Chemometrics and Intelligent Laboratory Systems 136 (2014) 24–35

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Soil heterogeneity characterization using PCA ($X_{variogram}$) - Multivariate analysis of spatial signatures for optimal sampling purposes

Zahra Kardanpour [a,b,*], Ole Stig Jacobsen [a], Kim H. Esbensen [a,b,c]

[a] Geological Survey of Denmark and Greenland (GEUS), Copenhagen, Denmark
[b] ACABS research group, University of Aalborg, campus Esbjerg (AAUE), Denmark
[c] Telemark University College, Porsgrunn, Norway

ARTICLE INFO

Article history:
Received 15 February 2014
Received in revised form 19 April 2014
Accepted 30 April 2014
Available online 10 May 2014

Keywords:
Variogram
Heterogeneity characterization
Heterogeneity management
Scale dependency
Representative sampling
Spatial data structure
Fate studies

ABSTRACT

This study explores how to combine variographic spatial characterization with multivariate data analysis, by showing how Principal Component Analysis (PCA) can be applied to unconventional types of data matrices, $X_{variogram}$. This is here performed on a specific data set from an agricultural field in western Jutland, Denmark, but the data analytical approach is *generic*. In order to characterize the heterogeneity of a typical sandy soil, a variographic experiment along a 1-D profile is performed on 38 different minerogenic variables (geochemical elements). While the variogram is defined for one variable only, it is shown how PCA is able to characterize a multitude of variograms simultaneously, facilitating subject-matter interpretation of the fingerprints of the process(es) responsible for the spatial heterogeneity encountered. PCA scores and loadings contain information pertaining to the specific matrix type consisting of variograms, $X_{variogram}$. Together with a companion paper, a complete approach for characterizing scale-varying spatial heterogeneity is presented with a view of developing sampling procedures for managing the intrinsic variability in natural soil and in *similar* systems (e.g. environmental characterization and monitoring, pollution in time and space, applied geochemistry, medical geology). Sampling in all of these contexts is shown to be much more than a 'materials handling' issue, by force involving the Theory of Sampling, TOS. The PCA ($X_{variogram}$) approach can be applied for tuning in of sampling procedures and 1-D and 2-D sampling plans in soil, environmental substrates, pollution and medical geology studies with a carrying-over potential to many other application fields with *similar* heterogeneity management needs.

1. Introduction

Development of representative soil sampling procedures addressing heterogeneity characterization (1-D, 2-D) forms the backdrop for the present methodological foray, which a.o. is responding to a growing need for scientifically based sampling procedures and principles for parameterization of the intrinsic variability in natural soil and similar systems (environmental characterization and monitoring, pollution in time and space, applied geochemistry, medical geology) [1–5].

In addition to vertical transport processes, a number of abiotic as well as biotic factors are prime determinants regarding the fate of environmental contaminants in soil and groundwater environments. Key transformation processes, and their kinetics, have been described for several important contaminants, aiming for an improved understanding of biogeochemical transformation processes of inorganics and organic compounds in the subsoil environment. However, most of this knowledge is based on laboratory-scale batch experiments involving homogenized samples representing a few grams of soil matrix at best. Even when procedures for species transport, kinetic studies and analysis follow current scientific and international standards, it is widely recognized that such small scale experiments do not reliably reflect inherent meso- and macro-scale complexibilities in the salient geological formations. The soil matrices involved are significantly heterogeneous and a number of subsequent scale-up issues are critically related to the true empirical variability in both vertical and horizontal dimensions; taking into account the time scale adds further system complexity. Recently, such issues have started to be brought into focus by the computer modeling community, seeking ways to address various uncertainties in pesticide leaching models [1,6–12]. The present study aims to contribute to this contemporary development.

Sampling in all of these contexts is much more than a 'materials handling' issue. Combining the tools developed here with specific domain knowledge on relevant processes and variable properties, it will be possible to address critical scale variability issues with confidence based on fully realistic parameter characterization. The background is a standing debate in a.o. the agricultural and environmental sciences (and related disciplines where 1-D and 2-D sampling is on the agenda) as to what constitutes an 'optimal' 2-D sampling pattern, or sampling plan. Fig. 1 (modified from [13]) illustrates the current options and alternatives.

* Corresponding author at: Geological Survey of Denmark and Greenland (GEUS) Copenhagen, Department of Geochemistry, Denmark, Oster voldgade 10, 1350, Copenhagen, Denmark.
E-mail address: zka@geus.dk (Z. Kardanpour).

http://dx.doi.org/10.1016/j.chemolab.2014.04.020
0169-7439/

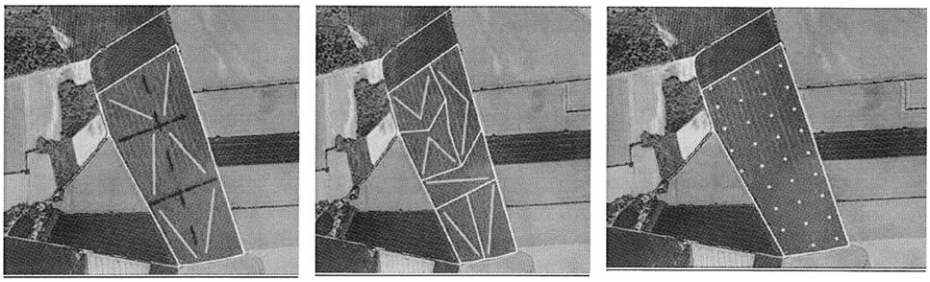

Fig. 1. Principal alternatives in 2-D sampling. Left: Random subdivision into a number of sub-fields, each characterized with a linear profile. Center: 'Intelligent subfield division' (i.e. based on previous knowledge), employing linear or 'V-shaped' profiles. Right: Regular grid-based sampling (white dots) and a *random direction*, variographic profile sampling (thick white line). In the latter situation, sometimes a direction is chosen based on previous knowledge.
Illustration, modified from [13].

