L. 2004. Methods for generating non-separable spatiotemporal covariance models with potential environmental applications. Advances in Water Resources, 27, 815–830.

Lark, R.M. 2002. Optimized spatial sampling of soil for estimation of the variogram by maximum likelihood. Geoderma, 105, 49–80.

Lark, R.M. 2009. A stochastic-geometric model of soil variation. European Journal of Soil Science, 60, 706–719.

Lark, R.M. 2010. Two contrasting spatial processes with a common variogram: inference about spatial models from higher-order statistics. European Journal of Soil Science, 61, 479–492.

Lark, R.M., Cullis, B.R., Welham, S.J. 2006. On spatial prediction of soil properties in the presence of a spatial trend: the empirical best linear unbiased predictor (E-BLUP) with REML. European Journal of Soil Science, 57, 787–799.

Matheron,G. 1962. Traité de Géostatistique Appliqué, Tome 1. Memoires du Bureau de Recherches Géologiques et Minières, Paris.

Marchant, B.P., Lark, R.M. 2006. Adaptive sampling for reconnaissance surveys for geostatistical mapping of the soil European Journal of Soil Science 57, 831–845

Milne. A.E., Haskard, K.A., Webster, C.P., Truan, I.A. Goulding, K.W.T, Lark, R.M. 2011. Wavelet analysis of the correlations between soil properties and potential

nitrous oxide emission at farm and landscape scales. European Journal of Soil Science 62, 467–478.

Milne, G. 1936. Normal erosion as a factor in soil profile development. Nature, 138, 548–549.

Pérez, F.L. 1995. Plant-induced spatial patterns of surface soil properties near caulescent Andean rosettes. Geoderma, 68, 101–121.

R Development Core Team. 2010. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. `http://www.R-project.org/`.

Ribeiro, P.J. & Diggle, P.J. 2001. geoR: a package for geostatistical analysis. R-NEWS, 1, 15–18.

Rosenbloom, N.A., Doney, S.C., Schimel, D.S. 2001. Geomorphic evolution of soil texture and organic matter in eroding landscapes. Global Biogeochemical Cycles, 15, 365–381.

Siersma, D. 1999. Voronoi diagrams and Morse theory of the distance function. In: *Geometry in Present Day Science* (eds O.E. Barndorff-Nielsen, E.B.V. Jensen), pp. 187–208. World Scientific, Singapore.

Scott, R.M., Webster, R., Lawrance, C.J. 1971. A land system atlas of Western Kenya. Military Vehicles and Engineering Establishment, Christchurch, Hampshire.

Strebelle, S. 2002. Conditional simulation of complex geological structures using multiple-point statistics. Mathematical Geology, 34, 1–21.

Visual Numerics, 2006. IMSL Fortran Numerical Library Version 6.0. Visual Numerics, Houston, Texas.

Walker, P.H. 1966. Postglacial environments in relation to landscape and soils on the Cary drift, Iowa. Iowa State University Exp. Station Research Bulletin, 549, 838–875.

Walker, P.H., Ruhe, R.V. 1968. Hillslope models and soil formation. 2. Closed systems. Trans. 9th Int. Congress of Soil Science, Adelaide, South Australia. 4, 561–568.

Walker, P.H., Hall, G.F., Protz. R. 1968. Soil trends and variability across selected landscapes in Iowa. Soil Science Society of America Proceedings, 32, 101–104.

Webster, R. 1965. A catena of soils on the Northern Rhodesia plateau. Journal of Soil Science, 16, 31–43.

Webster, R. 2000. Is soil variation random? Geoderma, **97**, 149–163.

Webster, R. & Oliver, M.A. 2007. Geostatistics for Environmental Scientists. 2nd Edition John Wiley & Sons, Chichester.

Whittle, P. 1954. On stationary processes in the plane. Biometrika, 41, 434–449.

Whittle, P. 1962. Topographic correlations, power-law covariance functions and diffusion. Biometrika, 49, 305–314.

Yaalon, D.H. 1975. Conceptual models in pedogenesis: can soil-forming functions be solved? Geoderma, 14, 189–205.

Youssef, R.A., Chino, M. 1989. Root-induced changes in the rhizosphere of plants. Soil Science and Plant Nutrition, 35, 461–468.

Table 1. Matérn parameters fitted to PCLT variograms.

Distance function, $\mathcal{D}(k)$	ν	ϕ	Effective range
k	5.2	8.9	73.4
$k^2/10$	5.0	10.4	83.9
$10\{k+1\}^{-1}$	0.7	10.7	36.9
$10\{k+1\}^{-2}$	0.5	4.1	12.4

Table 2. Summary statistics on gravimetric water content (/%) of soil from the transect on soils under arable and grassland over the Wicken Association.

Mean	42.4
Median	42.5
Standard deviation	7.39
Skewness	0.67
Quartile 1	36.8
Quartile 3	46.6

Figure captions

1. Block diagrams for (a) the Eldama Landsystem and (b) the Lolimo Landsystem in Westen Kenya. Taken from Scott et al. (1971).

2. Computed variograms (solid lines) for PCLT processes with intensity $\lambda = 0.25 \times 10^{-3}$ events per unit square of area and distance functions indicated on the graph. The mean experimental variograms from 5000 realizations of each PCLT process on a linear transect are shown as solid symbols.

3. Computed marginal distribution functions for standardized random variables generated by a PCLT processes with different distance functions.

4. Inferred variogram for thickness of the A horizon obtained by incorporating the topofunction of Walker *et al.* (1968) into a PCLT model, and with a nugget:sill ratio inferred from the coefficient of determination reported for that function. The variance is scaled to a sill of one and lag distances are scaled relative to the mean chord length of the underlying Voronoi tesselation.

