ELSEVIER

Geoderma 80 (1997) 1–44

GEODERMA

Discussion Paper

Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with Discussion) [1]

D.J. Brus *, J.J. de Gruijter

DLO Winand Staring Centre for Integrated Land, Soil and Water Research (SC–DLO), P.O. Box 125, 6700 AC Wageningen, The Netherlands

Received 26 October 1995; accepted 23 June 1997

Abstract

Classical sampling theory has been repeatedly identified with classical statistics which assumes that data are identically and independently distributed. This explains the switch of many soil scientists from design-based sampling strategies, based on classical sampling theory, to the model-based approach, which is based on geostatistics. However, in design-based sampling, independence has a different meaning and is determined by the sampling design, whereas in the model-based approach it is determined by the postulated model for the process studied. Design-based strategies are therefore also valid in areas with autocorrelation.

Design-based and model-based estimates of spatial means are compared in a simulation study on the basis of the design-based quality criteria. The simulated field consists of four homogeneous units that are realizations of models with different means, variances and variograms. Performance is compared for two sample sizes (140 and 1520) and two block sizes (8×6.4 km^2, 1.6×1.6 km^2). The two strategies are Stratified Simple Random Sampling combined with the Horvitz–Thompson estimator ($STSI$, t_{HT}), and Systematic Sampling combined with the block kriging predictor (SY, t_{OK}). Point estimates of spatial means by (SY, t_{OK}) were more accurate in all cases except the global mean (8×6.4 km^2 block) estimated from the small sample. In interval estimates on the other hand, p-coverages were in general better with the design-based strategy, except when the number of sample points in the block was small.

* Corresponding author. E-mail: brus@sc.dlo.nl

[1] References to the paper should include the Discussion and should be cited as: Brus, D.J., de Gruijter, J.J., 1997. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with Discussion). Geoderma 80, 1–59.
References to parts of the Discussion should be cited as, e.g., Laslett, G.M., 1997. In Discussion of: D.J. Brus and J.J. de Gruijter, Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil. Geoderma 80, 45–49.

PII S0016-7061(97)00072-4

Factors that determine the effectiveness and efficiency of the two approaches are the type of request, the interest in objective estimates, the need for separate unique estimates of the estimation variance for all points or subregions, the interest in valid and accurate estimates of the estimation or prediction variance, the quality of the model, the autocorrelation between observation and prediction points, and the sample size. These factors will be assembled in a decision-tree that can be helpful in choosing between the two approaches.

Models can also be used in the design-based approach. They describe the population itself, whereas in the model-based approach they describe the data generating processes. Errors in such models result in less accurate estimates, but the estimated accuracy is still valid.

Keywords: sampling; geostatistics; soil science; spatial interpolation

1. Introduction

The 1980s will go down in soil science history as an important decade. In this period the regionalized variable theory was introduced into soil science. Numerous papers recommended kriging, a set of interpolation techniques based on this theory, as optimal for the spatial estimation of soil properties. Examples include: Burgess and Webster (1980), Burgess et al. (1981), Webster (1985) and Oliver et al. (1989). Moreover, it was repeatedly stated that 'classical' methods of sample design and inference are inapplicable in earth sciences, because these methods assume data to be independent. As stated by Barnes (1988): "The classic development of nonparametric tolerance intervals begins with an assumption of independent, identically distributed random variables. This is unrealistic for the geologic environment — in general, geologic site characterization data are not independent."

As a result of such publications, many soil scientists seem to have abandoned the old 'classical statistics' and switched to the new 'geostatistics'. However, questions have remained about the role of bias in sampling, the quality of results when the model is misspecified (Diamond and Armstrong, 1984; Stein and Handcock, 1989), and the subjectivity of the choices made in geostatistical data analysis (Englund, 1990).

In another realm, lively discussions arose in publications on sampling theory about a fundamental issue: the choice between the so-called 'design-based' and 'model-based' approaches in sampling. Key references are: Smith (1976), Särndal (1978), Kalton (1981), O'Muircheartaigh and Wong (1981) and Hansen et al. (1983). A formal theoretical framework for both theories was presented by Cassel et al. (1977).

Although formulated in the general, non-spatial terminology of sampling theory, these discussions have everything to do with the role of randomization in sampling and the relative merits of 'classical' and 'geostatistical' methods of inference from spatial sample data. Despite its great relevance, the discussion

among survey statisticians went unnoticed by most geostatisticians, probably due to the fact that they used separate communication channels.

The purpose of the present paper is: (1) to draw the attention of soil scientists to the above discussion; (2) to disprove the assumption of invalidity of the design-based approach and to clarify the differences between the design-based approach and the model-based approach in the context of spatial sampling; (3) to compare the performance of the two approaches in a case study; and (4) to initiate a discussion on the relative merits of the two approaches in soil sampling and to make a first step in developing a decision-tree with rules for choosing between them.

There are obvious similarities between spatial sampling and field experiments, and the literature on experimental design contains discussions on randomization versus modelling that have remarkable parallels with those on design-based versus model-based sampling. Nevertheless, we prefer to keep this paper focussed on sampling; for the parallel discussion in experimental design we refer to Baird and Mead (1991), Cullis and Gleeson (1991), Grondona and Cressie (1991) and Zimmerman and Harville (1991).

Within sampling, we have imposed another broad restriction on the present paper on the following basis. As discussed below, sampling may involve two sources of randomness: the sample locations may or may not be selected at random and, given the locations, the values to be measured may or may not be considered as random. The corresponding 2×2 combinations are given in Table 1. Each of these four combinations represents a different approach to sampling. As the purpose of this paper is not to give an overall review but to invite discussion on principles, we confine this paper to contrasting the two most dissimilar approaches, namely those in which one or the other source of randomness is involved. Hence, we do not discuss the approaches with no or with both sources of randomness, i.e. the fully deterministic and fully random sampling strategies mentioned in Table 1.

We first present the basic notions and terminology that we will use. Then we describe the two approaches in relation to each other, using a simple example of point sampling in two dimensions. The theoretical part is concluded with a discussion of common misconceptions in this field. A case study is then

Table 1
Four types of sampling strategies defined by two sources of randomness

		Values at given locations	
		Fixed	Random
Sample locations	Fixed	Fully deterministic strategies	Model-based strategies: probabilistic kriging
	Random	Design-based strategies: classical sampling theory	Fully random strategies

presented by way of illustration, and the final section discusses aspects of choice for guidance in practical applications.

2. Basic notions and terminology

In the *design-based approach* stochasticity is introduced at the stage of sampling. The sample locations are selected by a pre-determined random selection procedure. Hence, in the design-based approach the stochastic process considered is a human-induced, random experiment comparable to the coin-tossing experiment. The rules of the random experiment (selection procedure) are accurately described by the 'sampling design', hence the term 'design-based approach'. Thus, 'sampling design' does not refer to the actual set of sample locations but to the procedure used to select it. A possible elementary outcome of this random experiment consists of a set of sample locations, and will be called a 'sample' for short; the number of locations is the 'sample size'. Hence, in classical sampling terminology, many different samples could be selected (drawn) from a given area (population) using the same random sampling design.

Two well-known sampling designs are, for instance, Simple Random Sampling (*SI*) and Stratified Simple Random Sampling (*STSI*). In *SI* a pre-specified number of sample locations is selected at random from the area, with equal probabilities of selection and independently from each other. This is done by taking the geographical coordinates of each sample location from a random number generator or from a table of random numbers. With *STSI* the area is first divided into a number of sub-regions, called strata, and then *SI* is applied to each of the strata separately. The sample sizes in the strata may be chosen such that the probabilities of the locations of being sampled differ between strata.

A given random design determines the probability of selection for every possible sample. In other words, the probabilities of occurrence of the elementary outcomes of the random experiment are exactly known. These are used when drawing conclusions, formulated as probability statements, from the sample for the population. In the statistical literature this is referred to as inference. As a consequence, if a sample is selected correctly according to the chosen design, the corresponding selection probabilities are strictly applicable and any inference based on them is valid by construction, not by assumption, whatever the spatial distribution in the area.

Apart from measurement error, sampling is the only source of stochasticity considered in the design-based approach. This implies that the unknown value at any given location and time is considered as fixed, not random.

The procedure used to calculate an estimate of a population parameter from the sample data is called the 'estimator'. It is useless to discuss the merits of a sampling design without considering the estimator to be used, or vice versa. It is the combination of the two that determines the accuracy and other properties of

the estimates. The pair (p, t) consisting of a design p and an estimator t is called a 'sampling strategy'. The search for design-based strategies that are cost-effective in specific circumstances by the optimal use of prior information has been a major task of sampling theorists since the 1940s (Neyman, 1934, 1938; Cochran, 1946; Horvitz and Thompson, 1952; Yates, 1960; Royall, 1971). This gave rise to devices like stratified, cluster, systematic, multistage, two-phase, and sequential sampling, which can be combined and specialized in many ways, and to estimators using measurements on auxiliary variables.

The theory of the design-based approach is often referred to as classical sampling theory, which should not be identified with classical statistics. Because classical sampling theory builds on probabilities inferred from randomization, it has developed separately from the mainstream of classical statistics. The latter builds mainly on probabilities inferred from stochastic models, for instance with the assumptions of independently, identically and normally distributed observations.

In the *model-based approach* the soil-forming process which has led to the field of values of a particular property in the study area is modelled as a stochastic process. This stochastic process is not a designed experiment as in the design-based approach, but is a mathematical abstraction used to describe reality. Consequently the probabilities of occurrence of the elementary outcomes of this process, the fields of values, are not known but have to be modelled. Hence, the field of values is assumed to be an outcome of the chosen random model. The probability distribution of all possible realizations of that model is the basic tool for inference in the model-based approach. This implies that the sample need not be selected by a random procedure, because the unknown value at any given location and time is already regarded as random, not fixed as in the design-based approach.

In model-based inference two types of quantities may be the goal: (1) functionals defined on the values of a single realization, such as spatial means, spatial variances and values at unvisited locations; or (2) parameters of the chosen model, such as model mean and model variance. The first are random quantities because they are defined on random values, while the second are fixed. This difference has important consequences for inference and interpretation, but is sometimes ignored. To emphasize the distinction, we speak of prediction of random quantities and of estimation of fixed quantities. In notations we underline random variables.

From a computational point of view, the most obvious difference between the two approaches is how the sample data are weighted to obtain estimates or predictions. In the design-based approach the weights are derived from the design. In the model-based approach the weights are derived from the chosen *model* and the actual configuration of *sample locations*. The same applies to the calculation of error estimates.

When survey statisticians began to develop model-based strategies they

naturally discussed the theory and merits of their proposals in relation to already existing design-based strategies. Hence, the general statistical literature on sampling has shown from the beginning a clear awareness of the two approaches and their relationship. By contrast, the earth sciences had developed a special statistical theory for spatial estimation, which is known as 'regionalized variable theory' (Matheron, 1963, 1971). At least initially, the development of this theory and the 'kriging' methodology based on it took place outside the regular statistical communication channels. It was largely a separate development, which created its own publication and conference circuits and terminology, collectively referred to as 'geostatistics'. Ironically, despite this separate development, regionalized variable theory has a closer connection with classical statistics than classical sampling theory has, in the sense that both are based on similar stochastic models. This is reflected, for instance, by the fact that various forms of kriging can be reformulated as special cases of generalized regression (Corsten, 1989). Classically oriented statisticians, once they have dug through the jargon, often have more affinity with regionalized variable theory than with classical sampling theory.

During this separate development, geostatisticians did not discuss their methodology in relation to classical sampling theory. An exception is Borgman and Quimby (1988). Only recently were kriging methods recognized and discussed as special cases of what had been defined before as model-based sampling strategies in the literature on sampling theory (De Gruijter and Ter Braak, 1990). Probably due to this isolation, misconceptions and confusion arose in geostatistical literature about issues related to random sampling. Before we go into these misconceptions, let us first clarify the differences between the two approaches with the help of a simple notional example of point sampling in two dimensions.

3. Design-based and model-based sampling strategies

Consider the sampling of soil profiles from a hypothetical area A, with a variable of interest z that can take only two values, 0 or 1. In this hypothetical example the value of z is supposed to be known everywhere in A (Fig. 1a). Suppose furthermore that we want to estimate the spatial mean of z within A, which equals the proportion of the area with $z = 1$. To estimate this, we select n locations and determine the value of z there. As is discussed below, either z or $\mathbf{x}_i$ could be random, so the values obtained will be indicated as $\underline{z}(\mathbf{x}_i)$ or $z(\underline{\mathbf{x}}_i)$, respectively.

For the sake of convenience we assume a finite population, which means that the number of different possible sampling locations is finite. If the population is assumed to be infinite, then probabilities should, strictly speaking, be replaced by probability densities, but in practice this has no influence on the results.

3.1. Design-based approach

In the design-based approach the locations are determined randomly. Fig. 1a shows a sample selected according to Simple Random Sampling with sample size $n = 25$. The value of z at any given location and time is regarded as fixed. Although these values are fixed, the locations are selected at random, so we denote the data by $z(\mathbf{x}_i)$.

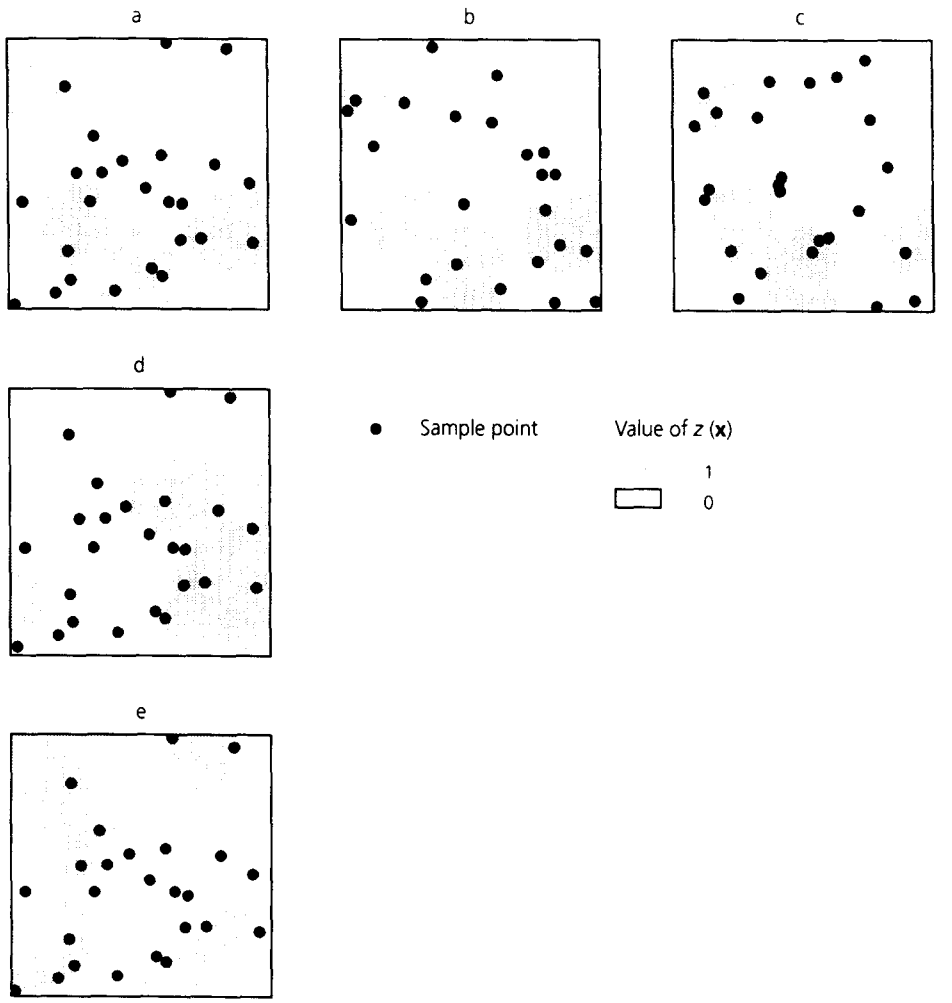


Fig. 1. Repeated sampling in the design-based approach (a, b, c) and in the model-based approach (a, d, e). In the design-based approach the population is fixed and the sampling locations are random. In the model-based approach the sampling locations are fixed and the population is random. The populations of (a), (d) and (e) are realizations of the same 'distance model' (see text), with true proportions of 0.30, 0.32 and 0.47, respectively.

