



ELSEVIER

Agriculture, Ecosystems and Environment 94 (2003) 31–47

**Agriculture
Ecosystems &
Environment**

www.elsevier.com/locate/agee

An overview of spatial sampling procedures and experimental design of spatial studies for ecosystem comparisons

Alfred Stein^{a,*}, Christien Ettema^{b,1}

^a Wageningen University, Mathematical and Statistical Methods Group, Dreijenlaan 4, 6703 HA Wageningen, The Netherlands

^b Wageningen University, Sub-Department of Soil Quality, Dreijenplein 10, 6703 HB Wageningen, The Netherlands

Received 10 July 2001; received in revised form 16 January 2002; accepted 29 January 2002

Abstract

Comparison of ecosystems and land use studies often require the use of non-classical statistics. This paper describes modern ways of approaching optimal sampling for ecological and environmental purposes. The first part of the paper deals with a description of different sampling procedures. A distinction is made between sampling surveys, optimal grid spacing and adaptive sampling. The second part of the manuscript uses a simulated example to illustrate the different sampling procedures. The third part contains an actual field study where various constraints were met that had to be incorporated in the lay out of the sampling scheme. Optimal statistical sampling procedures can be implemented and utilized to collect maximum information from limited resources.

© 2002 Elsevier Science B.V. All rights reserved.

Keywords: Optimal sampling; Adaptive sampling; Bio-diversity; Experimental design; Soil biota; Spatial simulations

1. Introduction

Spatial studies focusing on environmental, ecological and agricultural phenomena require a proper and carefully designed strategy for collecting data. Data can be difficult or expensive to collect, and both the sampling design and the quality of the data may affect the final qualities of an analysis (Cochran, 1977; Müller, 1998). In addition, comparison of ecosystems and land use studies are not often subject to classical statistics as replicated experiments. Both sampling and the experimental design address collecting of spatial data. In this paper both are addressed simultaneously.

The aim of sampling can be manifold: one may wish to assess the amount of yield, the spatial distribution of weeds or the amount of erosion. Much of the quality of modeling efforts depends upon the quality of collected data. These models can be simple relations between spatial properties or highly elaborate deterministic or stochastic models. To design a scheme for collecting data usually requires a proper formulation of at least two components: what is already known in the area and what is the objective for collecting data. Prior information may consist of already sampled data, the presence and shape of boundaries, and priority of sampling to sub-areas of particular importance. Sampled data can be either hard data like actually collected data in the past or soft, indirect data like knowledge that data are between some boundaries, e.g. based on organoleptic observations (Hendriks et al., 1998).

* Corresponding author. Tel.: +31-317-483551;
fax: +31-317-483554.

E-mail address: alfred.stein@mat.dpw.wau.nl (A. Stein).

¹ Tel.: +31-317-482981; fax: +31-317-483766.

With respect to different objectives for spatial sampling, we may distinguish between:

- sampling of the (spatial) mean, say the average (or total) number of nematodes in an area or the total yield on a field;
- sampling to identify effects of different treatments that are either imposed by a researcher or follow from different soil or ecological conditions;
- sampling to make a spatial investigation such as displayed on a map or characterized by a spatial distribution.

Additional requirements can be formulated as to how many samples can be afforded, whether there is any flexibility in this number and whether every location can be identified and is accessible. In a seminal paper, De Gruijter and Ter Braak (1990) discussed differences between model-based sampling and design-based sampling, a distinction that can be found as well in Särndal et al. (1992). In design-based sampling every point in an area has the same probability of being sampled, whereas in model-based sampling the single best scheme for a given objective emerges.

Research on sampling has a long history. Much of it is governed by the particular type of application: sampling for forestry requires selection from a bounded number of trees, sampling for soils and the environment honors differences in soil processes and homogeneity, sampling for hydrology and meteorology requires inclusion of temporal changes.

Some design aspects can be considered as well. Optimal sampling for linear regression models or experimental designs focuses on design questions and not so much on the use of previously collected data (a good overview is given by Mead et al., 1993). More recently developed adaptive sampling includes such information (Thompson and Seber, 1996), by allowing modification of a scheme if some criterion is fulfilled, such as the presence of a rare phenomenon. Such a criterion can also be based on multivariate aspects. Another general methodology, survey sampling, has more recently found a general setting (Müller, 1998; Van Groenigen and Stein, 1998), first without constraints, later including a range of constraints. In these studies, a quantitative objective is formulated that is approximated by possible data configurations as precisely as required by practical or theoretical considerations.

Information is needed on sampling, in particular for studying bio-diversity of soil fauna. Soil harbors exceptionally species-rich communities, which regulate terrestrial ecosystem fluxes of carbon, nutrients, leachates and trace gases (Anderson, 1995). In addition to tens of thousands of bacterial and fungal types, 1 m² of surface area may cover up to 1000 soil invertebrate species, such as mites and nematodes (Coleman and Crossley, 1996). Soil heterogeneity is crucial for maintenance of soil bio-diversity, as it provides diverse micro-habitats (allowing resource partitioning), and presents these habitats in such complex mosaics that competitors are often spatially or temporally separated (promoting patch dynamics) (Giller, 1996; Ettema et al., 2000).

The aim of this review paper is to: (1) identify the various techniques and procedures that are available for sampling of spatial properties; (2) highlight essential elements of these procedures by implementing these techniques on simulated data; (3) include a single application for spatial constrained sampling. This example describes the experimental and spatial sampling design for a recently started study on the structure and function of soil bio-diversity in grasslands.

2. Spatial sampling

2.1. Statistical sampling

Spatial sampling is equivalent to take observations in a predefined area. Observations may be obtained by means of measurements in the field, or on samples taken to laboratories. In this paper quantitative measurements are considered, which may be approximate (e.g. in rounded numbers) or precise (with many relevant decimals), but which pertain in principle to a fixed location. In inventory studies it is only rarely possible to observe the population to be sampled in its entirety. Therefore, the data are observed only at a limited number of locations. The population of all possible observations in such an area is in principle infinite, whereas the number of observations is finite.

Spatial attributes have been observed in the past following a great number of procedures. The main reason to use statistical sampling schemes is that such sampling guarantees scientific objectivity; if the same area is resampled, perhaps in the future, the results

will be comparable in a well-defined sense (Stuart, 1984). A statistical sampling scheme will avoid forms of bias like those caused by judgement sampling. This is especially valuable if the objective is to obtain data which are representative for the entire area of study, regardless of the size of this area. If however clearly different landscape units occur in an area, it is meaningful to statistically sample separately within these areas. But again sampling within these units should follow statistical procedures.

The basis of statistical sampling is the interpretation of differences and similarities between two or more measurements (Wollum, 1994; Wendroth et al., 1997). Drawing inference from a single measurement is highly risky. There is little reason to assume that another observation in the same area will resemble in any way the first observation. Drawing inference from two observations in the same area will exhibit some of the variation to be expected; the range of the two collected observations falls within the range of all possible observations. Taking a third, a fourth observation, etc., gives increasingly more information of the area under study, like an expression for the mean μ , the standard deviation σ and the standard error of the mean σ_e . It is decided by means of the objective of the study and the amount of prior information how many observations one should take. In this paper attention will be focused on how to determine in advance the number of observations to be taken, as well as their observation locations. Generally, both depend on the actual purpose with which the sampling is carried out in their spatial context.

2.2. Support size

Of importance in spatial sampling is the support size, i.e. the volume with which data are collected. It is common to let the support size correspond with the variation that has to be captured, i.e. different support sizes apply to nematodes, mesofauna and earthworms. Seinhorst (1988) compared densities of *Rotylenchus uniformis* in pairs of soil samples, one consisting of 60 cores ($\varnothing = 1$ cm), the other of 5 cores ($\varnothing = 5$ cm) and found a clear linear relationship between observed densities. Lal et al. (1981) compared different sizes of samplers for collection of Enchytraeidae in tropical soils. This study found an optimum value of 6.5 cm diameter sampler, yielding high numbers, a low co-

efficient of variation and the standard error of mean to mean. Common practice at the Sub-Department of Soil Quality (Wageningen University) is to use tubes of 2–3 cm for nematodes, 6 cm for mesofauna and cubes with a 25 cm edge for earthworms. The topic of finding the optimum support size is not pursued further in this paper not because it is unimportant, but because it falls outside the scope of this paper. From now on, it is supposed therefore that data are collected at the optimal support size.

