Incorporating models of spatial variation in sampling strategies for soil

Promotor: dr ir J. Bouma,
hoogleraar in de bodeminventarisatie en landevaluatie
speciaal gericht op de (sub)tropen

Co-promotor: dr ir J.J. De Gruijter,
hoofd van de afdeling Landinventarisatiemethoden,
Dienst Landbouwkundig Onderzoek, Staring Centrum,
Instituut voor Onderzoek van het Landelijk Gebied (SC-DLO)

Derk Jan Brus

Incorporating models of spatial variation in sampling strategies for soil

Proefschrift
ter verkrijging van de graad van doctor
in de landbouw- en milieuwetenschappen
op gezag van de rector magnificus,
dr C.M. Karssen,
in het openbaar te verdedigen
op vrijdag 19 november 1993
des namiddags te vier uur in de Aula
van de Landbouwuniversiteit te Wageningen

The research presented in this thesis was performed at

DLO Winand Staring Centre for Integrated Land,
Soil and Water Research (SC-DLO)
Postbus 125, 6700 AC Wageningen (The Netherlands)

CIP-data Koninklijke Bibliotheek, Den Haag

Brus, Derk Jan

Incorporating models of spatial variation in sampling strategies for soil / Derk Jan Brus. - [S.l. : s.n.]. - Ill.
Thesis Wageningen. - With ref. - With summary in Dutch.
ISBN 90-5485-183-X
Subject headings: spatial variation; soil surveys / sampling strategies; soil surveys / soil science.

Stellingen

I

De veronderstelling dat de klassieke-steekproefbenadering niet valide is in gebieden met ruimtelijke autocorrelatie en de veronderstelde optimaliteit van kriging-voorspellingen hebben geleid tot een hausse van toegepast geostatistisch onderzoek, waardoor de aandacht voor de wijze waarop de monsterlocaties worden geselecteerd ten onrechte is verslapt.

Dit proefschrift

II

De geostatistische benadering houdt geen rekening met de trekkingskansen van bodemmonsters, waardoor in deze benadering de mogelijkheden tot gebruik van voorinformatie bij de bemonstering beperkt zijn.

Dit proefschrift

III

Bij de interpolatietechniek kriging wordt de steekproeffout van het geschatte semivariogram ten onrechte verwaarloosd, waardoor de kriging-variantie vaak een te optimistische schatting is van de voorspellingsnauwkeurigheid.

Dit proefschrift

IV

Bij ruimtelijke interpolatie van bodemverontreiniging ten behoeve van risicobeoordeling verdient de model-ondersteunde benadering waarbij gebruik wordt gemaakt van niet-ergodische variogrammen, de voorkeur boven de geostatistische benadering.

Särndal, C-E, Swensson, B. & Wretman, J. 1992, *Model assisted survey sampling*. Springer-Verlag, New York.

Dit proefschrift

V

Onderzoek zoals o.a. uitgevoerd door Laslett *et al.* (1987) en Voltz & Webster (1990) waarin de verschillen tussen de voorspellingsnauwkeurigheid van interpolatiemethoden niet op hun statistische significantie worden getoetst, hebben een beperkte waarde omdat de conclusies over deze verschillen alleen betrekking hebben op de punten waar de voorspellingsnauwkeurigheid van de methoden is bepaald.

Laslett, G.M., McBratney, A.B., Pahl, P.J. & Hutchinson, M.F. 1987. Comparison of several spatial prediction methods for soil pH. *Journal of Soil Science* 38, 325-341.

Voltz, M. & Webster, R. 1990. A comparison of kriging, cubic splines and classification for predicting soil properties from sample information. *Journal of Soil Science* 41, 473-490.

Dit proefschrift

VI

Als iemand niet naar de kerk gaat kan daaruit niet worden geconcludeerd dat hij of zij niet in God gelooft. Zo kan ook uit het buiten beschouwing laten van ruimtelijke autocorrelatie binnen bodemkaartvlakken niet geconcludeerd worden dat verondersteld wordt dat er geen ruimtelijke autocorrelatie is binnen deze vlakken.

Heuvelink, G.B.M, 1993. *Error propagation in quantitative spatial modelling: applications in geographical information systems.* Proefschrift Universiteit Utrecht.

VII

De bewering dat de bemonsteringsstrategie die voorgeschreven is in het protokol fosfaatverzadigde gronden, zuivere schattingen van de gemiddelde fosfaatverzadigingsgraad van een perceel oplevert, is onjuist.

Van der Zee, S.E.A.T.M., Van Riemsdijk, W.H. & De Haan, F.A.M. 1990. *Het protokol fosfaatverzadigde gronden, deel II (Technische uitwerking).* Agricultural University, Wageningen.

VIII

Het is een wijdverbreid misverstand dat een statistische aanpak van bodembemonstering tot hogere kosten leidt.

IX

Geomorfologische gegevens kunnen een belangrijke rol spelen bij de ruimtelijke interpolatie van bodemkenmerken.

X

Bodemkundige onderzoeksinstituten dienen er voor te zorgen dat ze voldoende bodemgeografische veldkennis in huis houden.

XI

Een wettelijk recht op deeltijdarbeid is een noodzakelijke voorwaarde voor de emancipatie van de vrouw èn de man.

XII

De cursus 'Adviseren voor stafdiensten: invloed uitoefenen zonder hiërarchische macht' voor medewerkers van het Ministerie van Landbouw, Natuurbeheer en Visserij, geeft te denken over de wijze waarop bij dit ministerie invloed wordt uitgeoefend bij aanwezigheid van hiërarchische macht.

Ministerie van Landbouw, Natuurbeheer en Visserij. *Opleidingsgids 1994.*

XIII

Niets is zo praktisch als goede theorie.

Stellingen behorende bij het proefschrift 'Incorporating models of spatial variation in sampling strategies for soil'. Dick Brus, Wageningen,
19 november 1993.

Aan mijn vader en moeder,
Nelleke en Jeroen

Abstract

Brus, D.J., 1993. **Incorporating models of spatial variation in sampling strategies for soil.** Doctoral thesis, Wageningen Agricultural University, Wageningen, The Netherlands. (xii) + 211 pp.

The efficiency of soil sampling strategies can be increased by incorporating a spatial variation model. The model can be used in the random selection of sample points i.e. in the sampling design, or in spatial estimation (prediction). In the first approach inference is based on a sampling design, in the second on a probabilistic model. The advantages and disadvantages of these two approaches, referred to as the design-based and model-based approach, are dealt with from a theoretical and a practical point of view. Estimation by random sampling stratified by soil map unit, and kriging are taken as examples of the two approaches in several case studies.

The commonly accepted belief in geostatistical literature that the design-based approach is not valid in areas with autocorrelation is incorrect. Furthermore, the claimed optimality of the model-based approach is questionable. The two approaches use different criteria for assesment of the quality of estimates, consequently optimum estimation has a different meaning in each approach.

In a regional survey with small observation density (1 observation per 25 ha), estimates of values at points were generally not significantly improved by soil map stratification (α=0.10), neither by estimation with variograms as in kriging. Stratified random sample estimates of values at points were as accurate as those provided by kriging.

In the model-based approach the quality of the estimates depends on the quality of the model. To avoid this, a new approach for spatial estimation is proposed, the model-assisted approach, making use of non-ergodic variograms. This approach incorporates the sampling error of the non-ergodic variogram in the kriging error, making the estimation variance estimates always valid. A

set of new methods is presented for unbiased and robust estimation of the non-ergodic variogram and its sampling error.

Many factors determine the efficiency of an approach that incorporates spatial variation models, making the decision process rather complicated. A simple decision-tree is presented with seven questions related to the aim of the survey (local or global estimation, criteria for assessment of the quality of the estimates), the constraints (available budget and sampling costs) and prior information (soil map).

Additional index words: sampling strategy, soil map, phosphate saturation, spatial variation, model-based inference, design-based inference, model-assisted inference, kriging, stratified random sampling, non-ergodic variogram, bootstrap, robustness, unbiasedness.

Woord vooraf

In tegenstelling tot wat de omslag doet vermoeden, is dit boekje niet het werk van één persoon. Velen hebben bijgedragen aan de totstandkoming van dit proefschrift en ik wil hen daarvoor op deze plaats graag hartelijk bedanken.

In het bijzonder dank ik mijn co-promotor, Jaap de Gruijter. Jaap, jij hebt mij de beginselen van ruimtelijke statistiek bijgebracht, jij hebt veel ideeën aangedragen, en jij hebt mij fantastisch begeleid bij de uitwerking van deze ideeën. Ik prijs mij gelukkig met een zo deskundig afdelingshoofd.

Verder dank ik mijn promotor, Johan Bouma, voor de waardevolle en eerlijke gesprekken over de koers van het onderzoek. Jij zorgde er onder andere voor dat in deze studie niet alleen de statistische significantie maar ook de praktische relevantie van de resultaten wordt belicht. Ook dank ik Auke Breeuwsma voor zijn inbreng in dit onderzoek. Jij en andere medewerkers van jouw afdeling, met name Jos Reijerink, maakten het mogelijk dit onderzoek te richten op de problematiek van de fosfaatverzadigde gronden. Jos, ook jij dus hartelijk dank.

Jan Oude Voshaar van DLO - Groep Landbouwwiskunde wil ik bedanken voor zijn vele deskundige adviezen op het gebied van statistiek en de zeer plezierige en geduldige manier waarop dit is gedaan. Mijn mede-auteurs, Johan Bouma, Auke Breeuwsma, Arnold Bregt, Jaap de Gruijter, Michiel Jansen, Ben Marsman en Reind Visschers bedank ik voor hun kritische en opbouwende opmerkingen bij de concepten van de publikaties. Martin Knotters wil ik graag bedanken voor de goede opmerkingen bij de hoofdstukken 1 en 8.

Verder ben ik veel dank verschuldigd aan medewerkers van de ondersteunende afdelingen van DLO-Staring Centrum. Rini Schuiling bedank ik voor het maken van 'kaart-overlays' en al het andere 'GIS-werk' met ARC-INFO. Bram ten Cate van de afdeling Wetenschappelijke Informatie Overdracht wil ik graag bedanken voor de nauwgezette redactionele bewerking van dit proefschrift.

Voor het tekenen van de fraaie figuren dank ik Dasja ten Cate. Martin Jansen wil ik graag bedanken voor zijn hulp bij het ontwerp van de omslag.

De directie van DLO-Staring Centrum en Ben van der Pouw, hoofd van de hoofdafdeling Landinventarisatie en Landevaluatie, wil ik graag bedanken voor de mogelijkheden die zij mij boden voor dit onderzoek. Ben van der Pouw wil ik tevens bedanken voor het kritisch doorlezen van het hele proefschrift. Ook bedank ik Joop ten Cate, hoofd van de afdeling Veldbodemkunde, voor zijn hardnekkige verzet tegen toepassing van statistiek in het bodemgeografisch onderzoek. Het heeft er onder andere toe bijgedragen dat ik de ambachtelijke bodemkartering niet uit het oog ben verloren.

De psychische begeleiding van dit onderzoek was in handen van de cantorij van de Gereformeerde kerk in Ede o.l.v. Roel Staal. Met name de werken voor koor van J.S. Bach, H. Schütz en D. Buxtehude hadden en hebben voor mij een zeer hoge zangtherapeutische waarde.

Tot slot dank ik Nelleke, mijn vrouw, voor haar geduld en inlevingsvermogen. Abstract denken is jou op het lijf geschreven en hierdoor heb je een wezenlijke bijdrage kunnen leveren aan het uitstippelen en formuleren van de rode draad van het onderzoek.

Contents

Chapter 1

General introduction

General introduction

Motives

In recent years the need for soil survey information with quantified accuracy has grown considerably. Although soil maps exist for all parts of the world, the quality of these soil maps, expressed by the spatial variability of soil properties within map units and map purity, is seldom quantified. Moreover, soil maps focus on information relevant to land evaluation, i.e. for assessment of the land suitability for several land use types. They usually lack information on soil properties relevant to environmental protection studies. Consequently, often additional data have to be collected in the field. This thesis deals with statistical methods for collecting soil data (sampling) and for calculating estimates of areal means and values at points from the data (spatial estimation). By applying statistical methods the accuracy of these estimates can be quantified.

Collecting soil data is generally rather expensive, especially if laboratory measurements of soil physical and soil chemical properties are needed or measurements of difficult to determine properties such as characteristics of the frequency distribution of the water table. Therefore, it is important that efficient sampling strategies are developed; strategies giving maximum accuracy of the estimates at the lowest costs. One way of increasing the efficiency is to incorporate models of spatial variation in the sampling strategy. At the highest level, two approaches can be distinguished: (i) incorporation in the sampling design (sampling plan), where the model is used in the random selection of sample points and (ii) incorporation in the estimator, where the model is used in calculating the estimates from a sample.

An example of the first approach, dealt with in this thesis, is estimation by random sampling stratified by soil map unit. A soil map can be considered as a discrete model of spatial variation of the soil. By soil map stratification we hope to obtain more homogeneous subareas, the means of which can be estimated more accurately than the mean of the whole area for a given sample

size. This is an attractive approach for the Netherlands because there is a soil map at the scale of 1:50 000 for the whole of it and more detailed soil maps (scale 1:10 000) for large parts of it. Samples are selected from the strata by some chance mechanism such that all elements have a nonzero probability of inclusion in the sample, specified by the sampling design of that stratum (see below). These inclusion probabilities are used in the inference, i.e. in calculating the estimates. To estimate the values at points the estimated mean of a map unit is used as the estimator for all points within that map unit. Following Särndal *et al.* (1992) such inference will be referred to as design-based. However, there is a widespread belief in geostatistical literature that design-based inference is not valid if data are autocorrelated. This would severely restrict the applicability of this approach, because soil data often show autocorrelation, also within map units.

An example of the second approach, dealt with in this thesis, is kriging (Journel and Huijbregts, 1978). Kriging uses a probabilistic (geostatistical) model to describe spatial variation. An important part of this model is the variogram. It describes spatial dependence as a function of the lag vector separating two points. This model is believed to be a more natural and adequate description of spatial dependence than a soil map, which neglects spatial autocorrelation within map units. Kriging is qualified as optimal in the sense that its estimates are unbiased and have minimum variance. As opposed to the previous approach, inference is based on the model. Following Särndal *et al.* (1992) such inference will be referred to as model-based.

Purpose

The main purpose of this thesis is to elaborate on the merits of the design-based and model-based approach for incorporation of spatial variation models in sampling strategies, from a theoretical and a practical point of view. Important questions dealt with are:

(i) Is design-based inference indeed invalid in areas with autocorrelated data? (Chapter 2)

(ii) Are kriging estimates also optimal in the sense of classical sampling theory (Chapters 2 and 4) and what is the worth of the calculated kriging

variance as a measure of estimation variance? (Chapter 7)

(iii) Can estimates of global means (Chapter 3) and of values at points (Chapter 4) be improved by soil map stratification, and how strong is this effect? How strong is the effect of estimation with variograms on the accuracy of estimates at points? (Chapter 4) How strong is the effect of soil map stratification plus estimation with variograms on the accuracy of estimates at points? (Chapter 4)

(iv) Which approach gives the most accurate estimates of values at points: estimation by random sampling stratified by soil map, or kriging? (Chapter 4)

(v) Does it pay to revise a soil map before using it in spatial estimation? (Chapter 5)

(vi) What decision-rules can be used for choosing between the two approaches? (Chapters 2-7)

Definitions and scope

In this thesis, the survey region is considered as a finite population of soil auger cores, which implies that the number of different possible sampling locations is finite. This is because the main body of sampling theory is formulated in terms of finite populations. Results and conclusions presented in this thesis also hold if the population is assumed to be infinite. Sampling is the selection of a subset of elements from the population. Note that a sample refers to a set of elements and not to a single element. The probability of selection of any subset (sample) is determined by the sampling design. The elements in the sample are observed; the target variables are measured for each element in the sample. Unless otherwise stated, I assume that the variables are measured without error. The values of the target variables are used to calculate an estimate of the population parameters (e.g. means, totals) or the values at points and of the accuracy of these estimates. The combination of sampling design and formula used to calculate an estimate (estimator) is referred to as a sampling strategy.

The location of the soil auger cores is defined in two dimensions only. Variation of soil properties with depth is incorporated in the definition of the target

variable. Furthermore, I will focus on soil surveys at a regional scale with sampling densities varying from 1 to 10 observation points per 25 ha. Most of the soil properties observed are relevant to the problem of phosphate leaching from agricultural soils to the groundwater and surface water, which is a considerable environmental problem in the Netherlands. Important questions in this context are:

- how much phosphate can be sorbed by the soil in a block of a given size or at a given point before it leaches to the groundwater?
- what is the mean degree of phosphate saturation of a block or at a point?
- what is the fraction of the area saturated with phosphate?

Outline of this thesis

Chapter 2 outlines the perspectives and concepts of design-based and model-based inference of population parameters. The meaning of unbiased and minimum variance estimates obtained by these two types of inference, are compared. I deal with the relative merits of the two approaches and with rules for choosing. In a case study, stratified simple random sample estimates (stratified by a 1:50 000 soil map) of areal fractions saturated with phosphate and their accuracies are compared with those obtained by kriging, for blocks of various size.

In Chapter 3 I calculate the gain in precision of estimates of *the global means* of various phosphate sorption characteristics by using a 1:50 000 soil map and a land use map for stratification in simple random sampling. The precision with a simple random sample (without stratification) of the same size is taken as a point of reference. This is done in two areas with contrasting historical phosphate loads. Three stratifications are evaluated, namely by land use, soil map unit and by both, in combination with three methods of allocating numbers of sample points to the strata.

Chapter 4 assesses the effect of soil map stratification on the accuracy of estimates of four soil properties *at points*. Six estimation methods are examined: global mean, moving average, nearest neighbour, inverse squared distance, Laplacian smoothing splines and ordinary point kriging. I compare

the efficiency of the variogram and the 1:50 000 soil map as spatial variation models by comparing the accuracy of estimates at points obtained by unstratified kriging with that obtained by using the means of soil map units as estimators. Also the effect of the combination of both models is assessed by calculating the accuracy of estimates by kriging within three soil map units (stratified kriging).

The first sheets of the Soil Map of the Netherlands at the scale of 1:50 000 are possibly out-of-date and as a result are poor models of spatial variation. We might think of first revising the soil map and then using the revised map for stratification in spatial estimation. The alternative would be to leave the map unchanged and to spend all the money collecting additional data at points. In Chapter 5 the efficiency of four strategies for updating soil maps is compared: (i) revision, (ii) upgrading, (iii) revision plus upgrading and (iv) upgrading by two-phase sampling. In revision all funds are used for model improvement, whereas in upgrading and upgrading by two-phase sampling all funds are used for sampling. In revision plus upgrading the funds are distributed between the two activities. The efficiency is assessed for design-based estimation of the global mean and the values at points.

Kriging directly provides an estimate of the accuracy of the estimates, the kriging variance. However, the mean kriging variance often differs considerably from the empirical mean squared error of estimation: generally it is an underestimate. In this thesis I analyze the causes of this discrepancy and show how it can be eliminated by using non-ergodic variograms in kriging, i.e. variograms describing spatial dependence in the area actually sampled only (Isaaks and Srivastava, 1988). In Chapter 6 I propose a new set of sampling strategies for the estimation of these non-ergodic variograms. An important advantage of these strategies is that they yield unbiased and robust estimates of the sampling error of the estimated variogram.

Chapter 7 shows that incorporation of this sampling error in the error of estimates obtained by kriging eliminates the main cause of the discrepancy between the mean kriging variance and the mean squared error of estimation. Thus, a new approach for incorporation of spatial variation models in sampling

strategies emerges, the model-assisted approach. I explain how the role of the model differs fundamentally between the model-based and model-assisted approach.

Finally, the main results and major conclusions of this thesis are presented in Chapter 8. In this chapter I also present a decision-tree for choosing between the design-based, model-based and model-assisted approach.

Notation

Variables such as soil properties, and model parameters will be printed in italics. To distinguish between fixed and random variables, random variables will be underlined. Symbols for matrices and vectors will be printed in bold face italics.

References

Isaaks, E.H. & Srivastava, R.M. 1988. Spatial continuity measures for probabilistic and deterministic geostatistics. *Mathematical Geology* 20, 313-341.

Journel, A.G. & Huijbregts, C.J. 1978. *Mining Geostatistics*. Academic Press, London.

Särndal, C-E, Swensson, B. & Wretman, J. 1992. *Model assisted survey sampling*. Springer-Verlag, New York.

Chapter 2

Design-based versus model-based estimates of spatial means. Theory and application in environmental soil science

Environmetrics 4, 123-152 (1993)

with Jaap J. De Gruijter

Design-based versus model-based estimates of spatial means. Theory and application in environmental soil science

The perspectives and concepts of the classical sampling or design-based approach and the geostatistical or model-based approach are compared. We show that unbiasedness and minimum variance in the design-based approach is quite different from that in the model-based approach and that design-based strategies are always valid, whether or not there is spatial autocorrelation. Model-based predictions of spatial means will generally not have the desirable property of unbiasedness in the design-based sense. Model-based strategies contain a risk arising from biased selection of sample points and they do not allow the accuracy of predictions to be assessed objectively, i.e., based on the sample data alone. Model-based strategies are useful for local estimation i.e. for many small blocks and points, provided that there are enough data to estimate the variogram. In a case study on phosphate saturation, design-based and model-based estimates of the areal fractions saturated with phosphate were similar, but with smaller blocks the differences between the estimates provided by the two approaches, increased to magnitudes of practical importance.
KEY WORDS: *sampling strategy; spatial dependence; unbiasedness; phosphate saturation.*

Introduction

Since about 1960 random sampling designs have been successfully applied in soil geography, especially to estimate spatial means and variances of areas. For estimation of the values at points, these strategies seemed to be not very successful due to the large internal variance of the soil map units, whose estimated means were used as estimators (Morse and Thornburn, 1961; Webster and Beckett, 1968; Beckett and Webster, 1971; Marsman and De Gruijter, 1986). Since 1980 numerous papers have recommended kriging as a better technique for local estimation of soil properties. See for instance Burgess and Webster (1980), Burgess *et al.* (1981), Webster (1985) and Oliver *et al.* (1989). Kriging uses a probabilistic model

describing spatial autocorrelation or related quantities as a function of the lag vector separating two points and this model is believed to be a more natural and adequate description of spatial autocorrelation than a soil map which neglects spatial autocorrelation within map units. Kriging is qualified as optimal in the sense that its estimations are unbiased and have minimum variance. Kriging is also used for global estimation (block kriging), so there are two rather different approaches for estimating means of soil properties in areas, referred to as the design-based and the model-based approach. However, there is a rather wide-spread belief in the literature on applied geostatistics that design-based strategies are not valid in areas with autocorrelated data (see e.g. Russo and Bresler, 1981; Olea, 1984; Dahiya *et al.*, 1985; Yfantis *et al.*, 1987; Barnes, 1988; Burrough, 1991).