Given a sampling design p, the selection probabilities of the samples are known. By summing the selection probabilities of the samples that contain profile i we obtain the inclusion probability of soil profile i with sampling design p. These inclusion probabilities are the key to describing the sample-to-sample variation of a proposed estimator. In the Horvitz–Thompson estimator or π-estimator (Särndal et al., 1992) the values of the soil profiles are divided by their inclusion probabilities. In Simple Random Sampling, for instance, the inclusion probability of a soil profile is equal for all profiles and equals n/N, the sampling fraction, where N is the population size. This results in the following estimator of the spatial mean:

$$\underset{\sim}{\hat{m}}_A = \frac{\sum_{i=1}^{n} \frac{z(\underset{\sim}{\mathbf{x}}_i)}{\pi_i}}{N} = \sum_{i=1}^{n} \frac{1}{n} z(\underset{\sim}{\mathbf{x}}_i) \tag{1}$$

where $\underset{\sim}{\hat{m}}_A$ is the estimated spatial mean of area A, n is the sample size and π_i is the inclusion probability of soil profile i. Note that in this design all points get equal weight.

Once a sample has been taken and the spatial mean estimated, what would happen if another sample were taken from the same area using the same sampling design and the same estimator? This is the question posed in the design-based approach. Repeated sampling in the design-based approach is illustrated in Fig. 1a–c, showing three simple random samples of size 25. The three samples come from different locations, and so the estimated value also differs from sample to sample: 0.32 (= 8/25), 0.36 (= 9/25) and 0.32 for Fig. 1a, 1b and 1c, respectively. The corresponding estimation errors (the difference between estimated and true mean) are 0.02, 0.06 and 0.02. Although only one sample is usually taken from an area, in the design-based approach the mean and variance of the estimator over repeated sampling for a given design p plays an important role in statistical inference. These mean and variance values are indicated as the p-expectation (Eq. (2)) and the p-variance or sampling variance (Eq. (3)), and are defined as:

$$\mathrm{E}_p(\underset{\sim}{\hat{m}}_A) \equiv \sum_{s=1}^{S} p(s)\hat{m}_{A,s} \tag{2}$$

and

$$\mathrm{V}_p(\underset{\sim}{\hat{m}}_A) \equiv \mathrm{E}_p\left[\left\{\underset{\sim}{\hat{m}}_a - \mathrm{E}_p(\underset{\sim}{\hat{m}}_A)\right\}^2\right] \tag{3}$$

where E_p is the expectation over repeated sampling under design p, $p(s)$ is the selection probability of sample s, S is the number of possible samples under design p, $\hat{m}_{A,s}$ is the spatial mean estimated by sample s and V_p is the sampling variance under design p.

Given a design p, we may look for an estimator such that its p-expectation is equal to the true value of the spatial mean m_A:

$$\mathrm{E}_p(\underline{\hat{m}}_A) = m_A \tag{4}$$

In other words, the estimation error over all possible realizations of the sampling design p should be zero. If this equality holds the estimator is called p-unbiased. Apart from p-unbiasedness, another desirable property of the estimator is that its Mean Squared Error (MSE_p) is minimal:

$$MSE_p \equiv \mathrm{E}_p\left\{\left(\underline{\hat{m}}_A - m_A\right)^2\right\} \tag{5}$$

If an estimator is p-unbiased its MSE_p is equal to its sampling variance, otherwise MSE_p will exceed the sampling variance by the bias squared.

The estimated sampling variance can be used to estimate confidence intervals for m_A. In the design-based approach a confidence interval for m_A is interpreted as follows. Suppose that the true spatial mean were known and that we were to calculate the confidence interval for every possible sample drawn with the given design p. For each sample we observe whether the calculated interval covers m_A or not. The cumulative probability of those samples for which the interval includes m_A is called the confidence level or coverage probability $1-\alpha$ (Särndal et al., 1992, p. 55). Confidence intervals (levels) and coverage probabilities defined in this way are referred to as p-confidence intervals (levels) and p-coverage probabilities, respectively. A p-confidence interval for m_A at the approximate level $1-\alpha$ is often computed as:

$$\underline{\hat{m}}_A \pm z_{1-\alpha/2}\sqrt{\hat{\mathrm{V}}_p(\underline{\hat{m}}_A)} \tag{6}$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution. This interval contains m_A in an approximate proportion of $1-\alpha$ of repeated samples drawn using the given design p, if the sampling distribution of $\underline{\hat{m}}_A$ is approximately $N(m_A, \mathrm{V}_p(\underline{\hat{m}}_A))$ distributed and $\hat{\mathrm{V}}_p(\underline{\hat{m}}_A)$ is a consistent estimator of $\mathrm{V}_p(\underline{\hat{m}}_A)$. According to the Central Limit Theorem the first requirement will be satisfied for sufficiently large samples, regardless of the distribution of the population from which the sample is drawn. Note that the width of the estimated confidence interval will vary from sample to sample as the estimated sampling variance varies between samples.

3.1.1. Design-based estimation of values at points

The extension of the design-based approach to the estimation of values at points is not straightforward. Design-based estimation at points should not be confused with model-based prediction at points assuming a discrete model of spatial variation, i.e. pure nugget models within soil map units with different means and variances (see Sections 3.2 and 4.2). Most literature on prediction at

points from soil maps has opted for model-based viewpoints (see for instance Burgess and Webster, 1980; McBratney et al., 1981; Voltz and Webster, 1990). The design-based viewpoint is quite different.

In design-based estimation at points, the points are first grouped into domains. For a given domain, a design-based estimate of its mean is assigned to every point in this domain. Hereafter we shall assume that soil map units are used as domains. For the estimation of values at points we have proposed the spatial mean of squared errors, $m_A(\varepsilon^2)$, as a measure of accuracy for estimation (Brus et al., 1992):

$$m_A(\varepsilon^2) \equiv \frac{1}{N}\sum_{i=1}^{N}\left\{\hat{z}(\mathbf{x}_i) - z(\mathbf{x}_i)\right\}^2 \tag{7}$$

where N is the total number of soil profiles (possible sampling locations) in A, $\hat{z}(\mathbf{x}_i)$ is the estimated value of the property at point $\mathbf{x}_i$, $z(\mathbf{x}_i)$ is the true value of the property at point $\mathbf{x}_i$, and A is the area to which $\hat{z}(\mathbf{x}_i)$ is applied and where the accuracy is to be evaluated. In our specific case, since $\hat{z}(\mathbf{x})$ is constant within a map unit u, with value $\hat{m}_u$, the $m_u(\varepsilon^2)$ in u reduces to:

$$m_u(\varepsilon^2) = (\hat{m}_u - m_u)^2 + \frac{1}{N_u}\sum_{i=1}^{N_u}\left\{z(\mathbf{x}_i) - m_u\right\}^2 = (\hat{m}_u - m_u)^2 + v_u \tag{8}$$

where m_u is the true spatial mean of z in u, N_u is the total number of soil profiles (possible sampling locations) in u, and v_u is the spatial variance of z in u. For a given map unit the spatial variance v_u is fixed, and $m_u(\varepsilon^2)$ would be minimized by choosing $\hat{m}_u$ equal to the true mean m_u. Of course m_u is unknown and has to be estimated by sampling. This introduces sampling variation and the estimator may now be considered as random and, therefore, denoted by $\underline{\hat{m}}_u$. Now that $\underline{\hat{m}}_u$ is stochastic, the errors are stochastic too. Assuming that a random sampling design p is used, we can take the statistical expectation E_p of the $\underline{\varepsilon}^2$s over realizations of the sampling process defined by p and the spatial mean of these expected errors:

$$m_u\left\{\mathrm{E}_p\left(\underline{\varepsilon}^2\right)\right\} = \mathrm{E}_p\left\{\left(\underline{\hat{m}}_u - m_u\right)^2\right\} + v_u \tag{9}$$

The first term on the right hand side of Eq. (9) represents the Mean Squared Error (*MSE*) of estimating m_u under design p, and can be divided up into the squared bias and the variance:

$$m_u\left\{\mathrm{E}_p\left(\underline{\varepsilon}^2\right)\right\} = \left\{\mathrm{E}_p\left(\underline{\hat{m}}_u\right) - m_u\right\}^2 + \mathrm{V}_p\left(\underline{\hat{m}}_u\right) + v_u \tag{10}$$

With p-unbiased estimators the first term is zero and:

$$m_u\left\{\mathrm{E}_p\left(\underline{\varepsilon}^2\right)\right\} = \mathrm{V}_p\left(\underline{\hat{m}}_u\right) + v_u \tag{11}$$

The spatial mean of the p-expected, squared error for the whole area A, $m_A\{E_p(\varepsilon^2)\}$, can be calculated as the weighted sum of the $m_u\{E_p(\varepsilon^2)\}$, using the proportions of the total area in the map units (W_u) as weights:

$$m_A\left\{E_p\left(\underline{\varepsilon}^2\right)\right\} = \sum_{u=1}^{U} W_u m_u\left\{E_p\left(\underline{\varepsilon}^2\right)\right\} \tag{12}$$

where U is the number of map units.

3.2. Model-based approach

In the model-based approach the actual population from which we take the sample is considered to be just one of an infinite set of possible populations that are separate realizations of the same random model. Hence, this approach is also referred to as the superpopulation approach (Cassel et al., 1977). In the model-based approach the value of a soil property at each location is not fixed. It can take several possible values, each with a defined probability of occurring, thus forming a random variable $\underline{z}(\mathbf{x}_i)$. (Note that z is now underlined instead of $\mathbf{x}_i$.) The observed value at location $\mathbf{x}_i$ is regarded as the realized outcome of this random variable. Assuming random variables at points, the true spatial mean is also random, so that we predict rather than estimate it. To distinguish predictions from estimators, predictors will be denoted by a tilde ($\sim$).

Modelling the probabilities of the realizations of the stochastic process, i.e. the fields of values, boils down to modelling the N-dimensional joint distribution ξ of the random variables $\underline{z}(\mathbf{x}_i)$. In practice, this modelling is usually limited to the description of the mean and variance of $\underline{z}(\mathbf{x}_i)$ and the covariance or semivariance of $\underline{z}(\mathbf{x}_i)$ and $\underline{z}(\mathbf{x}_j)$ for any i and j. Populations are viewed as realizations of ξ. Fig. 1a, d and e, for instance, show three realizations of a 'distance model', in which $\underline{z}(\mathbf{x}_i) = 1$ if at least one centre is within a distance r from $\mathbf{x}_i$ (Matérn, 1960, pp. 37–39, 49). For r we took 0.15 units of length. The centres were generated by a Poisson process with intensity $\lambda = 6$, i.e. averaged over all realizations of the model there are six centres per unit area. For Fig. 1a, d and e this resulted into 6, 5 and 10 centres. Note that also the zones of the centres outside the unit area but within a distance of 0.15 unit length of the edge are shown. The true spatial means (proportions with value 1) of Fig. 1a, d and e are 0.30, 0.32 and 0.46, respectively.

Whereas in the above example the model mean and covariance function are known, they are commonly unknown in reality and should be estimated from a single realization of the model. To make things practical, it is assumed that the random process is second-order stationary or intrinsic, and ergodic. These assumptions are part of the model. For a second-order stationary process the expectation and variance of the process are constant, and the covariance of $\underline{z}(\mathbf{x}_i)$ and $\underline{z}(\mathbf{x}_j)$ depends only on the distance and direction separating $\mathbf{x}_i$ and $\mathbf{x}_j$. A somewhat weaker assumption is the intrinsic hypothesis, which means that the

first-order increments $z(\mathbf{x}) - z(\mathbf{x}+\mathbf{h})$ are assumed second-order stationary. A random process with this property is said to be intrinsic of order zero. For intrinsic random processes of order one, the second-order increments are second-order stationary, and so on.

In reality one can usually observe only a single realization of the model, and the model mean and covariance function (or variogram) should be estimated from this realization. To this end the process is assumed to be ergodic. This means that for increasing size of the domain, the spatial mean of a statistic, for instance the spatial mean of $z(\mathbf{x})$ or of the squared differences of all pairs of points separated by vector $\mathbf{h}$, converge (in mean square sense) to the mean over all realizations of the model (Cressie, 1991, p. 53; Papritz, 1993, p. 6). This enables us to estimate model parameters such as model mean and variogram, from the realization sampled.