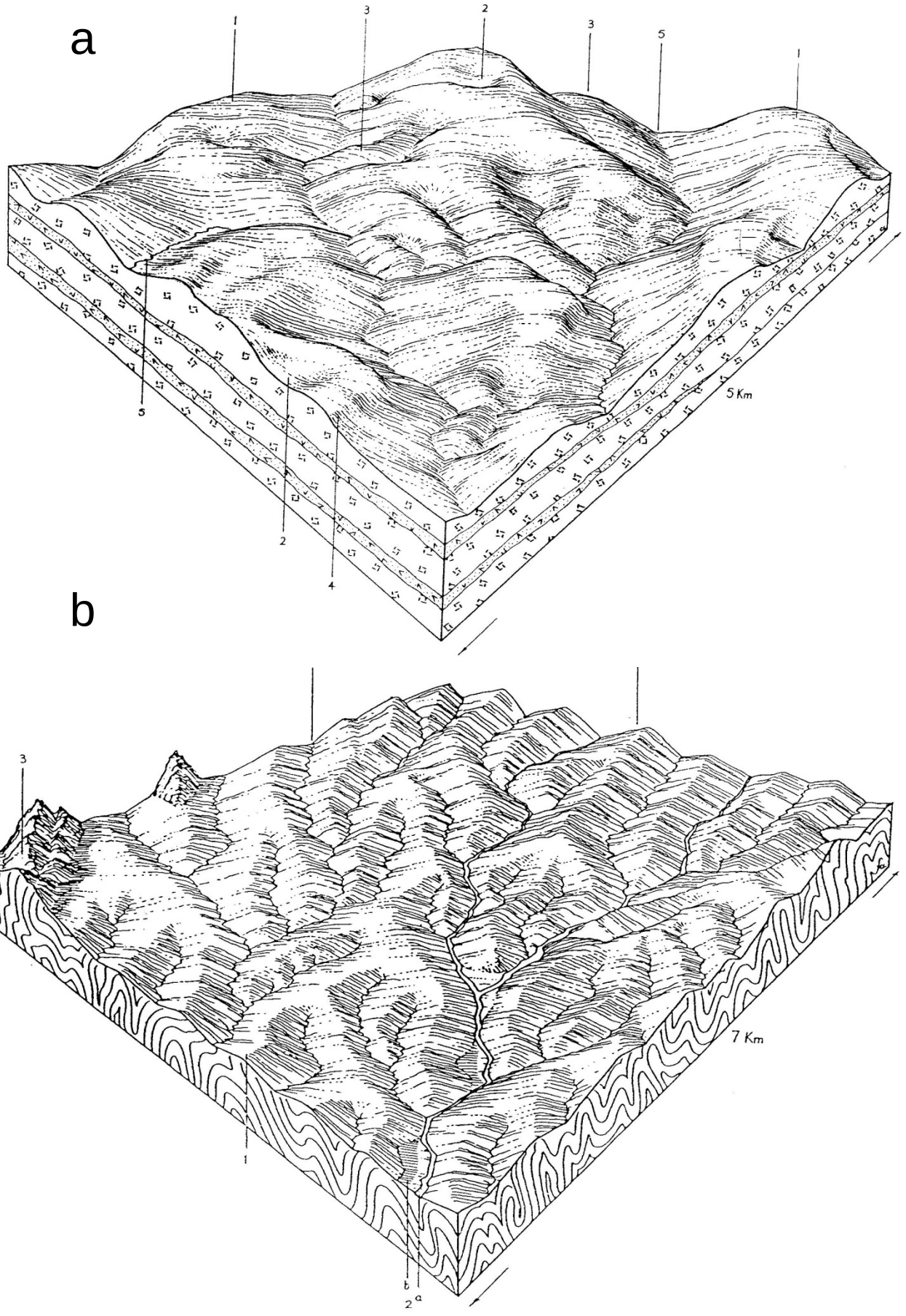
5. Empirical distribution function for gravimetric soil water content from data from the transect on soils under arable and grassland over the Wicken Association after standardization to zero mean and unit variance. The distribution function for a standardized PCLT process with distance function $\mathcal{D}(k) = k$ is superimposed.

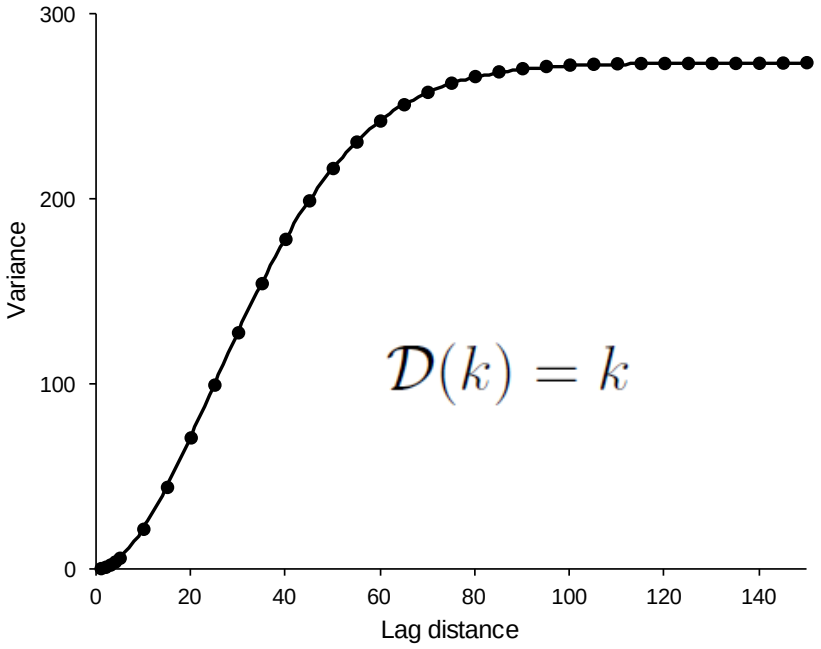
6. Profile plot of the weighted least squares for fits of a PCLT model, $\mathcal{D}(k) = k$, to the empirical variogram of gravimetric water content.