The purpose of this paper is:
- to make clear that classical sampling theory and design-based strategies are valid with autocorrelated data;
- to draw the attention of soil scientists to the different meanings of unbiasedness and minimum variance in model-based and design-based sampling strategies;
- to work out a case study to see whether differences in design-based estimates and model-based predictions of spatial means (areal fractions) are relevant to practice;
- to elaborate the relative merits of the two methods and to develop rules for choosing between them.

We estimated the fractions of square areas saturated with phosphate by stratified simple random sampling as an example of the design-based approach and by indicator block kriging as an example of the model-based approach. In stratified simple random sampling we used an existing soil map for stratification, which is a way of using an a priori model of spatial structure in the sampling design. In indicator block kriging, spatial structure was modelled by a single variogram for the entire area. Separate variograms for the soil map units might have given better results but it is usually impossible to estimate them well in practice for lack of data. We preferred to stick to methods that are practicable. Moreover it was *not* the purpose of our study

to determine which method gives the 'best' estimates. The true spatial means must be known for this. Also, one should either select the design-based or the model-based definition of 'best estimate' or choose a hybrid definition.

Hansen *et al.* (1983) discussed the use of models in sample surveys of finite populations in general. They chose the design-based quality criteria and they showed that small mistakes in the estimated parameters of a regression model describing the relation between the target variable and an auxiliary variable, may lead to substantial bias of the predicted population mean and statements about the sampling errors of those estimates may be very misleading. As the model is never known exactly, they preferred design-based estimators. This raises the question of whether kriging behaves similarly to regression. In this context the paper of Borgman and Quimby (1988) is very interesting. They discussed the advantages and disadvantages of classical random sampling and geostatistical sampling and they stated that probably the greatest shortcoming of the geostatistical approach is the difficulty in knowing when various model assumptions are acceptable or not acceptable.

Design-based and model-based sampling strategies

To clarify the differences between the design-based and model-based approach in spatial sampling we give an example of sampling soil profiles from a hypothetical area A with known values of property z at all points in this area (Fig. 2.1a). The soil profiles are the sampling units. Let us assume that variable z has only two possible values, 0 and 1, and that we want to estimate the spatial mean of z within A, defined as:

$$m_A \equiv \frac{1}{N} \sum_{i=1}^{N} z(\boldsymbol{x}_i) \qquad (2.1)$$

where:

m_A = true spatial mean of z in A;

N = total number of soil profiles (possible sampling locations) in A (population size);

$z(\boldsymbol{x}_i)$ = value of z at location $\boldsymbol{x}_i$, where $\boldsymbol{x}_i$ is the vector of co-ordinates of point i.

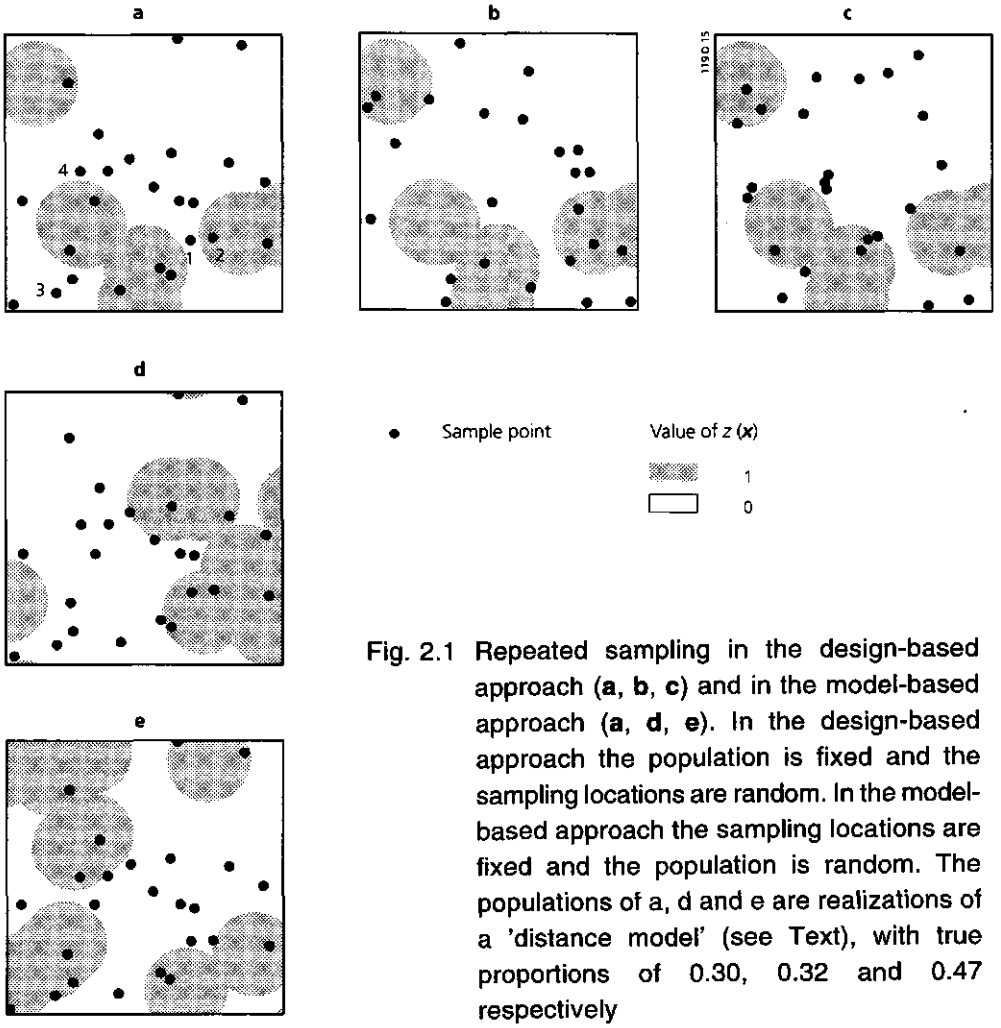


Fig. 2.1 Repeated sampling in the design-based approach (**a**, **b**, **c**) and in the model-based approach (**a**, **d**, **e**). In the design-based approach the population is fixed and the sampling locations are random. In the model-based approach the sampling locations are fixed and the population is random. The populations of a, d and e are realizations of a 'distance model' (see Text), with true proportions of 0.30, 0.32 and 0.47 respectively

For convenience, we assume that N is finite in Equation (2.1) and following equations, which implies that the number of different possible sampling locations is finite. If the population size is assumed to be infinite, then sums should be replaced by integrals, but this has no influence on the results.

The spatial mean of a variable with 0 and 1 as possible values equals the proportion of the area with $z = 1$. To estimate this we selected n locations and measured property z on the soil profiles there. The values obtained will be indicated as $z(\boldsymbol{x}_i)$. Either z or $\boldsymbol{x}_i$ could be random.

The design-based approach

In the design-based approach the n locations are selected randomly. Figure 2.1a shows a simple random sample with $n = 25$. At a given location the value of z is assumed to be fixed, but unknown. Although these values are fixed, the locations are selected randomly, and so we denote the data by $z(\underline{\boldsymbol{x}}_i)$.

The probability of selection for any given subset of soil profiles is determined by the sampling design p. These probabilities are known and are the key to describing the sample-to-sample variation of a proposed estimator (Särndal *et al.*, 1992). Summing the selection probabilities of all subsets with soil profile i in it, gives the inclusion probability of profile i. In the Horvitz-Thompson estimator or π-estimator (Särndal *et al.*, 1992) the values of the soil profiles are divided by their inclusion probabilities. For the spatial mean this results in the following estimator:

$$\underline{\hat{m}}_A = \frac{1}{N} \sum_{i=1}^{n} \frac{z(\underline{\boldsymbol{x}}_i)}{\pi_i} \tag{2.2}$$

where:

$\underline{\hat{m}}_A$ = estimated spatial mean of area A;

n = sample size;

π_i = inclusion probability of soil profile i.

With simple random sampling (Cochran, 1977), for instance, the probability of selection of any subset of size n is equal, or in other words all configurations of n points have equal probability. From this it follows that the inclusion probability of a soil profile is equal for all profiles and equals n/N, the sampling fraction. This results in the following estimator of the spatial mean:

$$\hat{\underline{m}}_A = \frac{1}{N} \sum_{i=1}^{n} \frac{z(\underline{x}_i)}{\pi_i} = \sum_{i=1}^{n} \frac{1}{n} z(\underline{x}_i) \qquad (2.3)$$

Note that with this design all points get equal weight.

Once a sample has been taken and the spatial mean has been estimated, one may wonder what would happen if another sample is taken from the area using the same design. This is what is done in the design-based approach; not in reality, but in the mind. Repeated sampling according to the design-based approach is illustrated in Figure 2.1a-c, showing three simple random samples of size 25. The three samples came from different locations, and so the estimated value from Equation (2.3) also differed from sample to sample: 0.32 (= 8/25), 0.36 (= 9/25) and 0.32 for Figure 2.1a, 2.1b and 2.1c respectively. So the estimation errors (the difference between estimated and true mean) were 0.02, 0.06 and 0.02. Although usually only one sample is taken from an area, in the design-based approach the mean and variance of the estimator over repeated sampling under a given design p plays an important role. This type of mean and variance is indicated as the p-expectation, Equation (2.4), and the p-variance or sampling variance, Equation (2.5), and are defined as:

$$E_p(\hat{\underline{m}}_A) \equiv \sum_{s=1}^{S} p(s) \hat{m}_{A,s} \qquad (2.4)$$

and

$$\sigma_p^2(\hat{\underline{m}}_A) \equiv E_p[(\hat{\underline{m}}_A - E_p(\hat{\underline{m}}_A))^2] \qquad (2.5)$$

where:

$E_p(\hat{\underline{m}}_A)$ = expectation of estimated spatial mean of A over repeated sampling under design p;

$p(s)$ = selection probability of sample s;

S = number of possible samples under design p;

$\hat{m}_{A,s}$ = spatial mean of A estimated by sample s;

$\sigma_p^2(\hat{\underline{m}}_A)$ = sampling variance of estimated spatial mean of A under design p.

The π-estimator of the sampling variance of the mean of a fixed size sampling design is equal to (Särndal *et al.*, 1992, p. 45):

$$\hat{\sigma}_p^2(\hat{\underline{m}}_A) = \frac{1}{2N^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\pi_i\pi_j - \pi_{i,j}}{\pi_{i,j}}\left(\frac{z(\underline{x}_i)}{\pi_i} - \frac{z(\underline{x}_j)}{\pi_j}\right)^2 \qquad (2.6)$$

where:

$\pi_{i,j}$ = probability that both of the soil profiles i and j will be included in the sample.

With simple random sampling this sampling variance equals (Särndal *et al.*, 1992, p. 46):

$$\hat{\sigma}_p^2(\hat{\underline{m}}_A) = \frac{\hat{\underline{V}}_A}{n}\left(1 - \frac{n}{N}\right) \approx \frac{\hat{\underline{V}}_A}{n} \qquad (2.7)$$

where:

$\hat{\underline{V}}_A$ = estimated spatial variance of A:

$$\underline{\hat{v}}_A = \frac{\sum_{i=1}^{n} (z(\underline{x}_i) - \underline{\hat{m}}_A)^2}{n-1} \qquad (2.8)$$

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

$$E_p(\underline{\hat{m}}_A) = m_A \qquad (2.9)$$

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. Cochran (1977, p. 22) showed that the estimator of the spatial mean of Equation (2.3) is p-unbiased with simple random sampling. Apart from p-unbiasedness, another desirable property of the estimator is that its Mean Squared Error (MSE_p) is minimal:

$$MSE_p \equiv E_p\{(\underline{\hat{m}}_A - m_A)^2\} \qquad (2.10)$$

If an estimator is p-unbiased, its MSE_p is equal to its sampling variance (Eq. 2.5), otherwise MSE_p will exceed the sampling variance.

Finally, the inclusion probabilities need not be equal for all soil profiles to be able to obtain unbiased estimates of the spatial mean. A well known example of a sampling design with unequal inclusion probabilities is stratified simple random sampling with optimum allocation. For a well defined design the inclusion probabilities are known and Equations (2.2) and (2.6) can be used to estimate the spatial mean and its sampling variance.

The model-based approach

In the model-based approach the actual population from which we take the sample is considered to be just one population out of an infinite set of possible populations having in common that they are realizations from the same N-dimensional joint distribution ξ. Consequently, this approach is also referred to as the superpopulation approach (Cassel *et al.*, 1977). One may think of these realizations as areas in which the same soil forming processes

have led to a more or less similar spatial pattern. So, in the model-based approach each soil profile (location) is not associated with one fixed value, as in the design-based approach, but with different possible values, each with a defined probability of occurring, thus forming a random variable. The actual value found at a given location i is interpreted as the realized outcome of the random variable $\underline{z}(\boldsymbol{x}_i)$. (Note the change in underlining.) Therefore, the true spatial mean is also assumed to be random:

$$\underline{m}_A \equiv \frac{1}{N} \sum_{i=1}^{N} \underline{z}(\boldsymbol{x}_i) \tag{2.11}$$

As the spatial mean is random we shall use the word 'prediction' instead of estimation. We shall use the superscript ~ in symbols for predicted variables. This approach is termed 'model-based' because its essential tool is a model describing the N-dimensional joint distribution ξ of the variables $\underline{z}(\boldsymbol{x}_i)$. (Often the symbol ξ is attached to the model.) In practice, this is usually done by characterizing the mean and variance of $\underline{z}(\boldsymbol{x}_i)$ and the covariance of $\underline{z}(\boldsymbol{x}_i)$ and $\underline{z}(\boldsymbol{x}_j)$ for any i and j. Populations are viewed as realizations from ξ. Figure 2.1a, d and e, for instance, show three realizations from a 'distance model' with true spatial means (proportions) of 0.30, 0.32 and 0.46 respectively. A distance model can be described by (Matérn, 1960, p. 37-39,49):

$$\underline{z}(\boldsymbol{x}_i) = \sum_{c=1}^{\underline{C}} \mathrm{f}(\boldsymbol{x}_i - \underline{\boldsymbol{x}}_c) \tag{2.12}$$

where:

$\underline{\boldsymbol{x}}_c$ = vector of co-ordinates of the cth centre (random);
$\underline{C}$ = number of centres (random).

In Figure 2.1a, 2.1d and 2.1e the centres are produced by a Poisson process with intensity $\lambda = 6$. For $\mathrm{f}(\boldsymbol{x}_i)$ we took simply:

$$f(\boldsymbol{x}_i) = 1 \quad \text{if } \underline{\delta}_{min}(\boldsymbol{x}_i) \leq D$$
$$f(\boldsymbol{x}_i) = 0 \quad \text{otherwise} \tag{2.13}$$

where:
$\underline{\delta}_{min}(\boldsymbol{x}_i)$ = smallest distance of point $\boldsymbol{x}_i$ to the centres;
D = predetermined, fixed distance.

Thus $\underline{z}(\boldsymbol{x}_i) = 1$ if at least one centre is within distance D from $\boldsymbol{x}_i$.

In reality one can observe only a single realization from ξ, and so the mean and variance of $\underline{z}(\boldsymbol{x}_i)$ are unknown. To circumvent this difficulty the random process is assumed to be ergodic i.e., its statistics can be determined from a single realization (Papoulis, 1965, p. 327). In practice it is often assumed that the covariance of $\underline{z}(\boldsymbol{x}_i)$ and $\underline{z}(\boldsymbol{x}_j)$ depends only on the distance and direction separating $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. Half the variance of $\{\underline{z}(\boldsymbol{x}_i) - \underline{z}(\boldsymbol{x}_j)\}$ is referred to as the semivariance, which is often modelled instead of the covariance.

This spatial covariance or semivariance is used in predicting. The spatial mean of a block is predicted by:

$$\underline{\tilde{m}}_A = \sum_{i=1}^{n} \lambda_i \underline{z}(\boldsymbol{x}_i) \tag{2.14}$$

where:
λ_i = weight attached to the ith point.

In general these weights will be different for different points. This is also true for a simple random sample! Let us see how these weights are calculated in this approach.

As in the design-based approach interest is in the mean and the variance of the predictor over repeated sampling. However, repeated sampling has a different meaning now and is possible only in theory: the sample is taken at the same locations but in different hypothetical areas from the same joint distribution ξ. This is illustrated by the vertical row of maps in Figure 2.1 (a,

d and e). These figures show that the values at some sampling locations differ between the three realizations, hence the predicted spatial means will also be different. The mean and variance of the predicted value is referred to as the ξ-expectation, Equation (2.15), and ξ-variance, Equation (2.16). Cassel *et al.* (1977, p. 82) defined these in terms of Lebesgue integrals (Papoulis, 1965, p. 141):

$$E_\xi(\underline{\tilde{m}}_A) \equiv \int_{R_N} \underline{\tilde{m}}_A d\xi \tag{2.15}$$

and

$$\sigma^2_\xi(\underline{\tilde{m}}_A) \equiv E_\xi[\{\underline{\tilde{m}}_A - E_\xi(\underline{\tilde{m}}_A)\}^2] = \int_{R_N} \{\underline{\tilde{m}}_A - E_\xi(\underline{\tilde{m}}_A)\}^2 d\xi \tag{2.16}$$

where:

E_ξ = expectation over realizations from distribution ξ (model mean);
σ_ξ^2 = variance over realizations from distribution ξ;
R_N = *N*-dimensional Euclidian space of possible realizations from the distribution ξ.

To find an optimal predictor for $\underline{m}_A$, the prediction error (the difference between the predicted value and the true value) of each realization is considered. The mean error over realizations of the model ξ should be zero (ξ-unbiasedness) and the Mean Squared Error should be minimal (minimal MSE_ξ):

$$\begin{aligned} &\text{Minimize } E_\xi\{(\underline{\tilde{m}}_A - \underline{m}_A)^2\} \\ &\text{subject to } E_\xi(\underline{\tilde{m}}_A - \underline{m}_A) = 0 \end{aligned} \tag{2.17}$$

In Equation (2.17) the true spatial mean, $\underline{m}_A$, is inside the brackets because this quantity is random too. The optimal weights are obtained by solving the equations:

$$\sum_{j=1}^{n} \lambda_j C(\boldsymbol{x}_i,\boldsymbol{x}_j) + \psi = \bar{C}(\boldsymbol{x}_i;A) \qquad \forall\ i = 1 \text{ to } n$$
$$\sum_{j=1}^{n} \lambda_j = 1 \tag{2.18}$$

where:

$C(\boldsymbol{x}_i,\boldsymbol{x}_j)$ = covariance of $\underline{z}(\boldsymbol{x}_i)$ and $\underline{z}(\boldsymbol{x}_j)$;

ψ = Lagrange multiplier;

$\bar{C}(\boldsymbol{x}_i,A)$ = mean covariance of $\underline{z}(\boldsymbol{x}_i)$ and A defined as:

$$\bar{C}(\boldsymbol{x}_i,A) \equiv \frac{1}{N}\sum_{j=1}^{N} C(\boldsymbol{x}_i,\boldsymbol{x}_j) \tag{2.19}$$

The ξ-variance of the predicted spatial mean equals:

$$\sigma_\xi^2(\underline{\tilde{m}}_A) = \bar{C}(A,A) + \sum_{i=1}^{n}\sum_{i=1}^{n} \bar{C}(\boldsymbol{x}_i,\boldsymbol{x}_j) - 2\sum_{i=1}^{n} \lambda_i\, \bar{C}(\boldsymbol{x}_i,A) = \bar{C}(A,A) - \sum_{i=1}^{n} \lambda_i\, \bar{C}(\boldsymbol{x}_i,A) - \psi \tag{2.20}$$

where:

$\bar{C}(A,A)$ = mean covariance of $\underline{z}(\boldsymbol{x}_i)$ in A, defined as:

$$\bar{C}(A,A) \equiv \frac{1}{N^2}\sum_{i=1}^{N}\ \sum_{j=1}^{N} C(\boldsymbol{x}_i,\boldsymbol{x}_j) \tag{2.21}$$

The Best Linear Unbiased Predictor of $\underline{m}_A$ and its variance can also be obtained as Generalized Least Squares solutions of the corresponding regression problems, avoiding the use of a Lagrange multiplier (Corsten, 1989):

$$\underline{\tilde{m}}_A = (\mathbf{1}'\boldsymbol{C}_S^{-1}\mathbf{1})^{-1}\mathbf{1}'\boldsymbol{C}_S^{-1}\underline{\boldsymbol{z}} + \bar{\boldsymbol{C}}_{S,A}\boldsymbol{C}_S^{-1}[\underline{\boldsymbol{z}} - \mathbf{1}(\mathbf{1}'\boldsymbol{C}_S^{-1}\mathbf{1})^{-1}\mathbf{1}'\boldsymbol{C}_S^{-1}\underline{\boldsymbol{z}}] \tag{2.22}$$

where:

$\boldsymbol{C}_S$ = matrix with covariances between sample points;
$\bar{\boldsymbol{C}}_{S,A}$ = vector with mean covariances between each sample point and all points in A

The covariance function of the distance model of Figure 2.1 is known (Matérn 1960, p. 38):

$$C(\boldsymbol{x}_i,\boldsymbol{x}_j) = e^{-2\lambda\pi r^2}\{e^{\lambda A(r,h)} - 1\} \qquad (2.23)$$

where:
r = radius of the circles;
λ = intensity of the Poisson proces;
$A(r,h)$ = area of the intersection of two circles with radius r and the centres $h = |\boldsymbol{x}_i - \boldsymbol{x}_j|$ apart.