The value at a particular point or the spatial mean of a block is predicted as a weighted average of the values at the sample points. In ordinary point kriging the weights are obtained by solving the equations:

$$\begin{aligned} &\sum_{j=1}^{n} \lambda_j C(\mathbf{x}_i,\mathbf{x}_j) + \psi = C(\mathbf{x}_i;\mathbf{x}_0) \quad \forall i = 1 \text{ to } n \\ &\sum_{j=1}^{n} \lambda_j = 1 \end{aligned} \tag{13}$$

where $C(\mathbf{x}_i,\mathbf{x}_j)$ is the covariance of the sample points $z(\mathbf{x}_i)$ and $z(\mathbf{x}_j)$, ψ is a Lagrange multiplier and $C(\mathbf{x}_i,\mathbf{x}_0)$ is the covariance of $z(\mathbf{x}_i)$ and the prediction point $z(\mathbf{x}_0)$. In block kriging the weights can be calculated by replacing $C(\mathbf{x}_i,\mathbf{x}_0)$ by $\bar{C}(\mathbf{x}_i, A)$, the mean covariance of $z(\mathbf{x}_i)$ and the block A. This predictor is a Best Linear Unbiased Predictor (BLUP), i.e. it is unbiased and has minimum variance. However, these quality criteria have a different meaning here, related to the different meaning of repeated sampling in this approach. Repeated samples are now taken at the same locations but in different hypothetical areas of the same joint distribution ξ. This is illustrated by the vertical row of maps in Fig. 1 (a, d and e). Basically, it is not the sampling which is repeated, but the drawing of a realization from the superpopulation. Fig. 1a, d and e show that the values at some sampling locations differ between the three realizations, so that the predicted spatial means will also be different. The mean and variance of the predicted value over realizations of the model ξ are referred to as the ξ-expectation and ξ-variance. For the BLUP the ξ-expectation of the prediction error (the difference between the predicted value and the true value) is zero (ξ-unbiasedness):

$$\mathrm{E}_\xi(\tilde{m}_A - m_A) = 0 \tag{14}$$

and the Mean Squared Error is minimal (minimal MSE_ξ):

$$MSE_\xi \equiv \mathrm{E}_\xi\left\{(\tilde{m}_A - m_A)^2\right\} \tag{15}$$

Because of the ξ-unbiasedness of the kriging predictor, the Mean Squared Error of the kriging predictor equals the ξ-variance of the kriging prediction error, which is shortly referred to as the kriging variance.

Prediction intervals can be calculated by substituting the calculated ξ-variance of the predicted spatial mean in Eq. (6). To interpret these ξ-prediction intervals realizations of the random model ξ are considered instead of random samples. For each realization we observe whether the calculated interval covers the m_A of that realization. Note that, unlike the p-confidence intervals, the width of the ξ-prediction interval is constant because the ξ-variance is independent of the values at the sampling locations.

The covariance function of the distance model of Fig. 1 is known (Matérn, 1960, p. 38):

$$C_\xi(h) = \mathrm{e}^{-2\lambda\pi r^2}\,\mathrm{e}^{\lambda A(r,h)} \tag{16}$$

where λ is the intensity of the Poisson process, r is the radius of the circles, and $A(r,h)$ is the area common to two circles with radius r and the centres h apart. Using this function, the optimal predictions were 0.38, 0.34 and 0.46 for Fig. 1a, 1d and 1e, respectively, and the associated prediction errors were 0.08, 0.02 and -0.01. Note the small error in Fig. 1e despite the large predicted value (0.46). This is because the true mean in Fig. 1e was also large compared to that in Fig. 1a (0.47). In practice the covariance function is unknown and must also be estimated from sample data.

Finally, we remark on the ξ-variance of prediction errors at points, assuming a discrete model of spatial variation. If such a model is postulated the area is divided into U mutually exclusive units. It is assumed that the value at any location in unit u is the sum of the mean of that unit and a residual noise term which shows no autocorrelation and whose variance $C_{0,u}$ is constant within each unit (Heuvelink, 1993, p. 32). In this case the ξ-variance of the prediction error is constant within unit u and equals $C_{0,u} + C_{0,u}/n_u$, where n_u is the number of sample points in unit u. Suppose the mean of a unit is estimated by a Simple Random Sample. The sample variance which is used as an estimate of $C_{0,u}$ is then also a p-unbiased estimator of the spatial variance v_u. Moreover, the sample variance divided by n_u is an unbiased estimator of the sampling variance for this design. Hence, the estimated ξ-variance of prediction errors for this design equals the estimated spatial mean of the p-expectation of the squared error (Eq. (10)). For more efficient sampling designs it overestimates the design-based estimation variance.

3.3. Relation between p-variance and ξ-variance of design-based estimator

Although the p-variance and ξ-variance are defined on different sources of randomness, there is a relation between the two for some combinations of sampling design and estimator. Suppose we have a stochastic model of spatial

variation, and realizations of this model are sampled according to design p (fully random strategy, Table 1). What we would like to know is the sampling variance of the design-based estimator of the global mean, averaged over all realizations of the model:

$$\mathrm{E}_\xi\left[\mathrm{V}_p(\hat{\underline{m}}_A)\right] = \mathrm{E}_\xi\left\{\mathrm{E}_p\left[\hat{\underline{m}}_A - \mathrm{E}_p(\hat{\underline{m}}_A)^2\right]\right\} \tag{17}$$

Let us consider now non-informative sampling designs. For these designs the selection probabilities of the samples are independent of the values at the sampling locations (Särndal et al., 1992, p. 33). For such designs E_ξ and E_p in Eq. (17) can be interchanged (Särndal, 1978, p. 33):

$$\mathrm{E}_\xi\left\{\mathrm{E}_p\left[\hat{\underline{m}}_A - \mathrm{E}_p(\hat{\underline{m}}_A)^2\right]\right\} = \mathrm{E}_p\left\{\mathrm{E}_\xi\left[\hat{\underline{m}}_A - \mathrm{E}_p(\hat{\underline{m}}_A)^2\right]\right\} \tag{18}$$

Only if such a design is combined with an estimator which is p-unbiased and ξ-unbiased we may write

$$\mathrm{E}_\xi\left[\mathrm{V}_p(\hat{\underline{m}}_A)\right] = \mathrm{E}_p\left[\mathrm{E}_\xi\left(\hat{\underline{m}}_A - m_A^2\right)\right] = \mathrm{E}_p\left\{\mathrm{E}_\xi\left[\hat{\underline{m}}_A - \mathrm{E}_\xi(\hat{\underline{m}}_A)^2\right]\right\} = \mathrm{E}_p\left[\mathrm{V}_\xi(\hat{\underline{m}}_A)\right] \tag{19}$$

In words, the sampling variance of the estimator, averaged over all realizations of the model, equals the model variance of the estimator, averaged over all samples drawn with design p.

Eq. (19) can be used to predict the performance of design-based sampling strategies in areas with a spatial structure consistent with a given model, by drawing many samples with design p, and calculating for each sample the ξ-variance of the design-based estimator. For an example we refer to Domburg et al. (1994).

3.4. Relation between p-variance of kriging predictor and ξ-variance of kriging prediction error

Now suppose that we have a realization of a known model of spatial variation, and a set of samples of size n drawn from this realization with design p. Each sample is used to predict the spatial mean by the BLUP, using the known model to calculate the kriging weights. This gives as many predicted means and corresponding kriging variances as (we have) samples. The variance of these predicted means is an estimate of the sampling variance of the kriging predictor, the mean of the kriging variances is an estimate of the p-expectation of the kriging variance. The question is what relation there is between these two parameters.

To derive this relation we first take the p-expectation of the kriging predictor, conditional on the realization actually sampled:

$$\mathrm{E}_p(\tilde{\underline{m}}_A) = \mathrm{E}_p\left[\sum_{i=1}^{n} \underline{\lambda}_i z(\underline{\mathbf{x}}_i)\right] = \sum_{i=1}^{n}\left\{\mathrm{E}_p\left[\underline{\lambda}_i\right]\mathrm{E}_p\left[z(\underline{\mathbf{x}}_i)\right] + \mathrm{C}_p\left[\underline{\lambda}_i, z(\underline{\mathbf{x}}_i)\right]\right\} \tag{20}$$

where $C_p[\underline{\lambda}_i, z(\mathbf{x}_i)]$ is the sampling covariance of the kriging weight and the value of z of point i. The kriging weights $\underline{\lambda}_i$ in Eq. (20) are stochastic because the locations $\underline{\mathbf{x}}_i$ are. If only points in area A are used for prediction and all points have an equal probability of being included in the sample, then $E_p[z(\underline{\mathbf{x}}_i)]$ equals the true spatial mean m_A. Moreover, $\Sigma E_p[\underline{\lambda}_i] = E_p[\Sigma \underline{\lambda}_i] = 1$, so:

$$\begin{aligned} E_p[\underline{\tilde{m}}_A] &= \sum_{i=1}^{n} \left\{ E_p[\underline{\lambda}_i] E_p[z(\underline{\mathbf{x}}_i)] + C_p[\underline{\lambda}_i, z(\underline{\mathbf{x}}_i)] \right\} \\ &= m_A \left\{ \sum_{i=1}^{n} E_p[\underline{\lambda}_i] \right\} \\ &\quad + \sum_{i=1}^{n} C_p[\underline{\lambda}_i, z(\underline{\mathbf{x}}_i)] \\ &= m_A E_p\left[\sum_{i=1}^{n} \lambda_i \right] + \sum_{i=1}^{n} C_p[\underline{\lambda}_i, z(\underline{\mathbf{x}}_i)] \\ &= m_A + \sum_{i=1}^{n} C_p[\underline{\lambda}_i, z(\underline{\mathbf{x}}_i)] \end{aligned} \tag{21}$$

Eq. (21) shows that the kriging predictor is not p-unbiased; the p-bias equals the sum of the p-covariances of Eq. (21). This result is used in the following derivation.

If p is non-informative and the predictor is ξ-unbiased we may write:

$$\begin{aligned} &E_p[V_\xi[\underline{\tilde{m}}_A - \underline{m}_A]] \\ &= E_p\left[E_\xi\left[(\underline{\tilde{m}}_A - \underline{m}_A) - E_\xi[\underline{\tilde{m}}_A - \underline{m}_A] \right]^2 \right] \\ &= E_p\left[E_\xi\left[(\underline{\tilde{m}}_A - \underline{m}_A)^2 \right] \right] = E_\xi\left[E_p\left[\left\{ \underline{\tilde{m}}_A - \left(E_p[\underline{\tilde{m}}_A] - \text{bias} \right) \right\}^2 \right] \right] \\ &= E_\xi\left[E_p\left[\left(\underline{\tilde{m}}_A - E_p[\underline{\tilde{m}}_A] \right)^2 + \text{bias}^2 + 2\left(\underline{\tilde{m}}_A - E_p[\underline{\tilde{m}}_A] \right)\text{bias} \right] \right] \\ &= E_\xi\left[V_p[\underline{\tilde{m}}_A] \right] + E_\xi[\text{bias}^2] \end{aligned} \tag{22}$$

For processes stationary in the mean, i.e. $E_\xi[\underline{z}(\mathbf{x})]$ is constant and independent of $\mathbf{x}$, the last term of Eq. (22) equals zero, and as a result the p-expectation of the kriging variance equals the ξ-expectation of the p-variance of the kriging predictor.

4. Misconceptions

Various misconceptions about sources of randomness, independence, assumptions in classical sampling theory and target quantities in geostatistical inference

occur more or less frequently in the literature on spatial sampling. The instances which we have encountered could be divided into two interrelated types of misconception. In the following we describe and discuss each type and illustrate it with examples from the literature.

4.1. Misconception 1: 'Independence of sample data is determined by the physics of the phenomenon being sampled'

This view fails to recognize that independent random selection of sample locations generates p-independence of sample data, regardless of the structure of the spatial variation. This follows directly from probability theory, for instance, Theorem 6A in Parzen (1960, p. 295). If $\underline{\mathbf{x}}_1$ and $\underline{\mathbf{x}}_2$ are random locations selected independently from each other, then the variables $z(\underline{\mathbf{x}}_1)$ and $z(\underline{\mathbf{x}}_2)$ are independent too. See also Theorem 2 in Ash (1970, p. 84). Imagine, for instance, independent random selection of two values z_1 and z_2 from any field of fixed z-values. Repeat this many times, each time retaining z_1 and z_2. A scatterplot of z_1 against z_2 will show no dependence, regardless of the structure of the field of z-values.

In his paper 'Principles of geostatistics' Matheron (1963) presented the original formulation of geostatistics, which was entirely deterministic: both the sample locations and the values at the locations were regarded as fixed, so that inference in a statistical sense was impossible. Instead, characteristics of spatial variation and estimation error were all defined as spatial integrals. Matheron (1963) warned explicitly that: "This formalism has inherited from its statistical origin a language in which one still speaks of variance and covariance, including however in those notions a new content. This similarity in vocabulary must not deceive."

It seems more than likely, nevertheless, that giving different meanings to terms that have existed for a long time in statistics, has caused confusion. This is particularly relevant with regard to the concept of independence. Matheron (1963) stated: "Two neighbouring samples are certainly not independent. They tend, in average, to be both high-grade if they originate from a high-grade block of ore, and vice-versa. This tendency, more or less stressed, expresses the degree of more or less strong continuity in the variation of grades within the mineralized space." Clearly, since no stochasticity was involved at all in his theory, Matheron (1963) did not mean independent in the usual probabilistic sense.

A recent example of this misconception can be found in a discussion paper by Journel (1994). At the end of his rejoinder Journel states: "... Second, the spatial characteristics of the sample stem from the physics of the phenomenon being sampled not from the sampling design. Drawing the sample locations $\mathbf{u}_\alpha s$ from a Poisson point process does not make the resulting data $Z(\mathbf{u}_\alpha)s$ independent; it is the physical process of, say deposition and transportation of the concentra-

tions $Z(\mathbf{u}_\alpha)$ that dictate the degree of space/time correlations between these data.''

Two examples from soil science are Burrough (1991) and Ditzler (1994). Burrough (1991) wrote: ''If we can safely assume that the variation of the attribute in question is normally distributed and spatially independent (fig. 7-1a) the number of samples needed to estimate the mean value of an attribute in a given map unit with given levels of confidence is given by conventional statistical theory''. Like Journel, Burrough disregards p-independence, which led him to suggest erroneously that the standard formula he gave for the sample size which is needed to obtain a confidence interval smaller than a prespecified width is only valid in the special case of a 'spatially independent attribute'.

And, finally, Ditzler (1994) wrote: ''While the development of confidence intervals (or other statistical tests such as analysis of variance) are very useful for conveying information to our clients, we must remember that an underlying assumption with these statistical procedures is that the observations are *independent* of one another. Only observations that are far enough apart to be beyond the range as shown on the variogram meet this requirement. Observations that are separated by smaller distances are correlated with one another and confidence intervals developed for the estimated mean will tend to be too narrow (Cressie, 1991).''

We encountered this misconception even in a frequently cited handbook on spatial sampling, viz. Gilbert (1987). In Section 4.5 ('Number of measurements: correlated data') it is stated that: ''the methods for estimating n (the size of a Simple Random Sample required to achieve a prespecified minimum accuracy, B and dG) in previous sections assume measurements are uncorrelated. In practice, a spatial correlation may be present so that part of the information contained in one measurement is also in other measurements taken close by in space.''

