Geoderma 213 (2014) 571–577

Contents lists available at ScienceDirect

Geoderma

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

Using bivariate multiple-point statistics and proximal soil sensor data to map fossil ice-wedge polygons

E. Meerschman *[,a], M. Van Meirvenne [a], G. Mariethoz [b], M.M. Islam [a], P. De Smedt [a], E. Van De Vijver [a], T. Saey [a]

[a] Research Group Soil Spatial Inventory Techniques, Department of Soil Management, Faculty of Bioscience Engineering, Ghent University, Coupure 653, 9000 Gent, Belgium

[b] National Centre for Groundwater Research and Training, School of Civil and Environmental Engineering, University of New South Wales, UNSW 2052, NSW, Sydney, Australia

ARTICLE INFO

Article history:
Received 25 April 2012
Received in revised form 3 December 2012
Accepted 27 January 2013
Available online 16 March 2013

Keywords:
Multivariate simulation
Pedometrics
Proximal soil sensor data
Multiple-point statistics
Training image

ABSTRACT

Multiple-point statistics (MPS) is a collection of geostatistical simulation algorithms that uses a multiple-point training image (TI) as structural model instead of a two-point variogram. MPS allows to simulate more complex random fields, like phenomena characterized by spatial connectivity. A very recent development is multivariate MPS in which an ensemble of variables can be simulated simultaneously using a multivariate TI. We investigated if multivariate MPS can be used for the processing of proximal soil sensor data, i.e. interpolating the sensor data and predicting the target variable. We measured a field with fossil ice-wedge polygons in the subsoil with an electromagnetic induction sensor and used these measurements to predict the location of wedge material in the subsoil. We built a bivariate TI with a categorical image of a random polygonal network as primary variable and a continuous image of the corresponding sensor values as secondary variable. Then, we performed a bivariate reconstruction with the recently developed Direct Sampling software. The resulting E-types provided an interpolated sensor data map and a probability map predicting the location of wedge material in the subsoil. This procedure was compared to the more traditional approach of interpolating the sensor data with ordinary kriging and performing a fuzzy *k*-means classification. Comparing the resulting maps with an aerial photograph revealing the location of the ice-wedges through polygonal crop marks, showed that MPS reconstructed the polygonal patterns much better. The local accuracy of the MPS maps was proven by an independent quantitative validation based on nine extra measurement lines and 94 bore hole samples. As a first application in soil science, our case study showed that multivariate MPS can be used for the processing of proximal soil sensor data. The flexibility of the technique opens perspectives for other new applications and therefore multivariate MPS can become a valuable part of the pedometrical toolbox.

1. Introduction

The key function in traditional geostatistics is the variogram, which is used as a model of the spatial structure. However, this two-point statistic is often not able to characterize complex random fields, such as phenomena showing spatial connectivity. To map complex random fields, multiple-point statistics (MPS) need to be considered (Guardiano and Srivastava, 1993). The fundamental idea of MPS is to replace the two-point variogram by a multiple-point training image (TI). A TI is a conceptual image of the expected spatial structure and is often built based on prior knowledge. A very recent development in MPS is multivariate MPS, in which an ensemble of variables can be simulated simultaneously using a multivariate TI (Mariethoz et al., 2010).

MPS was developed in petroleum geology and hydrogeology (Strebelle, 2002) and to date most of its applications can be found in these fields (e.g. Comunian et al., 2011; Huysmans and Dassargues, 2009; Le Coz et al., 2011; Ronayne et al., 2008; Strebelle et al., 2003; Zhang et al., 2006). Complex patterns, that are hard to model with traditional two-point geostatistics, also appear in soil science: a.o. dune patterns, paleochannels, limestone pavement, desiccation cracks, (relict) patterned ground, land-use patterns, sedimentary rock layers and soil pores. However, the use of MPS in the processing of soil data is still an open research question. In this paper, we investigated whether multivariate MPS can be used for the processing of proximal soil sensor data.

Proximal soil sensing is an increasingly used data source for soil inventory (McBratney et al., 2000). In a mobile setup, these sensors allow to rapidly collect indirect observations of the subsoil in a non-destructive way (Adamchuk et al., 2004). Processing proximal soil sensor data typically includes two steps: first the sensor data need to be interpolated to a regular grid and then this map can be used as a proxy to predict the target variable (de Gruijter et al., 2010).

Even though proximal soil sensor data are considered as high-resolution data, interpolating the data to a regular grid remains a crucial processing step. Sensor sampling is typically done with a sensor attached to a vehicle taking measurements at fixed intervals while driving

* Corresponding author. Tel.: +32 9 2646042; fax: +32 9 2646247.
E-mail address: eef.meerschman@ugent.be (E. Meerschman).

0016-7061/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.geoderma.2013.01.016

along parallel lines. With the instruments available today, the within-line sampling density is mostly no longer a limiting factor. The between-line distance, on the other hand, largely affects the costs of a field survey. Generally, this distance should be chosen based on the expected scale of the soil features one wants to map. Apart from interpolating the sensor data between measurement lines, spatial interpolation is required to complete the data set when some areas are inaccessible for the sensor survey.

To date, ordinary kriging (OK) is an often used method to interpolate sensor data because of its declustering ability. OK is a traditional geostatistical estimation technique based on a two-point variogram (Goovaerts, 1997). In our experience, OK is a successful method to interpolate sensor data. However, when the sensor data reflect subsoil phenomena that have a complex spatial pattern or are highly spatially connected, the two-point variogram is no longer sufficient. In practice, problems arise when the between-line distance is larger compared to the scale of the investigated soil features. Hence, it is worth investigating whether MPS can serve as a more suited interpolation technique for these situations.