Using this function, the optimal predictions were 0.38, 0.34 and 0.46 for Figure 2.1a, 2.1d and 2.1e respectively. So the prediction errors were 0.08, 0.02 and -0.01. Notice the small error of Figure 2.1e despite the large predicted value (0.46). This is because the true mean of Figure 2.1e was also large compared to that of Figure 2.1a (0.47). In practice the covariance function is unknown and should be estimated from sample data as well.

Dependent or independent?
In literature on applied geostatistics it is often stated that classical sampling theory is inapplicable to spatial sampling of soil properties because this theory assumes data to be independent. We shall show that this view is incorrect and we shall make clear that in the design-based approach independence is not *assumed* but *created* by the sampling design.

Stochastic dependence or independence is not a property of any population or region (De Gruijter and Ter Braak, 1990, 1992). It can be a property of a set of variables. Two random variables are independent if (Papoulis, 1965, p. 40):

$$P(v,w) = P(v)\ P(w) \qquad (2.24)$$

where:

$P(v,w)$ = probability that the first random variable takes the value v and the second random variable takes the value w;

$P(v)$ = probability that the first random variable takes the value v;

$P(w)$ = probability that the second random variable takes the value w.

The next question is what makes the variables random, or in other words, what is the source of stochasticity? In the design-based approach this is the sampling process. The locations are selected at random. If the locations $\underline{x}_i$ and $\underline{x}_j$ are selected at random and independently then the corresponding variables $z(\underline{x}_i)$ and $z(\underline{x}_j)$ are mutually stochastically independent (De Gruijter and Ter Braak, 1990). We noticed that this is hard to understand for some model-based thinkers. Therefore we illustrate it by an experiment, which is not a proof and superfluous for those who are familiar with design-based sampling. The experiment consists of taking a simple random sample with replacement of size two from area A of Figure 2.1a, repeating it 1000 times and counting the number of times we have equal values at the two points (both zeros or ones). Large or small numbers indicate dependency of variables: positive dependency for large and negative dependency for small numbers. Assuming independence the expected number equals $1000\{p^2 + (1-p)^2\}$, where p is the probability of getting value 1 which is equal to the areal proportion with value 1. In Figure 2.1a p equals 0.3 and therefore the expected number is 580. The realized number was 578. One may conceive this number as an estimate of the expected number *if data are independent*. The 95% confidence interval of this estimate equals [547, 608] which covers the expected value, and so this experiment indicates that the two variables are independent.

In the model-based approach stochasticity is introduced in a completely different way, namely via the model ξ. In this approach not the sampling locations but the model-realizations are drawn at random and independently from each other. The sampling locations stay where they are. To illustrate

this we executed an experiment for the points marked 1 and 2 in Figure 2.1a. We drew 1000 realizations of the distance model and counted the number of times with equal values at both points which was 804, so much larger than the expected and realized number in the previous experiment. The 95% confidence interval of this prediction is [778, 827]. However, in this approach one has many realizations and therefore the expected number over all model-realizations should be calculated. The mean of the areal proportions with value 1 over all model-realizations (ξ-expectation of areal proportion) is known and therefore also the probability of getting a 1. It is equal to 0.35. From this it follows that, assuming independence, the expected number is 545. The 95% confidence interval does not cover this value. Therefore independence is *not* likely for points 1 and 2 in Figure 2.1a. Remember that these points were selected at random and independently from each other. We repeated this experiment for points 3 and 4 of Figure 2.1a. The counted number is 539 with a 95% confidence interval of [508, 570] covering the expected number, so that independence is likely for points 3 and 4.

In conclusion, values measured at points selected randomly and independently from each other are independent in the design-based approach and at the same time can be dependent in the model-based approach.

Smoothing

Let us now suppose we want to estimate z or predict $\underline{z}$ at all points in A or the mean of z or $\underline{z}$ of m subareas (blocks) of A. For any point or block individually, the estimate should be p- or ξ-unbiased and the p- or ξ-variance should be minimal. On the other hand, one may also be interested in the quality of the estimates of the point- or block values as a whole. For example, one may wish that the spatial variance of the estimated values is close to that of the true values. Although simulation of random fields is the appropriate way to reach this, one may wonder how 'bad' results are if values are estimated or predicted optimally one by one.

In the design-based approach, the spatial variance of the true block-means, Equation (2.25), and of the estimated block-means, Equation (2.26), are defined as:

$$v_A(m_b) \equiv \frac{1}{N} \sum_{b=1}^{B} N_b (m_b - m_A)^2 \qquad (2.25)$$

and

$$v_A(\underline{\hat{m}}_b) \equiv \frac{1}{N} \sum_{b=1}^{B} N_b \, (\underline{\hat{m}}_b - \underline{\hat{m}}_A)^2 \qquad (2.26)$$

where:

$v_A(m_b)$ = spatial variance of true block-means in A;
$v_A(\underline{\hat{m}}_b)$ = spatial variance of estimated block-means in A;
m_b = spatial mean of block b;
$\underline{\hat{m}}_b$ = estimated spatial mean of block b;
B = number of blocks;
N_b = number of soil profiles (possible sampling locations) in block b.

Note that in Equations (2.25) and (2.26) possible differences in block size are taken into account.

Although the true block-means are unknown, the spatial variance of these true block-means can be estimated by:

$$\underline{\hat{v}}_A(m_b) = \underline{\hat{v}}_A - \frac{1}{N} \sum_{b=1}^{B} N_b \, \underline{\hat{v}}_b \qquad (2.27)$$

where:

$\underline{\hat{v}}_A(m_b)$ = estimated spatial variance of true block-means;
$\underline{\hat{v}}_A$ = estimated spatial variance of A;
$\underline{\hat{v}}_b$ = estimated spatial variance of block b.

In terms of variance components Equation (2.27) states that the estimated between-block variance of true block-means equals the estimated total variance minus the estimated pooled within-block variance. The spatial variance

of estimated block-means will be larger than that of the true block-means due to sampling errors. Therefore, in the design-based approach we have the reverse effect of smoothing. We quantified this effect by the relative variance, defined as the ratio of the variance of estimated block-means to the estimated variance of the true block-means:

$$r_v \equiv \frac{v_A(\underline{\hat{m}}_b)}{\underline{\hat{v}}_A(m_b)} \tag{2.28}$$

In the design-based approach the relative variance will be larger than 1.

In the model-based approach the spatial variance of true and predicted block-means are defined in a similar manner. The only difference is that the true spatial means are random. Therefore, the spatial variance of true block-means is random too:

$$v_A(\underline{m}_b) \equiv \frac{1}{N} \sum_{b=1}^{B} N_b \, (\underline{m}_b - \underline{m}_A)^2 \tag{2.29}$$

In the model-based approach the spatial variance of predicted block-means will be smaller than the estimated spatial variance of true block-means due to the well-known smoothing effect (Journel and Huijbregts, 1978, p. 451). So the relative variance, r_v, Equation (2.28), will be smaller than 1.

Smoothing may be reduced by using a limited number of sampling points in the neighbourhood of the block (Journel and Huijbregts, 1978, p. 451), but the choice of a neighbourhood is not part of the theory, hence it remains arbitrary.

Case study

To illustrate the differences in procedure and to see whether the differences between estimated and predicted means are relevant to practice, we used the data of a 'real-world' project on the susceptibility of soils to leaching of

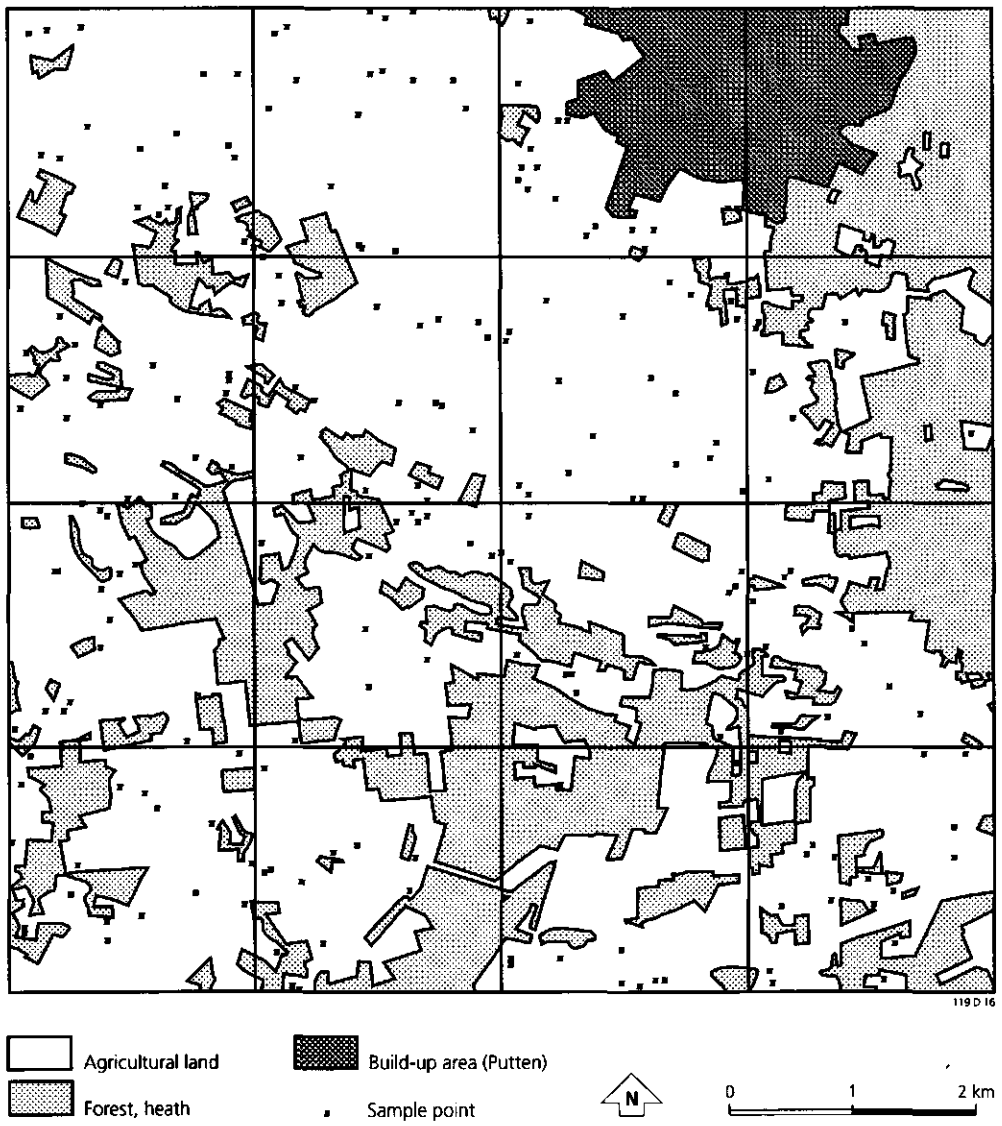


Fig. 2.2 *Stratified simple random sample from the study area Schuitenbeek*

phosphate (Breeuwsma *et al.*, 1989). Phosphate leaching to groundwater and surface water is a considerable environmental problem in large parts of the Netherlands. We estimated the areal fraction saturated with phosphate

by taking a stratified simple random sample from the area (Fig. 2.2). For the purpose of this study, we used the same data to predict this areal fraction by indicator block kriging.

Study area

The study area Schuitenbeek is 25 km north of Wageningen, near the village of Putten. It is part of the central coversand area. There is intensive livestock farming here and as a result there is a large surplus of manure. Only part of it is taken away. More phosphate is applied in manure than crops can take up, and this may lead to the soils becoming saturated with phosphate (Breeuwsma and Schoumans, 1987).

The study area is 8 km x 8 km, showing a great diversity in soil types, typical for coversand areas. Plaggepts, Plaggeptic Haplaquods and Typic Haplaquods (Soil Survey Staff, 1975) on coversand ridges alternate with Typic Humaquepts in shallow valleys and in the lower areas between coversand ridges. In the east, Entic Haplorthods and Typic Haplohumods occur on the coversand belt of an ice-pushed ridge formed during the Saalien glaciation and on the ice-pushed ridge itself, consisting of coarse, fluviatile sediments with gravel. The largest part of the ice-pushed ridge is covered with forest.

Areal fraction saturated with phosphate

Van der Zee (1988) showed that the concentration of phosphate in water at the top of the saturated zone after a long period of leaching depends largely on the areic mass of P_2O_5 sorbed by soil (P) and the maximum areic mass of P_2O_5 which can be sorbed by that soil (P_{max}). In this study P and P_{max} (both in kg P_2O_5 per m^2) were defined for the volume of soil above the mean highest water table ($\bar{W}_{min}$; m below surface). P_{max} was determined indirectly by measuring the oxalate-extractable aluminium and iron content in soil horizons above $\bar{W}_{min}$ and using these contents in the following regression equation (Breeuwsma and Silva, 1992):

$$P_{max} = \sum_{l=1}^{L} (4.6 + 0.39\, M_{ox,l})\, \delta_l \rho_l\, 0.71 \qquad (2.30)$$

where:

4.6 and 0.39	=	regression parameters (mol kg^{-1}, dimensionless);
L	=	number of horizons above $\bar{W}_{min}$;
$M_{ox,l}$	=	oxalate-extractable aluminium + iron content of soil horizon l (mol kg^{-1});
δ_l	=	for $l = 1...L-1$: thickness of soil horizon l; for $l = L$: thickness of Lth horizon above $\bar{W}_{min}$ (m);
ρ_l	=	volumic mass of soil horizon l (kg m^{-3}).
0.71	=	conversion coefficient (kg P_2O_5 mol^{-1}).

We measured ρ_l by taking volumetric samples by a gouge auger.

The ratio of P and P_{max}, referred to as the relative mass of phosphate sorbed by soil (P_{rel}), was used as a measure of phosphate saturation of the soil.

The areal fraction saturated with phosphate (A_c) is defined as the proportion of the area in which the relative mass of phosphate (P_{rel}) of soil profiles exceeds a critical value c (Van der Zee *et al.*, 1990).

We took 0.25 as a critical P_{rel}-value, which is mentioned in the protocol "Phosphate-saturated Soils", an official document specifying a standardized sampling design (Van der Zee *et al.*, 1990). To make our study more general, we also used 0.35 and 0.45 as critical values. These values appear to be the 16th, 33th and 53th percentiles, respectively, of the frequency distribution of the P_{rel}-values in the area.

Design-based approach: stratified simple random sampling

Forest and heath were excluded from the sampling frame because we wanted to estimate the areal fraction saturated with phosphate, A_c, of agricultural land only. The agricultural area was stratified according to the map units of an existing soil map at a scale of 1:50 000. We expected the map units

to have different mean P_{max} and therefore possibly also different mean A_c. This expectation was based on observed differences in $\bar{W}_{min}$ and a strong correlation between $\bar{W}_{min}$ and P_{max} (Brus *et al.*, 1992). Map units supposed to have an approximately equal P_{max} were grouped together. By stratifying according to soil map unit we expected to increase the efficiency of the sampling. To get separate estimates of 2 km x 2 km blocks, we also stratified according to 2 x 2 km blocks. Groups of map units within blocks were used as strata.

We allocated 222 points proportionally to the areal extent of the strata but with a minimum of two. The spatial mean, Equation (2.31), and its sampling variance, Equation (2.32), were estimated by (Cochran, 1977 p. 91):

$$\underline{\hat{m}}_A = \frac{1}{N} \sum_{h=1}^{H} N_h \underline{\hat{m}}_h \tag{2.31}$$

and

$$\hat{\sigma}^2_p(\underline{\hat{m}}_A) = \frac{1}{N^2} \sum_{h=1}^{H} N_h^2 \frac{\underline{\hat{v}}_h}{n_h} \tag{2.32}$$

where:
$\underline{\hat{m}}_h$ = estimated spatial mean of stratum h;
N_h = area of stratum h;
$\underline{\hat{v}}_h$ = estimated spatial variance of stratum h;
H = number of strata;
n_h = number of sample points in stratum h.

To estimate the A_c, first the P_{rel}-values at the sample points were transformed according to:

$$\begin{aligned} i_c &= 1 \text{ if } P_{rel} \geq c \\ i_c &= 0 \text{ otherwise} \end{aligned} \tag{2.33}$$

Model-based approach: indicator block kriging

Several model-based techniques can be used to predict the proportion of the area with $P_{rel} \geq c$: disjunctive kriging, multigaussian kriging, indicator kriging and probability kriging. We used indicator block kriging (*IK*) because in this technique P_{rel}-values are transformed likewise (Eq. 2.33) and only these transformed values are used in predicting. In doing so, results are comparable. Using indicator block kriging, the indicator value at a given location is interpreted as a random variable which has a Bernouilli distribution with ξ-expectation (Journel, 1983, p. 451):

$$E_\xi\{\underline{i}_c(\boldsymbol{x})\} = 1\ \text{Prob}[\underline{z}(\boldsymbol{x}) \geq c] + 0\ \text{Prob}[\underline{z}(\boldsymbol{x}) < c] = \text{Prob}[\underline{z}(\boldsymbol{x}) \geq c] = 1 - F(c) \tag{2.34}$$

Equation (2.34) shows that the expected value of the indicator variable can be interpreted in terms of probability of P_{rel}-values.

A_c is predicted as the spatial mean of the indicator variable. This mean has ξ-expectation:

$$E_\xi\{m_A(\underline{i}_c)\} \equiv \mu_A = E_\xi[\frac{1}{N}\sum_{i=1}^{N} \underline{i}_c(\boldsymbol{x}_i)] = \frac{1}{N}\sum_{i=1}^{N} E_\xi\{\underline{i}_c(\boldsymbol{x}_i)\} = \text{(for stationarity of } \underline{z}(\boldsymbol{x})\text{:) } 1 - F(c) \tag{2.35}$$

Analysis of spatial structure

We first analyzed the data for a trend in P_{rel} by regression taking the x and y coordinate and their second-order terms as predictors. Only 1.5 % of the spatial variance of P_{rel} could be explained by a linear trend and 6.6 % by a second order polynomial. Therefore we assumed stationarity in the mean, which means that $E_\xi\{\underline{z}(\boldsymbol{x}_i)\}$ is equal for all i. Besides for P_{rel} this also holds for all indicator variables because these are simply transformations of P_{rel}. In this case the semivariance equals half the ξ-expectation of the squared differences:

$$\gamma(\boldsymbol{x}_i,\boldsymbol{x}_j) \equiv \frac{1}{2}\sigma_\xi^2\{\underline{z}(\boldsymbol{x}_i)-\underline{z}(\boldsymbol{x}_j)\} = \frac{1}{2}E_\xi[\{\underline{z}(\boldsymbol{x}_i)-\underline{z}(\boldsymbol{x}_j)\}^2] \qquad (2.36)$$

Moreover we assumed that the semivariance is finite and depends only on the lag $\boldsymbol{x}_i - \boldsymbol{x}_j$, in other words the intrinsic hypothesis. We further assumed isotropy, so that $\boldsymbol{x}_i - \boldsymbol{x}_j$ could be represented by a scalar h.

To estimate the semivariances $\gamma(h)$, data pairs were grouped into 25 groups, according to their separating distance, and we calculated the means of the distances and of the squared differences of the pairs in each group. Figure 2.3 shows the sample variograms. To these we fitted spherical models:

$$\begin{aligned} \gamma(h) &= c_0 + c_1 \{\frac{3h}{2a} - \frac{1}{2}(\frac{h}{a})^3\} \quad \text{for } 0<h\leq a \\ \gamma(h) &= c_0 + c_1 \quad \text{for } h>a \\ \gamma(h) &= 0 \quad \text{for } h=0 \end{aligned} \qquad (2.37)$$

where:

a = range;
c_0 = nugget variance;
c_0+c_1 = sill.

The indicator variables have a Bernoulli distribution with expectation $p = 1-F(c)$. If two Bernoulli variables are independently and identically distributed, their squared difference is also Bernoulli distributed with expectation $r = 2p(1-p)$. The variance of the mean of n squared diffences equals $r(1-r)/n$. The reciprocal value of this variance was used as a weight in iterative fitting of the variogram model, using the fitted value as an estimate of r.

We modelled the indicator variograms for the three critical values of P_{rel} independently of one another. Journel and Posa (1990) recommend otherwise because of order relation problems that might arise by fitting independently. We encountered no such problems, however, and fitted independently to ensure comparability with the design-based strategy.

