Geoderma 213 (2014) 533–543

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

A stochastic–geometric model of the variability of soil formed in Pleistocene patterned ground

R.M. Lark [a,*], E. Meerschman [b], M. Van Meirvenne [b]

[a] British Geological Survey, Keyworth, Nottinghamshire NG12 5GG, UK
[b] Research Group Soil Spatial Inventory Techniques, Department of Soil Management, Faculty of Bioscience Engineering, Ghent University, Coupure 653, 9000 Gent, Belgium

ARTICLE INFO

Article history:
Received 19 December 2012
Received in revised form 13 June 2013
Accepted 21 July 2013
Available online 1 October 2013

Keywords:
Patterned ground
Apparent electrical conductivity
Electromagnetic induction
Stochastic geometry
Voronoi tessellation
Multiple point geostatistics

ABSTRACT

In this paper we develop a model for the spatial variability of apparent electrical conductivity, EC_a, of soil formed in relict patterned ground. The model is based on the continuous local trend (CLT) random processes introduced by Lark (2012b) (Geoderma, 189–190, 661–670). These models are non-Gaussian and so their parameters cannot be estimated just by fitting a variogram model. We show how a plausible CLT model, and parameters for this model, can be found by the structured use of soil knowledge about the pedogenic processes in the particular environment and the physical properties of the soil material, along with some limited descriptive statistics on the target variable. This approach is attractive to soil scientists in that it makes the geostatistical analysis of soil properties an explicitly pedological procedure, and not simply a numerical exercise. We use this approach to develop a CLT model for EC_a at our target site. We then develop a test statistic which measures the extent to which soils on this site with small values of EC_a, which are coarser and so more permeable, tend to be spatially connected in the landscape. When we apply this statistic to our data we get results which indicate that the CLT model is more appropriate for the variable than is a Gaussian model, even after the transformation of the data. The CLT model could be used to generate training images of soil processes to be used for computing conditional distributions of variables at unsampled sites by multiple point geostatistical algorithms.

1. Introduction

> *'Mais surtout nous insisterons sur la nécessité d'incorporer au maximum la physique du problème et le contexte géologique de la zone étudiée'.*
>
> [Chilès and Guillen (1984)]

In most geostatistical analyses of soil the data are assumed to be a realization of a multi-Gaussian random function, perhaps after they have been transformed so that their histogram represents a Gaussian distribution. Furthermore, the random function commonly has a spatial covariance function drawn from a limited subset of models (Webster and Oliver, 2007), which are used because of their convenient mathematical properties. In some of the earth sciences there has been progress in the development of random functions with parameters that are determined, or at least constrained, by parameters of underlying processes which have a physical meaning (e.g. Chilès and Guillen, 1984; Kolvos et al., 2004). This has advantages (Lark, 2012a); for example, the efficiency of spatial sampling to model the spatial covariance function could be improved if prior distributions for covariance parameters could be specified from process knowledge. However, this has not been achieved in soil science. Lark (2012a) suggested that this is probably because the variables that soil scientists study are commonly influenced by a more complex set of factors at more diverse spatial scales than is the case for the variables where it has proved possible to specify the covariance function from process information. For example, the covariance function for diffusion processes is well-established (Whittle, 1954, 1962), and diffusion is a source of spatial variation in the concentration of nutrients in soil, but it is just one of many sources of spatial variation, and is of limited importance at the spatial scales most generally studied for practical purposes.

Lark (2012a, 2012b) suggested that progress might be made by recognizing a number of distinct *modes* of soil variation, simple and generalizable rules that capture how the effects of factors of soil variation vary laterally, and which map naturally on to particular spatial random functions. For example, in conditions where soil variation is strongly determined by differences between discrete domains in the landscape (such as geological units, topographic units, fields etc.) then a subdivision of space into random sets such as Poisson Voronoi polygons may be appropriate (Lark, 2009) and properties of the spatial model (such as the mean chord length of the polygons) may be given a physical meaning.

Lark (2012b) proposed a mode of soil variation: continuous local trends. Under this mode of variation soil varies laterally in space, changing continuously rather than in a step-wise fashion; and these

* Corresponding author. Tel.: +44 115 9363242.
E-mail address: mlark@nerc.ac.uk (R.M. Lark).

0016-7061/$ – see front matter © 2013 Natural Environment Research Council. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.geoderma.2013.07.023

trends are local and repeating, so that they are essentially unpredictable (in contrast to a large-scale trend in a variable that might be observed across a study area). Examples of continuous local trends would be concentration gradients around the rhizosphere, or around individual plants, and catenary variation at landscape scale. Lark (2012b) proposed a general family of random functions to describe continuous local trends (CLT random functions). The value of a CLT variable at some location is given by a distance function, whose argument is the distance from the location of interest to the nearest event in a realization of a spatial point process. This makes the CLT a random function. The CLT variables considered by Lark (2012b), and in this paper, are Poisson CLT (PCLT) variables because the spatial point process is completely spatially random. Lark (2012b) estimated parameters of a PCLT process from data on a soil variable. It was also pointed out that the PCLT process might differ from a comparable Gaussian random function with respect to its multiple point statistics (Strebelle, 2002). This raises the possibility that PCLT models, as well as mapping closely on to a particular mode of soil variation, might be practically useful for applications where spatial connectivity plays a major role controlling processes in soil and so the multiple point statistics of the variable are important.