When random 2-D sampling is used, randomly chosen subfields are traditionally sampled individually and aggregate statistics are pooled to furnish an estimate for the total field. An alternative is the 'intelligent subfield division', or 'authoritative subdivision', which is based on previous knowledge (important when structural 2-D information is at hand), but individual subfield statistics are also aggregated here. An evergreen of considerable merit is total field characterization based on a regular grid, which furthers a composite estimate of the characteristics of the total field, the confidence of which scales directly with the number of increments (grid points) deployed.

The purpose of this paper is to introduce an alternative 1-D variographic approach and to illustrate it with a study of the heterogeneity of an exemplar material (sandy soil), from a well-studied experimental agricultural field (Fladerne Baek, Denmark): the variographic profile approach can be applied in all cases in which the sampling target (lot) displays no preferred directional characteristics, i.e. for all types of *isotropic* 2-D lots. For lots with special directional features, e.g. trends or other clues from the pertinent agricultural or geological history, which could possibly be related to the spatial distribution of the target analytical parameter, selection of the 1-D profile direction must be determined with this in mind.

2. Study area and sample collection

Fladerne Baek is situated on a peri-glacial outwash plain, approximately 9 km South West of Karup airport, Jutland, Denmark (56°N, 9°E). The substratum is an arable sandy soil (Typic Fragiorthod). The area has been tilled and cropped for more than 100 years, with main crops barley and potatoes. The Fladerne Baek field has been extensively plowed during all this time. Visual observation indicates a well mixed medium with almost no visual heterogeneity features. The main issues addressed in the present study are: does this visually apparent uniformity extend to all geochemical elements and soil/agricultural chemical compounds? Are there hidden chemical heterogeneities, and if so what are their spatial characteristics? How can one manage the simultaneous spatial heterogeneity descriptions of 38 chemical variables?

A 100 m long profile was laid out oriented parallel with the plow lines with sampling equidistance 1 m. This direction was selected after careful inspection of the field relationships (there is no special history for the field usage). In this case the minerogenic properties of the A horizon were not judged to have been *selectively* affected by plowing with respect to directionality. The Fladerne Baek field can therefore serve well as an illustration of the 1-D profiling technique, allowing to characterize the *long-range* variability of the target analytes.

The present soil samples cover a depth interval from 0 to 15 cm. Each individual sample totaled 250–350 g of moist soil. At the center of this profile, nine additional samples were taken in a square pattern, including two of the ordinary profile samples (with a center point added) and six samples located along the two parallel edges, totaling a nine point square grid, with a one meter equidistance. This central half-equidistance pattern was designed to allow a complementary expression of the *short-range* sampling variability. Thus the sampling rationale aimed at variographic characterization commensurate with the profile scale lengths between 1 m and 50 m (half the full length); the local short range replication sampling measured 1 × 1 m, intended as a basis for conventional statistical characterization.

The original moist soil was stored in darkness at 5 °C until sample preparation in the laboratory (after three days). Laboratory sample processing commenced with the removal of large non-soil objects (stones, roots). As the analytical aliquot here will be less than a gram, there is a critical need for representative mass reduction; the split-off remaining sample material was freeze-dried for future pesticide a.o. volatile parameter studies (paper III in this series). A description of all laboratory methods is given in a companion paper, [14], where compliance with the principles of the Theory of Sampling (TOS) is documented in full. In particular mass reduction employed a micro bed-blending stacking/transverse thin-slice reclaiming procedure [15], focused on obtaining representative 2 gram sub-samples. All primary samples were laid out in a narrow clean tray in a fashion producing a linear multi-layer composite bed from which several randomly selected transverse thin-slice increments were extracted. These were dimensioned in such a way that seven increments were needed to produce the intended mass for a composite sub-sample. After drying this procedure was repeated using appropriately scaled-down equipment for a further mass-reduction step, until the final analytical aliquot mass was reached.

3. Methods

3.1. Analytical method

A set of 38 inorganic minerogenic elements was analyzed: Al, Ba, Ca, Ce, Co, Cr, Cs, Cu, Fe,Ga, Ge, Hf, K, La, Li, Lu, Mg, Mn, Na, Nb, Ni, P, Pb, Pd, Pt, Rb, Sb, Sc, Sn, Sr, Ta, Th, Ti, U, W, Y, Zn, and Zr. Representative 0.1 g aliquots [15] were treated with hydrofluoric and nitric acid in a closed Savillex vessel (polytetrafluorethylene polymer) on a hotplate at 130 °C. After at least 24 h the sample is evaporated to dryness on the hotplate at 100 °C. Nitric acid is added and the sample is again evaporated to dryness. This last procedure is repeated a second time. Nitric acid and water are added and the vessel is closed and placed on the hotplate at 130 °C for at least 12 h. The sample is then diluted to 50 ml. Before measurement the sample is diluted a further 11 times. All elements were analyzed with a PerkinElmer-Elan 6100DRC ICP-MS, using PerkinElmer-Elan version 3.2 software for instrument control, data collection, calibration and quantification.

3.2. Theory of sampling (TOS)

The basics of the Theory of Sampling (TOS) can be stated as a set of three governing principles (GP) and four sampling unit operations (SUO), involving eight types of sampling errors (of which only five are