Precision Agriculture, 4, 193–201, 2003

Is the Soil Variability within the Small Fields of Flanders Structured Enough to Allow Precision Agriculture?

MARC VAN MEIRVENNE Marc.VanMeirvenne@ugent.be

Department of Soil Management and Soil Care, Ghent University, Coupure 653, 9000, Gent, Belgium

Abstract. The average area of agricultural fields in Flanders (Belgium) is about 1.7 ha, being very small compared to fields where precision agriculture is currently applied. Therefore this paper addresses the question whether the within-field variation of soil properties in such fields is structured enough to motivate precision agriculture. To answer this question, 9 soil properties determined on 380 soil samples located in 77 agricultural fields situated in the 5 most dominant pedoscapes of Flanders were used to analyze their spatial variation over intervals ranging from 5 to 900 m. The data set was subjected to a principal component analysis which identified two principal components (PCs) explaining more than 78% of the total variance. The first PC represented the chemical soil properties and the second the physical and biological properties. A variogram analysis of the scores on these two PCs showed that the micro-scale and random variation dominated (82%) the within-field variability of the first PC. The within-field variability of the second PC was dominantly spatially structured (only 37% micro-scale and random variation). Therefore, it was concluded that mainly for soil physical and biological properties (like soil textural fractions and organic matter), the average within-field variation in the small fields of the investigated landscapes is structured enough to allow precision agriculture.

Keywords: within-field variability, precision agriculture, variogram analysis, geostatistics

Introduction

Precision agriculture (PA), being the adaptation of management to site specific conditions, has triggered new attention to soil spatial variation since this is considered to be the key element to its successful implementation (Robert, 1999; Verhagen and Bouma, 1997). In particular the within-field scale is of importance for PA since traditional agriculture focuses mainly on the between-field variation of yield controlling properties. However, within-field scale does not define a particular order of dimension since the size of agricultural fields can vary considerably and it will be clear that large fields could be expected to benefit more from PA than small fields. Moreover, it is not sufficient to encounter an important within-field variation to motivate PA. This variation must be spatially structured to allow accurate mapping. Micro-scale and random variation cannot be mapped, they just add uncertainty to the cartographic information. Geostatistical tools, like variogram analysis, allow differentiation between structured variation and micro-scale and/or random variation (Cressie, 1991). Consequently these tools have been used intensively to map soil properties to guide PA. Mostly however, only a few soil properties sampled within one, or a few neighboring, fields have been considered (e.g. Geypens *et al.*, 1999; Mulla, 1993). Therefore, these results apply only locally. On the other hand, regional studies on soil properties (e.g. Van Meirvenne *et al.*, 1990) contain too few details to allow a quantification of the within-field variability. Rarely, studies have been reported which investigate the within-field variability of a large number of agricultural parcels over different pedoscapes. Yet such a study would be needed to allow a general insight into the within-field variation of soil properties within a region, and hence support the decision if the application of PA would be worthwhile.

In some countries, the longstanding and intensive agricultural activities have created strongly fragmented landscapes with very small fields. In Flanders, Belgium, the average area of an agricultural field is about 1.7 ha, which is very small compared to fields where PA is currently applied (e.g. Herbst *et al.*, 2001; Mulla, 1997; Shatar and McBratney, 1999). Consequently the following question can be formulated: "Is the within-field variation of soil properties sufficiently structured to allow PA in highly fragmented agricultural landscapes?". To answer this question, a (geo)statistical analysis was performed using 9 soil properties measured on 380 soil samples, located in 77 agricultural fields and situated in 5 pedoscapes of Flanders.

Theory

Principal component analysis (PCA)

Due to their multidimensionality, multivariate data sets (i.e. with more than 1 variable measured on each sample) can be difficult to interpret. One of the methods developed to overcome this problem is PCA. A PCA of a set of p variables generally aims to summarize—and hopefully improve the interpretation of—the available information by creating a few, say k, new variables that are orthogonal linear combinations of the original variables referred to as principal components (PCs). The analysis requires the computation of the eigenvalues and eigenvectors of the variance-covariance matrix (or the correlation matrix) of the p variables. The eigenvectors determine the directions of maximum variability while the eigenvalues specify the variances of the vectors. The mathematical details of a PCA can be found in standard text books (e.g. Johnson and Wichern, 1992).