3. Data collection

Data are located in space D . Therefore spatial variables as $Z(x)$ should be formulated, where Z denotes the variable and $x \in D$ the location. It is sometimes assumed that Z follows a specific distribution, like the normal distribution. Observations are denoted by $z(x_1), \dots, z(x_n)$. A sampling design, denoted by S_n , consists of the collection of data locations x_1, \dots, x_n .

As a first step, one may consider sampling if resources are limited, and observations from a few locations are to be collected. Several design issues have been considered, such as grid sampling, random sampling, clustered sampling, random transects and multistage sampling. If appropriate, stratification of an area may be done beforehand, requiring sampling to be repeated within each stratum. Several considerations can be made to choose the optimal design. In practice, this usually depends upon the amount of prior information, like available data, boundaries on the area and priority setting by the researcher. A recent criterion is to distribute the data as evenly as possible over an area of an arbitrary size and shape (Van Groenigen and Stein, 1998). A scheme further develops when collected observations are used, when knowledge of the system is used or when other preferences are utilized. This section first considers how to optimize a design with a given optimization criterion.

3.1. Survey designs

Survey designs are characterized by collecting data in such a way that a quantitative objective is satisfied. Here, a criterion denoted as $\phi(A|S)$ is considered, where dependence upon collected data is indicated by S and on the area by A . A can have an arbitrary

shape, whereas $|S|$, the number of elements of S , is usually determined in advance. A criterion is estimated by the fitness function $\hat{\phi}(A|S)$, being an estimator of $E(\phi(A|S))$. To evaluate it, the expectation is replaced by the average value for a set of N_e evaluation points:

$$E(\phi(A|S)) = \sum_{i=1}^{N_e} \hat{\phi}(x_i|S) \quad (1)$$

For the ordinary kriging standard deviation s_{OK} as a measure of map quality see Eq. (10) below, $\phi(A|S)$ equals the average kriging standard deviation in the area A given a sample S , and $\hat{\phi}(x_i|S)$ equals the originating standard deviation see Eq. (10) below, evaluated in N_e test locations. A criterion applied in this study is the Warrick and Myers (1987) criterion (WM-criterion), which distributes the $n(n-1)/2$ pairs of sampling points evenly over the distance classes for variogram calculations:

$$\sum_{i=1}^{n_c} (\zeta_i^* - \zeta_i)^2 \quad (2)$$

where n_c is the number of distance classes for the variogram, ζ_i^* is the required number of point pairs in the i th distance class and ζ_i is the realized number of point pairs in this distance class. The WM-criterion solely depends upon the distances between sampling points. Therefore, the fitness function can be directly calculated according to Eq. (1). The ideal distribution is defined as an even distribution of point pairs over the chosen distance classes. Below, another objective will be studied where the kriging variance will be minimized.

3.2. Adaptive sampling

Adaptive sampling (Thompson and Seber, 1996) is a relatively new sampling strategy where information gathered during data collection is essentially being used. An adaptive sampling scheme is not fully designed in advance, but evolves during the survey.

The design itself consists of two steps:

1. The sampling area A is divided into a set of contiguous blocks A_i , $i = 1, \dots, N$ of the same size. Each block is divided into an equal number of non-overlapping plots u_{ij} , $j = 1, \dots, M$. The neighborhood of a plot ∂u_{ij} is a contiguous

set of surrounding plots, consisting of the plot itself plus its four adjacent plots, depending only upon physical proximity. A numerical criterion \mathcal{C} is formulated that a plot should obey. If \mathcal{C} is fulfilled for a particular plot, then the four adjacent neighbors of the plot are sampled as well. To select the initial systematic sample, a plot is selected randomly in A_1 , the first block. Its location in the block defines the systematically arranged plots in all other blocks. This set of plots forms a primary unit. More than one primary unit may be randomly selected. Systematic sampling is applied, yielding an initial set of M sampling units (plots).

2. If the condition in step 1 is fulfilled, then neighboring plots are sampled as well. Secondary units (plots) are the sampling units included in the sampling after selection of a primary unit. Again the quantitative criterion is evaluated. If it is met, then further neighboring plots are added until no additional plots fulfill the criteria.

A major difference with conventional designs is therefore that the procedure for selection in adaptive designs depends upon population values observed in the field. Under the design, a network is defined as a set of one or more adjacent plots, all obeying the criterion \mathcal{C} . One feature of adaptive sampling is that the sample size is not known beforehand.

Adaptive sampling is known in photogrammetry as progressive sampling (Makarovic, 1973), where its main application is on elevation data. Starting from a grid spacing, the grid is sequentially densified (usually by a factor of 2) at positions where the surface deviates from a plane. To decide whether this occurs, the already calculated heights at the neighborhood of the mesh are used. This is essentially a hierarchical approach.

Three unbiased estimators for the mean of the variable and their variances will be used and compared in this study:

- The Hansen–Hurwitz (HH) estimator, which divides the observed values (the $Z(x)$ value) by their associated selection probability and multiplies it by the number of times a plot is selected. It is an unbiased estimator of the population mean, based on partial selection probabilities (Cochran, 1977; Thompson and Seber, 1996).

- The Horvitz–Thompson (HT) estimator, which uses partial inclusion probabilities of plots included in the sample. In adaptive cluster sampling however the inclusion probabilities of plots are unknown because the number of primary units not intersecting the group hit by plot u_{ij} but intersecting one or more plots fulfilling \mathcal{C} in its neighborhood may be unknown.
- An unbiased estimator of the population mean (μ) is the initial sample mean, obtained by ignoring all adaptively added plots to the sample. It offers the basis for comparison with adaptive and non-adaptive strategies.

More details are given in Thompson and Seber (1996). A case study to assess number of rare trees in a forest reserve is given in Acharya et al. (2000), and a case study to investigate animal abundance in a Kenyan wildlife reserve is given in Khaemba and Stein (in press).

3.3. Modeling spatial dependence

To model spatial dependence, a common way is to treat the data as realizations of regionalized variables. Spatial variability is expressed by covariance functions or variograms. The covariance function $C_Z(h)$ as a function of the distance h is defined as the covariance of the variable $Z(x)$ at two locations separated by h , i.e.

$$C_Z(h) = E[Z(x)Z(x+h)] - \mu_Z^2 \quad (3)$$

where μ_Z is the expected value of $Z(x)$. The variogram is defined as half the variance of the differences of $Z(x)$, as a function of the distance h , i.e.

$$\gamma_Z(h) = \frac{1}{2} E(Z(x) - Z(x+h))^2 \quad (4)$$

The variogram is estimated as half the average of the squared differences of all pairs of points that are separated by approximately the distance h (Burgess et al., 1981; McBratney and Webster, 1981):

$$\gamma_Z(h) = \frac{1}{2N_Z(h)} \sum_{i=1}^{N_Z(h)} (z(x_i) - z(x_i+h))^2 \quad (5)$$

where $z(x_i)$ and $z(x_i+h)$ denotes the i th pair of observations on $Z(x)$ separated by distance h , and $N_Z(h)$ is the total numbers of such pairs. Through the empirical variogram values thus obtained a variogram model

is fitted, like an exponential or a spherical model. For that purpose, pairs of points are then grouped into distance classes. Each distance class contains ζ_j pairs of points, with $j = 1, \dots, n_c$, n_c being the number of distance classes. The parameters of a linear or a non-linear variogram model are fitted, using a regression procedure. It has been advocated (Cressie, 1991) to use a weighted least squares fit with weights equal to ζ_j . With either a regular sampling scheme or a random scheme, the number of pairs of points may differ for each distance class. This has an effect on the precision with which the parameters are determined.