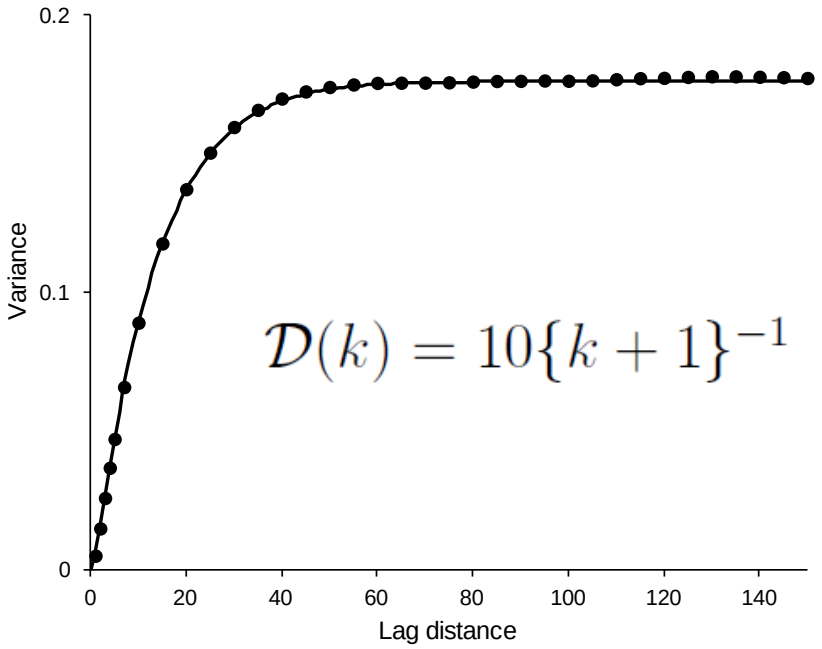
7. Empirical variogram of gravimetric water content (solid symbols) and fitted PCLT model, $\mathcal{D}(k) = k$.

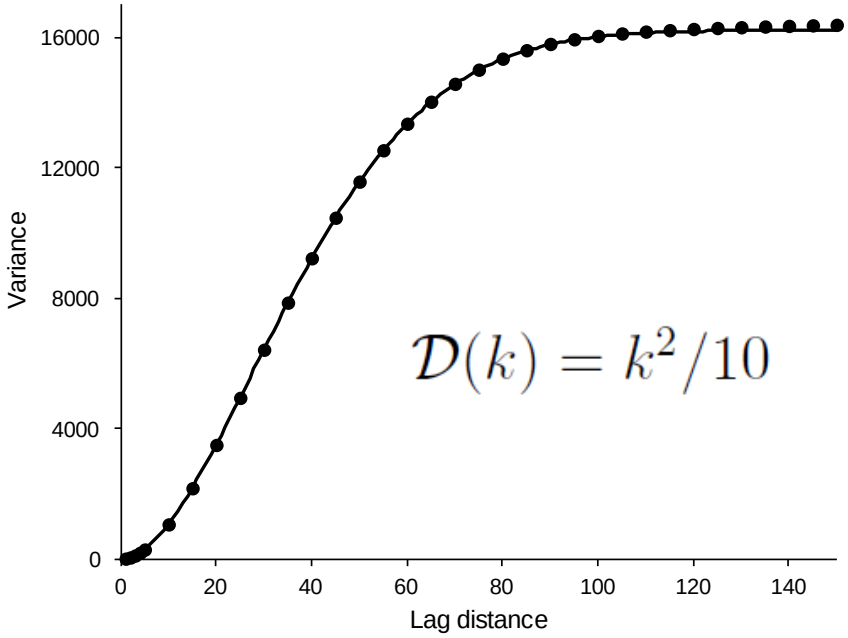
8. Distribution of gravimetric water content (nugget component excluded) at a location, $\mathbf{x}_0$ conditional on the gravimetric water content at $\mathbf{x}_0 + \{150, 0\}$ being larger than the empirical third quartile and the gravimetric water content at $\mathbf{x}_0 + \{-150, 0\}$ being smaller than the empirical first quartile. Distributions are shown for the PCLT model fitted to the Bedfordshire data (heavy line), and the equivalent distribution for a multigaussian process with the same covariance function (fine line).

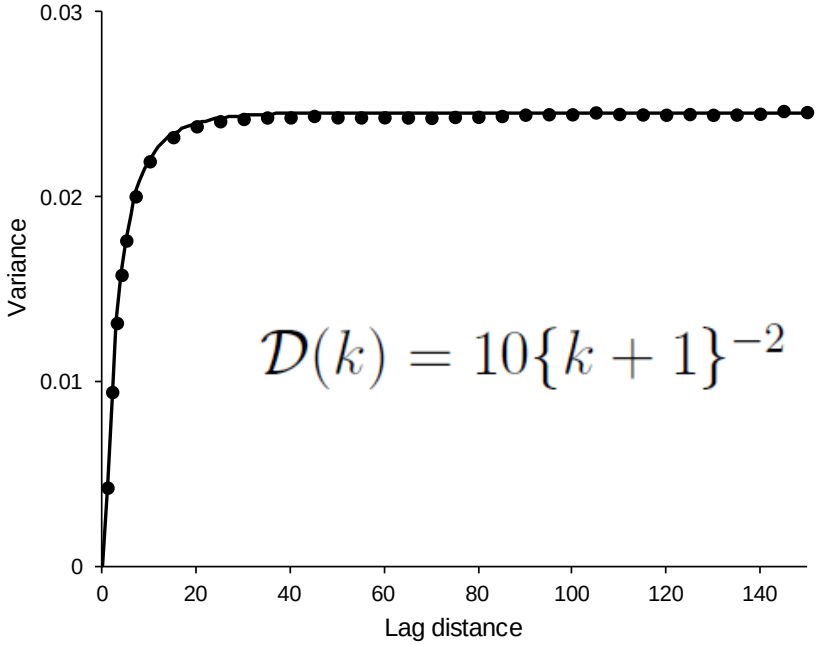
9. Space of the variate $[K, K_r]^{\mathrm{T}}$, with values of the integral of the joint density over four partly overlapping regions.