Variogram analysis

In geostatistics the semivariance $\gamma(\mathbf{h})$ is commonly used to describe and model the spatial variance of a regionalized property Z (Journel and Huijbregts, 1978), using:

$$\gamma(\mathbf{h}) = \frac{1}{2n(\mathbf{h})} \sum_{i=1}^{n(\mathbf{h})} \{z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})\}^2 \tag{1}$$

with $z(\mathbf{x}_i)$ the value of variable Z at location $\mathbf{x}_i$, $\mathbf{h}$ is a distance vector, or lag, and $n(\mathbf{h})$ is the number of pairs separated by $\mathbf{h}$. To these experimental values, a continuous curve is fitted. One often encountered model is the spherical model described by:

$$\begin{cases} \gamma(0) = 0 \\ \gamma(\mathbf{h}) = C_0 + C_1\left(\dfrac{3\mathbf{h}}{2a} - \dfrac{1}{2}\left(\dfrac{\mathbf{h}}{a}\right)^3\right) & \text{if } 0 < \mathbf{h} \leq a \\ \gamma(\mathbf{h}) = C_0 + C_1 & \text{if } \mathbf{h} > a \end{cases} \tag{2}$$

with C_0 the nugget effect (Y-intercept), a the range and $(C_0 + C_1)$ the sill. The sill represents the total variation encountered at and beyond a, while the nugget effect represents the micro-scale and/or random variation (Cressie, 1991). Therefore, the ratio between and $(C_0 + C_1)$ (called nugget to sill ratio, NSR) indicates the relative proportion of micro-scale and random variation versus the total variation within the study area. Equally, the ratio of C_0 to $\gamma(\mathbf{h})$ represents the proportion of micro-scale (at lags smaller than the smallest experimental lag) and random variation (unstructured variation) versus the variation encountered at a spatial scale $\mathbf{h}$.

Soil sampling and analysis

To investigate the spatial variability of soil properties in the highly fragmented agricultural areas of Flanders, five transects were located in five typical pedoscapes (Figure 1). These pedoscapes were (listed from north to south): the Polder area with loamy to clayey soils (Pol), the Sandy area (Sa), the Silty-sand area (Sisa), the Sandy-silt area (Sasi) and the Silt area (Si). Every transect had a length of 2 km and along it, 76 topsoil (0–30 cm) samples were taken at intervals ranging from 5 to 100 m (Figure 2).

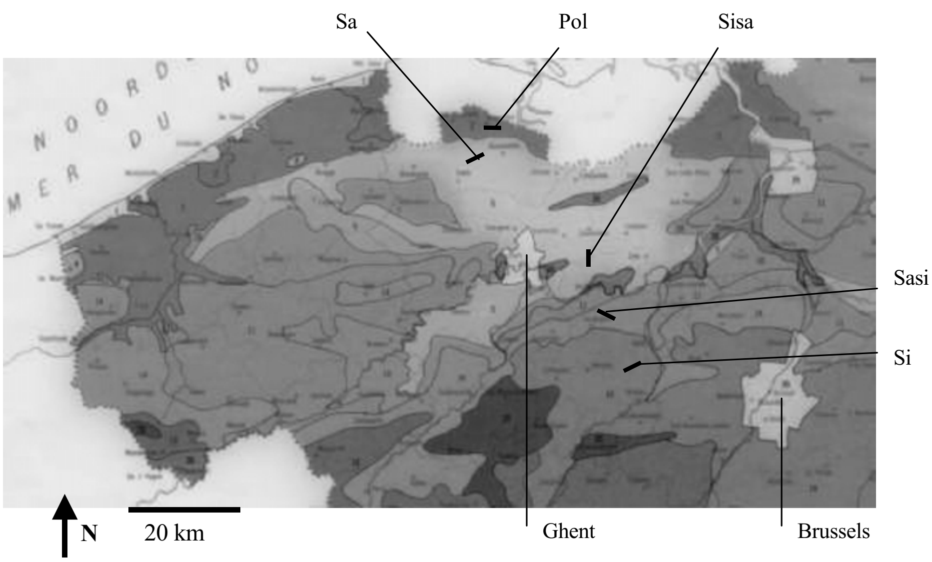


Figure 1. Western part of the soil association map of Flanders, Belgium, with location of the five transects (thick black lines): Polder area (Pol), Sandy area (Sa), Silty-sand area (Sisa), Sandy-silt area (Sasi) and Silt area (Si).