3.4. Experimental designs

In an experimental design a set of k treatments T_1, \dots, T_k is considered. These could be actual treatments, applied by a researcher to parts of a field, like the amount of a fertilizer, different management strategies like type of management or land use system, or natural phenomena like soil units or watersheds. These factors may have an effect on the value or the variability of $Z(x)$. For each location, we may then consider the i th treatment T_i . Application of treatments introduces non-stationarity on the data, as those locations that received the same treatment may have the same expectation μ_i , which may differ from that at locations with another treatment, say $T_{i'}$, having expectation $\mu_{i'}$. This leads to the general model:

$$Z(x) = \mu_i + \epsilon(x) + e(x) \quad (6)$$

where $E[\epsilon(x)] = E[e] = 0$, $\text{Cov}[\epsilon(x), \epsilon(x+h)] = C(h)$, the covariance function between the data and $\text{Cov}[e(x), e(x+h)] = 0$. The term $\epsilon(x)$ denotes the spatially dependent part of the error and $e(x)$ the spatially independent part. In experimental designs interest focuses on differences in average effects of particular combinations of levels of factors, their interactions and the significance of these differences.

Treatments can be applied to areas of land or fields, the so-called strata. Observations collected from within these strata are then well-known to be potentially different. In addition to the treatments therefore we have to consider spatial variation, i.e. to acknowledge the fact that observations close to each other are likely to be more similar than observations at a larger distance.

To investigate the significance of the differences in mean values between strata when the observations of a regionalized variable are (spatially) related, suppose k strata are investigated, and that for every stratum it is known that the spatial dependency structure is given by the variogram $\gamma_Z^{(i)}(h)$ for $h \geq 0$, $i = 1, \dots, k$. An estimator for the mean and the variance within the i th stratum equals:

$$\hat{\mu}_i = \frac{1'_n \Gamma_Z^{(i)-1} z}{1'_n \Gamma_Z^{(i)-1} 1_n} \quad (7)$$

where the matrix $\Gamma_Z^{(i)}$ contains values of the variogram between observations in the i th stratum, 1_n is a vector of n elements, all equal to 1, z is the vector with the n observations and the prime denotes the transpose of a vector. The variance of the mean is equal to

$$\text{var}(\hat{\mu}_i) = \frac{1}{1'_n \Gamma_Z^{(i)-1} 1_n} =: \frac{1}{\gamma_i} \quad (8)$$

The null hypothesis H_0 that no differences exist between the different strata and the alternative hypothesis H_1 can be formulated as:

- H_0 : $\mu_1 = \mu_2 = \dots = \mu_k$.
- H_1 : at least one μ_j differs from the other μ_i 's, $i \neq j$.

When the spatial structure is known, this hypothesis is tested with the test statistic:

$$T = \sum_{i=1}^k \gamma_i \hat{\mu}_i^2 - \frac{\left\{ \sum_{i=1}^k \gamma_i \hat{\mu}_i \right\}^2}{\sum_{i=1}^k \gamma_i} \quad (9)$$

which under H_0 has a χ^2 -distribution with $k - 1$ degrees of freedom. A similar expression holds if the variogram is used instead of the covariance function (Stein et al., 1988). Of course, in practical studies the spatial structure has to be estimated from the data. As the test value will only slightly change, the same χ^2 -distribution can be used.

3.5. Optimal sampling

The variograms between current and future observation points can be used to calculate the optimal grid spacing for sampling, necessary to monitor each variable in a regular grid to achieve a predetermined standard deviation of prediction error, called in this

paper the level of accuracy. The accuracy, in the same units as the variables, is defined as the maximum uncertainty occurring on an interpolated map (McBratney and Webster, 1983). It equals the kriging standard deviation in the most isolated location of prediction, i.e. the center point of a square grid cell with observations in each of the corners of this cell. The kriging standard deviation is

$$s_{OK} = \sqrt{\gamma'_0 \Gamma^{-1} \gamma_0 - x_a V x_a} \quad (10)$$

where γ_0 equals the vector with variogram values between observation location and the prediction location, Γ the $n \times n$ matrix with variogram values among the observation location, 1_n the vector of n elements all equal to 1, $x_a = 1 - \gamma'_0 \Gamma^{-1} 1_n$ and $V = (1'_n \Gamma^{-1} 1_n)^{-1}$. One may notice that Eq. (10) does not depend upon the vector of observations. By using its dependence upon the variogram, it may be used to optimize the configuration of the sampling scheme. Variograms with a nugget value have a minimum predetermined accuracy. An accuracy lower than the square root of the nugget value for these variables cannot be achieved by reducing the sampling spacing, but only by improving the measuring procedure. For monitoring purposes, i.e. sampling in space and time, it is interesting to compare optimal grid sizes of the same variable between the sample domains.

In the case study below, an iterative computer program was used to calculate the optimal grid spacing given a predetermined accuracy of each variable in both a triangular and a square grid. Starting with the largest spacing, grid spacing is reduced until the required accuracy is reached.

4. Simulated data

To explore and illustrate the different sampling procedures a field of 100×100 pixels was simulated using a sequential Gaussian routine (Deutsch and Journel, 1998). The left half of the field was generated using a spherical variogram, with a nugget parameter equal to 0.1, range parameter equal to 10 m and sill equal to 0.9. It was an unconditional simulation, with a mean equal to 5. The right side of the field was simulated with an exponential variogram with nugget equal to 0.5, range equal to 10.0 m and sill equal to 2.0 (Fig. 1). The mean of this part of the field was equal to 10.

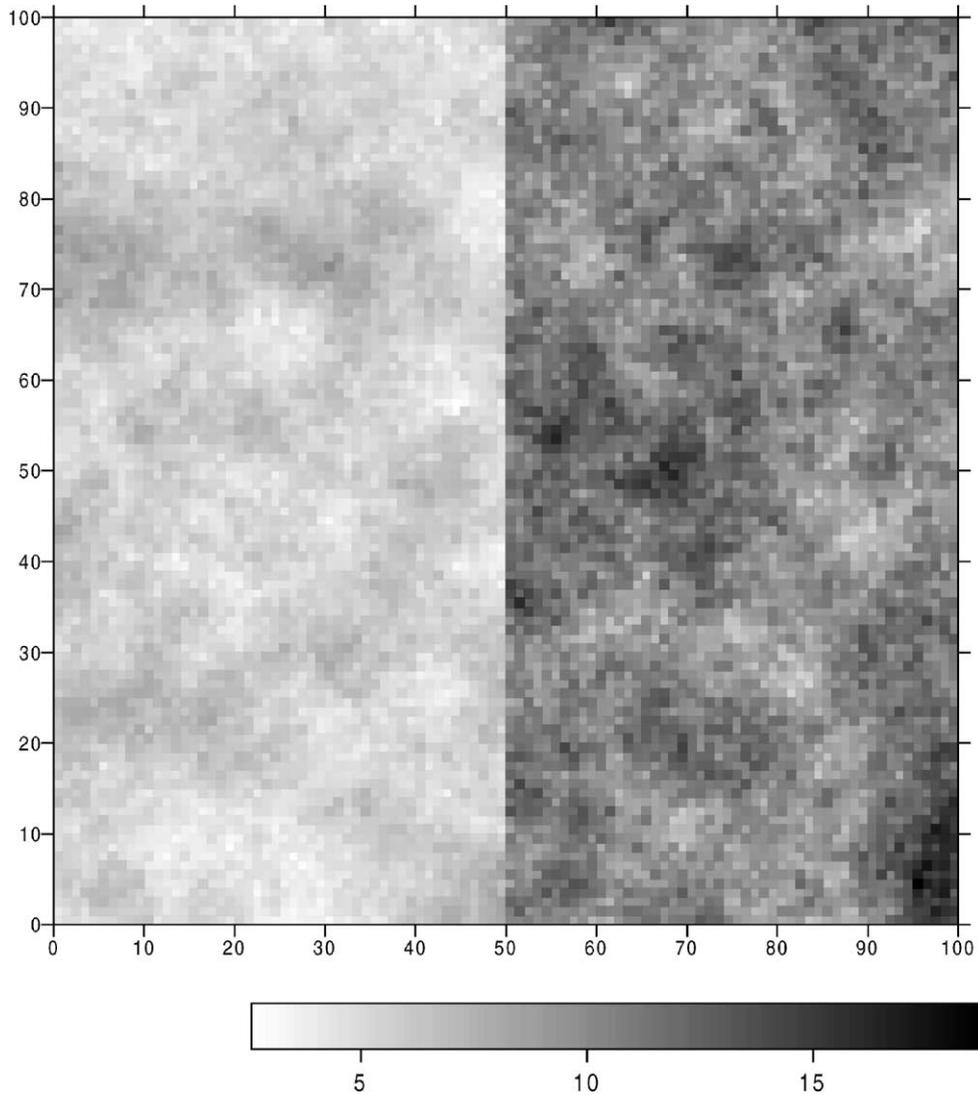


Fig. 1. A simulated field. The left half follows a spherical variogram model with nugget equal to 0.1 and sill equal to 0.9, and the right half an exponential model with nugget equal to 0.5 and sill equal to 2.0. Ranges are the same (10 m), the mean value at the left is 5 and at the right it is 10.