Whereas p-independence is related to a single realization of a stochastic model of spatial variation, ξ-independence is defined with respect to all realizations of such a model. Statements about ξ-dependence or ξ-independence are usually based on observations from a single realization. If for all sampled pairs of points in A separated by a vector $\mathbf{h}$, the value at the first point is correlated with the value at the second point, it is assumed that this also holds for all other realizations of the model. In other words, the process is assumed to be ergodic in covariance. The misconception about independence can thus be corrected as: 'ξ-independence of sample data is determined by the modeller; p-independence is determined by the sampling design'.

4.2. Misconception 2: 'Classical sampling theory assumes that sample data are independent'

If no distinction is made between ξ-independence and p-independence, and if independence is defined in model-based terms, a logical conclusion is that in

classical sampling theory it is assumed that data are ξ-independent. Examples from the literature are:

McBratney and Webster (1983): "Estimates of the number of observations needed for this purpose (to determine the mean, B and dG) have usually been based on classical sampling theory without regard to spatial dependence in the data... Finally, classical statistics assumes a nonincreasing variogram with a sill equal to the population variance. It is worth considering the possibility of replacing this variogram with more reasonable general variograms obtained from further investigations into the spatial variation of soil properties."

Dahiya et al. (1985): "In Part I (Dahiya et al., 1984), spatial variability of some nutrient constituents, viz. NO_3, K, Mg, and organic C, of a loess soil field was evaluated by applying classical statistical analysis (i.e., probability density function, mean, and variance). An implicit assumption in that analysis is that the observations of a given soil property are independent of one another regardless of their location in the field."

This misconception has had a serious effect in applications of spatial statistics, because it led soil scientists to dismiss the design-based approach on the erroneous ground that it makes an assumption which is usually invalid. As we have shown in the preceding, the design-based approach does not make any assumption of independence; it actually creates independence through randomization. The model-based approach, on the other hand, does make assumptions of debatable validity.

5. Efficiency of design-based and model-based sampling strategies

In the early period of kriging applications in soil science McBratney and Webster (1983) wrote an influential paper entitled 'How many observations are needed for regional estimation of soil properties?'. These authors compared the sample size needed to obtain a given precision in estimating a spatial mean using classical sampling theory with the sample size needed using regionalized variable theory. For each of three areas, Hole Farm, Broom's Barn and Plas Gogerddan, McBratney and Webster performed the following analysis of sample data obtained from a square grid. First, the population variance was estimated using the classical formula. This variance was divided by the sample size to obtain the estimation variance according to classical sampling theory. Second, block kriging was used to calculate the kriging estimation variance. The kriging estimation variance was smaller than the classical estimation variance by the spectacular factors 3.4, 8.8 and 6.6 for Hole Farm, Broom's Barn and Plas Gogerddan, respectively.

McBratney and Webster described their analyses and results clearly and in detail, but they formulated their conclusions in a way which easily leads to the suggestion that very large gains in efficiency, "three-and-half-fold to ninefold"

in their examples, may be expected from using regionalised variable theory instead of classical sampling theory. This interpretation of McBratney and Webster's results, however, is grossly misleading as we will argue below.

Two biases caused exaggeration: statistical and experimental bias. The first was introduced by applying the classical variance formula to sample data from a regular grid. This over-estimates the population variance and hence the classical estimation variance. To quantify this, we calculated the ξ-expectation of the estimation variance for Simple Random Sampling, given the sample sizes and the variograms in McBratney and Webster (1983), using the Monte-Carlo algorithm presented in Domburg et al. (1994). Divided by the kriging estimation variances, this reduced the efficiency ratios somewhat to 3.3, 7.2 and 5.6 for Hole Farm, Broom's Barn and Plas Gogerddan, respectively.

The experimental bias is found to be much more important, and arises from the fact that McBratney and Webster compared kriging from a square grid with Simple Random Sampling. Among the classical sampling strategies, Simple Random Sampling is hardly ever used in practice because of its notorious inefficiency; it is mainly of theoretical interest. Therefore, we calculated the ξ-expectation of the estimation variance as before, but now for Stratified Simple Random Sampling, with square strata and one sample point per stratum. This reduced the efficiency ratios drastically to 1.2, 1.6 and 1.0 for Hole Farm, Broom's Barn and Plas Gogerddan, respectively. For Stratified Simple Random Sampling with two points per stratum, which allows for model-free estimation of the standard error, the ratios were 1.4, 2.0 and 1.1.

McBratney and Webster quantified the accuracy of the design-based and model-based estimates of the global mean by different quality criteria, viz. the ξ-expected sampling variance and the ξ-variance. Papritz and Webster (1994) compared the efficiency on the basis of the same criterion, the sampling variance. They simulated two auto- and cross-correlated fields of pH, representing the pH at two different points in time, and estimated the global change in pH. They found that the geostatistical estimate was *less* precise than the paired stratified and paired systematic estimates. They found *p-MSE* ratios of 0.77 and 0.62 for the largest sample size. For a grid without additional points (for estimating the variogram) the model-based predictions and design-based estimates were equally precise.

Van Kuilenburg et al. (1982) compared the performance of design-based and model-based strategies for estimating values *at points* on the basis of equal effort in fieldwork. Design-based estimates based on a soil map and 150 sample points (in order to estimate the means of the map units) were 10% less accurate than those from kriging using no soil map and 530 sample points. Later on, Voltz and Webster (1990) compared kriging and soil map classification (formulated in model-based terms) with an equal number of data points. They concluded that soil map classification performed best where there were abrupt changes, while kriging and kriging combined with classification performed

better than classification where the soil varied gradually. Brus et al. (1996) found no significant differences between soil map classification, unstratified kriging and stratified kriging.

6. Case study

We hope that even the sceptical reader is convinced by now that both approaches are valid in spatial estimation or prediction. The next question to be answered is, which of the two methods is better. For instance, the spatial means of several blocks in an area can be estimated by block kriging or by stratified random sampling, using the blocks as strata. We compared the two approaches in a case study. We created an exhaustive data set because this means that the true block means and the values at all points are known, so that the data can be repeatedly sampled, and performances can be compared. The exhaustive data set was obtained by stochastic simulation (see below).

In the case of block kriging there are two sources of stochasticity, the random sampling and the stochastic model (fully random strategy, Table 1). Various quality criteria can be formulated when following such a strategy, differing in the quantity which is to be minimized, such as p-variance, ξ-variance or ξ-expected p-variance, or in the constraint, such as p-unbiasedness, ξ-unbiasedness or $p\xi$-unbiasedness (Särndal, 1978). We adopted the design-based quality criteria (p-variance of predictor, p-bias and p-coverage probability) to see which strategy gave the best estimates. This implies that we generated one realization only of the stochastic spatial variation model.

6.1. Stochastic simulation

At a regional scale the Dutch landscape often consists of several physiographic units with different pedogeneses. As a result, the means and variances of most soil properties differ considerably between these units and, moreover, in some places the values change abruptly at their edges. To reproduce these so-called heterogeneities, Deutsch and Journel (1992) recommend a two-step approach. The geometry of the physiographic units are modelled and simulated first, after which the values of the property under study within each homogeneous unit are simulated.

We did not model the geometry of the heterogeneities (the first step) and simply used a detailed soil map at scale 1:25,000 to obtain a map of the homogeneous, physiographic units of the Schuitenbeek area (Fig. 2). This map can be regarded as a realization of the unknown, categorical model. The soil map depicts an area typical of the Pleistocene part of the Netherlands. Its size is 8 km (from west to east) by 6.4 km (from north to south). The map units were

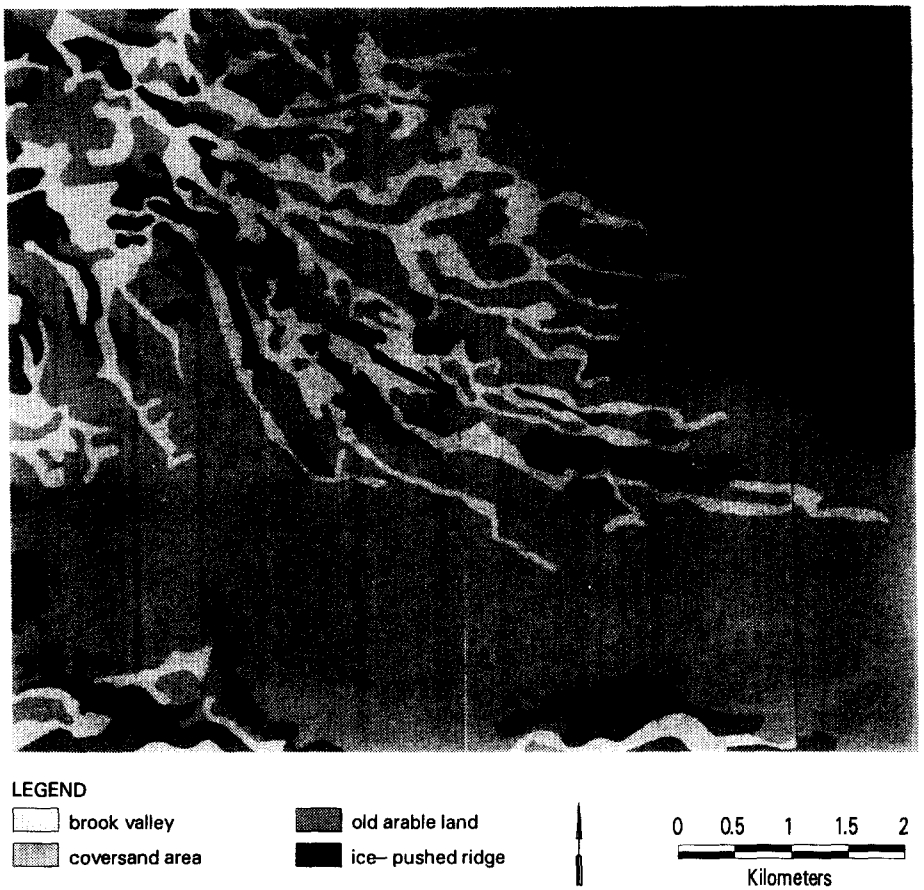


Fig. 2. Soil map, 1:25,000, with four physiographic units used in simulation.

grouped into four physiographic units: brook valleys, coversand area, old arable land, and ice-pushed ridge.

In the second step the values within each physiographic unit were simulated, independently from those in the other units, by sequential Gaussian simulation (Deutsch and Journel, 1992). As a target variable we chose the Mean Highest Watertable (W_{min}). In addition to measurements of W_{min} from the area shown on

Table 2
Normal-score variograms for four physiographic units used in simulation

Physiographic unit	Type of model	a_1 (m)	c_1	a_2 (m)	c_2
Brook valley	spherical	70	1	–	–
Coversand area	spherical + exponent.	65	0.69	3650	0.31
Old arable land	exponential	265	1	–	–
Ice-pushed ridge	exponential	745	1	–	–

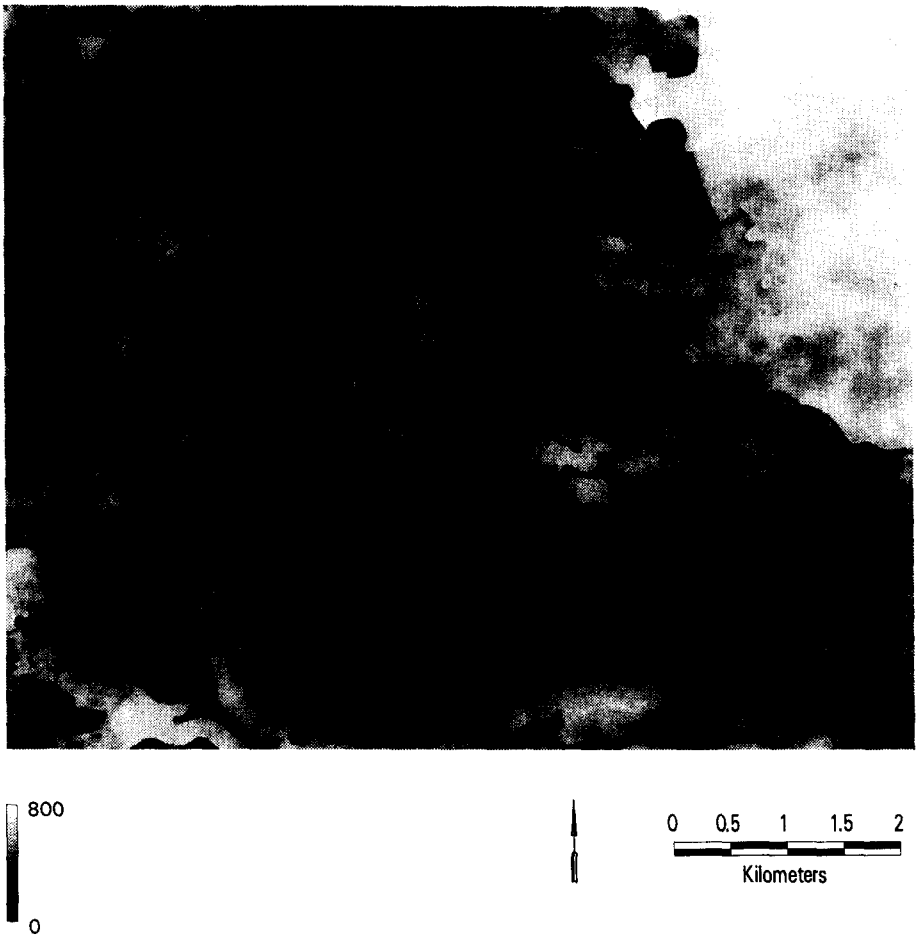


Fig. 3. Simulated field showing smooth variation inside the four physiographic units and locally abrupt changes at their edges.

the map (Schuitenbeek area), we used data from a similar area (the Wesepe area) to obtain variograms. Table 2 shows the type and parameter values of the normal score variograms of the four physiographic units used in simulation. We

Table 3
Spatial means and variances of conditioning data and simulated fields

	Conditioning data		Simulated data	
	mean (cm)	variance (cm^2)	mean (cm)	variance (cm^2)
Brook valley	24	141	24	132
Coversand area	48	847	47	756
Old arable land	100	2916	109	3152
Ice-pushed ridge	428	33822	408	34602