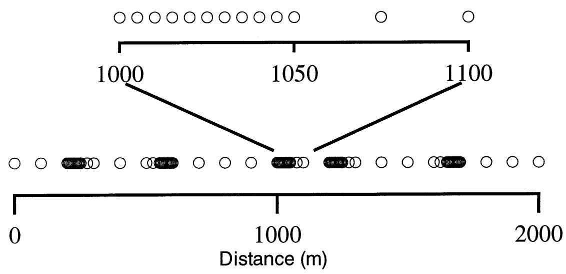


Figure 2. Sampling configuration along one transect.

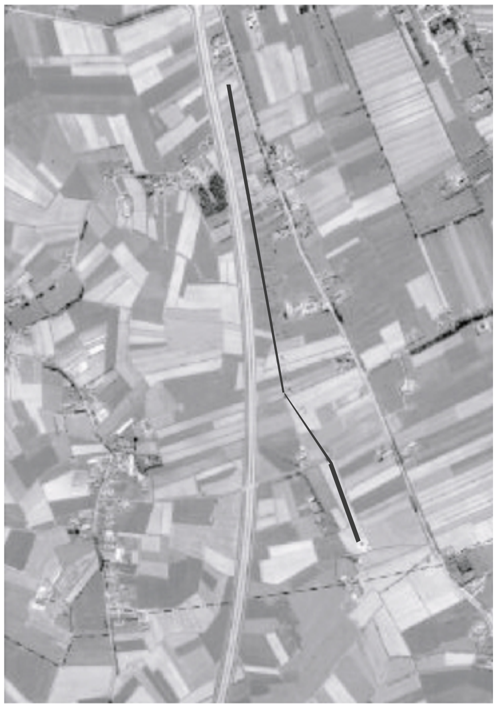


Figure 3. Orthophoto with transect (Sa) positioned on it (the transect has a total length of 2 km).

Only agricultural land was sampled. Information on parcel delineations and land use was recorded. The sampling was done according to a predetermined scheme and the starting point of the transect was located randomly within the first field. Although it was intended to sample along a line, sometimes the orientation had to be modified due to practical considerations (like the presence of non-agricultural land or the refusal by the land owner to grant permission to visit his land) (Figure 3).

The following soil properties were analyzed on every sample: three textural fractions determined by the conventional pipette-sieve method (clay, silt and sand), organic carbon

Table 1. Some statistical parameters of the nine variables considered (all samples pooled, $n = 380$)

	Clay (%)	Silt (%)	Sand (%)	OC (%)	pH-KCl	pF-WP (w/w%)	Ca (mg kg^{-1})	K (mg kg^{-1})	Na (mg kg^{-1})
Minimum	1.3	0.7	6	0.4	3.55	2.78	69	7.8	10.0
Maximum	25.4	83.5	97	3.6	7.51	15.16	39780	1160.2	131.6
Average	10.4	39.7	49.8	1.13	5.46	7.66	6387	137.6	36.2
Standard deviation	5.7	30.2	34.7	0.42	1.10	2.98	11070	108.6	24.6

(OC) determined by the Walkley and Black method, pH-KCl, gravimetric moisture content at 15 bar suction (considered to be the permanent wilting point—pF-WP) and plant available Ca, K and Na (determined on an ammonium-lactate extract at pH 3.75). These properties were chosen because they represented physical, chemical and biological soil conditions which are relatively stable in a medium to a long time frame. Table 1 contains some descriptive statistics of these variables and shows the wide range of values encountered by this transect sampling.