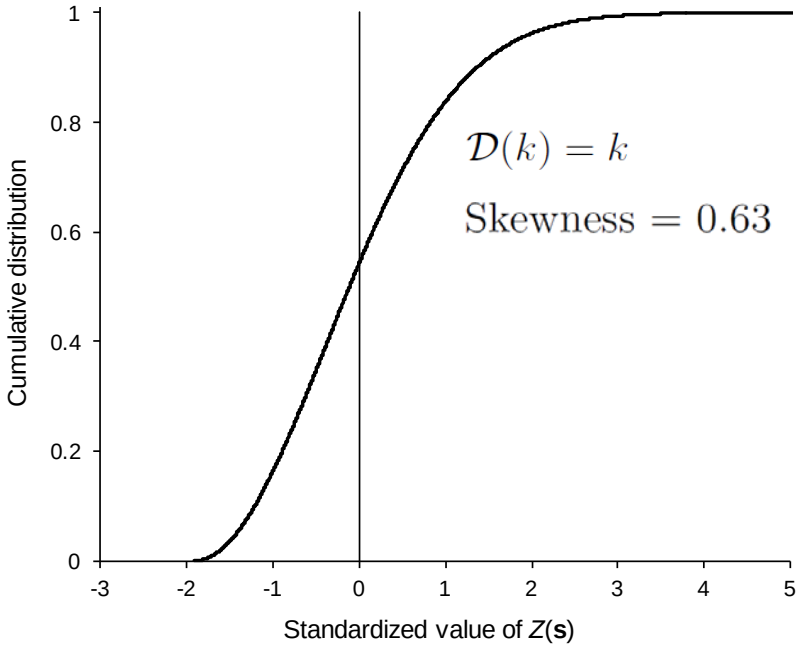

a
1
3
2
3
5
1
5
2
4
5 Km
b
3
1
b
2
a
7 Km

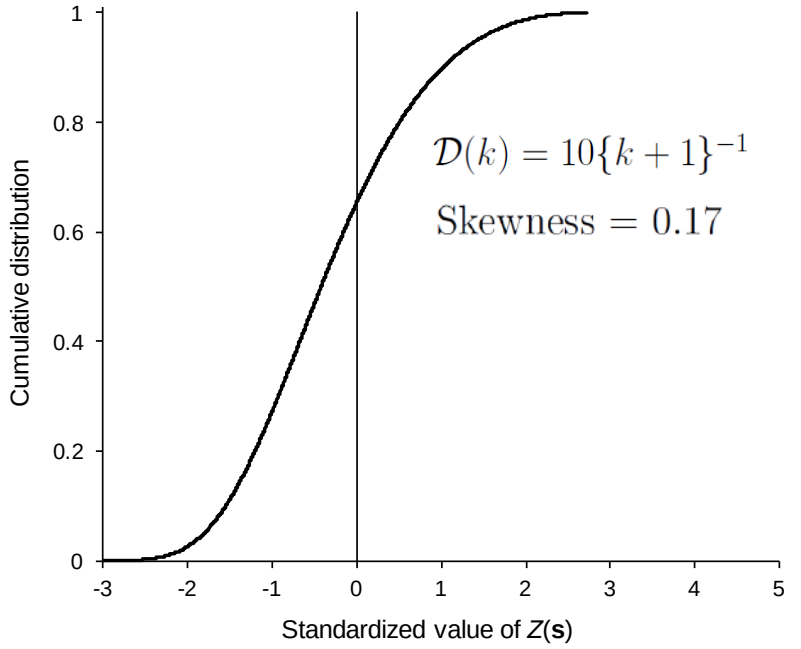

300
200
100
0
Variance
0
20
40
60
80
100
120
140
Lag distance
$\mathcal{D}(k) = k$

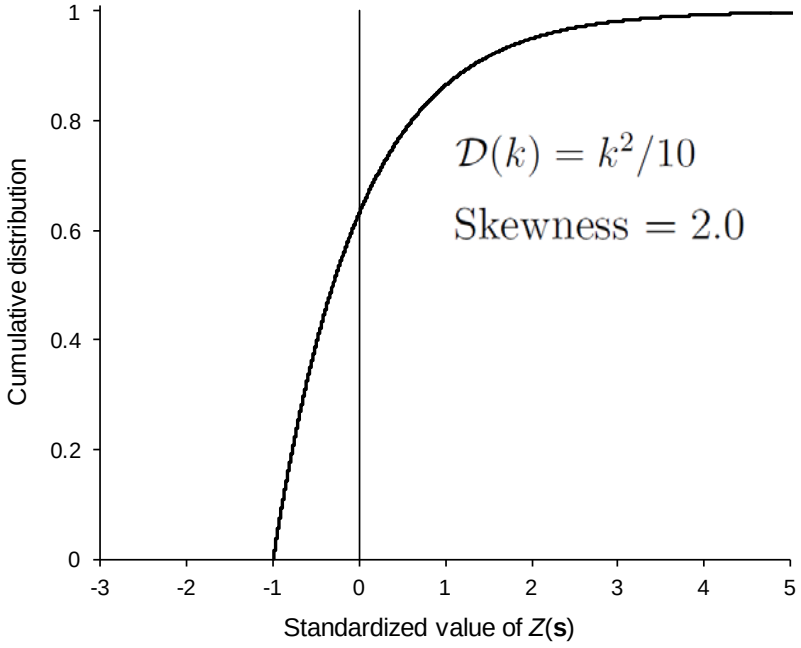

0.2
0.1
0
Variance
0
20
40
60
80
100
120
140
Lag distance
$\mathcal{D}(k) = 10\{k+1\}^{-1}$