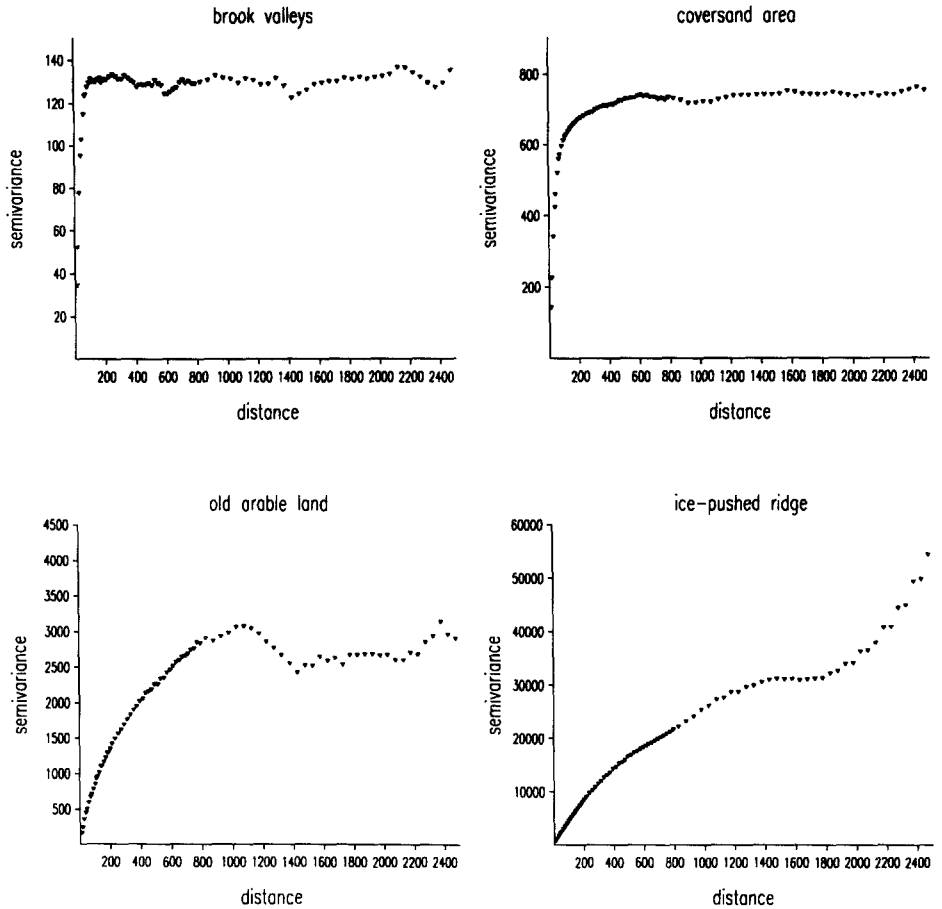


Fig. 4. Omnidirectional variograms of the four physiographic units, based on exhaustive variograms in four directions (N–S, E–W, NE–SW, NW–SE).

simulated 800×640 values at the nodes of a regular grid, i.e. the size of the squares was 100 m^2. The data from the Schuitenbeek area were used as conditioning data. Fig. 3 shows the result. The means and variances of the conditioning data were closely reproduced by simulation (Table 3). Fig. 4 shows the exhaustive variograms of the four physiographic units. These variograms allow only a rough comparison with the variograms used in the simulation, because the latter were based on data transformed to normal scores. Moreover, the variograms realized were determined partly by the spatial structure of the conditioning data. For instance, the simulated field of the old arable land unit clearly showed a hole effect, whereas we used a monotonically increasing function in simulation. Nevertheless, the ranges of the initial and resulting variograms were of the same order of magnitude for all units.

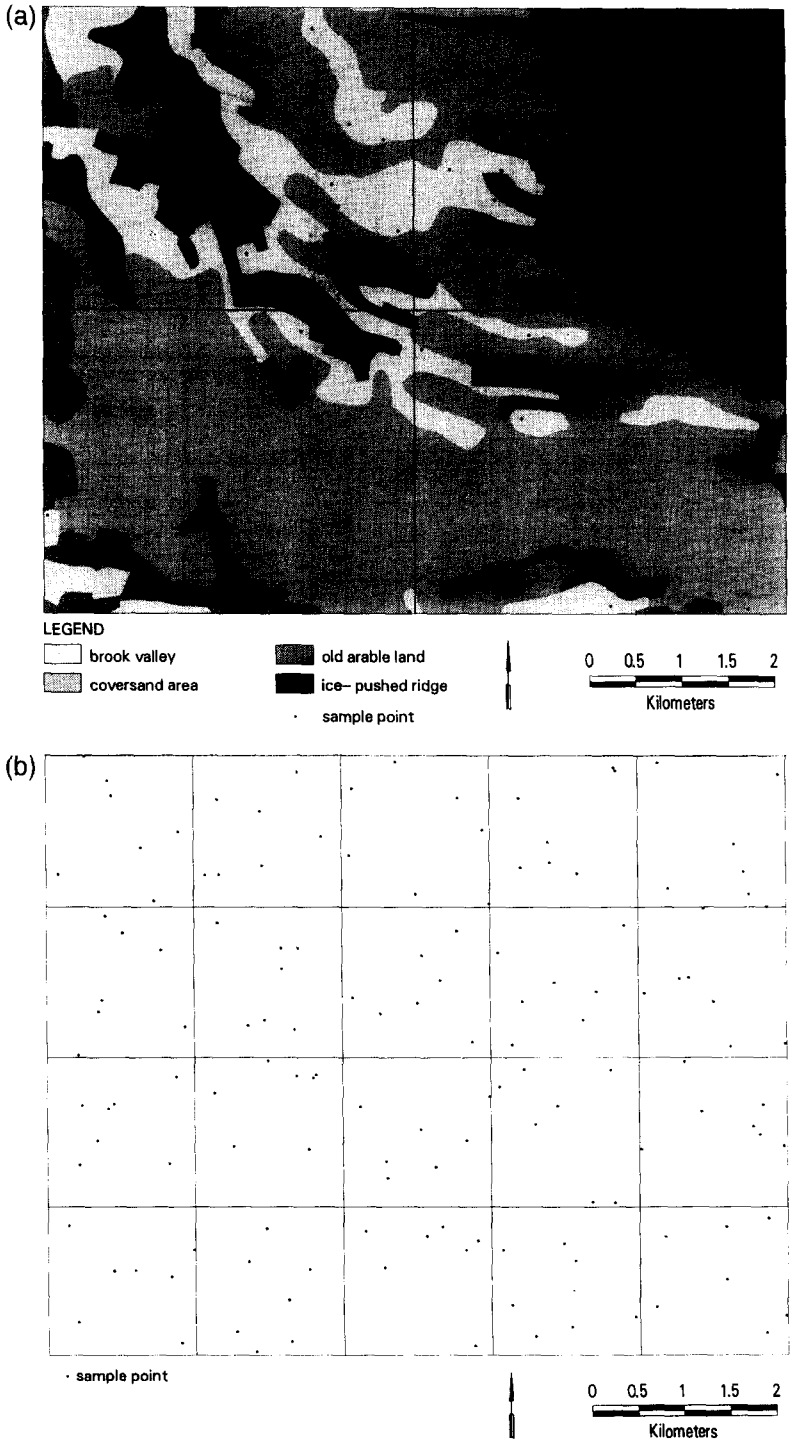


Fig. 5. Realizations of the design-based sampling designs: (a) sample size 140, estimation of global mean; (b) sample size 140, estimation of local means; (c) sample size 1520, estimation of global and local means.

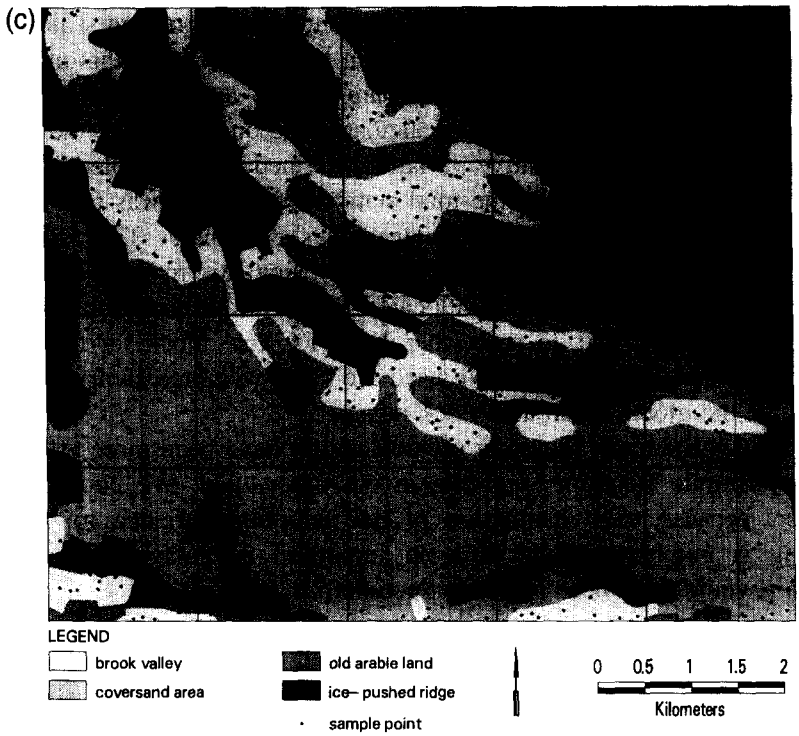


Fig. 5. (continued).

6.2. Sampling strategies

Two types of sampling strategies were used to estimate the global mean and the local means of square 1.6×1.6 km^2 blocks: Stratified Simple Random Sampling combined with the Horvitz–Thompson estimator (*STSI*, t_{HT}), and Systematic Sampling combined with the block kriging predictor (*SY*, t_{OK}). Each strategy was combined with two sample sizes: 140 and 1520. For each combination 100 samples were drawn, so we had 100 *SY* plus 100 *STSI* samples of size 140, and 100 *SY* plus 100 *STSI* samples of size 1520.

6.2.1. Stratified Simple Random Sampling

With the first type of strategy (*STSI*, t_{HT}) the sampling locations were selected by a Stratified Simple Random Sampling design. Stratification was adapted to sample size and block size (Fig. 5). In estimating the global mean from the small sample we used four rectangular blocks (4.0×3.2 km^2) for stratification, and 20 square blocks (1.6×1.6 km^2) for the large sample. Within these blocks, a soil map was used for further stratification, except for the case of estimating the local means from the small sample.

The soil map used in stratification was an outdated soil map at scale 1:50,000. The four physiographic units are depicted in far less detail on this map than on the map used in the simulation (compare Fig. 2 and Fig. 5a). Taking the simulated field for reality, we have calculated the quality of the old map. The purity was 75% and the relative variation, defined as the pooled within-class standard deviation divided by the total standard deviation, was 66%. Marsman and De Gruijter (1986) found similar values elsewhere. The efficiency of random sampling is potentially increased by using this map for stratification. Numbers of sample points were allocated to the strata proportionally to their size. This implied 7 and 76 points per 1.6×1.6 km^2 block for the small and large sample size, respectively. We used the usual weighted sample mean, t_{HT}, to estimate the global mean and block means.

Stratified Systematic Sampling possibly might have been more efficient than Stratified Simple Random Sampling in our case study. We chose the latter, however, because valid estimation of the sampling variance for systematic sampling is cumbersome, so that an important advantage of design-based strategies would have been lost (see Sections 7.3 and 7.6).

6.2.2. Block kriging

In the second type of strategy (SY, t_{OK}) the sampling locations were selected systematically. The locations formed square grids, with grid distances of 800 m (10×8 grid) and 200 m (40×32 grid). The orientation of the randomly placed grid was fixed, parallel to the sides of the area. Besides these grid points, additional points were selected to estimate the experimental variogram. For the small sample these were 60 points and 30 of them at 270 m (which is approximately 1/3 grid distance) from randomly selected grid points, and 30 at 10 m (1 node) distance. For the large sample 240 additional points were selected, 120 of them at 70 m (ca. 1/3 grid distance) and 120 at 10 m distance. After a grid point had been selected, we randomly selected a direction from four directions: N, E, S and W. It follows that the total sample size (grid points plus variogram points) was 140 and 1520 for the small and large sample, respectively. Fig. 6 shows one realization of each of the model-based sampling designs.

The global and local means were predicted by stratified block kriging. The strata were formed by the four units of the 1:50,000 soil map. Table 4 summarizes characteristics of the procedure: the variogram used in prediction, the units of which means have been predicted in order to predict the local and global means, and the neighbourhood. From the large samples we estimated variograms for each map unit separately. The additional variogram points were quartered, such that each map unit had 30 points at 100 m and 30 at 10 m distance. For the small sample the four experimental variograms would be too inaccurate. In this case we therefore assumed that the shape of the variograms was the same for the four map units but that its magnitude was different, i.e. a

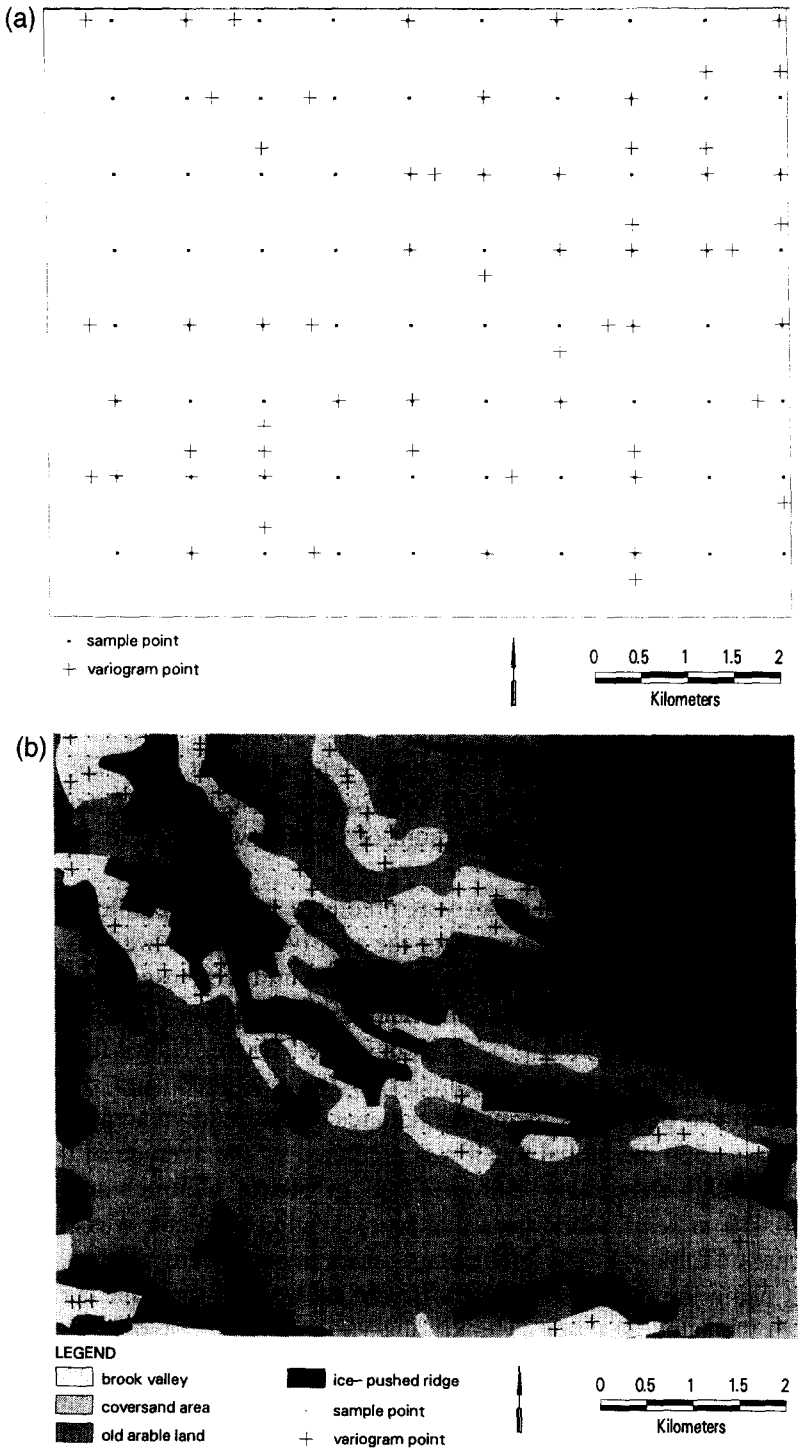


Fig. 6. Realizations of the model-based sampling designs: (a) sample size 140; (b) sample size 1520.