Defining field scale in Flanders

The five transects combined covered a length of 10 km and crossed 77 arable fields. Thus, the average dimension (which we define as being the average field scale $\mathbf{h}_{\text{field}}$) of one field is 130 m, and the average rectangular area of one field is about 1.7 ha. Some regional differences occurred: on average the fields in the Polder area are the largest (4 ha), those in the Silty area the smallest (1 ha). Clearly the rural area of Flanders is highly dissected, as Figure 3 illustrates.

Multivariate analysis

To overcome the limitations imposed by considering one soil property only, and to obtain a more general description of the most dominant soil processes, a multivariate data analysis of the data set was conducted. Therefore, a multivariate matrix was constructed containing all 380 samples (of the 5 transects combined) and the 9 soil properties determined on each of these samples. This matrix was subjected to a PCA performed on the correlation matrix (using SPSS). The correlation matrix (Table 2) indicates that some variables were strongly correlated (e.g. silt and sand, Ca and Na), as could be expected, but between other variables the correlation was sometimes moderate to weak. In general, a sufficient amount of correlation is present for PCA to be useful. The first PC explained 57.0% of the total variance of the data matrix, the second 21.4% and the third 9.9%. It was decided to retain only the first two PCs since they already explained more than 78% of the total variation, and the additional PCs contributed little extra information. These first two PCs were subjected to a Varimax rotation (Johnson and Wichern, 1992) to improve the interpretation of the two PCs. After rotation each PC represented each about

Table 2. Correlation matrix subjected to the PCA

	pF-WP	Clay	Silt	Sand	OC	pH-KCl	Na	K
Clay	0.954							
Silt	0.678	0.754						
Sand	−0.746	−0.820	−0.994					
OC	−0.352	−0.539	−0.548	0.565				
pH-KCl	0.697	0.641	0.235	−0.309	−0.373			
Na	0.762	0.667	0.222	−0.302	−0.216	0.818		
K	0.260	0.250	0.081	−0.112	−0.075	0.291	0.361	
Ca	0.716	0.619	0.117	−0.202	−0.199	0.822	0.922	0.196

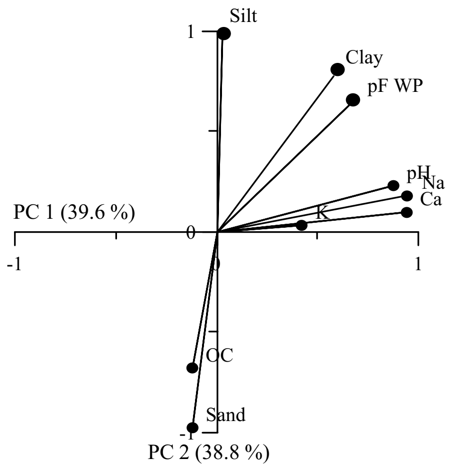


Figure 4. Loadings of the nine soil properties on the first two rotated principal components.

39% of the total variance. Figure 4 shows the obtained loadings (Johnson and Wichern, 1992, p. 397) of the nine variables on the rotated PCs.

The first PC was strongly dominated by pH, Na, Ca and somewhat less by K. Therefore, it was interpreted to represent the soil chemical properties. These properties can be modified by management (e.g. by fertilization or liming) and thus they are variable over a short to medium (several years) time period. The second PC was strongly influenced by the textural fractions silt and sand, and the biological property OC, which is on Figure 4 closely located to the variable "sand." This indicated that the OC is strongly (and positively) related to the sand content as a result of the larger organic matter applications by the farmers to sandy soils. So the second PC represented the physical and related biological soil properties, which are mainly related to soil genetical processes and, which are either invariable in time (like texture), or only modifiable over a medium time period (like OC). Clay was situated in between both PCs (but somewhat closer to the second PC) which reflects the colloidal nature of this soil property, influencing both physical and chemical soil conditions. The moisture content at wilting point (pF-WP) is known to be closely linked to soil texture (Hillel, 1980, p. 150) and in particular the clay fraction (Van Meirvenne and Hoffman, 1989), hence it appears in the close vicinity of the variable "clay" on the graph. So, PCA was successful in creating two new variables (the two PCs) which represented the major part (78%) of the information available in the nine observed soil properties. In this way, the need to select one or a few of the observed nine soil properties for the subsequent spatial analysis, was avoided.