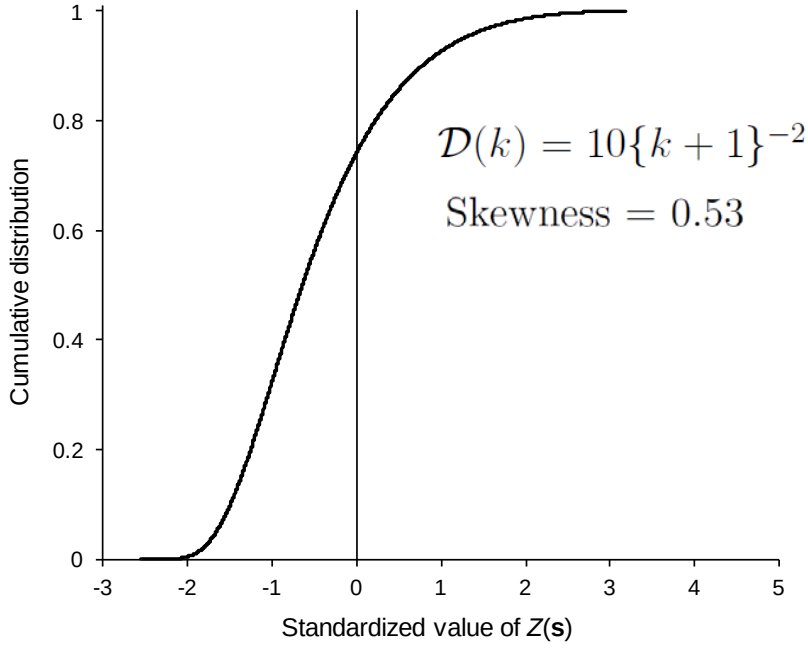

16000
12000
8000
4000
0
Variance
0
20
40
60
80
100
120
140
Lag distance
$\mathcal{D}(k) = k^2/10$

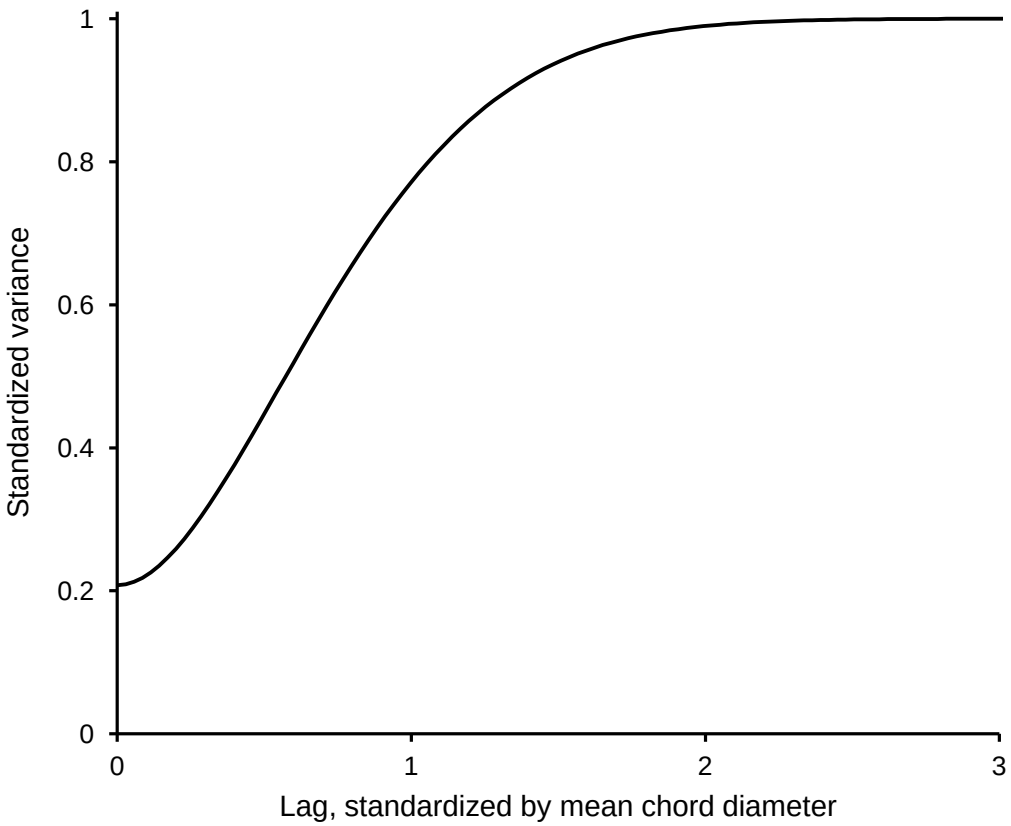

0.03
0.02
0.01
0
Variance
0
20
40
60
80
100
120
140
Lag distance
$\mathcal{D}(k) = 10\{k+1\}^{-2}$