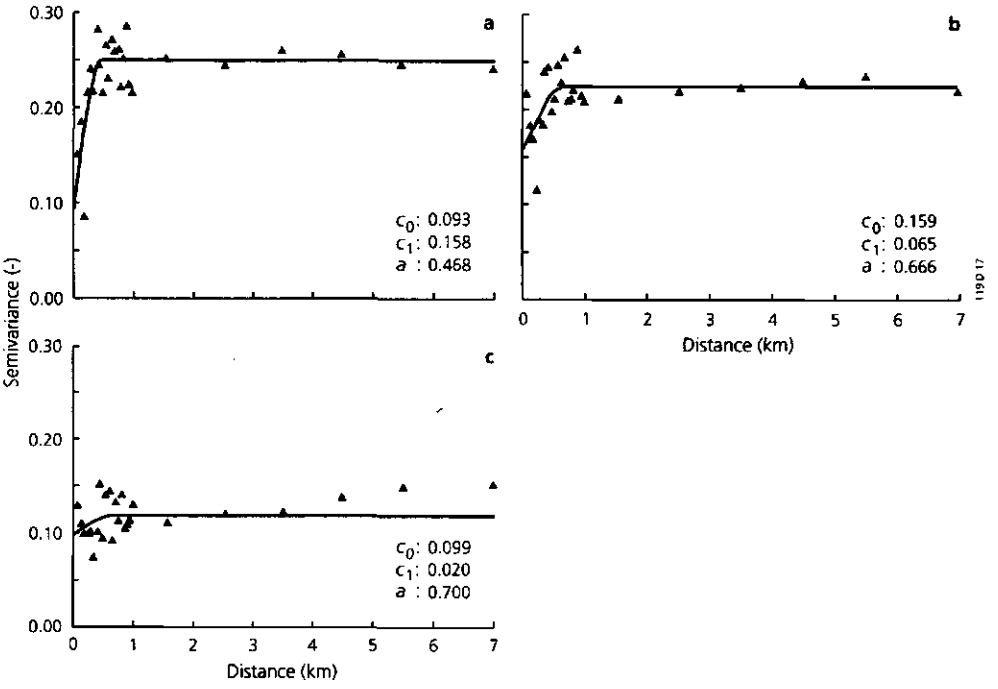


Fig. 2.3 Indicator variograms for three critical values of the relative mass of phosphate (**a**: 0.45, **b**: 0.35 and **c**: 0.25) with estimated parameters of spherical model: c_0 = nugget, $c_0 + c_1$ = sill, a = range

For $i_{0.25}$ the estimated sill was much larger than the theoretical maximum. This was caused by a slight increase in the semivariance beyond several kilometers (Fig. 2.3c). Therefore we used the design-based estimate of $A_{0.25}$, as an estimate of 1-F(0.25), to calculate the sill = F(0.25) [1-F(0.25)]. Next we estimated the parameters c_0 and a of the spherical model under the constraint that $c_0 + c_1$ was equal to the calculated sill.

The fitted models (Fig. 2.3) showed the following features:

1. The variograms reached a clear sill. For $i_{0.45}$ the sill equalled 0.250, which is in agreement with the fact that 0.45 was close to the median of P_{rel} (Fig. 2.3a). Moreover, this estimated sill value supported the assumption of stationarity.
2. The contribution of the nugget to the total variance increases from $i_{0.45}$, $i_{0.35}$ to $i_{0.25}$: the ratio c_0/sill was 37%, 71% and 83% respectively. This is known as the destructuration effect (Journel and Posa, 1990).
3. The range a increased in the same direction. From a theoretical point of view, one expects a constant range. The increase of a can possibly be explained by the positive correlation of the estimates of the range and nugget-effect.

Prediction

As the ξ-expectation can be derived from the variogram (Journel, 1983), simple kriging can be applied instead of ordinary kriging (Journel and Huybregts, 1978, p. 559). For the spatial mean of an indicator variable the simple kriging predictor is (ASCE, 1990):

$$\underline{\tilde{m}}_A(\underline{i}_c) = \sum_{i=1}^{n} \lambda_i \, \underline{i}_c(\boldsymbol{x}_i) + (1 - \sum_{i=1}^{n} \lambda_i) \, \hat{\underline{\mu}}_A = \hat{\underline{\mu}}_A + \sum_{i=1}^{n} \lambda_i \, [\underline{i}_c(\boldsymbol{x}_i) - \hat{\underline{\mu}}_A] \qquad (2.38)$$

The simple kriging predictor is ξ-unbiased for all λ_is, i.e. the weights need not necessarily sum to 1. Therefore the weights giving minimum ξ-variance can be obtained by solving Equation (2.18) after dropping the Lagrange multiplier and the last equation (sum of weights equal to 1). The ξ-variances of these predictions can be calculated by Equation (2.20). To estimate the mean ξ-covariance of a data point and a block, $\bar{C}(\boldsymbol{x}_i, A)$ (Eq. 2.19), and the mean ξ-covariance of a block, $\bar{C}(A,A)$ (Eq. 2.21) the blocks were discretized into a square grid of points at mutual distances of 500 m for the total area, 250 m for the 4 km x 4 km blocks and 200 m for the 2 km x 2 km blocks, respectively.

We used all data, so n = 222 for all blocks. This means that for the 2 km x 2 km and the 4 km x 4 km blocks points outside the blocks were also used. To fit the variogram and to krige, we wrote our own program in the Genstat language (Genstat 5 Committee, 1987).

Results

The areal fractions saturated with phosphate (A_c) of the agricultural land within - the total area; - within the four 4 km x 4 km blocks; - within the sixteen 2 km x 2 km blocks, estimated by *STSI* and predicted by *IK*, were compared.

Total area

From Table 2.1 we can see that the *STSI* estimates and *IK* predictions of the A_c of the total area were very similar. The standard errors of the *IK* predictions were slightly larger than those of the *STSI* estimates. It should be noted however that the design of the sample was not optimal for *IK*, i.e. the standard errors would have been somewhat smaller if the data points had been equally spaced.

Table 2.1 Areal fractions saturated with phosphate (A_c) of the total area estimated by stratified simple random sampling (*STSI*) and predicted by indicator block kriging (*IK*) for three critical values of the relative mass of phosphate (0.45, 0.35, 0.25). Between brackets: standard error; sample size is 222

Sampling strategy	Estimated/predicted areal fraction		
	0.45	0.35	0.25
STSI	0.441 (0.040)	0.659 (0.038)	0.861 (0.030)
IK	0.466 (0.047)	0.664 (0.046)	0.841 (0.035)

4 km x 4 km blocks

Table 2.2 shows the results for the 4 km x 4 km blocks. Except for $c = 0.45$, the mean absolute difference between the estimated and predicted A_c was larger than that of the total area (Table 2.4). As opposed to the total area, the *IK* standard errors were smaller than the *STSI* standard errors. The *STSI* standard error increased more than that of *IK* due to a decrease in the number of observations inside the estimation units.

Table 2.2 Areal fractions saturated with phosphate (A_c) of 4 km x 4 km blocks estimated by stratified simple random sampling (*STSI*) and predicted by indicator block kriging (*IK*) for three critical values of the relative mass of phosphate (0.45, 0.35, 0.25); n = number of sample points in block; between brackets: standard error

Block	n	Estimated/predicted areal fraction					
		0.45		0.35		0.25	
		STSI	*IK*	*STSI*	*IK*	*STSI*	*IK*
1	57	0.435 (0.082)	0.470 (0.054)	0.646 (0.070)	0.666 (0.053)	0.872 (0.067)	0.844 (0.038)
2	49	0.443 (0.098)	0.455 (0.058)	0.710 (0.092)	0.669 (0.056)	0.840 (0.077)	0.840 (0.040)
3	71	0.467 (0.066)	0.473 (0.050)	0.735 (0.069)	0.682 (0.050)	0.964 (0.033)	0.856 (0.036)
4	45	0.476 (0.084)	0.461 (0.061)	0.546 (0.092)	0.638 (0.060)	0.718 (0.090)	0.812 (0.043)
Mean standard error		0.085	0.056	0.079	0.055	0.065	0.039

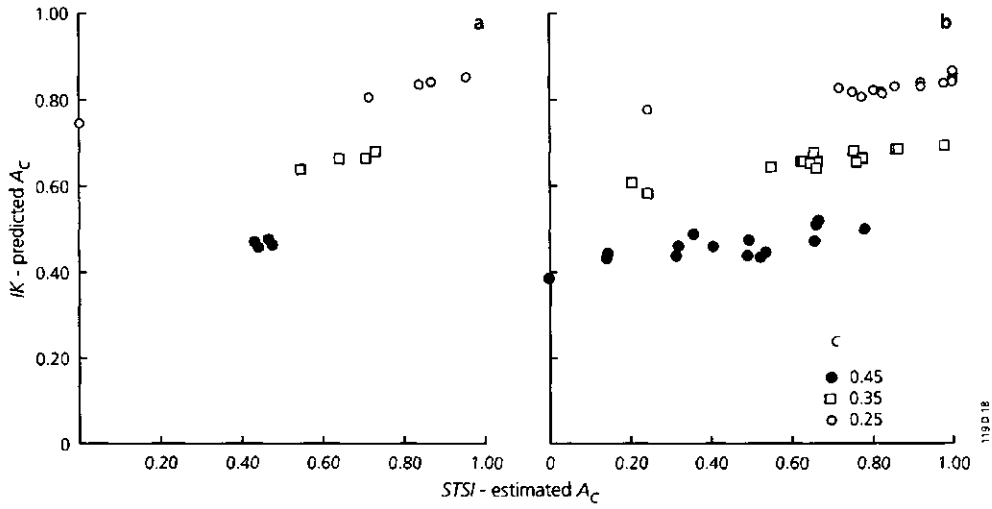


Fig. 2.4 Scatter plot of the areal fractions saturated with phosphate (A_c) estimated by stratified simple random sampling (*STSI*) and predicted by indicator block kriging (*IK*) for three critical values of the relative mass of phosphate (0.25, 0.35, 0.45). **a**: 4 km x 4 km blocks, **b**: 2 km x 2 km blocks

In Figure 2.4a the A_c's estimated by *STSI* have been plotted against those predicted by *IK*. The slope of the imaginary lines through the points, for a given critical P_{rel}-value, is much smaller than 1. This shows that only small differences between blocks were predicted by *IK* compared with those estimated by *STSI*. Nevertheless, for $c = 0.25$ and 0.35 positive correlations existed between the estimated and predicted A_c ($r = 0.99$ and 0.97 respectively). For $c = 0.45$ the correlation coefficient was very small: 0.07.

2 km x 2 km blocks

Table 2.3 shows the results for the 2 km x 2 km blocks. The mean absolute differences were much larger than that of the 4 km x 4 km blocks (Table 2.4). The mean standard errors from *STSI* were much larger than those from *IK*.

Table 2.3 Areal fractions saturated with phosphate (A_c) of 2 km x 2 km blocks estimated by stratified simple random sampling (*STSI*) and predicted by indicator block kriging (*IK*) for three critical values of the relative mass of phosphate (0.45, 0.35, 0.25); n = number of sample points in block; between brackets: standard error. Mean standard error for *IK* based on blocks 1-15. * = no estimate

Block	n	Estimated/predicted areal fraction					
		0.45		0.35		0.25	
		STSI	*IK*	*STSI*	*IK*	*STSI*	*IK*
1	18	0.136 (0.008)	0.441 (0.088)	0.201 (0.069)	0.614 (0.086)	0.819 (0.121)	0.827 (0.060)
2	12	0.403 (0.166)	0.464 (0.099)	0.867 (0.127)	0.696 (0.095)	1.000 (0.000)	0.857 (0.065)
3	10	0.522 (0.160)	0.440 (0.101)	0.759 (0.143)	0.662 (0.096)	0.820 (0.130)	0.830 (0.065)
4	16	0.663 (0.189)	0.527 (0.090)	0.752 (0.177)	0.689 (0.088)	0.752 (0.177)	0.829 (0.061)
5	15	0.491 (0.229)	0.445 (0.092)	0.646 (0.232)	0.667 (0.089)	0.716 (0.223)	0.840 (0.061)
6	12	0.777 (0.157)	0.507 (0.095)	0.977 (0.000)	0.704 (0.093)	0.988 (0.012)	0.855 (0.064)
7	12	0.312 (0.243)	0.440 (0.098)	0.663 (0.038)	0.652 (0.094)	0.855 (0.145)	0.842 (0.063)
8	11	0.135 (0.118)	0.435 (0.095)	0.641 (0.305)	0.662 (0.093)	0.993 (0.008)	0.854 (0.064)
9	18	0.534 (0.132)	0.451 (0.084)	0.775 (0.115)	0.674 (0.083)	0.921 (0.089)	0.844 (0.058)
10	17	0.351 (0.140)	0.490 (0.082)	0.655 (0.167)	0.689 (0.081)	1.000 (0.000)	0.870 (0.056)
11	21	0.493 (0.130)	0.479 (0.080)	0.554 (0.131)	0.652 (0.079)	0.776 (0.127)	0.816 (0.055)
12	7	0.000 (0.000)	0.388 (0.112)	0.240 (0.248)	0.590 (0.107)	0.240 (0.248)	0.782 (0.073)
13	18	0.315 (0.034)	0.463 (0.082)	0.635 (0.167)	0.664 (0.081)	0.918 (0.061)	0.851 (0.056)
14	18	0.655 (0.231)	0.477 (0.083)	0.864 (0.178)	0.695 (0.080)	1.000 (0.000)	0.863 (0.055)
15	17	0.659 (0.144)	0.517 (0.097)	0.666 (0.145)	0.666 (0.095)	0.830 (0.145)	0.834 (0.066)
16	0	* *	0.451 (0.291)	* *	0.662 (0.275)	* *	0.833 (0.201)
Mean standard error		0.139	0.092	0.150	0.089	0.099	0.061

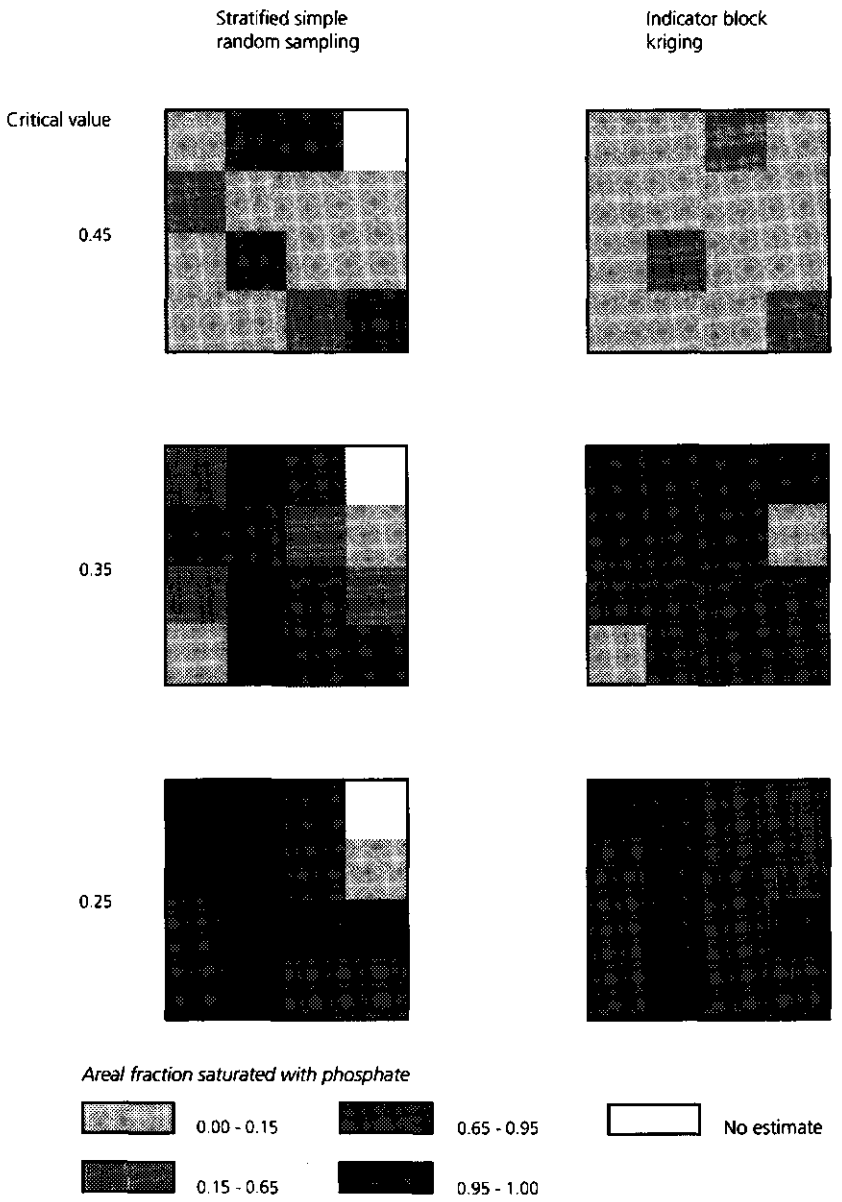


Fig. 2.5 Areal fractions saturated with phosphate (A_c) of 2 km x 2 km blocks estimated by stratified simple random sampling and predicted by indicato block kriging for three critical values of the relative mass of phosphate (0.25 0.35. 0.45)

Table 2.4 Mean absolute difference between *STSI* estimated and *IK* predicted areal fractions saturated with phosphate (A_c) of blocks for three critical values of the relative mass of phosphate (0.25, 0.35, 0.45)

Block	Difference (-)		
	0.45	0.35	0.25
Total area	0.025	0.005	0.020
4 km x 4 km blocks	0.017	0.052	0.058
2 km x 2 km blocks	0.161	0.123	0.110

Figures 2.4b and 2.5 show that also for the 2 km x 2 km blocks differences between *IK* predicted block-means were much smaller than between *STSI* estimated block-means. This was affirmed by the values of the smoothing parameters: 0.04, 0.11 and 0.09 for *IK* and 2.04, 5.40 and 3.11 for *STSI* for c = 0.25, 0.35 and 0.45 respectively. We found similar results for the spatial means of the relative mass of phosphate, P_{rel} (Brus and De Gruijter, 1993). Nevertheless, the *STSI* estimates and *IK* predictions were strongly correlated: r = 0.927, 0.932 and 0.790 for c = 0.25, 0.35 and 0.45, respectively. For *STSI* the value for the upper right block is missing because there were no sampling points in this block. However, it was possible to obtain this value by *IK*.

Discussion[1]

Case study

The results show that *STSI* and *IK* produced different estimates and predictions, respectively, of the areal fractions saturated with phosphate (A_c). These differences were so small for the total area that they could be ignored for practical purposes. The small differences arise from the large number of sampling points used in estimation and prediction. In general, the more observations one makes, the smaller the differences one may expect between any two strategies.

[1]This section will be published in Brus, D.J. & De Gruijter, J.J. 1993. Does kriging really give unbiased and minimum variance predictions of spatial means?. *Journal of Soil Science* 44, no. 4.

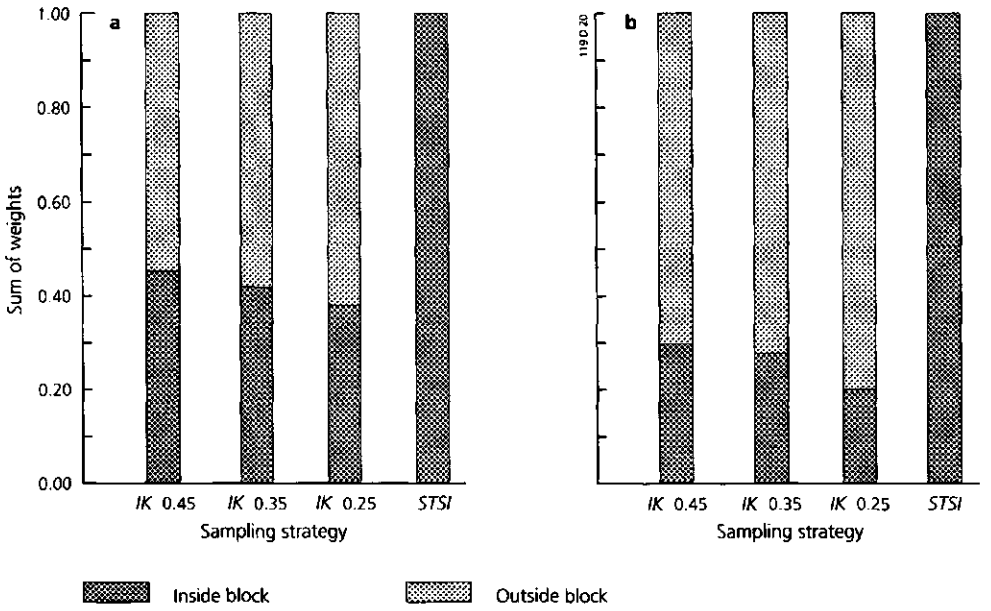


Fig. 2.6 Mean sum of weights of points inside and outside 4 km x 4 km blocks (**a**) and 2 km x 2 km blocks (**b**), for indicator block kriging (*IK*) and stratified simple random sampling (*STSI*). 0.45, 0.35, 0.25: critical values of the relative mass of phosphate

The difference between block-means estimated by *STSI* and predicted by *IK* increased with decreasing block size. This can be explained by the increasing differences between the *STSI* weights and *IK* weights attached to the data points. In *IK* data points outside the block were also used for prediction, whereas in *STSI* these points have zero weight. In *IK* points inside the block had the larger weights, but the sum of the weights of external points was considerable. Figure 2.6 shows these sums for the 4 km x 4 km and the 2 km x 2 km blocks. Going from 0.45 to 0.25, the sum of the weights of the internal points decreased, which is due to the increase of the nugget-variance (Fig. 2.3). For the 2 km x 2 km blocks the sums of the weights of internal points were somewhat smaller because they had fewer internal points.

Although *p*-unbiasedness is not part of the quality criterion used in *IK*, one may wonder whether the *IK* predictions are still *p*-unbiased with the design used. We first consider the case of the ordinary kriging predictor (Journel and Huijbregts, 1978, p. 563) combined with any design. The *p*-expectation of the ordinary block kriging predictor, conditional on the realization actually sampled is equal to:

$$E_p(\underline{t}_{OK}) = E_p\{\sum_{i=1}^{n} \underline{\lambda}_i z(\underline{\boldsymbol{x}}_i)\} = \sum_{i=1}^{n} [E_p(\underline{\lambda}_i)\ E_p\{z(\underline{\boldsymbol{x}}_i)\} + \sigma_p^2\{\underline{\lambda}_i, z(\underline{\boldsymbol{x}}_i)\}] \qquad (2.39)$$

where:

$\underline{t}_{OK}$ = ordinary block kriging predictor;

$\sigma_p^2\{\underline{\lambda}_i, z(\underline{\boldsymbol{x}}_i)\}$ = sampling covariance of the weight $\underline{\lambda}_i$ and $z(\underline{\boldsymbol{x}}_i)$.