Table 4
Specification of kriging procedures applied

Sample size	Block size	
	1.6×1.6 km^2	8×6.4 km^2
140	(a) local relative variogram (b) all map units within block (c) same map unit in prediction block + adjacent blocks	local relative variogram 4 map units same map unit in all blocks
1520	(a) 1 variogram per map unit (b) all map units within block (c) same map unit in prediction block	1 variogram per map unit 1280 elementary blocks elementary block + adjacent elementary blocks in same map unit

(a) type of variogram(s); (b) units of which means have been predicted in order to predict local or global mean; (c) neighbourhood.

proportional effect (Journel and Huijbregts, 1978, p. 189). As a scaling factor we used the square of the local mean. Isaaks and Srivastava (1989) refer to this variogram type as the local relative variogram. An omnidirectional variogram was estimated from each sample. We fitted spherical and exponential models to the sample variograms, with additional double spherical models for the small sample. All models were fitted without and with a nugget term. The numbers of pairs divided by the fitted values were used as weights in iterative fitting (Cressie, 1985). The residual mean sum of squares was used to select the best model.

With $n = 140$ the global mean was predicted by the weighted mean of the four predicted map unit means, using the relative areas as weights. To predict the mean of a map unit all the data points in this unit were used. To calculate the kriging variance we assumed that the map unit means were independent, which accords with the simulation procedure described above. Under this assumption the kriging variance can be calculated as the weighted sum of the kriging variances of the map units, using the squares of the relative areas as weights.

With $n = 1520$ the global mean was predicted by the average of the predicted means of 1280 elementary 200 m $\times$ 200 m blocks (Journel and Huijbregts, 1978, p. 410 *et seq.*), while the usual block kriging predictor was applied in all other cases (Table 4). The elementary block means were predicted by block kriging, using the data inside the block and those from the adjacent elementary blocks in the same map unit. The kriging variance of the predicted global mean was approximated by combining the elementary errors as described by Journel and Huijbregts (p. 415).

The means of the 1.6×1.6 km^2 blocks were predicted by stratified block kriging, using the data from the block itself only ($n = 1520$), or those from the adjacent blocks as well ($n = 140$).

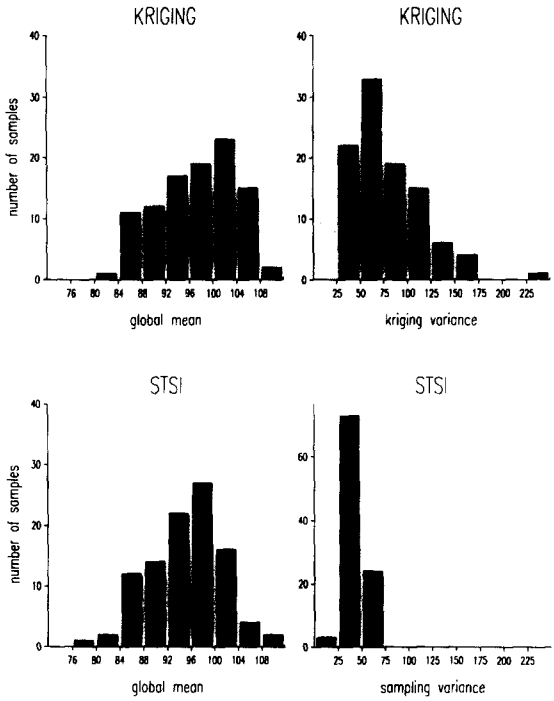


Fig. 7. Histograms of 100 design-based estimates and 100 model-based predictions of the global mean and their estimation and prediction variance, for sample size 140.

6.3. Results

Figs. 7 and 8 show histograms of the 100 ($STSI$, t_{HT}) estimates and the 100 (SY, t_{OK}) predictions of the global mean and their estimation variances. Statistics of these histograms are shown in Table 5: the second column contains the variances of the 100 estimated and predicted means. The third column contains the estimated p-bias which was obtained by subtracting the true global mean from the averages of the 100 estimated or predicted means. The estimate of the p-MSE (fourth column) was obtained simply by combining the values of the second and third column. The fifth column shows the average of the 100 sampling variances and kriging variances. Confidence (prediction) intervals were calculated for each sample for $\alpha = 0.05$, 0.10 and 0.20, assuming a t-distribution for the design-based estimator of the local means (with 6 and 75 degrees of freedom for $n = 120$ and 1520, respectively), or a normal distribution in all other cases. The last three columns show the fractions of the samples of which the confidence interval contained the true value. Ideally these p-coverages should equal 95%, 90% and 80%, respectively. The p-coverage will be smaller than these values if the estimator (predictor) is p-biased or its standard error underestimated, while it will be larger if the standard error is overestimated and the estimator is p-unbiased. We tested the null-hypotheses H_0: p-coverage

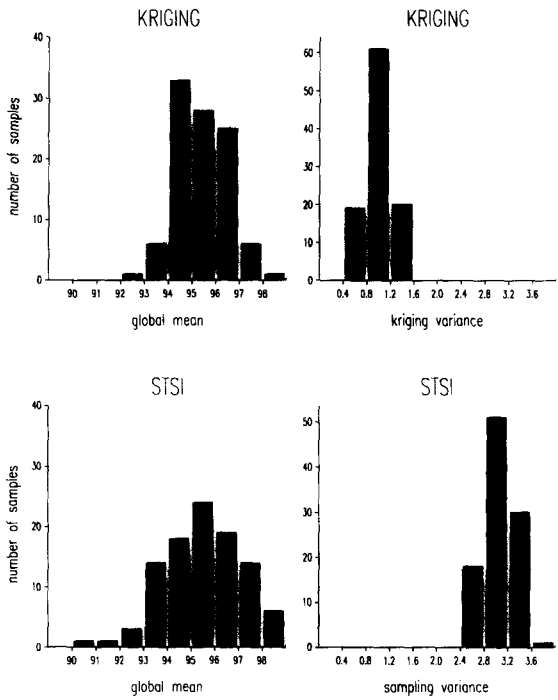


Fig. 8. Histograms of 100 design-based estimates and 100 model-based predictions of global mean and their estimated estimation and prediction variance, for sample size 1520.

$= 0.95$, 0.90 and 0.80 respectively, with nominal significance level $\alpha = 0.05$. For the twenty 1.6×1.6 km^2 blocks the above-mentioned statistics are presented as histograms (Figs. 9 and 10).

6.3.1. Global mean

For $n = 140$ ($STSI$, t_{HT}) was slightly more accurate than (SY, t_{OK}): the sampling variance as well as the p-bias were somewhat smaller (Table 5). For

Table 5
Statistics of 100 design-based estimates and 100 model-based predictions of the global mean for sample sizes (n) 140 and 1520

	n	p-variance	p-bias	p-MSE	Mean p-variance	p-coverage		
					Mean ξ-variance	0.95	0.90	0.80
Design-based	140	38.2	−0.439	38.3	43.7	0.97	0.94	0.82
Model-based	140	43.0	1.714	45.9	80.2	*1.00*	*0.96*	*0.90*
Design-based	1520	2.605	−0.084	2.612	3.066	0.98	0.92	0.84
Model-based	1520	1.280	−0.196	1.318	1.000	0.94	0.87	*0.68*

Italics: p-coverage differs significantly ($\alpha = 0.05$) from nominal value.

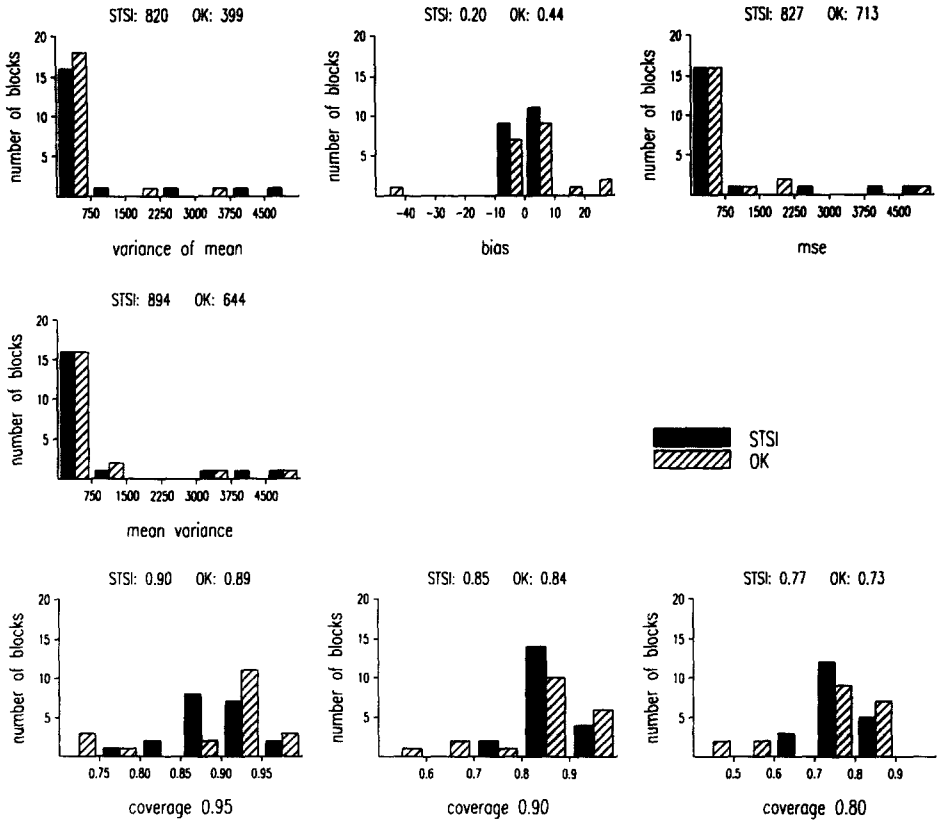


Fig. 9. Histograms of statistics of 100 design-based estimates (*STSI*) and 100 model-based predictions (*OK*) of local means for sample size 140; mean values are indicated at the top.

(*SY*, t_{OK}) the contribution of the *p*-bias to the *p-MSE* was only small. The *p*-coverage was significantly larger than $1-\alpha$ for (*SY*, t_{OK}), mainly due to the conservative estimate of the sampling variance by the kriging variance: 80.2 versus 43.0.

For the larger sample, however, (*SY*, t_{OK}) outperformed (*STSI*, t_{HT}). Its *p*-variance was considerably smaller, and as the *p*-bias was small too, *p-MSE* was also relatively small. There were remarkably good *p*-coverage values for $1-\alpha=0.95$ and 0.90: 0.94 and 0.87, respectively. For $1-\alpha=0.80$ the *p*-coverage was significantly too low (0.68). Just as for the small sample size, *p*-coverage values for (*STSI*, t_{HT}) were somewhat too high, but these differences were not significant ($\alpha=0.05$).

6.3.2. Local means

For $n=140$ the *p*-variance, averaged over all 1.6×1.6 km^2, for (*STSI*, t_{HT}) was considerably larger than for (*SY*, t_{OK}): 819.6 versus 398.5 (Fig. 9). If we

ignore the outliers (blocks 15, 19 and 20 for ($STSI$, t_{HT}) and blocks 15 and 20 for (SY, t_{OK})), the difference in average p-variance is greatly reduced (280.2 versus 156.0), but (SY, t_{OK}) is still the best. The p-bias for (SY, t_{OK}) was considerable for some blocks: the maximum value was -62 cm. Despite this, (SY, t_{OK}) was on average superior to ($STSI$, t_{HT}) as regards p-MSE. After deletion of the outliers, the average of the mean kriging variance was close to that of the mean sampling variance (143.9 versus 156.0). On average, both (SY, t_{OK}) and ($STSI$, t_{HT}) showed undercoverage. With both strategies undercoverage or overcoverage was statistically significant ($\alpha = 0.05$) in several blocks: for $1 - \alpha = 0.95$, 0.90, 0.80, six, nine and seven blocks for (SY, t_{OK}), respectively, and eleven, seven and five blocks for ($STSI$, t_{HT}), respectively. Moreover, the coverage values for (SY, t_{OK}) varied greatly between the blocks. For some blocks coverage was extremely low (Fig. 9).

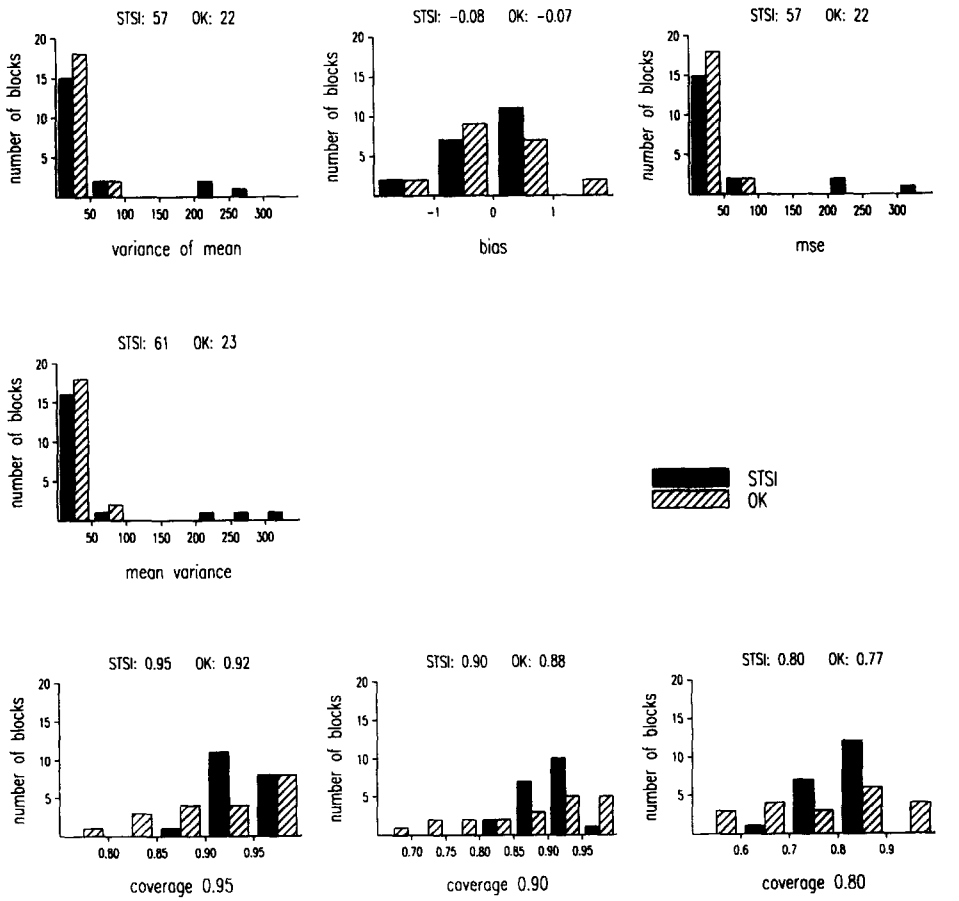


Fig. 10. Histograms of statistics of 100 design-based estimates and 100 model-based predictions of local means for sample size 1520; mean values are indicated at the top.