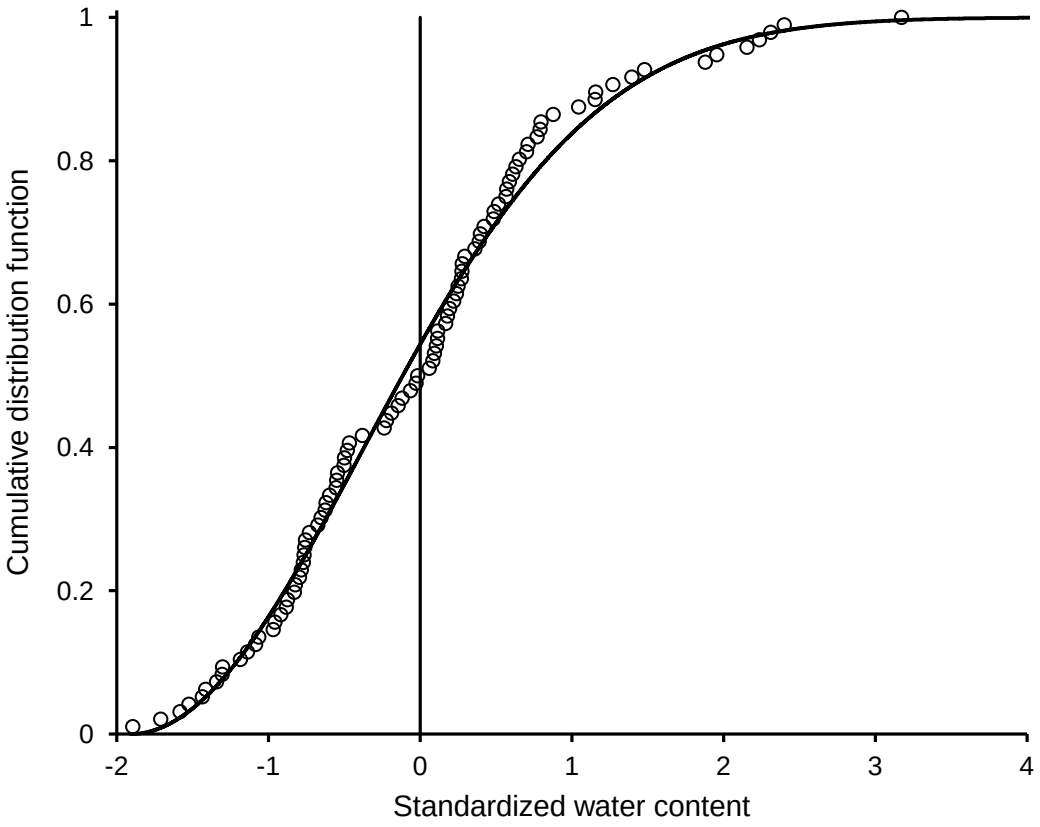

$\mathcal{D}(k) = k$
Skewness = 0.63
Cumulative distribution
Standardized value of Z(s)

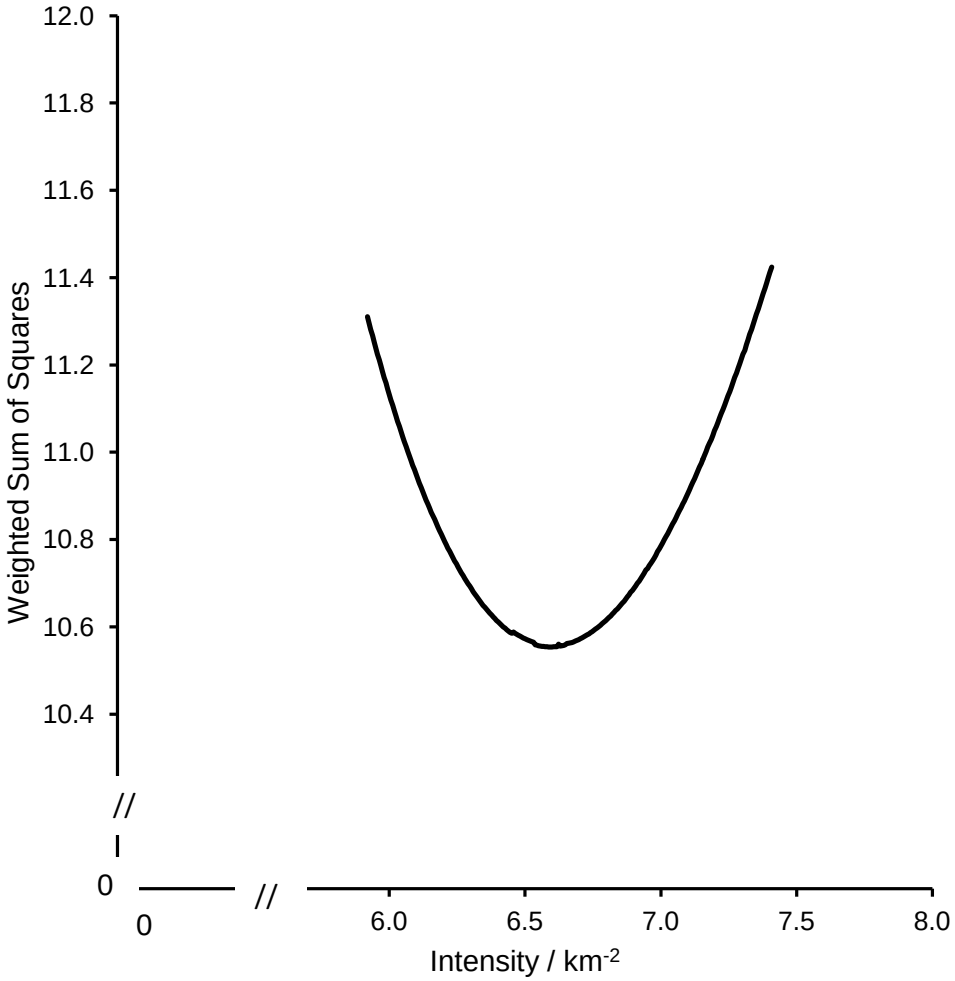

$\mathcal{D}(k) = 10\{k+1\}^{-1}$
Skewness = 0.17
Cumulative distribution
Standardized value of Z(s)