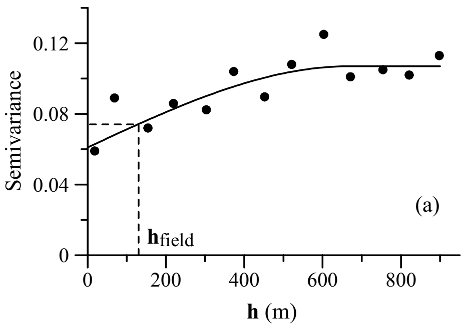
Variogram analysis of the scores on the two PCs

To find an answer to the research question, the one-dimensional variograms of the sample scores on the two identified PCs were calculated using Eq. (1) for each transect. These results were pooled into one variogram for score 1 and another for score 2. Pooling resulted in a large number of pairs for every experimental point ($n(\mathbf{h})$ ranged between 349 and 1315). These two variograms were each modeled by a single spherical model (Eq. (2)) and Table 3 gives the model parameters. Both variogram models were bounded, that is, they reached an upper horizontal bound after the initial slope. It showed that the processes under investigation could be considered to be stationary at the scale of investigation (Webster and Oliver, 2001). The single models indicated that one spatial pattern or process was dominant (Burrough and McDonnell, 1998). The variogram of the scores on the first PC (Figure 5(a)) contained a large NSR (57%) and a range of 670 m. The first parameter indicated that the first PC displayed a considerable micro (<5 m) and random variation. The spatially structured part (i.e. the spatially autocorrelated part) however extended over a large distance, being several times the average size of a field. This suggested a regional pattern in respect to land use, that is, neighboring field frequently belong to the same farmer using a similar type of soil management. The NSR of the scores variogram of the second PC (Figure 5(b)) was only 30%, but the range was considerably shorter (250 m) than the range of the first PC variogram. The spatial behavior of the second PC is therefore quite different from the first PC: a smaller micro and random variation, allowing more accurate interpolation, but its variation occurs over smaller distances (less than twice the average field scale), requiring more frequent observations to allow mapping.

If the average dimension of an agricultural field ($\mathbf{h}_{\text{field}} = 130$ m) is introduced into the variograms, it can be noticed that both variograms have about similar amounts of the total variation (as represented by the sill) present at this field scale: $\gamma(\mathbf{h}_{\text{field}})/(C_0 + C_1) \cdot 100 = 69\%$ for the first PC and 80% for the second. So, the largest part of the total variation of soil properties encountered in these pedoscapes is present within

Table 3. Parameters of the spherical models fitted to the experimental variograms of the scores of the samples on the first and second PC (see Figure 5)

	Scores on PC1	Scores on PC2
C_0	0.061	0.017
C_1	0.046	0.039
a (m)	670	250
NSR (%)	57.0	30.4



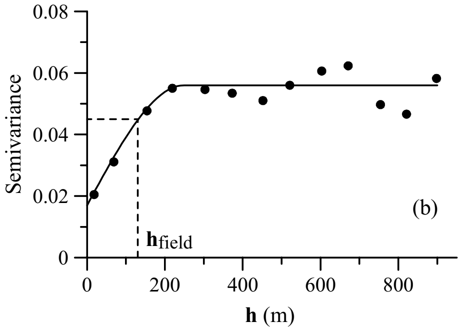


Figure 5. Experimental (dots) and modeled (black curves) variograms of the scores of the samples on the (a) first PC and the (b) second PC.

a field scale. This suggests that, in terms of total within-field variability, the application of PA could be worthwhile, even in these small fields.