In this paper we use a PCLT random function to model the variation of apparent electrical conductivity, EC_a, of soil at a site where this variable is strongly influenced by spatial patterns in the parent material. These patterns arose from the development of ice wedges in Eocene clay under permafrost conditions, and subsequent infilling by coarser material which leads to strong textural contrasts in the soil. The objective is to show how soil knowledge: general knowledge about soil formation in the particular environment and its relationship to EC_a, and some simple descriptive statistics of the data (summary statistics and empirical variograms), allow us to select and fit a PCLT model. We then compare the PCLT model with a trans-Gaussian (TG) model of the data, i.e. a model fitted by conventional geostatistical analysis after the data have been transformed to approximate normality. Specifically we compare the models with respect to a statistic that summarizes the spatial connectivity of the coarser material, which might be relevant to simulations of transport processes in the soil. We then evaluate which model appears best to represent the spatial pattern in the data.

2. Case study

2.1. The study area and data collection

We surveyed an area of Pleistocene patterned ground in the sandy silt region of Belgium. The patterned ground was identified by polygonal crop marks on an aerial photograph and interpreted to be the result of ice wedge formation during the last glacial period. The study area and data collection were discussed in detail by Meerschman et al. (2011), therefore we limit ourselves here to a brief presentation of it. More general information on ice-wedge polygons constitutes part of the soil-knowledge base that we use in this study, and is presented in Section 2.3.2. below as it is required.

The study area (0.6 ha) was located in an agricultural field in Deinze, Belgium (central coordinates: 51°01′16″N, 3°29′41″E). Excavation of a small part of the study area (6 × 6-m) to a depth of 0.9 m uncovered an ice-wedge pseudomorph with a diameter of about 6 m. The wedges were formed in clay-rich Tertiary marine sediments that were covered with a 0.6 m layer of silty-sand Quaternary deposits. Texture analysis on 94 subsoil samples (0.6–0.8 m) showed a clear contrast between the Eocene host material (on average 21% clay) and the superficial material (on average 6% clay).

Previous studies (Cockx et al., 2006; Saey et al., 2009) have shown that EC_a is a useful covariate to study textural variability at profile and polygon-scale in soils formed in these conditions. The study area was surveyed with a mobile proximal soil sensor measuring the EC_a (mS m^{-1}) of an underlying soil volume down to approximately 1.5 m. The sensor was mounted on a sled pulled by an all terrain vehicle. The vehicle drove along parallel lines with an in-between distance of on average 0.75 m. The within-line distance between sensor response registrations was 0.15 m.

2.2. Initial data analysis

Meerschman et al. (2011) noted that the EC_a measurements clearly reflected the polygonal patterns: small EC_a values indicated the former ice wedges filled with lighter material. In addition to the short-range variation in EC_a, there were large values of EC_a near an old field track in the north-east of the surveyed region. To avoid any assumptions about the form of this trend we decided to restrict our analyses to the lower left quadrant of the surveyed area, a region of approximately 40 × 40-m, with 17 792 observations, which excludes this area with elevated EC_a. Fig. 1 shows a post-plot of these data.

Fig. 2 shows the histogram of the data. Summary statistics are presented in Table 1. Note that the data are mildly skewed. In the analyses reported below the PCLT model was fitted in all cases to the raw data, and all analyses with the TG model were done with the data after a transformation which is described in Section 2.3.1 below.

2.3. Spatial analysis

In this section we describe the analysis of the EC_a data to fit a TG model and a PCLT model. The first task (Section 2.3.1) was straightforward after a data transformation, which is described. In Section 2.3.2 we describe how soil knowledge was used to fit the PCLT model.

2.3.1. Trans-Gaussian model

The objective of the case study is to compare a continuous local trend (PCLT) model of the data with a trans-Gaussian (TG) model, as might be used in standard geostatistical analysis. Although the data are only mildly skew, since the objective of this exercise is to compare a Gaussian or Trans-Gaussian model with a stochastic geometric alternative, it was decided to transform the data so that the histogram and summary statistics were as close as possible to those expected for data drawn from a Gaussian random variable. We therefore used a Box-Cox transformation of the data to normality for the TG modelling:

$$\begin{aligned} y &= \frac{z^{\zeta}-1}{\zeta} \quad \zeta \neq 0, \\ &= \log_e(z) \quad \zeta = 0, \end{aligned} \tag{1}$$

where z is a value on the original scale and y is a transformed value. We used the BOXCOX procedure from the MASS package (Venables and Ripley, 2002) for the R platform (R Development Core Team, 2012) to find the likelihood profile of the ζ parameter, and selected the value with maximum likelihood. The data were then transformed with the maximum likelihood estimate of ζ, substituted into Eq. (1) and then standardized to zero mean and unit variance. The estimate of ζ and summary statistics for the data after transformation, and standardization, are presented in Table 2.

An isotropic empirical variogram of the transformed and standardized data was then computed using the method of moments estimator due to Matheron (1962) as implemented in the FVARIOGRAM directive in GenStat (Payne, 2010). An authorized model was then fitted to the estimated variogram by weighted least squares (Cressie, 1985) using the MVARIOGRAM procedure in GenStat (Harding et al., 2010). Alternative models were considered and the stable or powered exponential model was selected on the basis of the Akaike information criterion (McBratney and Webster, 1986). This variogram model takes the form

$$\gamma(r) = c_0 + c_1(1-\exp(-\{r/a\}^{\kappa}), \tag{2}$$