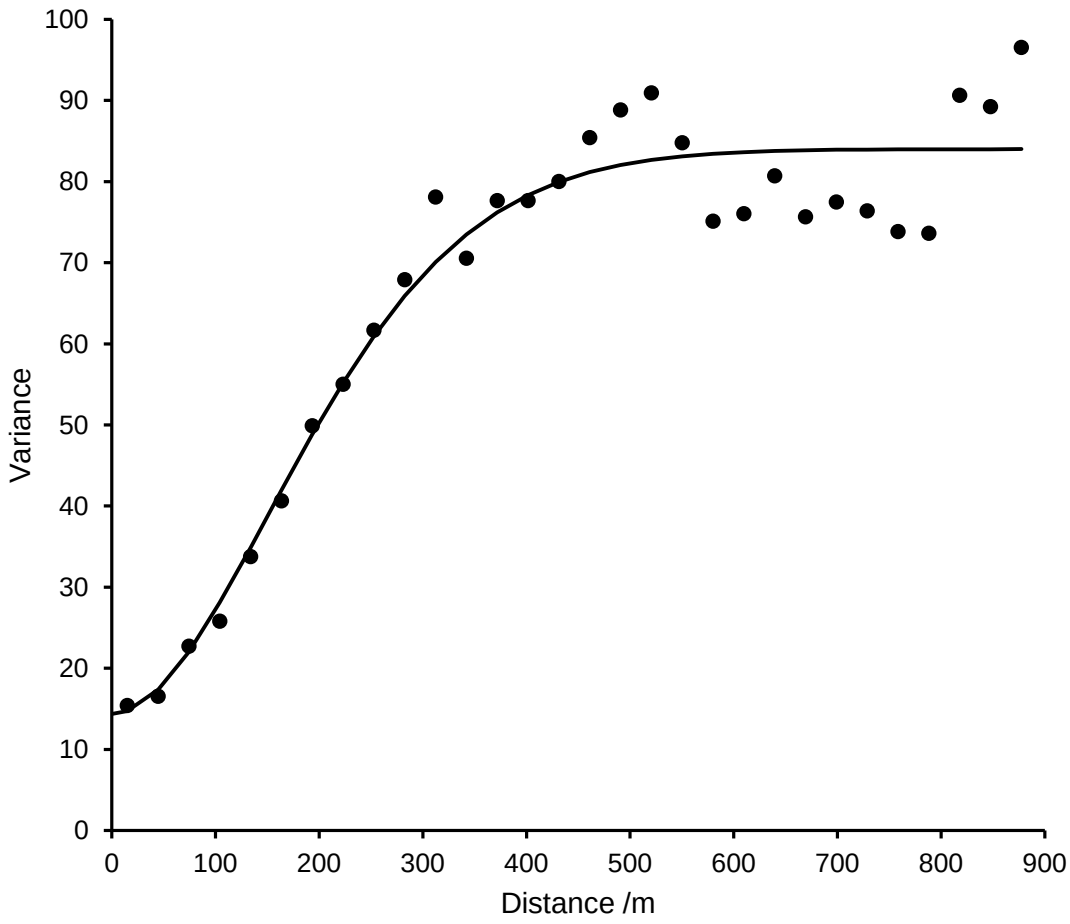

$\mathcal{D}(k) = k^2/10$
Skewness = 2.0
Cumulative distribution
Standardized value of Z(s)