However, a large within-field variation is not sufficient to motivate PA. In order to manage this within-field variability spatially, it must be structured enough at a spatial scale sufficiently large to allow management to take it into consideration. With present day tillage and fertilization equipment such a scale would be a few square meters. Therefore, we have to consider the ratio of the nugget effect to the variance at the field scale: $\gamma(\mathbf{h}_{\text{field}})/C_0 \cdot 100$. For the first PC this ratio is 82.2% whereas it is 37.8% for the second PC. So at a field scale the first PC has a dominant micro-scale (<5 m) and random within-field variation whereas the second PC shows dominantly a spatially structured (i.e. autocorrelated) within-field variation. Therefore, the latter offers a much better opportunity for PA to account for spatial variability.

Conclusions

A PCA of the nine soil properties considered allowed identification of two groups of soil properties: (i) the chemical (represented by the first PC) and (ii) the physical and biological (represented by the second PC). The scores on these two PCs were used to perform a variogram analysis. From this analysis the following conclusions could be drawn:

1. The spatial variance of both chemical and physical soil properties is dominantly (70–80%) present at a within-field scale (being on average 130 m in our study area).
2. At a within-field scale, the chemical soil properties display dominantly a micro-scale (<5 m) and random variation (82% of the total within-field variation). This type of variation was much less important (37%) for the soil properties associated with the second PC (physical and biological properties).Hence, in the small fields of the landscapes investigated, when PA is to be guided by soil properties, it should be mainly by those properties dominating the second PC, such as soil textural fractions and OC.

References

Burrough, P. A. and McDonnell, R. A. 1998. *Principles of Geographical Information Systems* (Oxford University Press, Oxford, UK), p. 333.

Cressie, N. 1991. *Statistics for Spatial Data* (John Wiley & Sons, New York, USA), p. 900.

Geypens, M., Vanongeval, L., Vogels, N. and Meykens, J. 1999. Spatial variability of agricultural soil fertility parameters in a Gleyic Podzol of Belgium. Precision Agriculture **1**, 319–326.

Herbst, R., Lamp, J. and Reimer, G. 2001. Inventory and spatial modelling of soils on PA pilot fields in various landscapes of Germany. In: *Third European Conference on Precision Agriculture*, edited by G. Grenier and S. Blackmore (agro Montpellier, France), pp. 395–400.

Hillel, D. 1980. *Fundamentals of Soil Physics* (Academic Press, New York), p. 413.

Johnson, R. A. and Wichern, D. W. 1992. *Applied Multivariate Statistical Analysis* (Prentice-Hall International, Englewood Cliffs, USA), p. 642.

Journel, A. G. and Huijbregts, Ch. J. 1978. *Mining Geostatistics* (Academic Press Inc., London, UK), p. 600.

Mulla, D. J. 1993. Mapping and managing spatial patterns in soil fertility and crop yield. In: *Soil Specific Crop Management*, edited by P. C. Robert, R. H. Rust and W. E. Larson (American Society of Agronomy, Madison, USA), pp. 15–26.

Mulla, D. J. 1997. Geostatistics, remote sensing and precision farming. In: *Precision Agriculture: Spatial and Temporal Variability of Environmental Quality*, edited by Ciba Foundation (John Wiley & Sons, Chichester, UK), pp. 100–115.

Robert, P. C. 1999. Precision agriculture: research needs and status in the USA. In: *Proceedings of 2nd European Conference on Precision Agriculture*, edited by J. V. Stafford (Sheffield Academic Press, UK), pp. 19–33.

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

Van Meirvenne, M. and Hofman, G. 1989. Spatial variability of soil texture in the polder area: II. Cokriging. Pedologie **39**, 209–226.

Van Meirvenne, M., Hofman, G., Van Hove, J. and Van Ruymbeke, M. 1990. A continuous spatial characterization of textural fractions and $CaCO_3$ content of the topsoil of the Polder Region of northwest East-Flanders, Belgium. Soil Science **150**, 710–716.

Verhagen, J. and Bouma, J. 1997. Modeling soil variability. In: *The State of Site Specific Management for Agriculture*, edited by F. J. Pierce and E. J. Sadler (American Society of Agronomy, Madison, USA), pp. 55–67.

Webster, R. and Oliver, M. 2001. *Geostatistics for Environmental Scientists* (John Wiley & Sons, Chichester, UK), p. 272.