The weights, $\underline{\lambda}_i$, are now stochastic because the locations $\boldsymbol{x}_i$ are.

If (i) only points in the block are used for prediction, and (ii) all points have an equal probability of being included in the sample, then $E_p\{z(\underline{\boldsymbol{x}}_i)\}$ equals the true spatial mean m_A. Moreover, $\Sigma E_p(\underline{\lambda}_i) = E_p(\Sigma\underline{\lambda}_i) = 1$, so:

$$\begin{aligned} E_p(\underline{t}_{OK}) &= \sum_{i=1}^{n} [E_p(\underline{\lambda}_i)\ E_p\{z(\underline{\boldsymbol{x}}_i)\} + \sigma_p^2\{\underline{\lambda}_i, z(\underline{\boldsymbol{x}}_i)\}] = m_A\{\sum_{i=1}^{n} E_p(\underline{\lambda}_i)\} + \sum_{i=1}^{n} \sigma_p^2\{\underline{\lambda}_i, z(\underline{\boldsymbol{x}}_i)\} \\ &= m_A E_p(\sum_{i=1}^{n} \lambda_i) + \sum_{i=1}^{n} \sigma_p^2\{\underline{\lambda}_i, z(\underline{\boldsymbol{x}}_i)\} = m_A + \sum_{i=1}^{n} \sigma_p^2\{\underline{\lambda}_i, z(\underline{\boldsymbol{x}}_i)\} \end{aligned} \qquad (2.40)$$

So even if requirements (i) and (ii) are met, the ordinary block kriging predictor is *p*-unbiased only if the sum of the covariances of $\underline{\lambda}_i$ and $z(\underline{\boldsymbol{x}}_i)$ is 0, which will generally be not the case, except when a pure nugget model has been accepted. For a pure nugget model the covariances are 0 for all *i*, regardless the configuration of points, because $\lambda_i = 1/n$ for all *i*. The above reasoning also holds for the simple kriging predictor which is used in indicator kriging, but then an extra requirement is needed to achieve *p*-unbiasedness, namely the ξ-expectation of the spatial mean (which is assumed to be known in advance) should be equal to the true spatial mean.

It follows from the discussion above that the *IK* predictor is not *p*-unbiased for the 4 km x 4 km blocks and the 2 km x 2 km blocks, regardless of the design and the covariance term, because the predictor also uses external sampling points. For the total area, the strategy ($STSI,\underline{t}_{IK}$) is not *p*-unbiased because the inclusion probability of the sampling points was not equal for all the points.

Choice of sampling strategy

The choice of the strategy should depend amongst others on what criterion of quality is required and on the sampling costs.

Quality criteria

Regarding the quality criterion there are two important questions to be answered: (i) are criteria required based on *p*-expectations or on ξ-expectations?[2] and (ii) what is most important, to obtain an estimate as close as possible to the true spatial mean (minimum estimation variance) or to obtain an unbiased and robust estimate of the estimation variance?

Question (i). We believe ξ-unbiasedness is a weak criterion. This can be shown by the following example. Suppose, one selects locations with a probability inversely proportional to P_{rel}-values thought likely from prior information. No theoretical objections can be made 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 P_{rel} are positively correlated, then there would be a strong tendency to under-estimate the mean of P_{rel} or A_c. Such misleading predictions are still ξ-unbiased. The reason for this is that the sampled area is just one realization from the model and larger values would occur at the same locations in other realizations.

With respect to the criterion minimum ξ-variance, it is important to realize that a small ξ-variance does not necessarily imply a small sampling variance for

[2]Quality criteria can also be based on both expectations, i.e both the sampling locations and the model-realizations from which the sample is taken, vary. We minimize the expected *p*-variance or expected MSE_p, $E_\xi E_p\{(\underline{\tilde{m}}_A-\underline{m}_A)^2\}$. This can be done subject to three unbiasedness conditions: *p*-unbiasedness, ξ-unbiasedness and *p*ξ-unbiasedness. For further details we refer to Cassel *et al.* (1977) and Särndal (1978).

the area sampled. For reasonably large samples (and as a result a precise estimate of the ξ-variance), a small ξ-variance means that in most of the model-realizations the sampling variance will be small too (Hansen *et al.*, 1983; Domburg *et al.* in press).

If *p*-unbiasedness is chosen as a criterion, design-based sampling strategies are the only option.

Question (ii). In the model-based approach several subjective assumptions are made about the spatial structure, e.g., about stationarity, isotropy and model type of variogram, that influence the results directly (Englund, 1990). The quality of the estimate of the ξ-variance of the predictor depends on the quality of the model of the spatial structure. In the design-based approach no such assumptions are made. As shown above, the idea that spatial data are assumed to be uncorrelated in the design-based approach is a misunderstanding. Several subjective decisions are made in the design-based approach either, e.g. in choosing the stratification criteria, the number of strata and the size of clusters (in cluster sampling). However, these are accounted for in the inference, still allowing for *p*-unbiased and robust estimation of the sampling variance of the estimator, independent of the spatial structure of the population.

If the true model is known, model-based predictions will be closer to the true value in situations with clear spatial autocorrelation between the data points. However, in practice the true model is unknown. This introduces an extra error which should be added to the ξ-variance. Moreover, models of spatial structure can also be used in design-based strategies, for example by soil map stratification. This has the advantage that *p*-unbiased and robust estimates of the sampling variance can still be obtained. The efficiency of variograms and soil maps as models of spatial structure for estimation of spatial means should be further compared in future.

In conclusion, we recommend a design-based strategy if a *p*-unbiased and robust estimate of the estimation variance is essential, e.g. if a risk analysis is to be made. If it is most important to obtain an estimate as close as possible to the true value only and no (unbiased and robust) estimate of the estimation

variance is needed, also model-based strategies are suitable. However, we doubt whether (estimated) variograms are better models of spatial structure than soil maps and lead to smaller errors indeed in soil surveys at a regional scale.

The subjectivity in the choice of the model (and the selection of the sampling locations) as such is a drawback of the model-based approach because it makes the survey results vulnerable to criticism (Borgman & Quimby, 1988). Objectivity may be especially important if results are used in legal or regulatory activities as in the case study described above.

Sampling costs
Besides the quality criterion, the choice of the sampling strategy should be determined by the available budget. If there are many blocks and measurement costs per sampling point are large it is rarely feasible to sample each block adequately resulting into inaccurate design-based estimates of the mean. In this case model-based strategies seem to be advantegeous. However, quite a few sampling points are needed to estimate the model. According to Webster & Oliver (1992) a variogram computed from a sample of 150 points will generally be satisfactory and one from 225 data will usually be reliable. In general these sample sizes would enable the means of several blocks (5 - 10) to be estimated adequately using a design-based strategy.

Conclusions

Model-based and design-based sampling strategies resulted in different estimates of the areal fractions of land saturated with phosphate. There were only slight differences for the total area, but the differences between the estimates provided by the two approaches increased with decreasing block size to practically relevant magnitudes. This is because of the increasing difference in the number of sampling points used in prediction or estimation. In kriging points outside the block are used, and the total weight of these points, which depends on the variogram, the block size and the configuration of the data points, was considerable in the case described here.

Strong smoothing was obtained from the model-based method, i.e., the spatial variance of the predicted values was much less than that of the true values, whereas the design-based method produced the reverse.

In general model-based strategies will not produce *p*-unbiased predictions of the spatial mean. Moreover, the quality of the estimate of the variance of the predictor remains unknown and depends on the quality of the model of the spatial structure.

If enough funds are available to take a sample from each block we recommend design-based sampling strategies, because they ensure *p*-unbiasedness. If the number of blocks is large relative to the affordable sample size, model-based strategies are suitable.

Acknowledgements

We thank J. Bouma and A. Breeuwsma for their helpful comments on an earlier draft of this paper. We also thank R. Webster and an anonymous referee for their constructive remarks and text proposals. Finally, we thank A. Papritz who pointed out that in general equal inclusion probability and use of only internal points are not sufficient conditions to obtain *p*-unbiased predictions of spatial means by kriging.

References

ASCE Task Committee on Geostatistical Techniques in Geohydrology of the Ground Water Hydrology Committee of the ASCE Hydraulics Division. 1990. Review of Geostatistics in geohydrology I: Basic concepts, II: Applications. *Journal of Hydraulic Engineering* 116, 612-658.

Barnes, R.J. 1988. Bounding the required sample size for geologic site characterization. *Mathematical Geology* 20, 477-490.

Beckett, P.H.T. & Webster, R. 1971. Soil variability; a review. *Soils and Fertilizers* 34, 1-15.

Borgman, L.E. & Quimby W.F. 1988. Sampling for tests of hypothesis when data are correlated in space and time. In *Principles of Environmental Sampling* (ed. L.H.

Keith), American Chemical Society, Professional Reference Book. 25-43.

Breeuwsma, A., Reijerink, J.G.A., Schoumans, O.F., Brus D.J. & Van het Loo, H. 1989. *Fosfaatbelasting van bodem, grond- en oppervlaktewater in het stroomgebied van de Schuitenbeek.* Report 10, DLO The Winand Staring Centre for Integrated Land, Soil and Water Research (SC-DLO), Wageningen.

Breeuwsma, A. & Schoumans, O.F. 1987. Forcasting phosphate leaching from soils on a regional scale. In *Vulnerability of soil and groundwater to pollutants.* Proceedings and Information 38. (eds W.A. Duijvenboden and H.G. van Waegeningh), 973-981. National Institute of Public Health and Environmental Hygiene, The Hague.

Breeuwsma, A. & Silva, S. 1992. *Phosphorus fertilisation and environmental effects in the Netherlands and Italy.* Report 57, The Winand Staring Centre for Integrated Land, Soil and Water Research (SC-DLO), Wageningen.

Brus, D.J., De Gruijter, J.J. & Breeuwsma, A. 1992). Strategies for updating soil survey information: a case study to estimate phosphate sorption characteristics. *Journal of Soil Science* 43, 567-581.

Brus, D.J. & De Gruijter, J.J. 1993. Does kriging really give unbiased and minimum variance predictions of spatial means. Journal of Soil Science 44, no 4.

Burgess, T.M. & Webster, R. 1980. Optimal interpolation and isarithmic mapping of soil properties. I. The semivariogram and punctual kriging. *Journal of Soil Science* 31, 315-331.

Burgess, T.M., Webster, R. & McBratney, A.B. 1981. Optimal interpolation and isarithmic mapping of soil properties. IV. Sampling strategy. *Journal of Soil Science* 32, 643-659.

Burrough, P.A. 1991. Sampling designs for quantifying map unit composition. *Spatial Variabilities of Soils and Landforms.* SSSA Special Publication 28, 89-125. Cassel, C-M., Särndal, C-E. & Wretman J.H. 1977. *Foundations of inference in survey sampling.* Wiley, New York.

Cochran, W.G. 1977. *Sampling techniques.* 3rd edn. Wiley, New York.

Corsten, L.C.A. 1989. Interpolation and optimal linear prediction. *Statistica Neerlandica* 43, 69-84.

Domburg, P., De Gruijter, J.J & Brus, D.J. 1993. A structured approach to designing soil survey schemes with prediction of sampling error from variograms. *Geoderma* (in press).

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. *Zeitschrift für Pflanzenernaehrung und Bodenkunde.* 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. *Mathematical Geology* 22, 407-415.

De Gruijter, J.J. & Ter Braak, C.J.F. 1992 Design-based versus model-based sampling

strategies: Comment on R.J. Barnes' Bounding the required sample size for geologic site characterization. *Mathematical Geology* 24, 859-864.

Englund, E.J. 1990. A variance of geostatisticians. *Mathematical Geology* 4, 417-455.

Genstat 5 Committee 1987. Genstat 5 Reference Manual. Clarendon Press. Oxford.

Hansen, M.H., Madow, W.G. & Tepping, B.J. 1983. An evaluation of model-dependent and probability-sampling inferences in sample surveys. *Journal of the American Statistical Association* 78, 776-793.

Journel, A.G. 1983. Nonparametric estimation of spatial distributions. *Mathematical Geology* 15, 445-468.

Journel, A.G. & Huijbregts, C.J. 1978. *Mining Geostatistics.* Academic Press, London.

Journel, A.G. & Posa, D. 1990. Characteristic behavior and order relations for indicator variograms. *Mathematical Geology* 22, 1011-1025.

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.

Matérn, B. 1960. Spatial variation. Stochastic models and their application to some problems in forest surveys and other sampling investigations. *Meddelanden Från Statens Skogsforskningsinstitut* 49, no. 5.

Morse, R.K. & Thornburn, T.H. 1961. Reliability of soil maps. In: *Proceedings of the 5th International Conference on Soil Mechanics and Foundation Engineering.* Paris, 259-262.

Olea, R.A. 1984. Sampling design optimization for spatial functions. *Mathematical Geology* 16, 369-392.

Oliver, M., Webster, R. & Gerrard, J. 1989. Geostatistics in physical geography. Part I: theory. *Trans. Inst. Br. Geogr. N.S.* 14, 259-269.

Papoulis, A. 1965. *Probability, random variables and stochastic processes.* McGraw-Hill, Kogakusha.

Russo, D. & Bresler, E. 1981. Soil hydraulic properties as stochastic processes: I. An analysis of field spatial variability. *Soil Science Society of America Journal* 45, 682-687.

Särndal, C-E., Swensson, B. & Wretman, J. 1992. *Model assisted survey sampling.* Springer-Verlag, New York.

Särndal, C-E. 1978. Design-based and model-based inference in survey sampling. *Scandinavian Journal of Statistics* 5, 27-52.

Soil Survey Staff 1975. *Soil Taxonomy.* A basic system of soil classification for making and interpreting soil surveys. Agriculture Handbook 436. Soil Conservation Service. US Department of Agriculture, Washington DC.

Van der Zee, S.E.A.T.M. 1988. *Transport of reactive contaminants in heterogeneous soil systems.* Doctoral thesis, Agricultural University, Wageningen.

Van der Zee, S.E.A.T.M., Van Riemsdijk, W.H. & De Haan, F.A.M. 1990. *Het protokol fosfaatverzadigde gronden, deel I (Toelichting) en deel II (Technische uitwerking).* Agricultural University, Wageningen. Department of Soil science and plant nutrition.

Webster, R. & Beckett, P.H.T. 1968. Quality and usefulness of soil maps. *Nature* 219, 680-682.

Webster, R. 1985. Quantitative spatial analysis of soil in the field. *Advances in soil science*, 3, 1-70.

Webster, R. & Oliver, M.A. 1992. Sample adequately to estimate variograms of soil properties. *Journal of Soil Science* 43, 177-192.

Yfantis, E.A., Flatman, G.T. & Behar, J.V. 1987. Efficiency of kriging estimates for square, triangular, and hexagonal grids. *Mathematical Geology* 19, 183-205.

Chapter 3

Improving design-based estimation of spatial means by soil map stratification. A case study of phosphate saturation

Geoderma, in press

Improving design-based estimation of spatial means by soil map stratification. A case study of phosphate saturation

The usefulness of soil maps and maps of land use was evaluated to estimate the spatial means of several phosphate sorption characteristics in two areas with contrastive historical phosphate loads. The maps were used to stratify the areas for random sampling. This is a way of incorporating knowledge of spatial structure into a design-based sampling strategy. Three stratifications were evaluated, viz. by land use, soil map unit and by both, in combination with three methods of allocating sample points to the strata: proportional, optimum and near-optimum. The efficiency of various stratified simple random sampling designs was calculated from data of one sample from each area.

The phosphate sorption characteristics were: (i) the areic mass of P_2O_5 sorbed by soil, i.e the mass of P_2O_5 per m^2 actually sorbed by soil above a reference depth; (ii) the maximum areic mass of P_2O_5 sorbed by soil, i.e the areic mass which can potentially be sorbed by soil above a reference depth; (iii) the relative mass of phosphate sorbed by soil, i.e. the ratio of (i) and (ii); (iv) the areal fraction of soil saturated with phosphate, i.e. the fraction of an area with a relative mass of phosphate sorbed by soil larger than a critical value.

For the maximum areic mass of P_2O_5 and the areic mass of P_2O_5 sorbed by soil, stratification by soil map unit will be worthwhile in both areas. For the relative mass of phosphate sorbed by soil and the areal fraction of soil saturated with phosphate there will be a gain only where the historical phosphate load is small. The gain for the areal fraction of soil saturated with phosphate depends strongly on the critical value of the relative mass of phosphate sorbed by soil. This gain may be further increased by stratifying also according to land use.

Introduction

Eutrophication of the surface water is serious in the Netherlands and agriculture is a principal cause. Agricultural soils in the Netherlands, especially

in the sandy districts, are intensely manured. The surplus phosphate not taken up by plants accumulates in the soil by sorption, and because the sorption capacity is limited the soil gradually becomes saturated. As a result, phosphate leaches to groundwater and surface water (Van der Zee, 1988). To prevent further degradation of the environment, manuring in susceptible areas is regulated by the 'Regulation of phosphate-saturated Soils' which is included in the Soil Conservation Law.

To assess the susceptibility of phosphate leaching or the actual degree of phosphate saturation of a given area, the soil must be surveyed. The sampling strategy, i.e. sampling design plus estimator, is important. Sampling strategies can be divided into model-based and design-based strategies. (Cassel *et al.*, 1977; De Gruijter and Ter Braak, 1990). We concluded elsewhere (Brus and De Gruijter, 1993) that if probability sampling is feasible and enough funds are available to take a sample of a pre-determined, minimum size from each block, design-based strategies are better because: (i) their estimates are model-free and therefore more robust; (ii) there are no subjective decisions on sampling locations; (iii) it makes use of quality-criteria based on expectations over realizations of the sampling design (p-expectations) which are more relevant to practice than the model-based criteria based on expectations over realizations of the stochastic model describing spatial structure (ξ-expectations).

A model of spatial structure can be both an advantage and a disadvantage for spatial sampling and estimation. If the model is correct, sampling can be more efficient. In design-based sampling strategies information about spatial structure can be used to divide an area into more homogeneous subareas, strata. Thus, knowledge of spatial structure is incorporated in the sampling design, whereas in model-based strategies it is used in prediction. Stratification as a way of using knowledge of spatial structure, has some important advantages. Estimates keep their robustness, and p-unbiasedness and minimum p-variance can be used as criteria of quality.

The purpose of this study was to assess the increase in precision of estimates of the spatial mean of various phosphate sorption characteristics by

using a soil map and a land use map as models of spatial structure for stratification. Several sampling designs with different stratification-types or allocation-types were compared, but only one random sample was taken.

Data collection

Survey areas

Data were collected in two separate areas in the eastern part of the Province of Overijssel: Bentelo and Ootmarsum (Hack-Ten Broeke *et al.*, 1990). Only the agricultural parts of the areas were sampled. These cover 1233 ha and 2252 ha in Bentelo and Ootmarsum, respectively. Much more manure is produced in Bentelo than in Ootmarsum and so its phosphate load is larger. Moreover the water table in Bentelo is nearer to the surface and therefore the maximum mass of phosphate that can be sorbed by soil before phosphate leaches to the water table is smaller. Based on the comparatively large phosphate load and shallow water table we expected more of the area of Bentelo to be satuarated with phosphate.

Target variables

We estimated the sampling variance of the mean of three phosphate sorption characteristics: (i) the maximum areic mass of P_2O_5 sorbed by soil (P_{max}); (ii) the areic mass of P_2O_5 sorbed by soil (P); (iii) the relative mass of phosphate sorbed by soil (P_{rel}). P_{max} was measured indirectly via the oxalate-extractable iron and aluminium content:

$$P_{max} = 0.5\ M_{ox}\ \delta\ \rho\ 0.71 \qquad (3.1)$$

where:

0.5 = regression parameter (-);
M_{ox} = oxalate-extractable aluminium + iron content of soil (mol kg^{-1});
δ = depth of soil profile (m);
ρ = volumic mass of soil (kg m^{-3});
0.71 = coefficient to convert the dimension from mol P per m^2 to kg P_2O_5 per m^2 (kg mol^{-1}).

There is a strong correlation between P_{max} and M_{ox} (Schoumans et al., 1987; Van der Zee, 1988).

For δ we took the mean highest water table ($\bar{W}_{min}$) or 1 m if $\bar{W}_{min}$ is larger than 1 m. This depth corresponds to the one used in the protocol 'Phosphate-saturated Soils', an official document specifying, for example, a standardized sampling design (Van der Zee *et al.*, 1990).

Analogous to P_{max}, P was calculated by:

$$P = P_{ox}\ \delta\ \rho\ 0.71 \qquad (3.2)$$

where:
P_{ox} = oxalate-extractable P_2O_5 content of soil (mol kg^{-1}).

For the measurement of M_{ox} and P_{ox} we took one bulk sample of the soil profile to the depth defined above. For ρ means of soil horizons, measured in previous soil surveys, were used (Hoekstra & Poelman, 1982).

The relative mass of phosphate sorbed by soil (P_{rel}) is the ratio of P and P_{max}:

$$P_{rel} = \frac{P}{P_{max}} \qquad (3.3)$$

In addition to the sampling variance of these spatial means, we estimated the sampling variance of the areal fraction of soil saturated with phosphate (A_c) defined as the fraction of the area with a P_{rel}-value greater than or equal to a given critical value (c). For c we took 0.25, 0.35 and 0.45; the first one is being used in the protocol 'Phosphate-saturated Soils'.

Actual sampling design

Stratification
The two areas were stratified using a soil map, land use map and a map of drainage. For Bentelo only the first two maps were used. The existing

1:50 000 soil map (Stichting voor Bodemkartering, 1979; Ebbers & Van het Loo, 1992) was used. Previous studies showed clearly different P_{max} for some of the units of this map (Brus *et al.*, 1992). Soil map units supposed to have an approximately equal P_{max} were grouped together (Table 3.1, Fig. 3.1a).