4.1. Optimal survey

As the first stage the effects of various sample sizes, i.e. of various distances between sample points, was investigated, using Eq. (10) (Fig. 2). Here, the distance between sampling points on a triangular grid was plotted as a function of the required precision as expressed by the kriging variance. This graph is read as follows: if the precision is set to 1, then a sampling distance

for the field in the left side of Fig. 1 equal to 12 m is required, whereas on the right side of Fig. 1 this would be equal to approximately 4 m. It was further noticed from Fig. 2 that at the left part of the area no precision (kriging standard deviation) below 0.1 can be reached, a consequence of the size of the nugget effect. Also, a precision equal to 1.05 and more is always obtained, even with a single observation, irrespective of the sampling plan. In the right side of the

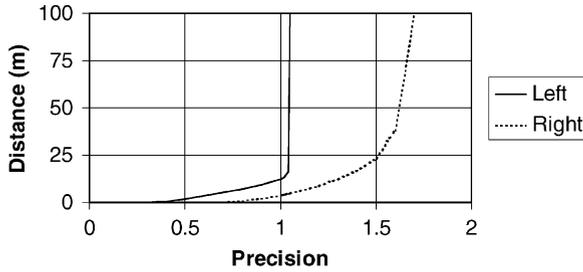


Fig. 2. Distance between sampling points as a function of the precision obtained by ordinary kriging.

field, with the larger variability, a precision (kriging standard deviation) below 0.7 can never be obtained.

Taking the sampling plan for a precision equal to 1, Fig. 3 emerges. The left part of the field, with little variation, requires 45 points and the right part of the field with high spatial variation requires 377 points. A possible pattern to locate these points is given in Fig. 3, but is already evident that spreading can be done much more effectively, when field boundaries are carefully considered.

4.2. Optimal spreading

In the next stage an optimization stage in the sampling carried out above was considered, using

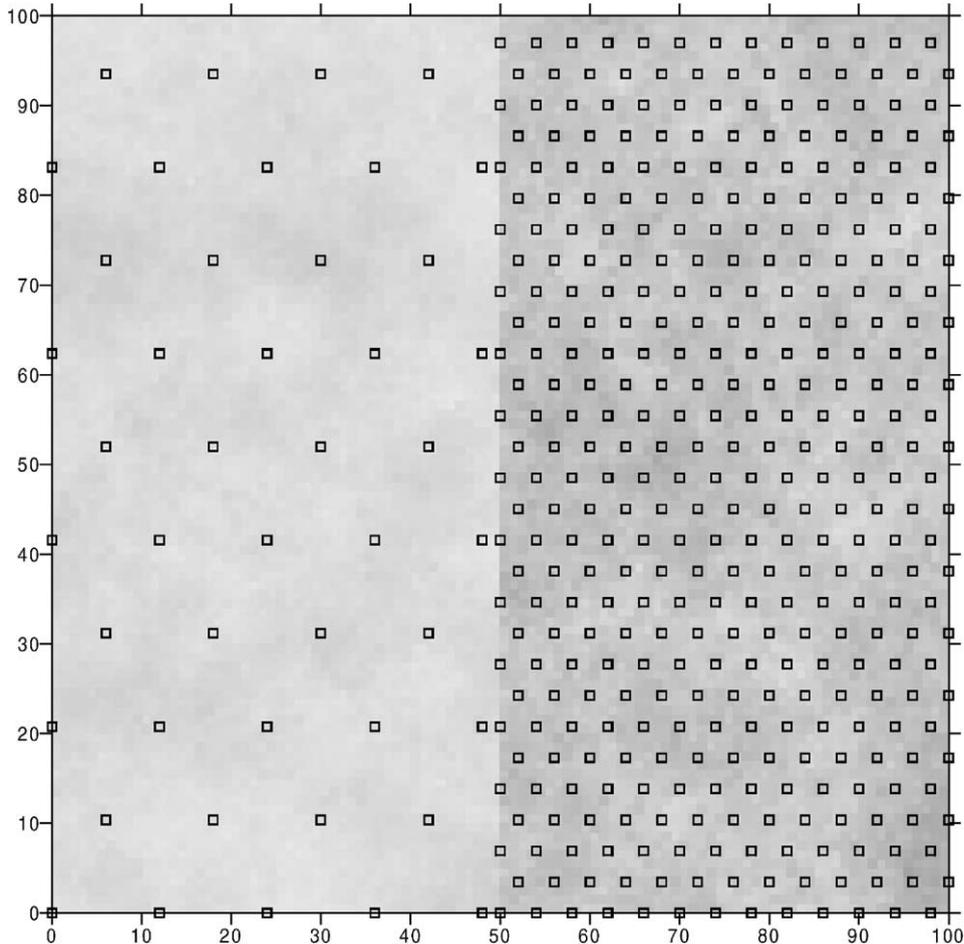


Fig. 3. Optimized sampling plan following a triangular grid. The resulting map will be of the uniform precision that the kriging variance is at most 1.

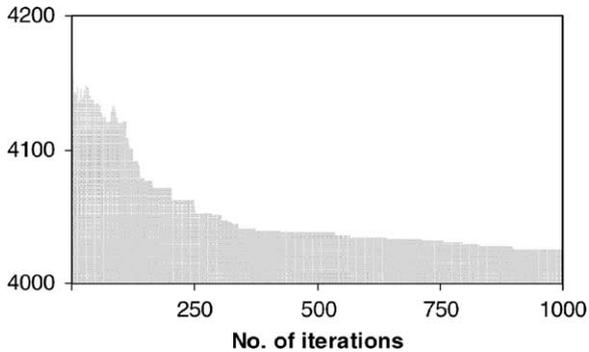


Fig. 4. Development of the objective function during optimization of the triangular grid.

Eq. (1). For computational purposes the calculation is restricted to the left part of the field only, where 45 points were allocated. Optimality is defined in terms of the average kriging standard deviation in the 5000 grid cells that make up this part of the field, i.e.

$$\hat{\phi} = \sum_{i=1}^{5000} s_{OK}^2(x_i|S) \quad (11)$$

where $s_{OK}^2(x_i|S)$ equals the kriging variance in the center of a grid cell. A standard simulated annealing procedure was carried out, with the starting temperature set equal to 1, the cooling down factor equal to 0.9 and the number of steps equal to 1000 (Van Groenigen and Stein, 1998). Results for the objective function are given in Fig. 4, showing that the solution is close to optimal. The resulting sampling grid is shown in Fig. 5, which illustrates that the observation points have moved away from the field edges, that a pattern emerges that is close to a triangular scheme, but that the scheme is not very regular. This is mainly due to the size of the range parameter (10 m), smaller than the required distance between the sampling points. Many sampling schemes therefore lead to the same value of the optimality criterion.

4.3. Testing for significance

Testing for significance in the presence of spatial dependence was done with Eq. (9), using a randomly collected set of 50 data from each side of the field. The spatial means and standard deviations were found equal to 4.396 and 0.9248 for the left side and 10.374

and 2.311 for the right side, respectively. The test statistic T was equal to 24.466, following a χ_2^2 distribution. This is highly significant ($\text{Prob}(T > 10.60) < 0.005$).

As a comparison, an ordinary t -testing was applied to the same data, i.e. without incorporating spatial dependencies. Means and standard deviations were equal to 5.289 and 0.997, respectively, for the left part of the field and 10.374 and 2.218, respectively, for the right part of the field, showing minor differences with the previous values. A test value of 14.784 was found that was again highly significant ($p < 0.001$). Therefore, in this case, testing with inclusion of dependence gives a similar result as testing with independent data. This is not always likely to be the case.