If the sensor variable differs from the target variable (i.e. the soil variable of interest), a model is needed to predict the target variable from the sensor variable, which then serves as an ancillary or secondary variable (de Gruijter et al., 2010). For example, if the sensed variable is electrical resistivity and the variable of interest is porosity, the modeling of the relationship between these two attributes is critical. Depending on the specific situation and the type of target variable, a variety of pedometrical techniques can be used for this aim, ranging from numerical classification to CLORPT and hybrid techniques (McBratney et al., 2000). For instance, fuzzy *k*-means is an often used predictive classification technique to delineate zones with homogeneous soil properties based on proximal soil sensor data (Cockx et al., 2006, 2007; Islam et al., 2011; Vitharana et al., 2008b). Examples of CLORPT techniques are predicting the depth to contrasting soil layers from proximal soil sensor data with inverse modeling techniques (De Smedt et al., 2011; Saey et al., 2008, 2009) or predicting the soil clay content based on neural network approaches (Cockx et al., 2009). Vitharana et al. (2008a) used regression kriging to predict the depth to clay substratum and Triantafilis et al. (2001) compared different hybrid techniques to predict soil salinity from proximal soil sensor data.

Multivariate MPS is promising for both the interpolation of sensor data and the prediction of the target variable. This technique is mainly developed for situations where one variable is (partially) known and the other is to be simulated (the collocated simulation paradigm). Using a bivariate TI is especially interesting when the relationship between the variables is known through training data but cannot simply be expressed as a mathematical relationship (Mariethoz et al., 2010; Meerschman et al., 2013). To investigate the use of multivariate MPS, we applied it to a case study aiming to predict the location of fossil ice-wedge polygons in the subsoil based on electromagnetic induction (EMI) data.

Fossil ice-wedges polygons are a clear example of spatially connected subsoil features. They are remnants of thermal contraction cracks that were formed during glacial periods (Kolstrup, 1986). At the end of the glaciation these soil cracks were filled up and covered with wind and water transported sediments (French, 2007). Hence, today fossil ice-wedge polygons can be recognized as polygonal networks in the subsoil that are filled with soil material (wedge material) that is younger than the surrounding material (host material). Mapping these cryogenic features is of interest since they cause abrupt changes in the subsoil composition, possibly inducing preferential flow paths for e.g. agro-contaminants or nutrients. Furthermore, the morphology of this polygonal network is important for paleoclimatological reconstructions (Plug and Werner, 2002, 2008). It has recently been shown that EMI sensors are an effective aid in the mapping of fossil ice-wedge polygons, especially when the textural contrast between the wedge material and the host material is sufficiently strong (Cockx et al., 2006; Meerschman et al., 2011).

In this paper, we used an EMI sensor to measure a field with fossil ice-wedge polygons in the subsoil. Then, we applied bivariate MPS to interpolate the proximal soil sensor data to a regular grid and to simultaneously derive a map estimating the location of the fossil ice-wedge polygons in the subsoil. To set a comprehensive framework for the evaluation of the new method's prediction performance, we compared it with the often applied procedure of interpolating the sensor data with OK and then performing a fuzzy *k*-means classification to derive the possibility of finding wedge material in the subsoil.

2. Material and methods

2.1. Study area and data collection

Fig. 1a shows an oblique aerial photograph of an agricultural field in Belgium (central coordinates: 51°01′16″ N, 3°29′41″ E). The photograph was taken on 4 August 1996 when sugar beets were cultivated on the field. At that moment polygonal crop marks revealed the presence of an underlying network of fossil ice-wedges. Besides this

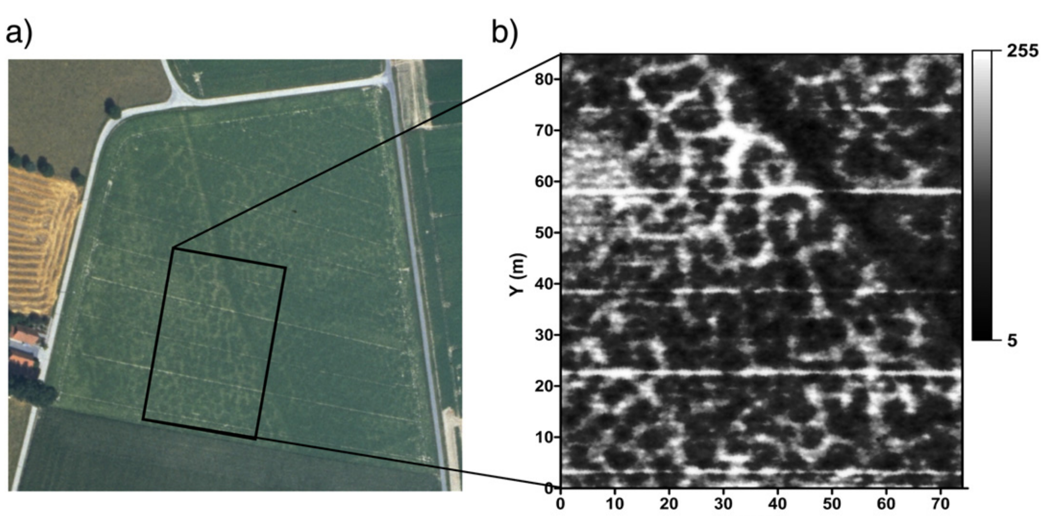

Fig. 1. (a) Aerial photograph taken on 4 August 1996 showing polygonal crop marks and a former field track (north-southeast oriented) with delineation of the study area (large rectangle) and (b) same aerial photograph after georectification, clipping, and color stretching. Coordinates are according to the Belgian metric Lambert-72 projection.
a, J. Bourgeois, Department of Archaeology and Ancient History of Europe, Ghent University, Belgium, photo: J. Semey.