Table 3.1 Clustering of units of 1:50 000 soil map into strata for Bentelo and Ootmarsum

Bentelo

No.	Map units
1	*ABv*(III), *pZg*23(II), *pZg*23(III)
2	*Hn*21(III), *Hn*21(V), *Hn*21(V*), *Hn*21*x*(V), *cHn*21(V*), *pZn*23(III)
3	*Hn*21(VI), *cHn*23(VI), *Hn*21(III/V/VI), *Hn*21(III/VI), *Hn*21(V/VI)
4	*kpZg*23(III), *kpZg*23(III*), *kpZg*23*x*(III)
5	*zEZ*21(VI), *zEZ*23(VI)
6	*zEZ*21(VII), *zEZ*23(VII), *zEZ*21(VI/VII)

Ootmarsum

No.	Map units
1	*ABv*(II), *ABvF*(II), *ABv*(III), *ABv*(II/V), *pZg*23(III), *pZg*23(V), *pZg*23t(V), *pZg*23*x*(V), *pZg*23(VI)
2	*Hn*21(V), *Hn*21(V*), *Hn*21*tF*(v), *Hn*21*t*(V), *Hn*21*x*(V*), *Hn*23(V), *Hn*23(V*), *Hn*23*t*(V), *Hn*23*x*(V), *gHn*21*t*(V), *gHn*30*t*(V), *mHn*23*x*(V), *pZn*21(V*)
3	*bEZ*23*t*(V), *bEZ*23*t*(VI)
4	*KT*(V), *KX*(V), *KT*(VI)
5	*cHn*21(VI), *cHn*23(VI), *Hn*21(VI), *Hn*23(VI), *Hn*23*x*(VI), *gHn*23*x*(VI), *Hn*21/*KT*(V/VI), *Hn*21/*Hn*21*x*(V/VI/VII)
6	*cHn*21(VII), *cHn*21(VII*), *cHn*23(VII), *Hn*21(VII), *Hn*21(VII*), *gHn*21(VII), *gHn*23(VII), *gHn*23(VII*), *gHn*30(VII), *gHn*30(VII*), *gY*21(VII*), *gY*30(VII*), *Hd*21(VII*), *gHd*21(VII*), *gHd*21/*gHn*23*x*(VII/VII*)
7	*bEZ*23(VII*), *bEZ*23t(VII*), *zEZ*21(VII*), *zEZ*23(VII), *zEZ*23(VII*)

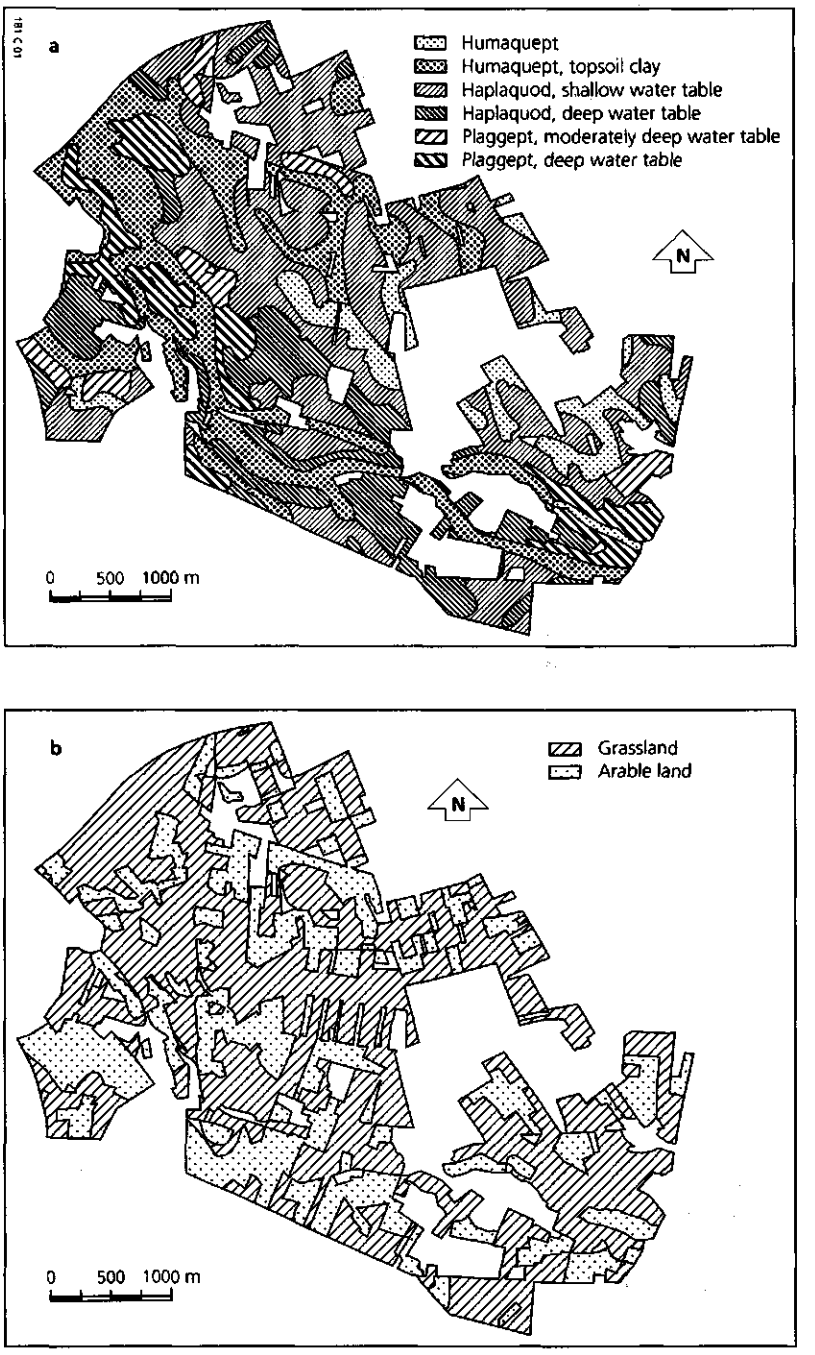


Fig. 3.1 Generalized soil map (**a**) and land use map (**b**) of Bentelo used for stratification (After Hack-Ten Broeke *et al.*, 1990)

There are only two kinds of land use: grassland and arable land (Fig. 3.1b). Maize is the main arable crop and the land is heavily manured. By stratifying according to land use we hope to reduce the variance in P. Only the soil map units with a shallow water table (unit 1 to 4 of Table 3.1) of Ootmarsum were divided according to drainage. The aim was to obtain separate estimates of A_c within these wet areas. Combining the subdivisions from the maps produced 12 strata in Bentelo and 26 strata in Ootmarsum.

In Bentelo 66 profiles were sampled, in Ootmarsum 116. This corresponds with a density of one sample point per 19 ha.

Allocation
In Bentelo the sampling points were allocated in proportion to the area of the strata with a minimum of two (see Table 3.3 p. 65, column right-hand side). The minimum of two points per stratum was required to obtain unbiased estimates of the variance within strata. In Ootmarsum proportional allocation would have resulted in small sample sizes and imprecise estimates for the strata in the wet areas. These soils are relatively vulnerable to phosphate leaching to the water table. It was especially important therefore, to obtain precise estimates of A_c here and so they had twice the number of sample points as they would have had with proportional allocation. Otherwise, points were allocated in proportion to size (area) of the strata, with a minimum of two.

Data analysis

Studied sampling designs
To calculate the increase in precision of estimates of the spatial mean by stratification, we estimated the sampling variance of the spatial mean for nine sampling designs (Table 3.2).

In proportional allocation the sample size of a stratum is proportional to its size (area). In optimum allocation, in addition to differences in size, dif-

Table 3.2 Analysed sampling designs

Code	Stratification	Allocation
SI	no	-
L/P	land use map	proportional
L/O	land use map	optimum
S/P	soil map	proportional
S/NO	soil map	near-optimum
S/O	soil map	optimum
LS/P	land use + soil map	proportional
LS/NO	land use + soil map	near-optimum
LS/O	land use + soil map	optimum

ferences in spatial variation within the strata are also taken into account. For optimum allocation, a stratum with the same size (relative area) but with a larger internal spatial variation, receives more sample points.

If the size or variance of a stratum is very small, its optimum sample size may be zero or one. For such strata we would get no estimates of the spatial mean and therefore we would be unable to obtain unbiased estimates of the spatial mean of the total area. One sample point in each stratum would suffice, but then we would not be able to calculate unbiased estimates of the sampling variance of the spatial mean. Therefore, we calculated the sampling variance of the mean for near-optimum allocation, i.e. optimum allocation with a minimum of two points in each stratum.

Estimation of sampling variance of mean under non-executed sampling designs

To estimate the sampling variance of the global means for the nine designs, we need take only one sample, as I shall show. The sampling variance is estimated in two steps:

(i) estimation of the spatial variance of the phosphate sorption characteristics within the total area and within the strata used in the non-executed designs;

(ii) estimation of the sampling variance.

Estimation of the spatial variance (step i) is simple because the 'new' strata are combinations of strata used in the executed sampling design. In Bentelo the strata of the executed sample are the same as those of the *LS/P* and *LS/NO* designs. We used the procedure followed by Marsman and De Gruijter (1986):

$$\underline{\hat{v}}_t(z) = \underline{\hat{m}}_t(z^2) - \{\underline{\hat{m}}_t(z)\}^2 + \hat{\sigma}_p^2\{\underline{\hat{m}}_t(z)\} \qquad (3.4)$$

where:

$\underline{\hat{v}}_t(z)$ = estimated spatial variance of characteristic z inside the new stratum t;
$\underline{\hat{m}}_t(z^2)$ = estimated spatial mean of z^2 of the new stratum t;
$\underline{\hat{m}}_t(z)$ = estimated spatial mean of z of the new stratum t;
$\hat{\sigma}_p^2\{\underline{\hat{m}}_t(z)\}$ = estimated sampling variance of the estimator $\underline{\hat{m}}_t$ under the executed sampling design p.

This equation can also be used to estimate the spatial variance inside the total area, $v_A(z)$, which we need for the simple random sampling design. A similar procedure was used by Cochran for the specific case of estimating the sampling variance of the mean of a simple random sample given the results of a stratified simple random sample (Cochran, 1977, p. 136). The components of Equation (3.4) can be estimated by the usual estimators for stratified simple random sampling as follows. The means of z and z^2 in a new stratum t (m_t) were estimated by:

$$\underline{\hat{m}}_t = \sum_{h=1}^{H} W_h \, \underline{\hat{m}}_h \qquad (3.5)$$

where:

H = number of (original) strata inside the new stratum t;
W_h = weight of stratum h calculated as the proportion of the area of stratum h inside the new stratum t;
$\underline{\hat{m}}_h$ = estimated spatial mean of stratum h.

The sampling variance of the estimated mean of a new stratum t under the executed sampling design p can be estimated by (Cochran, 1977, p. 92):

$$\hat{\underline{\sigma}}_p^2(\underline{\hat{m}}_t) = \sum_{h=1}^{H} W_h^2 \, \hat{\underline{\sigma}}_p^2(\underline{\hat{m}}_h) = \sum_{h=1}^{H} W_h^2 \, \frac{\hat{\underline{v}}_h}{n_h} \, (1-f_h) \qquad (3.6)$$

where:

$\hat{\underline{\sigma}}_p^2(\underline{\hat{m}}_h)$ = estimated sampling variance of the sample mean of (original) stratum h under the executed sampling design p;

$\hat{\underline{v}}_h$ = estimated spatial variance of stratum h;

n_h = sample size of stratum h;

f_h = sampling fraction of stratum h.

The sampling fractions are close to zero for all strata so that the last term of Equation (3.6) can be dropped. If the spatial variance $\hat{\underline{v}}_h$ is expressed in terms of proportions, Equation (3.6) becomes:

$$\hat{\underline{\sigma}}_p^2(\underline{\hat{p}}_t) = \sum_{h=1}^{H} W_h^2 \, \frac{\underline{\hat{p}}_h \, (1-\underline{\hat{p}}_h)}{n_h} \, (1-f_h) \qquad (3.7)$$

where:

$\underline{\hat{p}}_h$ = estimated proportion of (original) stratum h.

Once the spatial variances within the new strata of the non-executed designs or within the total area are known, the sampling variances of the mean can be estimated simply for a given sample size and, with stratified simple random sampling, for a given allocation of the sample points to the strata. For proportional and optimum allocation the sampling variance can be estimated without calculating the sample sizes of the strata, whereas for near-optimum allocation these sample sizes have to be calculated first.

For simple random sampling the variance of the mean is simply:

$$\hat{\underline{\sigma}}_p^2(\underline{\hat{m}}_A) = \frac{\hat{\underline{v}}_A}{n} \, (1-f) \approx \frac{\hat{\underline{v}}_A}{n} \qquad (3.8)$$

where:

$\hat{\sigma}_p^2(\underline{\hat{m}}_A)$ = estimated sampling variance of the estimated mean of the total area under the new design *p*.

For stratified simple random sampling (*STSI*) we have:

$$\hat{\sigma}_p^2(\underline{\hat{m}}_A) = \sum_{t=1}^{T} W_t^2 \frac{\hat{v}_t}{n_t} \tag{3.9}$$

where:
T = number of new strata;
n_t = sample size of the new stratum *t*.

Proportional allocation
Using proportional allocation, we substitute:

$$n_t = n\ W_t \tag{3.10}$$

in Equation (3.9). The sampling variance reduces to:

$$\hat{\sigma}_p^2(\underline{\hat{m}}_A) = \frac{1}{n} \sum_{t=1}^{T} W_t\ \hat{v}_t \tag{3.11}$$

Cochran (1977, p. 100) shows that if terms in $1/N_t$ are negligible then:

$$\sigma^2_{STSI(pr)} = \sigma^2_{SI} - \frac{1}{n} \sum_{t=1}^{T} W_t\ (m_t - m_A)^2 \tag{3.12}$$

Optimum allocation
For a fixed total sample size the sampling variance of the mean is minimized if (Cochran, 1977, p. 98):

$$n_t = n \frac{W_t\, s_t}{\sum_{t=1}^{T} W_t\, s_t} = n \frac{N_t\, s_t}{\sum_{t=1}^{T} N_t\, s_t} \qquad (3.13)$$

where:
s_t = spatial standard deviation within a new stratum t.

Substituting this in Equation (3.9), gives the minimum sampling variance of the mean for a fixed sample size:

$$\hat{\sigma}_p^2(\underline{m}_A) = \frac{(\sum_{t=1}^{T} W_t\, \underline{\hat{s}}_t)^2}{n} \qquad (3.14)$$

However, in practice this minimum cannot be achieved for two reasons. Firstly, the allocation is based on estimates of the spatial variance within the strata. These estimates might be biased, for example if they are based on data from outside the study area. But even if we have unbiased estimates, the estimation error will not be zero. The second reason is that the calculated sampling variance is based on non-integer numbers of sample points in the strata.

Relating the variance for optimum allocation to that of simple random sampling gives:

$$\sigma^2_{STSI(op)} = \sigma^2_{SI} - \frac{1}{n} \sum_{t=1}^{T} W_t\, (m_t - m_A)^2 - \frac{1}{n} \sum_{t=1}^{T} W_t\, (s_t - s_A)^2 \qquad (3.15)$$

Near-optimum allocation
First we established the near-optimum sample sizes of the strata by calculating the optimum sample sizes. Strata with initially less than two points received two. The remaining sample points were allocated optimally to the other strata. The sampling variance of the mean was finally estimated using Equation (3.9).

Results and discussion

Sample sizes of strata

The distribution of the 66 sample points in Bentelo over the strata for the different types of allocation is given in Table 3.3. The proportional sample sizes do not depend on the target variable, but the optimum sample size does. For example, stratum 4 of Bentelo should have nine points to optimize the estimation of the mean P_{max}, but it should have only three points to optimize the estimation of the mean P. This means that the best allocation for a given variable may not be best for another. However, the optimum sample sizes for the six variables were positively correlated (Table 3.4). In general, the correlation coefficients were fairly large (> 0.7).

Table 3.3 Distribution of sample points over LS-strata for different types of allocation for Bentelo. pr = proportional allocation; op = optimum allocation; no = near-optimum allocation; ex = executed allocation

Stratum	Number of sample points													
	pr	op						no						ex
		P_{max}	P	P_{rel}	$A_{0.25}$	$A_{0.35}$	$A_{0.45}$	P_{max}	P	P_{rel}	$A_{0.25}$	$A_{0.35}$	$A_{0.45}$	
1	4	3	7	6	7	5	5	3	6	6	6	5	5	4
2	1	0	0	1	0	0	0	2	2	2	2	2	2	2
3	9	9	11	11	14	11	10	9	11	11	11	11	10	9
4	6	9	3	6	9	8	8	9	3	6	7	7	7	6
5	17	11	16	18	19	15	16	10	15	16	16	14	15	16
6	8	4	4	8	0	7	6	4	4	7	2	6	5	7
7	7	8	5	8	9	8	8	8	5	7	7	7	7	7
8	4	5	3	3	8	6	4	5	3	3	7	5	4	4
9	1	3	4	1	0	2	2	3	4	2	2	2	2	2
10	2	5	3	0	0	0	0	4	3	2	2	2	2	2
11	4	7	7	2	0	0	4	7	7	2	2	2	4	4
12	3	2	3	2	0	4	3	2	3	2	2	3	3	3

Results for Ootmarsum were similar. The optimum sample sizes were also strongly correlated with the proportional sample sizes. This is to be expected because the optimum sample size is directly related to the size of the strata (Eq. 3.12) and because the variance within the strata also increases with increasing size.

For *LS*-stratification the optimum sample size for estimating A_c of several strata will be zero because their estimated spatial variance was zero. Using this stratification, the relative precision of estimates of A_c for near-optimum allocation will be notably less than that for optimum allocation.

Relative precision

We used the ratio of the estimated sampling variance of a simple random sample to that of the strategy under consideration as a measure of the relative precision:

$$\hat{\sigma}^2_{rel} = \frac{\hat{\sigma}^2_{SI}(\hat{m}_A)}{\hat{\sigma}^2_{p}(\hat{m}_A)} \qquad (3.16)$$

From Equations (3.8), (3.10) and (3.13) it is obvious that the relative precision for proportional and optimum allocation is independent of the sample

Table 3.4 Correlation coefficients for sample size of LS-strata for optimum and proportional allocation for Bentelo. n_{pr} = proportional sample size; n_{op} = optimum sample size

n_{pr}	1						
$n_{op}(P_{max})$	0.77	1					
$n_{op}(P)$	0.85	0.72	1				
$n_{op}(P_{rel})$	0.98	0.72	0.85	1			
$n_{op}(A_{0.25})$	0.85	0.78	0.77	0.87	1		
$n_{op}(A_{0.35})$	0.93	0.72	0.77	0.95	0.92	1	
$n_{op}(A_{0.45})$	0.97	0.82	0.86	0.97	0.90	0.96	1
	n_{pr}	P_{max}	P	P_{rel}	$A_{0.25}$	$A_{0.35}$	$A_{0.45}$
		n_{op}					

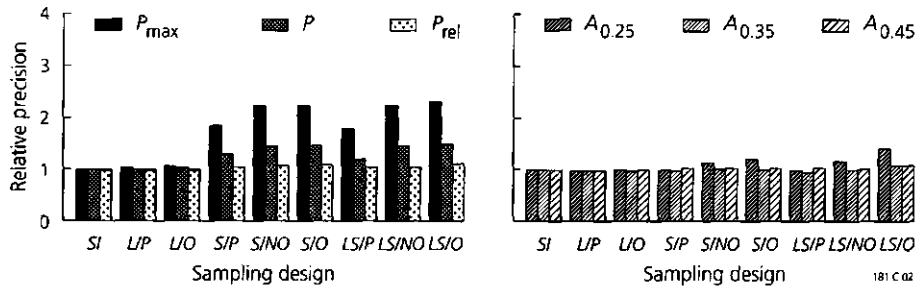


Fig. 3.2 Relative precision of the estimated spatial mean of six soil properties of Bentelo, for nine sampling designs; P_{max}: maximum areic mass of P_2O_5 sorbed by soil, P: areic mass of P_2O_5 sorbed by soil, P_{rel}: relative mass of phosphate sorbed by soil, A_c: areal fraction of soil saturated with phosphate for c as critical level of P_{rel}

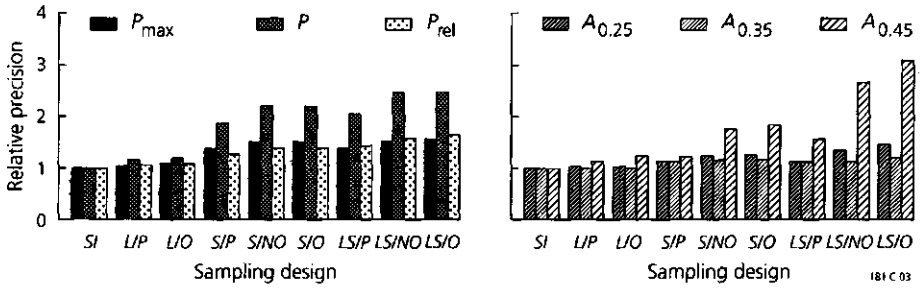


Fig. 3.3 Relative precision of the estimated spatial mean of six soil properties of Ootmarsum, for nine sampling designs; P_{max}: maximum areic mass of P_2O_5 sorbed by soil, P: areic mass of P_2O_5 sorbed by soil, P_{rel}: relative mass of phosphate sorbed by soil, A_c: areal fraction of soil saturated with phosphate for c as critical level of P_{rel}

size n. However, the relative precision for near-optimum allocation depends on n. As n increases, the relative precision for near-optimum allocation approaches that of optimum allocation. The relative precision is presented in Figures 3.2 and 3.3.