4.4. Adaptive sampling

The final part of this example concerns looking data that exceed the value of 15. This limit was chosen arbitrarily, but could be motivated by a consideration that values above the mean and twice the standard deviation were to be identified, being spatial outliers started with distributing 100 data over the area following a regular grid with an arbitrary origin in the 10×10 lower left sub-grid. Whenever a value exceeding 15 was observed, the immediate neighbors to four sides were sampled as well. The sampling scheme is shown in Fig. 6. The original sampling plan had to be extended by 66 additional samples. The scheme hits the cluster of high values in the lower right of the figure, whereas the smaller clusters in the center are missed. Notice that the final number of samples cannot be determined beforehand. Increasing the detection level from 15 to, say, 16, would lead to a larger probability of missing the cluster, whereas decreasing the level from 15 down to 14 would lead to possibly larger samples.

5. Case study: experimental and spatial sampling design for a soil bio-diversity study

5.1. Methods

5.1.1. Study site and treatments

The case study to be reported here was set up in 1999 to study the relation between above- and be-

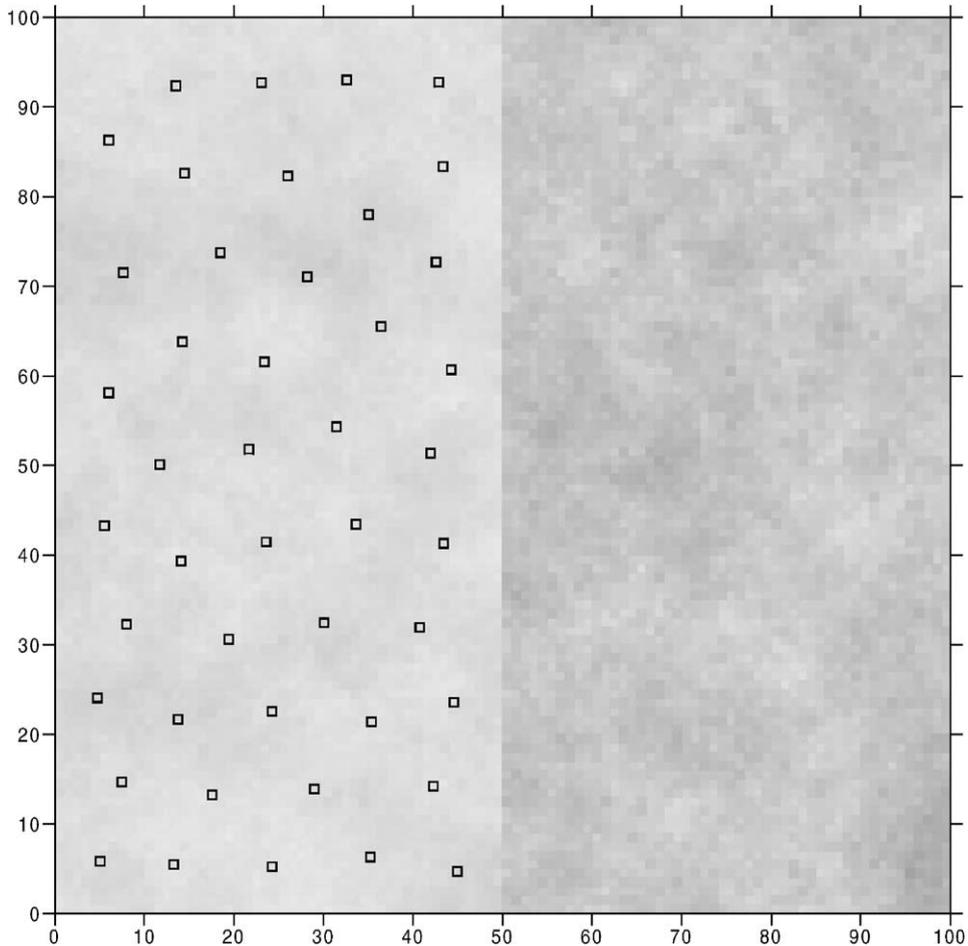


Fig. 5. The triangular grid after optimization according to minimizing the average kriging variance in the cells.

lowground bio-diversity. A practical aim was to quantify and compare spatial effects of plant diversity on soil resource heterogeneity, soil bio-diversity patterns and nutrient cycling rates. The main hypothesis is that plants, at the plant-to-plot scale (ca. $0.1\text{--}25\text{ m}^2$), are the primary architects of belowground diversity and function, by increasing the spatial heterogeneity of soil resources that can be exploited by soil biota. If the hypothesis is not rejected, then these effects are expected to be larger with increasing plant diversity and to accumulate over time as the vegetation develops.

To test these hypotheses, a field trial was set up, which combined a spatial sampling design with a

randomized block design with three plant diversity treatments, including observations on spatial patterns of diverse soil biota and soil resources. The field trial was conducted in the “Bovenbuurtse Weilanden”, located between Wageningen and Bennekom, The Netherlands. In September 1999, an arable site of $27 \times 27\text{ m}^2$ was cleared of vegetation (the last crop was potato, *Solanum tuberosum* L.) and rotor-tilled, so that evolution of spatial patterns, and divergence between plant diversity treatments, may be tracked from an (as much as possible) homogeneous starting point. The procedure is to be repeated over three growing seasons. In this paper the spatial arrangement of the observation points is discussed.

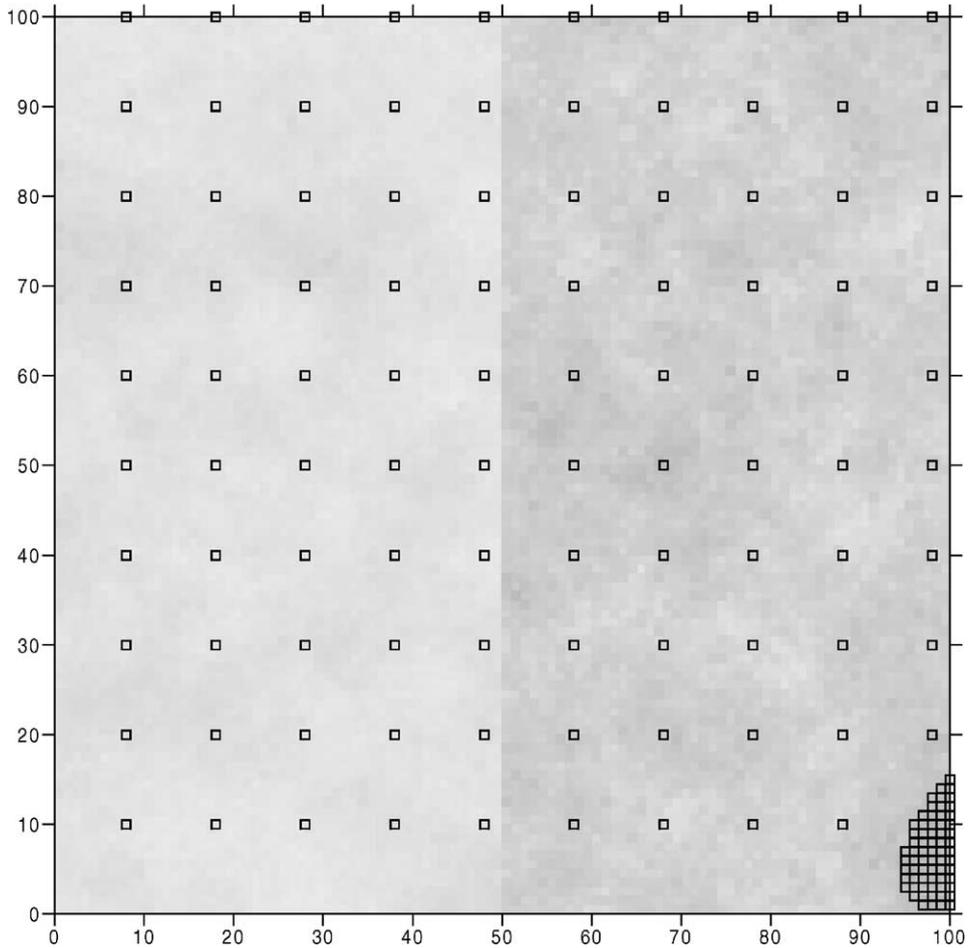


Fig. 6. An adaptive scheme put on the random field.

The area was divided into nine $9 \times 9 \text{ m}^2$ plots, which were randomly assigned to three treatments: T_1 —no vegetation, T_2 —monospecific grassland (*Agrostis capillaris*) and T_3 —polyspecific grassland (a mixture of 15 grass and herbaceous species). The first sampling event took place in September 2000, 1 year after sowing treatments T_2 and T_3 .