Just as for the small sample, (SY, t_{OK}) outperformed $(STSI, t_{HT})$ with respect to the average p-variance and p-MSE with the large sample (Fig. 10). The average coverage values with $(STSI, t_{HT})$ were equal to $1-\alpha$ for all α's, and undercoverage or overcoverage was just significant ($\alpha = 0.05$) for one, three and three blocks for $1-\alpha = 0.95$, 0.90 and 0.80, respectively. For (SY, t_{OK}) eleven, eleven, and sixteen blocks showed significant undercoverage or overcoverage.

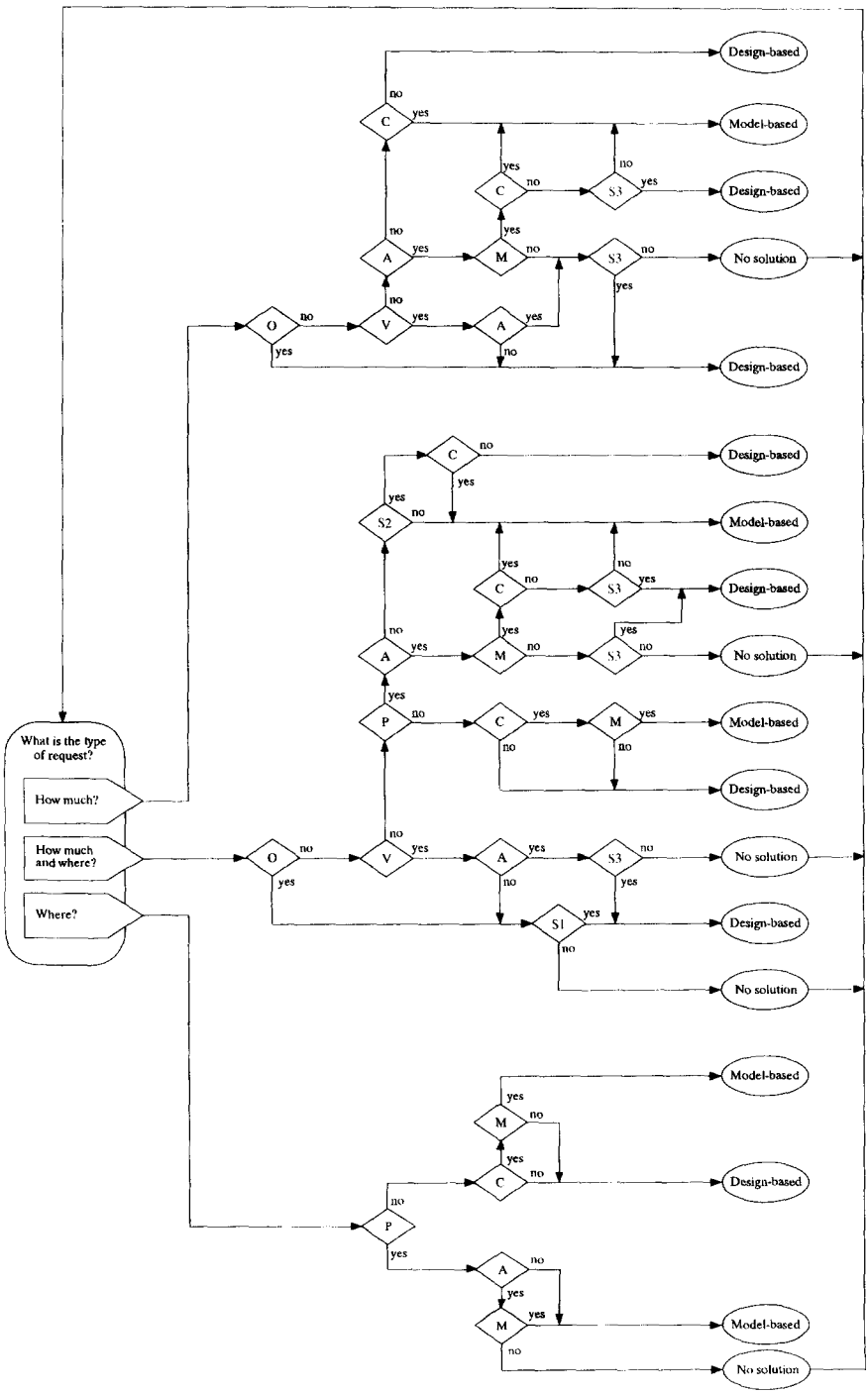
6.4. Discussion

Model-based estimates of the global mean with $n = 140$ were not more accurate because of the poor quality of the model and the large distance between the observation points. The model-based predictions of the local means were more accurate, but in this case these are compared with design-based estimates from Simple Random Samples of size 7, which is inefficient. For the global and local means, the relative accuracy of (SY, t_{OK}), defined as the ratio between the p-MSE of $(STSI, t_{HT})$ and that of (SY, t_{OK}), was larger for the large sample. This is possibly due to the better fit of the model used in kriging and the stronger autocorrelation between the data points and the block. However, we should bear in mind that the difference in efficiency between the two types of strategy is caused by the combined effect of a different predictor *and* a different sampling design.

As regards p-coverage, $(STSI, t_{HT})$ outperformed (SY, t_{OK}). The significant overcoverage or undercoverage of the model-based prediction intervals can be explained by the p-bias of the kriging predictor and the difference between the mean kriging variance and the sampling variance.

For the small sample size the local means were estimated by using the data from the adjacent blocks as well. This largely explains the relatively large p-bias in this case (Fig. 10). In all other cases the estimated p-bias was very small, which shows that the contribution of the covariance term in Eq. (21) was negligible (see Section 3.4).

Regarding the discrepancy between the mean kriging variance (kriging variance averaged over the 100 samples) and the p-variance of the kriging predictor, suppose that the model ξ is known and that this model is used to predict m_A. As was explained in Section 3.4 for a given realization of the model ξ, the p-expected kriging variance may differ from the sampling variance of the kriging predictor (Eq. (22)). The larger the area relative to the range of the variogram, the smaller this difference will be. In reality the true model is unknown and has to be estimated. For an estimated model ξ' the p-expected kriging variance may differ from the p-expected kriging variance for the true model ξ. The more sample points are used in estimating the model the smaller this difference will be. This explains the small difference between the mean kriging variance and the sampling variance of the predicted mean for $n = 1520$


What is the type of request?
How much?
How much and where?
Where?
Design-based
Model-based
No solution
C
S3
M
A
O
V
S2
P
S1
yes
no

relative to that for $n = 140$. Together with the relatively small p-bias for the large sample, this also explains the better coverage values of (SY, t_{OK}) for the large sample relative to those for the small sample.

For the small sample the design-based confidence interval of the local mean also showed significant ($\alpha = 0.05$) undercoverage or overcoverage for many blocks. This can be explained by the small sample size (7 points per block) together with the skew distribution of $W_{\min}$ inside the blocks. Therefore the assumption of a t-distributed sample mean with d.f. = 6 was not realistic for most of the blocks. For the large sample (76 points per block) map units within blocks were used as strata. However, $W_{\min}$ still was very skewed in many strata, partly due to impurities, and as a result the distribution of the estimator was still somewhat asymmetric for some blocks. This may explain the undercoverage or overcoverage for these blocks.

7. Selection of statistical approach

In the previous sections we showed that both the model-based and the design-based approach are valid for spatial sampling and estimation, and that the model-based approach is not necessarily optimal if only one realization is considered. Thus we now face the problem of selecting the 'best' approach in various circumstances.

Many factors determine the effectiveness and efficiency of a statistical approach for spatial sampling and estimation, making the decision process rather complicated. As a support for these decisions a decision tree has been developed with ten questions to be answered during the selection (Fig. 11). In some cases there is no solution for the problem posed. Then the aim of the survey should be changed, e.g. by reducing the number of subregions, or the constraints should be revised, e.g. by raising the budget so that more points can be sampled.

The decision tree has a limited range of application; it does not support the choice of fully random and fully deterministic approaches (Table 1) and it cannot be used for surveys aiming at the estimation of model parameters.

Fig. 11. Decision tree for choosing between design-based and model-based sampling strategies. O: is objectivity required? V: is strict validity required? P: are separate estimates of the estimation variance for all points or all subregions required? A: are accurate estimates of the estimation or prediction variance required? M: can the spatial variation be modelled adequately? C: will the observation and prediction points be autocorrelated? S1: is the sample large enough to obtain p-unbiased estimates of all local means? S2: is the sample large enough to obtain p-unbiased estimates of the sampling variance of all local means? S3: is the sample large enough to obtain accurate estimates of the sampling variance of the global or local means?

7.1. What is the type of request?

The aim of the survey should strongly influence the statistical approach. Domburg et al. (1994) distinguish three types of request, '*how much*', '*where*' and '*how much and where*'. In surveys concerned with 'how much', target quantities such as totals, means, proportions, medians or quantiles are estimated. Surveys with the emphasis on 'where' deal with the question where specific soil properties are present, and usually result in a map. If a 'where' request is answered, the 'how much' question is implicitly also answered. *Ceteris paribus*, the 'where' request requires more data than the 'how much' request. Surveys dealing with the 'how much and where' request aim at estimates for the whole survey region and for several subregions.

The field of application of the design-based approach is the estimation of population parameters or, in other words, the 'how much' request. The strength of the model-based approach, on the other hand, is that it can predict values at points. Nevertheless, the two approaches can in principle deal with either request.

7.2. Is objectivity required?

This question is about the objectivity of the estimates of the population parameters. In the model-based approach there are no restrictions on the selection of the sampling locations. Selection by subjective judgements of the 'representativity', and even preferential sampling of high or low grade zones is allowed. Moreover, Englund (1990) showed that the assumptions made in modelling spatial variation are highly subjective. Objectivity can be realized by, (1) selecting locations with equal probability or, if unequal probabilities are used these should be accounted for in the inference, and (2) using design-based weights in the inference. In statistical terminology, estimates should be p-unbiased.

The subjectivity in the selection of the sampling locations and the choice of the model as such is a drawback of the model-based approach because the model might be wrong which makes the survey results vulnerable to criticism (Borgman and Quimby, 1988). Objectivity may be especially important if results are used for legal or regulatory purposes.

7.3. Is strict validity required?

The previous question was about the objectivity of the estimates of the means or values at points, whereas this question is about the objectivity of the estimates *of the estimation variance and confidence intervals*. We define validity as follows. An application of a given strategy in a given situation is valid, and leads to valid results, if the premises and assumptions underlying the strategy

comply with reality. The probability statements thus obtained are correct, but not necessarily as accurate as possible. For instance, a valid application of a strategy results in a confidence interval with the correct (i.e. equal to the nominal) coverage, but the interval may be wider than obtained by other, more efficient strategies.

In the model-based approach several assumptions need to be made, e.g. about stationarity, isotropy and function type. Consequently, we are never sure about the validity of the results. In other words, the worth of the probability statements is unknown. If the validity is unknown, it is important that the strategy is robust against deviations of the assumptions (see Section 7.6).

In the design-based approach no assumptions are made on spatial variation. For interval estimation we have to make an assumption on the type of the distribution of the estimator indeed, but for reasonably large samples a normal distribution is not a strong assumption (see Section 7.10). Moreover, in some cases the type of distribution is known. For instance if fractions are estimated from a Simple Random Sample the estimator is binomially distributed.

We conclude that if we want strict validity a design-based strategy is the only option. Like objectivity, validity is important if results are used for legal or regulatory purposes.

7.4. Are separate unique estimates of the estimation variance for all points or all subregions required?

Values at points can be estimated or predicted by model-based and design-based strategies. However, unlike the model-based approach, the design-based approach does not provide a separate unique estimate of the estimation variance for a given point. It provides the mean estimation variance for all points in a given domain only (Eq. (11)). Similarly, if we want an estimate of the estimation variance for all subregions, the model-based approach is the only option, unless we can afford a sample size of two or more from each subregion (see Section 7.9).

7.5. Are accurate estimates of the estimation or prediction variance required?

Estimates of the estimation or prediction variance may be valid, but very inaccurate as well. For instance a Simple Random Sample of size two provides a p-unbiased estimate of the sampling variance, but this estimate is usually very inaccurate. Accurate estimates of the estimation or prediction variance are essential if, for instance, confidence (prediction) intervals have to be calculated in a risk analysis.

Sometimes interest is more in the relative values of the estimation or prediction variance than in the absolute values. For instance, in network

optimization new locations are often situated at sites with the lowest relative accuracy. In this case we do not require a reliable model (See section 7.6).

7.6. Can the spatial variation be modelled adequately?

It is well known that the calculated kriging variance is not robust against model misspecification. Therefore, if we want accurate estimates of the prediction variance, we need a reliable model of spatial variation. Webster and Oliver (1992) state that a variogram computed from a sample of 150 points will generally be satisfactory and one from 225 data will usually be reliable. Gascuel-Odoux and Boivin (1994) found similar results. Prior information, if present, can be used to postulate a model. If major heterogeneities are suspected, a map of these heterogeneities is of great importance. If an equal shape of the variogram for all units is too strong an assumption, different variograms should be allowed for, which will, however, drastically raise the required sample size.