Land use map stratification
In general, the relative precision of designs with land use stratification was close to 1. So in general land use stratification will not be worthwile. Only in Ootmarsum we expect a small gain for P and $A_{0.45}$ (Fig. 3.3). The gain for P can be explained by differences in historical phosphate-load of grassland and arable land. In Bentelo these differences were much smaller as a result of crop rotation and intense manuring of both land use types in recent years.

Soil map stratification
For P_{max} and P the relative precision of designs with soil map stratification was considerable larger than 1 in both areas (Figs 3.2 and 3.3). For optimum allocation the relative precision was considerably larger than for proportional allocation. From this we can conclude that the spatial means and spatial standard deviations of P_{max} and P of soil map units differ considerably (Eqs. 3.12 and 3.15). This can partly be explained by the strong correlation between P_{max} and $\bar{W}_{min}$ (correlation coefficients for Bentelo 0.85 and for Ootmarsum 0.80) and between P and soil map unit in Ootmarsum. In Bentelo the gain for P was smaller than in Ootmarsum perhaps because recent manuring of all soil evened out differences between soil map units.

Given the design, for P_{rel} (the ratio of P and P_{max}) the relative precision was smaller than that of P and P_{max} in both areas and for all designs. In Bentelo the relative precision was even close to 1 (e.g. 1.07 for *S/O*-design), whereas those of P and P_{max} for this design were 1.45 and 2.22 respectively. This can possibly be explained by the moderate correlation of P and P_{max} in Bentelo (correlation coefficient: 0.68).

Given the critical P_{rel}-value (c) for all designs, the relative precision for A_c in Ootmarsum was larger than that in Bentelo. This is consistent with the

relatively large gain in precision for P_{rel} in Ootmarsum. For A_c the relative precision differed markedly for c. For example, in Ootmarsum (Fig. 3.3) the relative precision for $c = 0.45$ was 1.81 for the *S/O* design, whereas it was 1.15 for $c = 0.35$, for the same design. In Ootmarsum 0.45 was at the margin of the frequency distribution of P_{rel}: there were only few points with $P_{rel} > 0.45$ and these points were concentrated inside a few strata. In other words, for $c = 0.45$ there were many strata with very small A_c and therefore a very small internal spatial variance, and few strata with a larger A_c and spatial variance. By taking these differences in proportions and variances into account when allocating, considerable gain in precision may be expected. Analogous to this, the gain for $c = 0.25$ in Bentelo can be explained: there were only few points with $P_{rel} < 0.25$.

Stratification according to land use map and soil map
In Ootmarsum the relative precision for designs with land use plus soil map stratification was larger than for designs with soil map stratification only (given the type of allocation), especially for P and $A_{0.45}$. This is consistent with the result that for these characteristics land use stratification would also be worthwile. For P the gain will approximately equal the sum of the single gains; for A_c the gain is greater.

Conclusions

1. Soil map stratification improved the estimation of the spatial means of the maximum areic mass of P_2O_5 sorbed by soil (P_{max}) and the areic mass of P_2O_5 sorbed by soil (P) in Bentelo and Ootmarsum.
2. For the relative mass of phosphate sorbed by soil (P_{rel}) and the areal fraction of soil saturated with phosphate (A_c) soil map stratification produced an increase in precision only in Ootmarsum, where historical phosphate load is small. This gain can be increased further by stratifying according to land use.
3. For A_c gain depends mainly on the critical value of P_{rel} (c). If c is close to the margin of the frequency distribution of P_{rel}, strata are homogeneous

and a gain in precision may be expected, especially with optimum allocation.

4. In many cases, optimum and near-optimum allocation will give better results than proportional allocation. This stresses the importance of knowing the internal variance of map units.
5. The optimum sample sizes of the strata were different for the six characteristics. Therefore, for a soil survey with more than one target variable we should either select one of them as being the most important or make some compromise (Cochran, 1977, p. 119). Another simple solution is to allocate proportionally.

Acknowledgements

I thank G. Ebbers and H. Van het Loo for providing the soil map of Ootmarsum and M.J.D. Hack-Ten Broeke, H. Kleijer and J. Reijerink for providing the random sample data. I thank J. Bouma, A. Breeuwsma and J.J. De Gruijter for critical review of the paper. Last but not least, I would like to thank R. Webster for improving the English.

References

Brus, D.J., De Gruijter, J.J. & Breeuwsma, A. 1992. Strategies for updating soil survey information: a case study to estimate phosphate sorption characteristics. *Journal of Soil Science* 43, 567-581.

Brus, D.J. & De Gruijter, J.J. 1993. Does kriging really give unbiased and minimum variance predictions of spatial means? *Journal of Soil Science* 44, no 4.

Cassel, C-M., Särndal, C-E. & Wretman J.H. 1977. *Foundations of inference in survey sampling.* Wiley, New York.

Cochran, W.G. 1977. *Sampling techniques.* Wiley, New York.

De Gruijter, J.J. & Ter Braak, C.J.F. 1990. *Model-free estimation from spatial samples: a reappraisal of classical sampling theory.* Mathematical Geology 22, 407-415.

Ebbers, G. & Van het Loo, H. 1992. *Bodemkaart van Nederland, Schaal 1:50 000.* Toelichting bij kaartblad 28 Oost - 29 Almelo-Denekamp. DLO-Staring Centrum, Wageningen.

Hack-Ten Broeke, M.J.D., Kleijer, H., Breeuwsma, A., Reijerink, J.G.A. & Brus, D.J. 1990. *Fosfaatverzadiging van de bodem in twee gebieden in Overijssel.* Rapport 108, DLO- Staring Centrum, Wageningen.

Hoekstra, C. & Poelman. J.N.B. 1982. *Dichtheid van gronden gemeten aan de meest voorkomende bodemeenheden in Nederland.* Stichting voor Bodemkartering, Wageningen, Rapport nr. 1582.

Marsman, B.A. & De Gruijter, J.J. 1986. *Quality of soil maps. A comparison of soil survey methods in a sandy area.* Soil Survey Institute. Wageningen, the Netherlands. Soil Survey Papers, 15.

Schoumans, O.F., Breeuwsma, A. & De Vries, W. 1987. *Use of Soil Survey Information for assessing the phosphate sorption capacity of heavily manured soils.* In: W.A. Duijvenboden and H.G. Van Waegeningh (eds): Vulnerability of soil and groundwater to pollutants. National Institute of Public Health and Environmental Hygiene. The Hague. Proceedings and information, 38: 1079-1088.

Stichting voor Bodemkartering 1979. *Bodemkaart van Nederland, Schaal 1:50 000.* Toelichting bij de kaartbladen 34 West (Enschede), 34 Oost (Enschede), 35 (Glanerbrug). Pudoc, Wageningen.

Van der Zee, S.E.A.T.M. 1988. *Transport of reactive contaminants in heterogeneous soil systems.* Doctoral thesis. Agricultural University, Wageningen.

Van der Zee, S.E.A.T.M., Van Riemsdijk, W.H. & De Haan, F.A.M. 1990. *Het protokol fosfaatverzadigde gronden, deel I (Toelichting) en deel II (Technische uitwerking).* Landbouwuniversiteit, Wageningen. Vakgroep Bodemkunde en Plantevoeding.

Chapter 4

Stratification by soil map units to improve estimates of spatially varying soil properties at points

Journal of Soil Science, submitted

with *Jaap J. De Gruijter*
Ben A. Marsman
Reind Visschers
Arnold K. Bregt
Auke Breeuwsma
Johan Bouma

Stratification by soil map units to improve estimates of spatially varying soil properties at points

A specially designed case study investigated the effect of soil map stratification on the accuracy of estimates of four soil properties at points. Six estimation methods were examined, using different weighting functions: global mean, moving average, nearest neighbour, inverse squared distance, Laplacian smoothing splines and ordinary point kriging. The soil properties estimated were thickness of A1 horizon, maximum areic mass of phosphate sorbed by soil, mean highest water table and mean lowest water table.

The performance of the methods was measured by estimating the spatial means of squared and absolute error (quality criteria not conditional on the sample of test points) by a stratified simple random sample of test points. The mean squared error was very large in proportion to the spatial variation in the total area for all methods and properties.

Differences between methods were small. In general, no statistically significant stratification or weighting effects on the quality of estimates were found. The effect of weighting plus stratification was generally not significant either, except for kriging and inverse squared distance weighting. Weighting with inverse squared distance was as satisfactory as weighting by ordinary point kriging. However, the latter was superior near data points. Also, when combined with soil map stratification, kriging was more reliable in the sense that it estimated all properties well. Estimates obtained using the means of *six* soil map units were better than those obtained from unstratified kriging and as good as kriging within *three* map units.

Introduction

Conventional soil maps show the spatial pattern of multivariate soil classes, defined in terms of many soil properties such as presence or absence of diagnostic soil horizons, texture, organic carbon content, depth to water table. These properties are assumed to be strongly correlated, so that the spatial pattern of the classes is not too intricate to be mapped on a single map. Such maps can be used in many applications, ranging from landuse suitability studies (Van

Lanen *et al.*, 1992) to soil protection studies (Breeuwsma *et al.*, 1986). For this, thematic maps showing the spatial pattern of single soil properties, relevant to the problem under consideration, are derived from the general-purpose map. However, these thematic maps are not always sufficiently accurate. The inaccuracy is caused by:

- the imperfect correlation between the properties. As a result, the boundaries on the general-purpose map are a compromise. These compromise boundaries may differ considerably from the boundaries of the target property depicted on the thematic map;
- the abrupt change of the value of the target variable at the boundaries of the map units, whereas in reality lateral changes will often be more gradual;
- spatial autocorrelation within soil map units is neglected.

Giltrap (1978) and Burgess & Webster (1980a,b) were the first to propose to map soil properties by kriging to overcome these problems. In kriging use is made of a variogram which describes spatial autocorrelation as a continuous function of the lag vector $\boldsymbol{h}$ separating two points. However, kriging ignores sharp boundaries which do exist in reality. Further, it is unrealistic to assume that the variogram is similar for all soil classes.

Stein *et al.* (1988) and Voltz & Webster (1990) addressed this shortcoming by combining soil map classification and kriging. They used the soil map to stratify the survey area, and then they interpolated the soil properties in each stratum by kriging, using data from that stratum only. Stein *et al.* (1988) used separate variograms for groups of map units, whereas Voltz & Webster (1990) used a single pooled variogram within map units, because separate variograms would require too many data. Stratified kriging using a pooled variogram gave a 7% decrease of mean square error of prediction in an area with sharp boundaries between map units. Stein *et al.* (1988) and Voltz & Webster (1990) did not select the test points by probability sampling and were therefore unable to quantify the accuracy of their results and to test the statistical significance of the decrease.

One serious drawback of kriging is the large number of observation points needed to estimate the variogram (Webster and Oliver, 1992). If kriging is im-

practicable, more empirical estimation methods such as nearest neighbour, moving average, inverse squared distance or splines may be attractive. Using a density of observation points of 1.5 per ha, Van Kuilenburg *et al.* (1982) found that the estimates of moisture supply capacity in a coversand area by inverse squared distance weighting and ordinary point kriging were almost equally accurate and those of nearest neighbour were statistically significantly worse. Laslett *et al.*, (1987) compared the performance of several estimation methods at a field scale with a grid spacing of 10 m. They found very small differences in the accuracy of estimates of pH by inverse squared distance, ordinary point kriging and Laplacian smoothing splines. Bregt (1992) compared the estimation methods local mean, global mean, inverse distance and kriging at several grid densities ranging from 8 to 200 observations of the depth to the pyritic layer per km^2. He found no statistically significant differences between local mean, inverse distance and kriging at any density.

All these heuristic estimation methods may be combined with soil map stratification as in kriging and this is what we did in this study, which had four main aims:

- to assess the effect of soil map stratification on the accuracy of estimates at points for six estimation techniques (global mean, moving average, nearest neighbour, inverse squared distance, Laplacian smoothing splines and ordinary point kriging);
- to assess the effect of the estimator on the accuracy of estimates;
- to draw attention to the true spatial mean of the estimation error as a criterion to assess the performance of estimation methods, because this criterion is not conditional on the sample of test points, and to stress the importance of quantifying the precision of an estimate of this spatial mean of estimation error;
- to see whether the precision of the estimated spatial mean of estimation error can be increased by stratifying the sample of test points according to distance to observation point and soil map unit.

The global mean and ordinary point kriging methods have the advantage that the mean estimation error can be obtained directly by applying theory. Test points are not needed. We compared these theoretically derived estimation

errors with those obtained from the test points.

In our study the soil map was used for stratification only. By doing this, not all information embodied in the soil map is used. Heuvelink & Bierkens (1992) and Van Meirvenne *et al.* (1993) showed how estimates of the mean of map units can be combined with estimates obtained by kriging and stratified kriging. We included estimation from representative profile descriptions as a separate estimation technique.

Survey region & data collected

Data were collected from a representative, weakly undulating coversand area around the village of Wesepe (north-east of Deventer), in the Province of Overijssel, The Netherlands (Figure 4.1). The study covered a rectangular area, 6 km from east to west and 8 km from north to south, with a great variety of soil types. The main soil type is Typic Haplaquod (Soil Survey Staff, 1975) with water table classes ranging from III (wet) to VII (dry) (Van der Sluijs & De Gruijter, 1985). On top of the coversand ridges, Plaggepts and Plaggeptic Haplaquods with deep water tables occur. The valleys are dominated by Typic Humaquepts. Here and there these Humaquepts are strongly enriched by iron (bog iron ore) and have a clay cover of less than 40 cm.

Soils were sampled by auger at 188 observation points and 96 test points. The observation points form a square grid of 12×16 points, 500 m apart, with four missing points (non-soil). To test the accuracy of the estimation methods, we selected test points by a stratified simple random sampling design (see **Estimation error**).

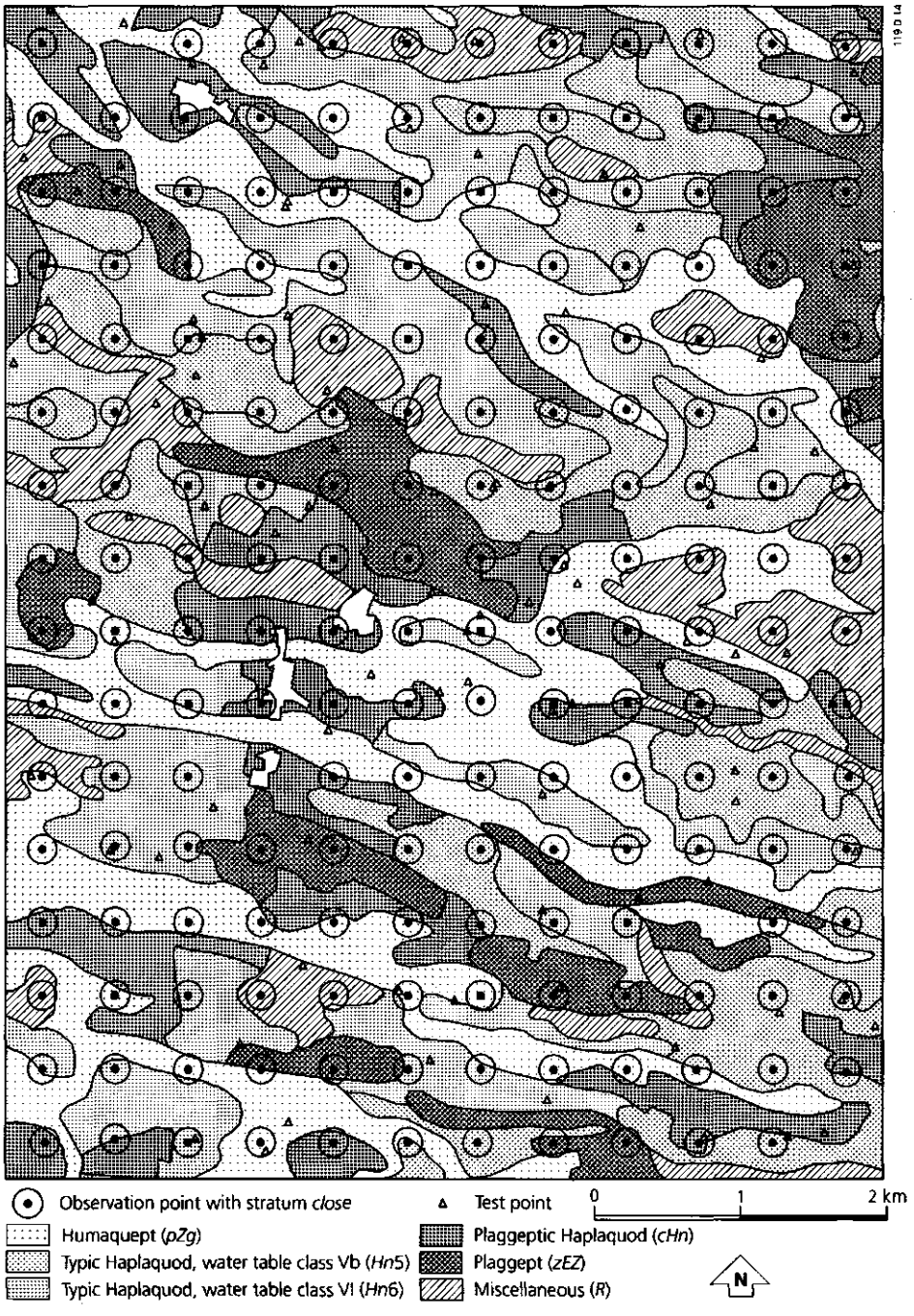


Fig. 4.1 Study area with location of observation points and test points

We measured the following soil properties:

- thickness of mollic or umbric or plaggen epipedon (d(A1), in cm);
- maximum areic mass of P_2O_5 sorbed by soil above $\bar{W}_{min}$ (P_{max}, in kg per m^2);
- mean highest water table ($\bar{W}_{min}$, in cm below surface);
- mean lowest water table ($\bar{W}_{max}$, in cm below surface).

The spatial means and standard deviations of these properties within the map units are shown in Table 4.1.

Table 4.1 Spatial mean ($\hat{\underline{m}}$) and spatial standard deviation ($\hat{\underline{s}}$) of soil properties inside total area (A) and inside groups of soil map units estimated from the stratified simple random sample of test points (*pZg...*, see Fig. 4.1)

Soil map unit	d(A1) (cm)		P_{max} (kg m^{-2})		$\bar{W}_{min}$ (cm)		$\bar{W}_{max}$ (cm)	
	$\hat{\underline{m}}$	$\hat{\underline{s}}$	$\hat{\underline{m}}$	$\hat{\underline{s}}$	$\hat{\underline{m}}$	$\hat{\underline{s}}$	$\hat{\underline{m}}$	$\hat{\underline{s}}$
Total area (A)	29	18	2.1	1.5	62	44	132	48
pZg	25	7.2	2.5	2.2	40	14	99	15
Hn5	23	10	1.2	0.5	45	16	120	17
Hn6	26	28	1.6	1.0	59	29	133	32
cHn	36	12	2.8	1.4	80	44	157	52
zEZ	51	21	3.3	1.4	134	79	207	76
R	27	4.7	1.7	0.6	54	22	119	23

P_{max} was calculated by the following regression equation (Van der Zee *et al.*, 1990):

$$P_{max} = \sum_{l=1}^{L} 0.5\ M_{ox,l}\ \delta_l\ \rho_l\ 0.71 \qquad (4.1)$$

where:

0.5 = regression parameter (-);
$M_{ox,l}$ = oxalate-extractable aluminium + iron content of soil horizon *l* (mol kg^{-1});
δ_l = thickness of horizon *l* as far as lying above $\bar{W}_{min}$, or if $\bar{W}_{min}$ > 1 m, as far as lying within 1 m (m);
ρ_l = volumic mass of horizon *l* (kg m^{-3});
L = number of horizons;
0.71 = conversion coefficient (kg mol^{-1}).

We measured $M_{ox,l}$ and ρ_l of all horizons at each point.

Both $\bar{W}_{min}$ and $\bar{W}_{max}$ were estimated in the field, mainly on the basis of hydromorphic characteristics in the soil profile (see Van der Sluijs & De Gruijter, 1985 for definitions and estimation technique). Both estimates were corrected for systematic measurement error by measuring the water table depth at 21 points and simultaneously at reference points with known $\bar{W}_{min}$ and $\bar{W}_{max}$. These measurements were done at a time the water table reached $\bar{W}_{min}$ or $\bar{W}_{max}$ at the reference points. The depths of the water table at the 21 points on these day were used as errorless measurements of $\bar{W}_{min}$ and $\bar{W}_{max}$ when calibrating the field estimates by regression analysis.

Volumic mass of soil (ρ_l) was measured by taking volumetric samples with a gouge auger. These measurements were calibrated against those from samples taken with rings. To this end two samples were taken from 30 horizons, 10 cm apart, and one gouge auger sample was taken in the middle (Visschers en Marsman, 1991).

Estimation methods

The soil properties were estimated at the test points by six main estimation methods: global mean (*GM*), moving average (*N*3), nearest neighbour (*N*1), inverse squared distance (*ISD*), Laplacian smoothing splines (*LSS*) and ordinary point kriging (*OK*). All methods were used without stratification (e.g. *OK*-0) and within three strata (e.g. *OK*-3). The global mean estimation method was

also used within six strata (*GM*-6). Including the representative profile description method (*RP*), this gives fourteen estimation methods.

On the basis of expert knowledge from previous studies we clustered the original soil map units into three or six strata. The clustering varied with the soil property.

Representative profile description (*RP*)
In conventional soil survey, a representative soil profile description is given for each map unit. The values of the soil properties derived from this description should be close to the mean or the median and can be used as estimators for all points within this map unit. In general, such estimates of the mean are biased and introduce a systematic error.