5.1.2. Spatial sampling design

The primary objective of the first sampling was to estimate experimental variogram parameters, to determine the scale of patchiness in the distributions of the variables of interest (note that this information can be used at later sampling dates, when the primary

objective may switch to reducing kriging variance). Based on logistical and financial limitations, the maximum number of spatial samples to be analyzed for soil biota and soil resources, per (yearly) sampling event, was set to 252, viz. 28 samples in each of nine ($9 \times 9 \text{ m}^2$) plots. Considering the interest in small-scale variation and the rule of thumb that at least 30 point pairs per distance class are needed to adequately estimate semivariance (with the variogram covering up to 50% of the maximum separation distance), for 10 distance classes of 0.5 m each were chosen (i.e. 0.0–0.5 m, ..., 4.5–5.0 m) and a 11th class for all pairs separated by a distance $>5.0 \text{ m}$. Thus, to obtain an even distribution of point pairs

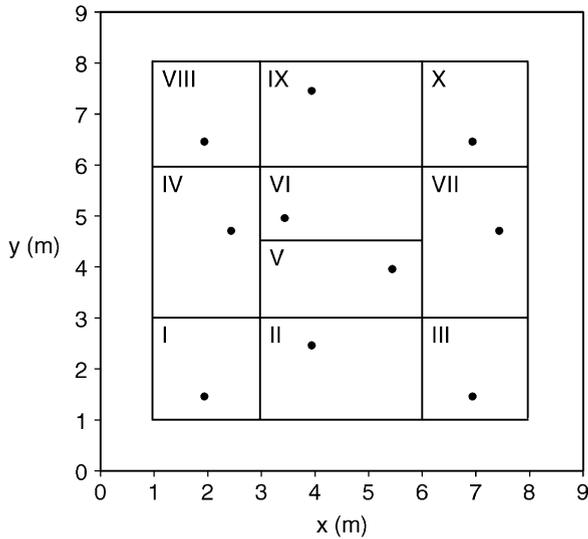


Fig. 7. Each plot is divided into 10 sub-plots. In each sub-plot, one point was randomly chosen and kept fixed during the optimization of the 18 remaining, variable points, to ensure adequate coverage of the sampling area.

over the chosen distance classes, ζ_i^* in Eq. (2) equals $28 \times (27/2)/11 = 34.4$.

A straightforward application of the WM-criterion leads to extreme clustering of sampling points (Van Groenigen and Stein, 1998), resulting in large kriging variances. Although kriging was not a primary objective of the first sampling event, the possibility of kriging was kept open by choosing 10 random points in each of 10 strata (Fig. 7) and keeping those fixed during optimization.

The rules for selecting the 10 fixed points (Fig. 7) are as follows:

1. Always stay 1 m from the edge of the plot.
2. Choose point (x, y) in sub-plot I at random. Points in similarly dimensioned sub-plots III, VIII and X are selected at the same relative position.
3. Choose a point (x, y) in sub-plot II at random. A point in the similarly dimensioned sub-plot IX is selected at the same relative position.
4. Choose a point (x, y) in sub-plot IV at random. A point in the similarly dimensioned sub-plot VII is selected at the same relative position.
5. Choose a point (x, y) in sub-plot V at random. A point in the similarly dimensioned sub-plot VI is selected at the same relative position.

selected upon reflection with respect to the middle point of the dividing line between the two sub-plots.

6. As a pragmatic solution to facilitate localization of points in the field during sampling, coordinates of the (x, y) points are rounded to the nearest multiple of 0.25 m.

To optimize the locations of the 18 remaining points, spatial simulated annealing (SSA) (Van

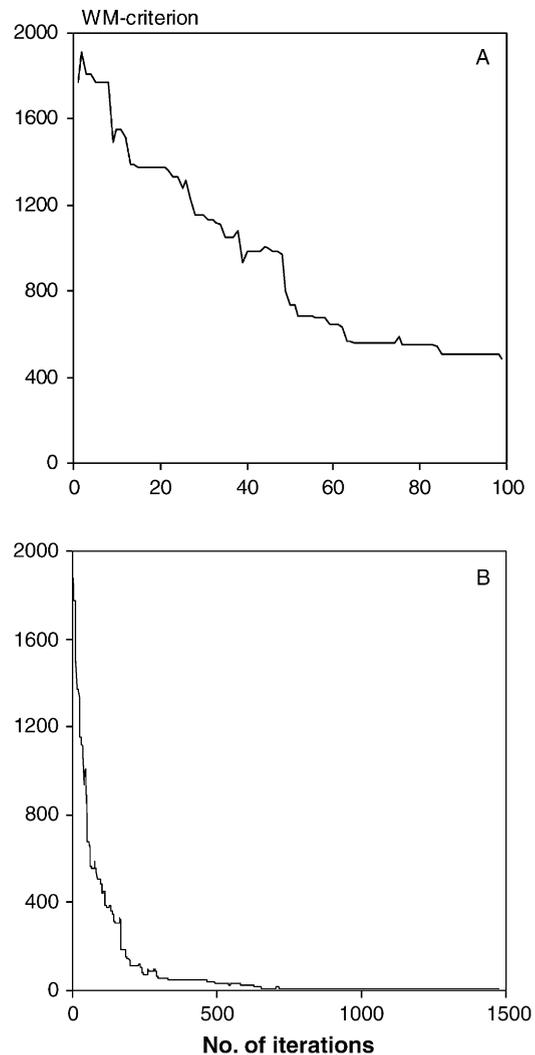


Fig. 8. Development of the WM-optimization criterion of the sampling scheme for the bare plot, during the first 100 iterations (A) and the first 1600 iterations (B).

Groenigen and Stein, 1998) was used, utilizing the WM-criterion (Eq. (2)). Starting with a sampling scheme of 18 randomly chosen points and the 10 fixed points selected above, the optimization algorithm generated random distortions in the scheme by randomly moving one of the 18 variable points. The two constraints were that the new location should fall within the $5 \times 5 \text{ m}^2$ center of the plot, and that it should be at least 0.20 m from other points. The latter constraint was introduced to allow enough room for soil coring in the field. After generating the random distortion, the quality of the resulting scheme was evaluated using the WM-criterion (Eq. (2)). Improve-

ments in the sampling scheme (lower WM-criterion, more even point pair distribution) were accepted, but deteriorations were sometimes allowed to avoid local minima. The latter event had a decreasing probability as the optimization advanced (the so-called cooling schedule, see Van Groenigen and Stein, 1998).

For the vegetated plots introduced, a third constraint was introduced to allow for later analysis of a possible problem arising from the crop grown in the area before the field experiment was started. In spring 2000, the sown plots (T_2 and T_3) showed alternating rows of better and lesser growth, which was interfering with the experimental objectives. After soil