7.7. Will the observation and prediction points be autocorrelated?

An adequate spatial variation model is one thing, autocorrelation between the observation points and the prediction points (blocks) is another. It is only if points are autocorrelated that one can profit from the model. The strength of autocorrelation depends on the density and pattern of sampling, and on the variogram characteristics such as the range (if any) and the relative nugget. The literature shows that autocorrelation should be strong to increase the efficiency. Gascuel-Odoux and Boivin (1994) reported a kriging standard deviation of 52% of the standard deviation in the sample in a case study with an average autocorrelation coefficient between neighbouring observation points of ca. 0.4. Voltz and Webster (1990) found somewhat larger values for autocorrelation coefficients of 0.3 and 0.6: 68% and 72%, respectively.

7.8. Is the sample large enough to obtain p-unbiased estimates of all local means?

This question plays a role in the 'How much and where' request only. The answer to this question is yes if we can afford at least one, possibly composite sample from each subregion.

7.9. Is the sample large enough to obtain p-unbiased estimates of the sampling variance of all local means?

Again this question is relevant only for the 'How much and where' request. The answer is yes if at least two, possibly composite samples from each subregion can be taken.

7.10. Is the sample large enough to obtain accurate estimates of the sampling variance of the global or local means?

If we want valid and accurate estimates of the estimation variance or confidence intervals, the design-based approach is appropriate unless the sample is small, say smaller than about 15. For smaller samples the estimates of the sampling variance often become very inaccurate, and for interval estimation the distribution of the estimator of the spatial mean is not close enough to normality.

8. Theses

An important aim of this paper is to initiate a discussion on the role of design-based and model-based sampling strategies in soil science. Our views on this are largely reflected by the decision tree (Fig. 11), which we hope can give structure to the discussion. To encourage further discussion we now present two theses.

8.1. The model-based criteria ξ-unbiasedness and minimum ξ-variance are not very useful

We believe that ξ-unbiasedness is a weak criterion. It only guarantees that the calculations are consistent with the model postulated. Suppose one selects locations with an inclusion probability proportional to the values of the target variable z thought likely from prior information (πps-sampling; Särndal et al., 1992, p. 90). No theoretical objections can be made against this selection, because in the model-based approach there are no restrictions on the selection of the sampling locations. If the prior estimate and the true value of z are positively correlated, then there would be a strong tendency to over-estimate the mean of z. Such misleading predictions, however, are still ξ-unbiased. The reason for this is that the area sampled is just one realization of the model, and smaller values would occur at the same locations in other realizations.

The problems of 'biased sampling' are well-known. In mining and environmental applications preferential sampling in high grade (highly polluted) zones is common. Several declustering techniques have been developed to correct for such preferential sampling (Isaaks and Srivastava, 1989, p. 237 *et seq.*). However, the success of this practical solution depends on the actual degree of spatial clustering in reality. Using the Walker Lake data, Journel found that the cell declustering technique failed fully to remove the bias (Journel, 1994, p. 70).

Regarding the minimum ξ-variance criterion, we have already pointed out that the quality of the ξ-variance estimate is as good as the quality of the model: if a poor model is used, the calculated kriging variance is useless.

Note that this criticism of the model-based criteria does not imply a dismissal of stochastic models in the inference. If the design-based criteria *p*-unbiasedness, *p*-variance and *p*-coverage are chosen as quality criteria, we can still use a stochastic model in prediction, as we did in the above case study (fully random strategy). If valid estimates of *p-MSE* are required, a model-free approach should be considered as the ξ-*MSE* can differ strongly from the *p-MSE* (see Section 6).

8.2. The use of design-based sampling strategies should be given more attention in soil survey

In the past two decades highly sophisticated kriging techniques such as disjunctive cokriging (Finke and Stein, 1994) have been introduced and applied in soil survey. On the other hand, design-based sampling in soil survey is still in its infancy. The potential of more advanced methods in soil survey still remains to be explored. To mention two: methods using descriptive models and methods for small-domain estimation.

8.2.1. Methods using descriptive models

The use of models in the inference is not restricted to the model-based approach. For instance, measurements of an auxiliary variable at random locations can be used in a design-based regression estimator (Cochran, 1977). Särndal et al. (1992) refer to design-based sampling strategies that make use of a model as model-assisted strategies. They point out, on p. 227, that the role of the model in design-based inference differs from that in the model-based approach. In the latter it describes a process by which the data have been generated, whereas in the design-based approach it describes the population itself. In the model-based approach it is assumed that the population was generated by the model, and inference is based on this assumption. In the design-based approach we hope that the model describes the population reasonably well, but inference is not conditioned on this. This implies that the validity of the variance estimates does not depend on the correctness of the model, whereas it does in the model-based approach. Errors in the model result in less accurate estimates, but the estimates of the *p-MSE* are still valid. Despite this advantage, we know of no examples from soil science applying the model-assisted approach.

8.2.2. Methods for small domain estimation

As was pointed out above, model-based strategies have great potential if the means of many subregions have to be predicted from a small sample. However, there are also design-based methods which have been developed to cope with this problem of so-called small domains, such as strategies using 'synthetic' estimators. We refer to Chaudhuri (1994) for a comprehensive review. 'Synthetic' estimators make use of measurements of the target variable in the domain

of interest and in similar domains. A simple example is a regression estimator in which similar domains are grouped to estimate the β's of the regression model. This estimator is p-biased and the estimator is useful only if the domains are really alike.

9. Conclusions

(1) During the past two decades, the design-based approach has been dismissed by many soil scientists on the false grounds that this approach assumes independence and that model-based predictions are optimal. Independence in the design-based approach is not assumed but created by the sampling design, and model-based predictions are not necessarily optimal if performance is measured with the design-based quality criteria.

(2) In our simulation study, model-based predictions of the global and local means were generally more accurate (smaller p-MSE) than design-based estimates, except for the global mean estimated from the small sample ($n = 140$). By contrast, p-coverages of confidence intervals were generally closer to the desired value using the design-based strategy, except for the local means estimated from the small sample.

(3) In deciding between the model-based and the design-based approach various aspects should be taken into account, such as the type of request, the interest in objective estimates, the need for separate unique estimates of the estimation variance for all points or subregions, the interest in valid and accurate estimates of the estimation or prediction variance, the quality of the model, the strength of autocorrelation between observation points, and the sample size.

(4) Research into design-based strategies deserves more attention from soil surveyors. Moreover, we need more comparative studies on the performance of model-based and design-based sampling strategies.

References

Ash, R.B., 1970. Basic Probability Theory. Wiley, New York, 337 pp.

Baird, D., Mead, R., 1991. The empirical efficiency and validity of two neighbour models. Biometrics 47, 1473–1487.

Barnes, R.J., 1988. Bounding the required sample size for geologic site characterization. Math. Geol. 20, 477–490.

Borgman, L.E., Quimby, W.F., 1988. Sampling for tests of hypothesis when data are correlated in space and time. In: Keith, L.H. (Ed.). Principles of Environmental Sampling. American Chemical Society, Professional Reference Book, pp. 25–43.

Brus, D.J., De Gruijter, J.J., Breeuwsma, A., 1992. Strategies for updating soil survey information: a case study to estimate phosphate sorption characteristics. J. Soil Sci. 43, 567–581.

Brus, D.J., De Gruijter, J.J., Marsman, B.A., Visschers, R., Bregt, A.K., Bouma, J., 1996. The performance of spatial interpolation methods and choropleth maps to estimate properties at points. A soil survey case study. Environmetrics 7, 1–16.

Burgess, T.M., Webster, R., 1980. Optimal interpolation and isarithmic mapping of soil properties, I. The semivariogram and punctual kriging. J. Soil Sci. 31, 315–331.

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.

Burrough, P.A., 1991. Sampling designs for quantifying map unit composition. In: Spatial Variabilities of Soils and Landforms. SSSA Spec. Publ. 28, 89–125.

Cassel, C.M., Särndal, C.E., Wretman, J.H., 1977. Foundations of Inference in Survey Sampling. Wiley, New York, 192 pp.

Chaudhuri, A., 1994. Small domain statistics: a review. Stat. Neerl. 48, 215–236.

Cochran, W.G., 1946. Relative accuracy of systematic and stratified random samples for a certain class of population. Ann. Math. Stat. 17, 164–177.

Cochran, W.G., 1977. Sampling Techniques. Wiley, New York, 3rd ed., 428 pp.

Corsten, L.C.A., 1989. Interpolation and optimal linear prediction. Stat. Neerl. 43, 69–84.

Cressie, N., 1985. Fitting variogram models by least squares. Math. Geol. 17, 563–586.

Cressie, N., 1991. Statistics for Spatial Data. Wiley, New York, 900 pp.

Cullis, B.R., Gleeson, A.C., 1991. Spatial analysis of field experiments: an extension to two dimensions. Biometrics 47, 1449–1460.

Dahiya, I.S., Kersebaum, K.C., Richter, J., 1984. Spatial variability of some nutrient constituents of an Alfisol from loess, I. Classical statistical analysis. Z. Pflanzenernaehr. Bodenkd. 147, 695–703.

Dahiya, I.S., Anlauf, R., Kersebaum, K.C., Richter, J., 1985. Spatial variability of some nutrient constituents of an Alfisol from loess, II. Geostatistical analysis. Z. Pflanzenernaehr. Bodenkd. 148, 268–277.

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., 1992. GSLIB. Geostatistical Software Library and User's Guide. Oxford University Press, New York, 340 pp.

Diamond, P., Armstrong, M., 1984. Robustness of variograms and conditioning of kriging matrices. Math. Geol. 16, 809–822.

Ditzler, C., 1994. Geostatistics: a brief look at its application in soil survey. Soil Surv. Hor. 35, 69–73.

Domburg, P., De Gruijter, J.J., Brus, D.J., 1994. A structured approach to designing soil survey schemes with prediction of sampling error from variograms. Geoderma 62, 151–164.

Englund, E.J., 1990. A variance of geostatisticians. Math. Geol. 22, 417–456.

Finke, P.A., Stein, A., 1994. Application of disjunctive cokriging to compare fertilizer scenarios on a field scale. Geoderma 62, 247–263.

Gascuel-Odoux, C., Boivin, P., 1994. Variability of variograms and spatial estimates due to soil sampling: a case study. Geoderma 62, 165–182.

Gilbert, R.O., 1987. Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold, New York, 320 pp.

Grondona, M.O., Cressie, N., 1991. Using spatial considerations in the analysis of experiments. Technometrics 33, 381–392.

Hansen, M.H., Madow, W.G., Tepping, B.J., 1983. An evaluation of model-dependent and probability-sampling inferences in sample surveys. J. Am. Stat. Assoc. 78, 776–807.

Heuvelink, G.B.M., 1993. Error Propagation in Quantitative Spatial Modelling. Applications in Geographical Information Systems. Thesis, University of Utrecht, 151 pp.

Horvitz, D.G., Thompson, D.J., 1952. A generalisation of sampling without replacement from a finite universe. J. Am. Stat. Assoc. 47, 663–685.

Isaaks, E.H., Srivastava, R.M., 1989. Applied Geostatistics. Oxford University Press, New York, 561 pp.

Journel, A.G., 1994. Resampling from stochastic simulations. Environ. Ecol. Stat. 1, 63–91.

Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics. Academic Press, London, 600 pp.

Kalton, G., 1981. Models in the practice of survey sampling. In: Proceedings of 43rd Session of the International Statistical Institute, Buenos Aires.

Marsman, B.A., De Gruijter, J.J., 1986. Quality of Soil Maps. A Comparison of Soil Survey Methods in a Sandy Area. Soil Survey Papers 15, Soil Survey Institute, Wageningen, 103 pp.

Matérn, B., 1960. Spatial variation. Stochastic models and their application to some problems in forest surveys and other sampling investigations. Medd. Statens Skogsforskningsinst. 49 (5), 144.

Matheron, G., 1963. Principles of geostatistics. Econ. Geol. 58, 1246–1266.

Matheron, G., 1971. The Theory of Regionalized Variables and its Applications. Ecole Nationale Supérieure des Mines de Paris, 211 pp.

McBratney, A.B., Webster, R., 1983. How many observations are needed for regional estimation of soil properties?. Soil Sci. 135, 177–183.

McBratney, A.B., Webster, R., Burgess, T.M., 1981. The design of optimal sampling schemes for local estimation and mapping of regionalized variable, I. Theory and method. Comput. Geosci. 7, 331–334.

Neyman, J., 1934. On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. J. R. Stat. Soc. 97, 558–606.

Neyman, J., 1938. Contribution to the theory of sampling human populations. J. Am. Stat. Assoc. 33, 101–116.

Oliver, M., Webster, R., Gerrard, J., 1989. Geostatistics in physical geography, Part I. Theory. Trans. Inst. Br. Geogr. N.S. 14, 259–269.

O'Muircheartaigh, C., Wong, S.T., 1981. The impact of sampling theory on survey sampling practice: a review. In: Proceedings of 43rd Session of the International Statistical Institute, Buenos Aires.

Papritz, A., 1993. Estimating Temporal Change of Soil Properties. Doct. Thesis, Swiss Federal Institute of Technology, Zürich, 166 pp.

Papritz, A., Webster, R., 1994. Estimating temporal change in soil monitoring, II. Sampling from simulated fields. Eur. J. Soil Sci. 46, 13–27.

Parzen, E., 1960. Modern Probability Theory and its Applications. Wiley, New York, 464 pp.

Royall, R.M., 1971. Liniair regression models in finite population sampling theory. In: Godambe, V.P., Sprott, D.A. (Eds.), Foundation of Statistical Inference. Holt, Rinehart and Winston of Canada, Ltd., pp. 259–279.

Särndal, C.E., 1978. Design-based and model-based inference in survey sampling. Scand. J. Stat. 5, 27–52.

Särndal, C.E., Swensson, B., Wretman, J., 1992. Model Assisted Survey Sampling. Springer-Verlag, New York, 694 pp.

Smith, T.M.F., 1976. The foundations of survey sampling: a review. J. R. Stat. Soc. A 139, 183–204.

Stein, M.L., Handcock, M.S., 1989. Some asymptotic properties of kriging when the covariance function is misspecified. Math. Geol. 21, 171–190.

Van Kuilenburg, J., De Gruijter, J.J., Marsman, B.A., Bouma, J., 1982. Accuracy of spatial interpolation between point data on soil moisture supply capacity, compared with estimates from mapping units. Geoderma 27, 311–325.

Voltz, M., Webster, R., 1990. A comparison of kriging, cubic splines and classification for predicting soil properties from sample information. J. Soil Sci. 41, 473–490.

Webster, R., 1985. Quantitative spatial analysis of soil in the field. Adv. Soil Sci. 3, 1–70.

Webster, R., Oliver, M.A., 1992. Sample adequately to estimate variograms of soil properties. J. Soil Sci. 43, 177–192.

Yates, F., 1960. Sampling Methods for Censuses and Surveys. Charles Griffen and Co., London.

Zimmerman, D.L., Harville, D.A., 1991. A random field approach to the analysis of field-plot experiments and other spatial experiments. Biometrics 47, 223–239.