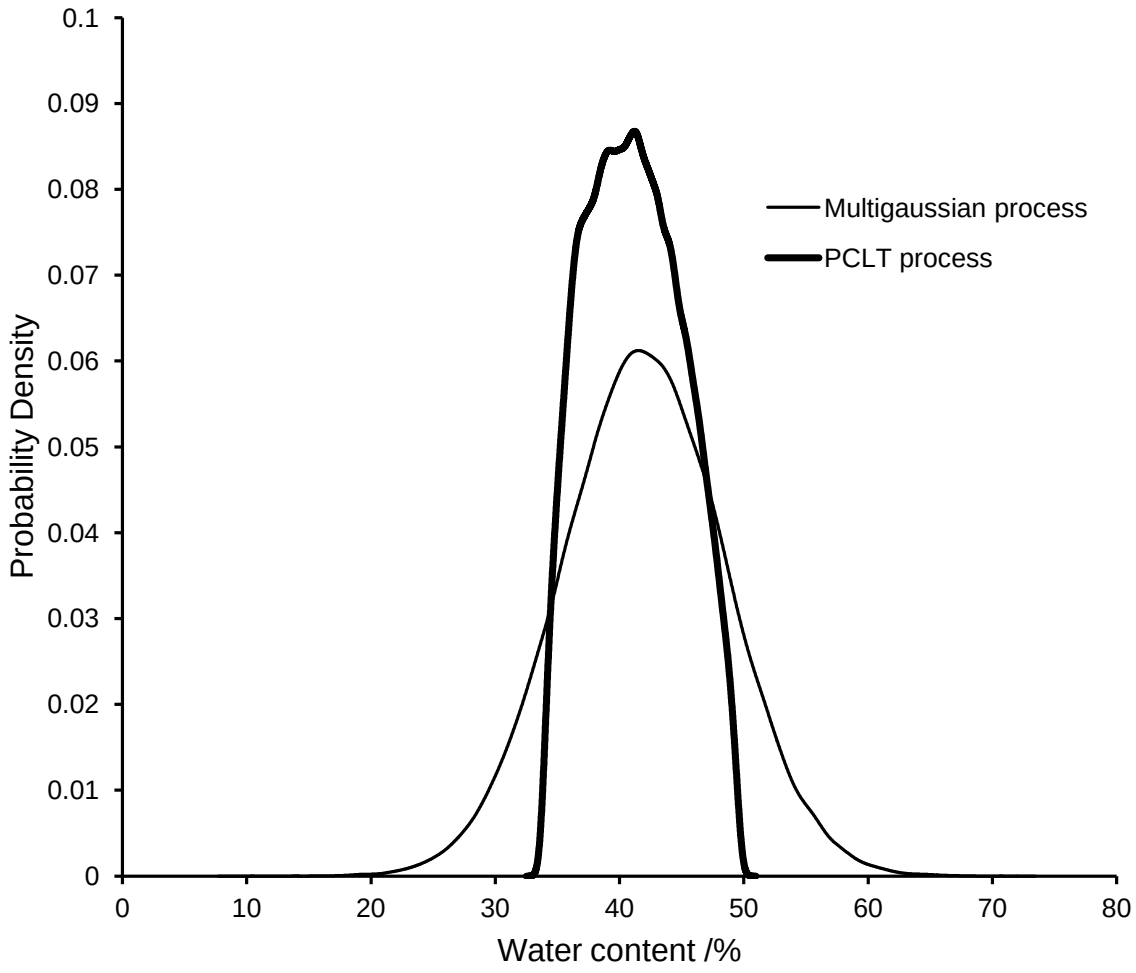

$\mathcal{D}(k) = 10\{k+1\}^{-2}$
Skewness = 0.53
Cumulative distribution
Standardized value of Z(s)

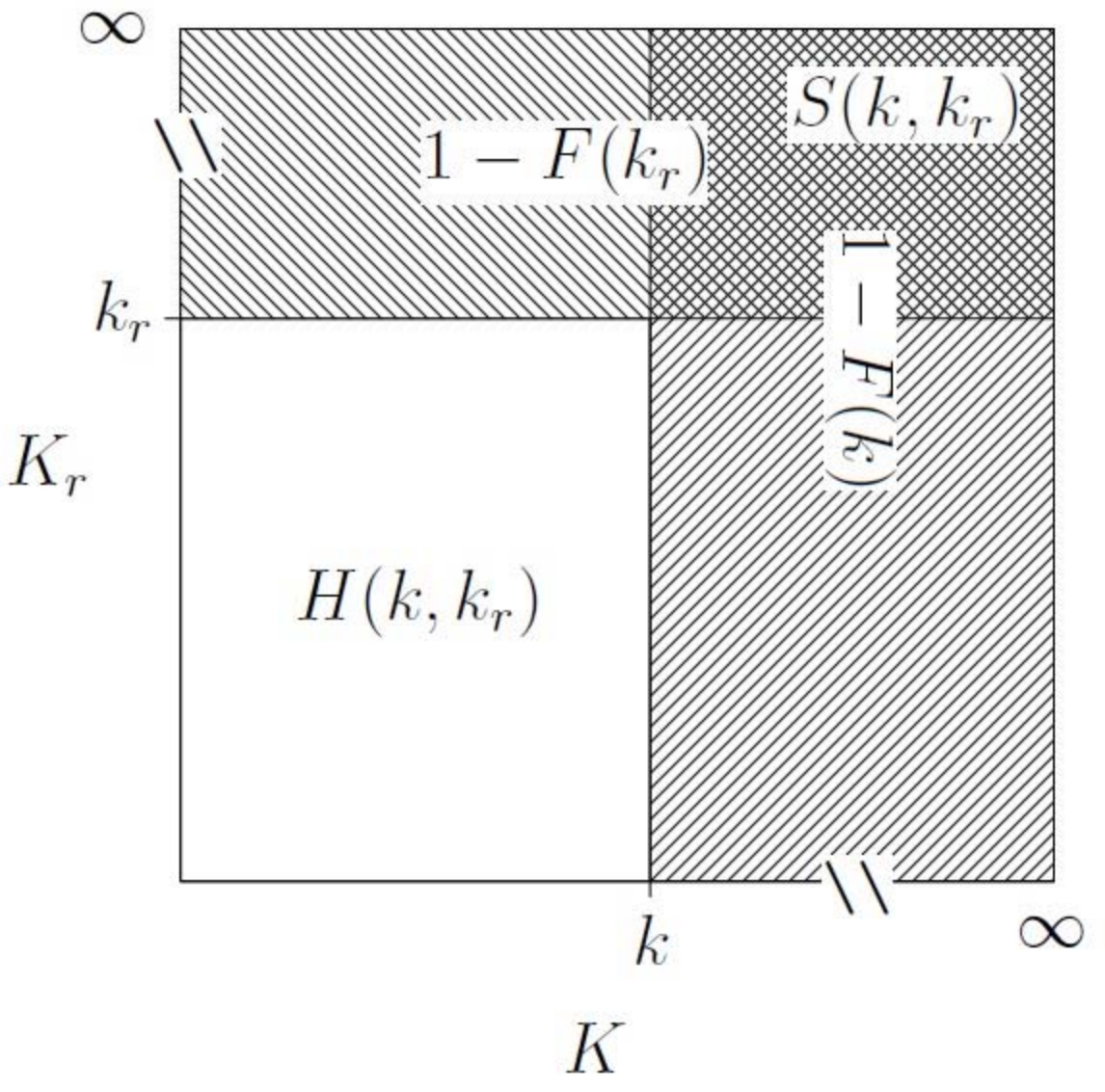

1
0.8
0.6
0.4
0.2
0
Standardized variance
0
1
2
3
Lag, standardized by mean chord diameter


Cumulative distribution function
1
0.8
0.6
0.4
0.2
0
-2
-1
0
1
2
3
4
Standardized water content


12.0
11.8
11.6
11.4
11.2
11.0
10.8
10.6
10.4
Weighted Sum of Squares
//
0
//
0
6.0
6.5
7.0
7.5
8.0
Intensity / km-2


Variance
Distance /m
0
10
20
30
40
50
60
70
80
90
100
0
100
200
300
400
500
600
700
800
900


0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
Probability Density
0
10
20
30
40
50
60
70
80
Water content /%
Multigaussian process
PCLT process


∞
$1 - F(k_r)$
$S(k, k_r)$
$1 - F(k)$
k_r
K_r
$H(k, k_r)$
k
∞
K