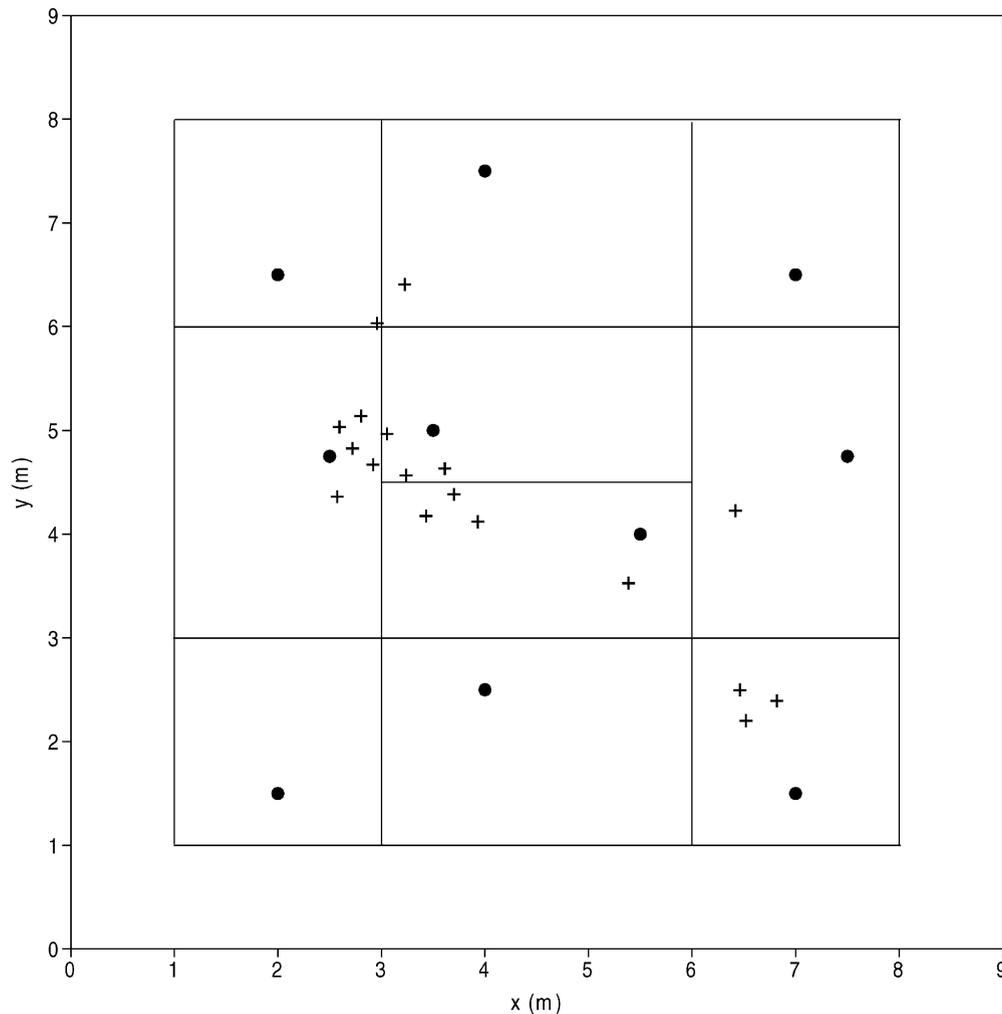


Fig. 9. The optimized sampling scheme for the bare plot, with 10 fixed points (●) and 18 variable points (+).

analysis, it was concluded that soil tillage had failed to adequately homogenize soil nutrient levels, with the patterns reflecting the potato rows of last year's crop. After addition of 50 kg N and 50 kg P ha⁻¹, and repeated mowing, the row pattern had disappeared by the time of the first spatial soil sampling in September 2000. However, to allow detection of possible remaining effects below ground, a final constraint on the sampling scheme was introduced to locate 14 sampling points into the potato rows of the previous year and 14 into spaces between these rows.

5.2. Results

5.2.1. Bare plots

In the bare plots (T_1) obviously no row pattern could be observed. Therefore only the first two constraints (minimum distance between points 0.20 m, points located within the 5×5 m² center) were applied in the optimization of the sampling scheme. During the optimization, local minima were avoided as inferior solutions were regularly accepted (Fig. 8A). Within 1500 iterations, the WM-criterion was minimized to 2.55 (Fig. 8B) (it could not be minimized to zero because $\zeta_i^* = 34.4$, i.e. not an integer). The resulting point pairs had an even distribution over the distance classes (Table 1), and the resulting sampling scheme was clearly clustered (Fig. 9). This sampling scheme was applied to each of the three replicate bare plots.

Table 1

Point pair distributions, WM-criterion and minimum distance in sampling schemes of example plots

Distance class (m)	Bare plot (Fig. 8)	Vegetated plot (Fig. 10)
0.0–0.5	34	34
0.5–1.0	34	34
1.0–1.5	35	34
1.5–2.0	34	34
2.0–2.5	34	35
2.5–3.0	35	34
3.0–3.5	35	34
3.5–4.0	35	35
4.0–4.5	34	35
4.5–5.0	35	35
>5.0	34	35
WM-criterion	2.545	2.545
Minimum distance (m)	0.233	0.209

5.2.2. Vegetated plots

In all vegetated plots (T_2 and T_3) row patterns were observed in spring 2000. Before this pattern was removed using fertilization and mowing (see Section 5.1), the beginning and end of each row was mapped. The coordinates and the average slope coefficient of the row lines were the data input for the constraint that half of the 28 sampling points should fall into former better rows and half should fall into former lesser rows. As the row patterns had different coordinates in each plot, optimization was carried out for each plot separately. Here we present the results for one plot only.

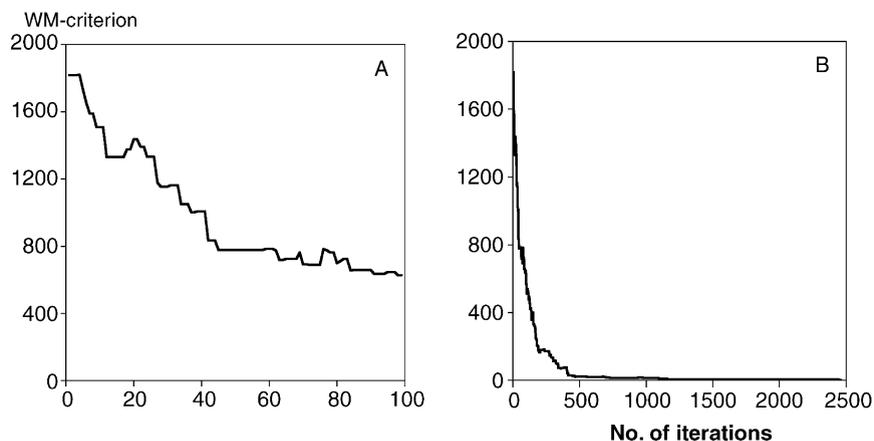


Fig. 10. Development of the WM-optimization criterion of the sampling scheme for the vegetated plot, during the first 100 iterations (A) and the first 2500 iterations (B).

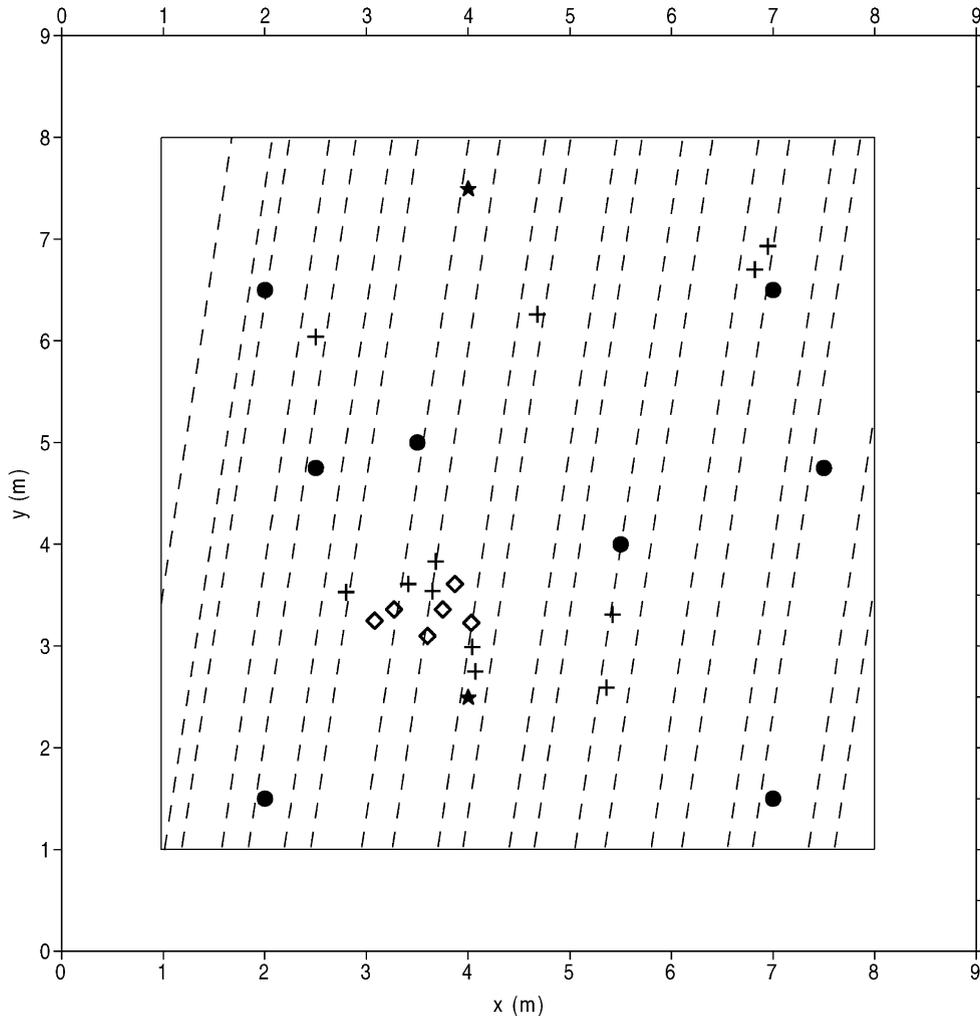


Fig. 11. The optimized sampling scheme for the vegetated plot. The dotted lines indicate the borders of the former potato rows, in which better and lesser growth patterns were observed in spring 2000. At the time of sampling, the row pattern had disappeared (see Section 5.1). Symbols indicate: eight fixed points in lesser rows (●), two fixed points in better rows (★), 12 variable points in lesser rows (+) and six variable points in better rows (◇).

Similarly to the bare plot optimization, local minima were avoided as inferior solutions were regularly accepted (Fig. 10A). Within 2500 iterations, the WM-criterion was minimized to 2.55 (Fig. 10B). The resulting point pairs have an even distribution over the distance classes (Table 1), and the resulting sampling scheme is clustered similarly to the bare plot (Fig. 11 versus Fig. 9). At the start of the optimization, it appeared that of the 10 fixed points, eight were positioned in a lesser growth row and two in a better growth row (Fig. 11). Therefore, the 18 vari-

able points were optimized such that six were placed in lesser growth rows and 12 in better growth rows, resulting in a sampling scheme with an even (14:14) distribution of the 28 points over the lesser and better growth rows, respectively.

5.3. Discussion

The results show that SSA is very effective in designing sampling schemes for variogram estimation. In the two examples, a complete solution was

found for the WM-criterion. However, it must be noted that the optimization algorithm is very sensitive to the initial random sampling scheme, such that several runs (of thousands of iterations each) were needed to come to a complete solution. For some of the plots, it was not possible to come to a complete solution even after many trial runs. In those cases, solutions with a WM-criterion <10.55 were accepted, as long as there were at least 30 point pairs in each distance class. In effect, the number of point pairs per distance class always ranged between 32 and 36.

The results for the vegetated plots also illustrate that SSA is especially useful for designing schemes with complex sampling constraints, such as the requirement for an even point distribution over former lesser and better rows in the vegetated plots.

6. Conclusions

The conclusions of this review indicate that at present a mosaic of spatial sampling procedures for ecological data is available. Several were distinguished in this study: (1) survey designs, (2) experimental designs and (3) adaptive designs. All procedures allow collection of spatial data, where some of procedures also allow to include prior information and sampling constraints. The main differences concern the proper choice of an optimization criterion for (1), acknowledgment of treatment effects in (2) and the use of collected information during sampling in (3). The example showed how also subjective information was included into spatial sampling, where artificial lines in the field were to be honored. In addition, local differences in the field could easily be overcome, by developing the proper objective of sampling.

Acknowledgements

This research was carried out as part of the project Soil Heterogeneity Under Developing Grassland, with a grant from the Royal Dutch Academy of Sciences. Several discussions with Dr. Brussaard and collaborators of the Sub-Department of Soil Quality of

Wageningen University were useful during the field work as described in this paper.

References

- Acharya, B., Bhattarai, G., De Gier, A., Stein, A., 2000. Systematic adaptive cluster sampling for the assessment of rare tree species in Nepal. *For. Ecol. Manage.* 137, 65–73.
- Anderson, J., 1995. Soil organisms as engineers: micro-site modulation of macro-scale processes. In: Jones, C., Lawton, J. (Eds.), *Linking Species and Ecosystems*. Chapman & Hall, New York, pp. 94–106.
- Burgess, T.M., Webster, R., McBratney, A.B., 1981. Optimal interpolation and isarithmic mapping of soil properties IV. Sampling strategy. *J. Soil Sci.* 32, 643–659.
- Cochran, W.G., 1977. *Sampling Techniques*, 3rd Edition. Wiley, New York.
- Coleman, D., Crossley, D.A., 1996. *Fundamentals of Soil Ecology*. Academic Press, San Diego, CA.
- Cressie, N.A.C., 1991. *Statistics for Spatial Data*. Wiley, New York.
- De Gruijter, J.J., Ter Braak, C.J.F., 1990. Model free estimation from spatial samples: a reappraisal of classical sampling theory. *Math. Geol.* 22, 407–415.
- Deutsch, C.V., Journel, A.G., 1998. *GSLIB: Geostatistical Software Library and User's Guide*, 2nd Edition. Oxford University Press, New York.
- Ettema, C., Rathbun, S., Coleman, D., 2000. Spatiotemporal patchiness and the coexistence of five species of *Chronogaster* (Nematoda: Chronogasteridae) in a riparian wetland. *Oecologia* 125, 444–452.
- Giller, P., 1996. The diversity of soil communities, the poor man's tropical rainforest. *Biodivers. Conserv.* 5, 135–168.
- Hendriks, L.A.M., Leumanns, H., Stein, A., De Bruijn, P.J., 1998. Use of soft data in a GIS to improve estimation procedures for environmental purposes. *Water Air Soil Poll.* 101, 217–234.
- Khaemba, W., Stein, A., in press. Improved sampling for airborne surveys to estimate wildlife population parameters in the African Savannah. *Wildlife Research*.
- Lal, V.B., Singh, J., Prasad, B., 1981. A comparison of different size of samples for collection of Enchytraeidae (Oligochaeta) in tropical soils. *J. Soil Biol. Ecol.* 1, 21–26.
- Makarovic, B., 1973. Progressive sampling for digital terrain models. *ITC J.* 3, 397–416.
- McBratney, A.B., Webster, R., 1981. The design of optimal sampling schemes for local estimation and mapping of regionalized variables: 1. Theory. *Meth. Comput. Geosci.* 7, 331–334.
- McBratney, A.B., Webster, R., 1983. Optimal interpolation and isarithmic mapping of soil properties. V. Co-regionalization and multiple sampling strategies. *J. Soil Sci.* 34, 137–162.
- Mead, R., Curnow, R.N., Hasted, A.M., 1993. *Statistical Methods in Agriculture and Experimental Biology*, 2nd Edition. Chapman & Hall, London.

- Müller, W.G., 1998. Collecting Spatial Data—Optimum Design of Experiments for Random Fields. Physical-Verlag, Heidelberg.
- Särndal, C.E., Swensson, B., Wretman, J., 1992. Model Assisted Survey Sampling. Springer, New York.
- Seinhorst, J.W., 1988. The Estimation of Densities of Nematode Populations in Soil and Plants. Sveriges Lantbruksuniversitet, Växskyddsrapporter Jordbruk 51. Swedish University of Agricultural Sciences, Research Information Centre, Uppsala.
- Stein, A., Hoogerwerf, M., Bouma, J., 1988. Use of soil-map delineations to improve (co-)kriging of point data on moisture deficits. *Geoderma* 43, 163–177.
- Stuart, A., 1984. *The Ideas of Sampling*. Griffin, London.
- Thompson, S.K., Seber, G.A.F., 1996. *Adaptive Sampling*. Wiley, New York.
- Van Groenigen, J.W., Stein, A., 1998. Constrained optimization of spatial sampling using continuous simulated annealing. *J. Environ. Qual.* 27, 1078–1086.
- Warrick, A.W., Myers, D.E., 1987. Optimisation of sampling locations for variogram calculations. *Water Resour. Res.* 23, 496–500.
- Wendroth, O., Reynolds, W.D., Vieira, S.R., Reichardt, K., Wirth, S., 1997. Statistical approaches to the analysis of soil quality data. In: Gregorich, E.G., Carter, M.R. (Eds.), *Soil Quality for Crop Production and Ecosystem Health*. Elsevier, Amsterdam, pp. 247–276.
- Wollum, A.G., 1994. Soil sampling for microbiological analysis. In: Weaver, R.W., et al. (Eds.), *Methods of Soil Analysis. Part 2. Microbiological and Biochemical Properties*. Soil Science Society of America, Madison, pp. 1–14.