Global mean (*GM*-0, *GM*-3, *GM*-6)
In *GM*-0, a *p*-unbiased estimate of the spatial mean of the whole study area ($\hat{m}_A$) is used as an estimator. This method would be appropriate if there were no spatial autocorrelation in the area. This is not very realistic and this method was incorporated for reference purposes only. In *GM*-3 and *GM*-6 the estimated means of three (*GM*-3) and six (*GM*-6) soil map units ($\hat{m}_u$) are used as estimators.

Moving average (*N*3-0, *N*3-3)
The value at a test point is estimated as the unweighted mean of the three nearest observation points. In *N*3-0 no stratification was applied. In *N*3-3 there were three strata and the mean of the three nearest points in the stratum of the test point was used as an estimator.

Nearest neighbour (*N*1-0, *N*1-3)
The value at the nearest observation point is used as an estimator. In *N*1-3 there were three strata, and the value at the nearest point in the same stratum as the test point was used.

Inverse squared distance (*ISD*-0, *ISD*-3)
In this method the value at a test point is estimated as a weighted average

of the values at the observation points. The weights are proportional to the inverse of the squared distance:

$$\lambda_i = \frac{1}{d_i^2} \left[\sum_{j=1}^{n} \frac{1}{d_j^2}\right]^{-1} \qquad (4.2)$$

where:
d_i = distance from test point $\boldsymbol{x}_0$ to observation point $\boldsymbol{x}_i$.

In *ISD*-3 only the observation points in the same map unit as the test point were used.

Laplacian smoothing splines (*LSS*-0, *LSS*-3)
By smoothing splines in two dimensions the following quantity is minimized:

$$n^{-1} \sum_{i=1}^{n} [z(x_i,y_i) - f(x_i,y_i)]^2 + \alpha \iint [(\delta^2 f/\delta x^2)^2 + 2(\delta^2 f/\delta x \delta y)^2 + (\delta^2 f/\delta y^2)^2] dx dy \qquad (4.3)$$

where:
$z(x_i,y_i)$ = value at observation point i;
$f(x_i,y_i)$ = value of the spline function at observation point i;
α = non-negative parameter.

The first term measures the goodness-of-fit to the data, the second term is a measure of the rate of change of slope and therefore a measure of roughness of the estimation surface. The parameter α reflects the relative importance of these two terms; in this study it was determined by cross-validation. For further details, see Wahba & Wendelberger (1980) and Hutchinson & Gessler (1993). In *LSS*-3 Equation (4.3) is minimized for the observation points in the same map unit as the test point.

Ordinary point kriging (*OK*-0, *OK*-3)
As in inverse squared distance, the predicted value is a weighted average of the values at the observation points. In kriging use is made of a stochastic model ξ describing the joint distribution of the variables $\underline{Z}(\boldsymbol{x}_i)$ to calculate the

weights (Journel & Huijbregts, 1978). The covariance function is an important part of this model. It describes the relation between the covariance of the values measured at two points and $\boldsymbol{h}$, where $\boldsymbol{h}$ is the lag vector between the two points. In practice, instead of the covariance, one often models the semivariance, which is defined as half the variance of $\{\underline{z}(\boldsymbol{x}_i) - \underline{z}(\boldsymbol{x}_j)\}$.

The optimal weights are obtained by solving the equations:

$$\begin{aligned} &\sum_{j=1}^{n} \lambda_j \gamma(\boldsymbol{x}_i, \boldsymbol{x}_j) + \psi = \gamma(\boldsymbol{x}_i, \boldsymbol{x}_0) \qquad \forall\ i = 1 \text{ to } n \\ &\sum_{j=1}^{n} \lambda_j = 1 \end{aligned} \tag{4.4}$$

where:

$\gamma(\boldsymbol{x}_i, \boldsymbol{x}_j)$ = semivariance of $\underline{z}(\boldsymbol{x}_i)$ and $\underline{z}(\boldsymbol{x}_j)$;
ψ = Lagrange multiplier;
$\gamma(\boldsymbol{x}_i, \boldsymbol{x}_0)$ = semivariance of $\underline{z}(\boldsymbol{x}_i)$ and the random variable $\underline{z}$ at the test point $\boldsymbol{x}_0$.

By adding the test points in variogram estimation, we obtained estimates of the semivariance at short distances (< 500 m). Nonlinear models were fitted in Genstat by maximizing the likelihood according to the modified Newton method. We used the inverse of the numbers of pairs as weights. In *OK*-3, we estimated semivariances and fitted models for the strata separately. Figure 4.2 and Table 4.2 show the results. For *d*(A1) there was clear anisotropy, for the other properties this was not significant. For all properties, the variograms for the three strata were clearly different. The difference in sill was most striking, but the range, nugget and model-type also differed. For example, we fitted pure nugget models to the sample variogram of P_{max} in stratum *pZg* (Humaquepts), whereas spherical models without nugget were fitted for stratum *cHn, zEZ* (plaggeptic Haplaquods, Plaggepts).

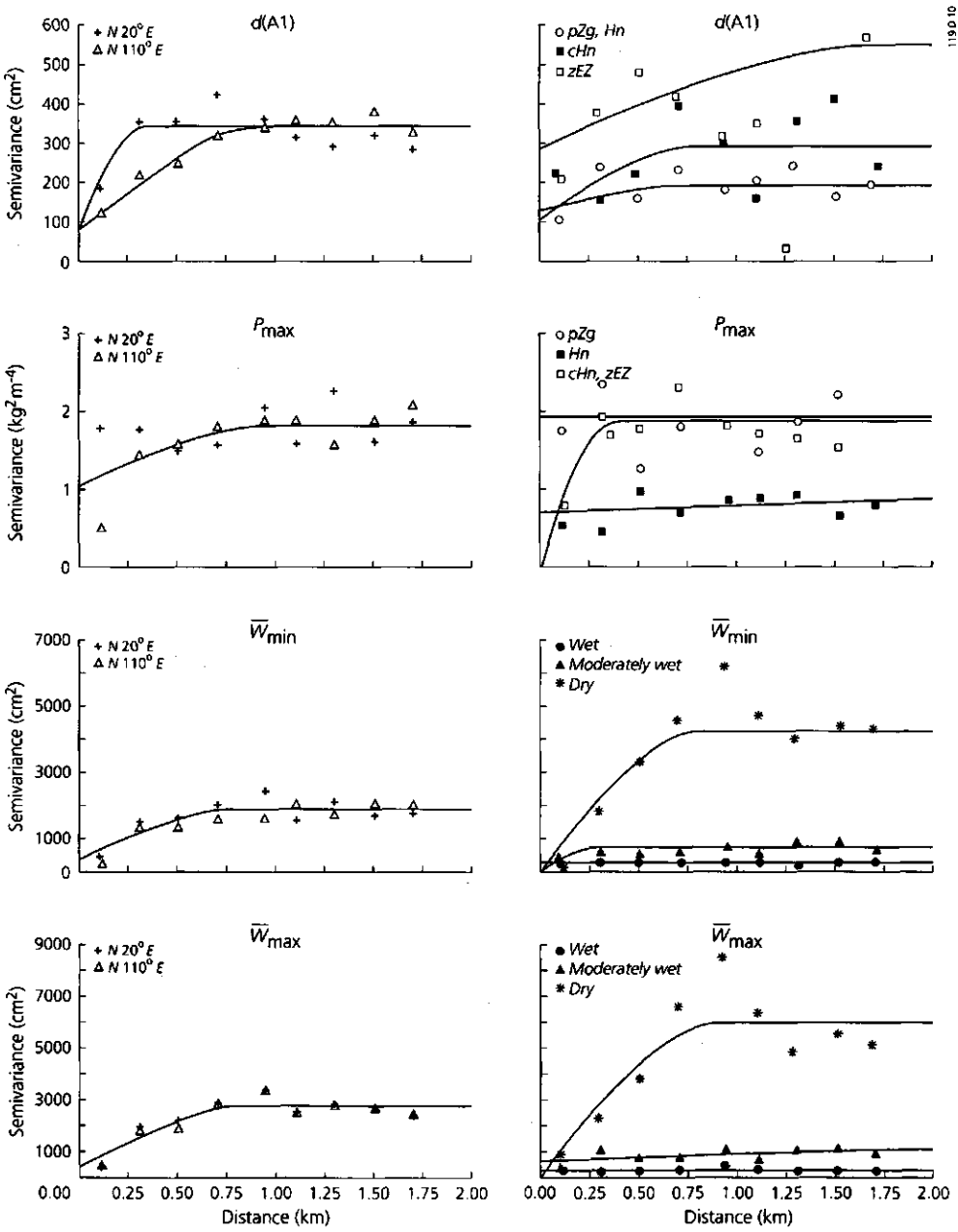


Fig. 4.2 Sample variograms and fitted models for d(A1), P_{max}, $\bar{W}_{min}$ and $\bar{W}_{max}$ for total area (left-hand side) and for soil map units (right-hand side). For values of parameters, see Table 4.2

Table 4.2 Type and parameter values of variogram models

Soil property	Stratum	Model	c_0	c	a(km)
No stratification					
d(A1)	-	spherical + nugget anisotropic	81	260	1.050-0.220
P_{max}	-	spherical + nugget	1.06	0.77	1.026
$\bar{W}_{min}$	-	spherical + nugget	283	1626	0.859
$\bar{W}_{max}$	-	spherical + nugget	340	2401	0.921
With stratification					
d(A1)	*pZg, Hn*	spherical + nugget	125	60	0.735
	cHn	spherical + nugget	101	179	0.772
	zEZ	spherical + nugget	281	253	1.780
P_{max}	*pZg*	pure nugget	1.88	-	-
	Hn	linear	0.68	0.079	
	cHn, zEZ	spherical	-	1.80	0.338
$\bar{W}_{min}$	wet	pure nugget	242	-	-
	mod.wet	spherical	-	715	0.293
	dry	spherical	-	4174	0.851
$\bar{W}_{max}$	wet	exponential	236	124	1.09
	mod.wet	spherical + nugget	649	325	1.64
	dry	spherical	-	5844	0.941

Estimation error

As measures of estimation accuracy we used the spatial means of the squared error ($m_A(\varepsilon^2)$) and of the absolute error ($m_A(|\varepsilon|)$). Unlike the Mean Squared Error (*MSE*) and Mean Absolute Error (*MAE*) which are usually defined as the arithmetic means of the errors at test points, $m_A(\varepsilon^2)$ and $m_A(|\varepsilon|)$ are not conditional on the sample of test points. Van Kuilenburg *et al.* (1982) were the first who proposed to use $m_A(\varepsilon^2)$ as a measure of estimation accuracy. The spatial mean of squared error is defined as:

$$m_A(\varepsilon^2) = \frac{1}{N}\sum_{i=1}^{N}\{\hat{z}(\boldsymbol{x}_i) - z(\boldsymbol{x}_i)\}^2 \qquad (4.5)$$

where:
N = total number of possible sampling locations in A.

The definition of $m_A(|\varepsilon|)$ is simply obtained by substituting the absolute error in Equation (4.5).

Empirical estimation
These spatial means were estimated by a stratified simple random sample (*STSI*) of test points. The two stratification criteria we used were the distance to the observation points, and the soil map units (Fig. 4.1). It is well known that the estimation error is related to the distance to the observation points. Therefore we expected this stratification to yield more precise estimates of the spatial mean of the estimation error. We distinguished two strata using this distance criterion: close to the nearest observation point (≤ 100 m) and further away from the nearest observation point (> 100 m).

In previous studies it was shown that the spatial variance of our soil properties (d(A1), $\bar{W}_{min}$, $\bar{W}_{max}$ and P_{max}) within soil map units differed statistically significantly (Marsman and De Gruijter, 1986; Brus *et al.*, 1992). Given the sample size, the mean estimation error in map units with a large internal variance will generally also be relatively large. Therefore we also stratified according to soil map unit. The units of the 1:50 000 soil map were clustered into six groups. Combining the two distance strata with the six map strata resulted in a total of twelve strata. The test points were allocated as follows:
- to the map unit strata, proportionally to their area;
- to the strata 'close to observation points' and 'distant from observation points', 1/4 and 3/4 of the points of a map unit respectively.

In stratified simple random sampling the spatial mean of the squared error is estimated by:

$$\underline{\hat{m}}_A(\varepsilon^2) = \sum_{h=1}^{H} W_h \underline{\hat{m}}_h(\varepsilon^2) = \sum_{h=1}^{H} W_h \frac{\sum_{i=1}^{n_h} \{\hat{z}(\underline{x}_i) - z(\underline{x}_i)\}^2}{n_h} \qquad (4.6)$$

where:

H = number of strata;

W_h = weight of stratum h calculated as the proportion of the area of stratum h;

$\underline{\hat{m}}_h(\varepsilon^2)$ = estimated spatial mean of the squared error of stratum h.

The sampling variance of the estimated mean under any stratified simple random sampling design p can be estimated by (Cochran, 1977; p. 92):

$$\hat{\sigma}_p^2\{\underline{\hat{m}}_A(\varepsilon^2)\} = \sum_{h=1}^{H} W_h^2 \, \hat{\sigma}_p^2\{\underline{\hat{m}}_h(\varepsilon^2)\} = \sum_{h=1}^{H} W_h^2 \frac{\hat{v}_h(\varepsilon^2)}{n_h} (1 - f_h) \qquad (4.7)$$

where:

$\hat{\sigma}_p^2\{\underline{\hat{m}}_h(\varepsilon^2)\}$ = estimated sampling variance of the sample mean of the squared error of stratum h;

$\hat{v}_h(\varepsilon^2)$ = estimated spatial variance of the squared error in stratum h;

n_h = sample size of stratum h;

f_h = sampling fraction of stratum h.

The sampling fractions are close to zero for all strata, so the last term of Equation (4.7) can be dropped. These sampling variances can be used to calculate confidence intervals of $\underline{\hat{m}}_A(\varepsilon^2)$. The confidence interval of the square root of the estimated spatial mean of squared errors can be calculated by taking the square root of the upper and lower bounds of the confidence intervals of $\underline{\hat{m}}_A(\varepsilon^2)$.

Relative precision of estimated spatial mean of error

The strategy $(SI,\bar{z})$ consisting of the simple random sampling design (SI) and the unweighted sample mean $(\bar{z})$ is often taken as a reference when considering alternative strategies for estimating the spatial mean. Let p denote some other design, with the same sample size to ensure a fair comparison, and $\underline{\hat{m}}_A$

the estimator of the spatial mean for that design. The relative precision of this strategy $(p,\underline{\hat{m}}_A)$ defined as:

$$\underline{\hat{\sigma}}^2_{rel}(p,\underline{\hat{m}}_A) = \frac{\underline{\hat{\sigma}}^2_{SI}(\bar{z})}{\underline{\hat{\sigma}}^2_{p}(\underline{\hat{m}}_A)} \tag{4.8}$$

expresses how well strategy $(p,\underline{\hat{m}}_A)$ performs compared to the reference strategy $(SI,\bar{z})$. For the sampling design p we took *STSI* with proportional allocation, $STSI_{pr}$, and the estimator of Equation (4.6), and *STSI* with optimum allocation, $STSI_{op}$ (Särndal *et al.*, 1992). The relative precision for proportional and optimum allocation are independent of the sample size n:

$$\begin{aligned}
\underline{\hat{\sigma}}^2_{rel}(STSI_{pr},\underline{\hat{m}}_A) &= \frac{\underline{\hat{V}}_A(\varepsilon^2)}{\sum_h W_h \underline{\hat{V}}_h(\varepsilon^2)} \\
\underline{\hat{\sigma}}^2_{rel}(STSI_{op},\underline{\hat{m}}_A) &= \frac{\underline{\hat{V}}_A(\varepsilon^2)}{(\sum_h W_h \underline{\hat{s}}_h(\varepsilon^2))^2}
\end{aligned} \tag{4.9}$$

where:

$\underline{\hat{s}}_h(\varepsilon^2)$ = estimated spatial standard deviation of the squared error in stratum h.

The estimation surface of a given soil property differs between methods, and consequently so does the surface of estimation errors. From this it follows that the mean and variance of estimation error of the strata depend on the soil property interpolated *and* on the estimation method. As a result, the relative precision may differ for each combination of soil property and estimation method.

Deriving estimation errors from theory

In *GM* and *OK* an estimate of the spatial mean squared error can be derived directly, without test points, from theory. In *GM*-0, $m_A(\varepsilon^2)$ is equal to the spatial variance of z within the total area, $v_A(z)$, plus the squared difference between the estimated and the true global mean of z (Brus *et al.*, 1992). In *GM*-3 and *GM*-6, $m_A(\varepsilon^2)$ equals a weighted average of the $m_u(\varepsilon^2)$ of the map units. As the

true spatial means of z are unknown, the difference between estimated and true mean is unknown. To avoid this problem we consider the expectation of the estimation error at points over all realizations of the sampling design. Each realization, i.e. each set of observation points, gives an estimate of the spatial mean of z in A and consequently an estimate of the value at a point i in A. The expectation of this error due to the difference between estimated spatial mean and true spatial mean of z is 0. Therefore, we estimated the spatial mean of the p-expected, squared errors, $m_A\{E_p(\underline{\varepsilon}^2)\}$ over repeated sampling of observation points. (Note that the errors are stochastic, not the true spatial mean.) For *GM*-0 this spatial mean equals:

$$m_A\{E_p(\underline{\varepsilon}^2)\} = [E_p\{\underline{\hat{m}}_A(z)\} - m_A(z)]^2 + \sigma_p^2\{\underline{\hat{m}}_A(z)\} + v_A(z) \qquad (4.10)$$

As stated before, the first term equals zero. The second term, the sampling variance of the estimated mean of z in A, $\sigma_p^2\{\underline{\hat{m}}_A(z)\}$, and the third term, the spatial variance of z in A, $v_A(z)$, should be estimated from the sample. However, we used a systematic sampling design, and in that case the model of spatial correlation must be known so that the sampling variance and spatial variance can be estimated. To circumvent this, we estimated the sampling variance and spatial variance by their simple random sample estimators. These estimates are generally conservative, i.e. the sampling variance and spatial variance are overestimated (Särndal *et al.*, 1992, p. 80). From this it follows that the estimate of the spatial mean of $E_p(\underline{\varepsilon}^2)$ is conservative too.

Using *OK* the variance of the prediction error at point $\boldsymbol{x}_0$ can be calculated by:

$$\sigma_\xi^2\{\underline{\tilde{z}}(\boldsymbol{x}_0) - \underline{z}(\boldsymbol{x}_0)\} = \sum_{i=1}^{n} \lambda_i \, \gamma(\boldsymbol{x}_i, \boldsymbol{x}_0) \qquad (4.11)$$

As indicated by the subscript ξ this kriging variance denotes the variance over realizations of a stochastic model ξ and therefore differs entirely from the sampling variance (Brus and De Gruijter, 1993). The spatial mean of this kriging variance ($m_A(\sigma_\xi^2)$) equals:

$$m_A(\sigma_\xi^2) = m_A\{E_\xi(\underline{\varepsilon}^2)\} = E_\xi\{m_A(\underline{\varepsilon}^2)\} \quad (4.12)$$

Thus because $m_A(\sigma_\xi^2)$ is calculated over all realizations from ξ, it will generally be unequal to $m_A(\varepsilon^2)$. One should be aware of this when comparing these two quantities.

Results and discussion

General results

Tables 4.3 and 4.4 show the square root of the estimated spatial means of the squared errors ($\sqrt{\underline{\hat{m}}_A(\varepsilon^2)}$) and of the absolute errors ($\underline{\hat{m}}_A(|\varepsilon|)$) respectively. Most striking is the very large values (at least approximately 80 %) of $\sqrt{\underline{\hat{m}}_A(\varepsilon^2)}$ in proportion to the spatial standard deviation in the total area for all methods and properties. Apparently, a large part of the spatial variation remained within the map units and in between the observation points. This raises the question of whether a quality criterion based on the error of estimates at points is realistic, given the scale of the soil map and the density of the observation points. Soil surveyors generally agree that 1:50 000 soil maps should not be used for estimation at points, or for estimating the means of small (< 5 ha) plots.

The small differences between methods in $\sqrt{\underline{\hat{m}}_A(\varepsilon^2)}$ and in $\underline{\hat{m}}_A(|\varepsilon|)$ are very remarkable too, especially when related to the wide 90% confidence intervals. This already suggests there are only few statistically significant differences between methods (see hereafter). Despite this, we ranked the methods from best (rank 1) to worst (rank 14). The tables show that *GM*-6 and *OK*-3 were ranked low (≤ 4) and *N*1-0 was ranked high (≥ 13) for the squared error as well as the absolute error. *GM*-6 performed better than *OK*-0, although not significantly: the difference in $\sqrt{\underline{\hat{m}}_A(\varepsilon^2)}$ and their 90% confidence intervals were 1.2 ± 3.0, 0.05 ± 1.3, 2.8 ± 6.2 and 4.7 ± 7.5 for d(A1), P_{max}, $\bar{W}_{min}$ and $\bar{W}_{max}$ respectively. This shows that the soil map is as good a model of spatial variation or even better than the variogram. The ranking for the squared error differed only marginally from that for the absolute error; therefore we will focus on the squared error.

Table 4.3 Square root of the estimated spatial mean of squared error, $\sqrt{\underline{\hat{m}}_A(\varepsilon^2)}$, and rank for fourteen estimation methods and four soil properties; between brackets: 90% confidence interval

Estimation method	d(A1)		P_{max}		$\bar{W}_{min}$		$\bar{W}_{max}$	
	$\sqrt{\underline{\hat{m}}_A(\varepsilon^2)}$ (cm)	rank	$\sqrt{\underline{\hat{m}}_A(\varepsilon^2)}$ (kg m^{-2})	rank	$\sqrt{\underline{\hat{m}}_A(\varepsilon^2)}$ (cm)	rank	$\sqrt{\underline{\hat{m}}_A(\varepsilon^2)}$ (cm)	rank
RP	19.9 (11.4-25.6)	13	-	-	37.0 (27.3-44.6)	6	43.2 (30.2-53.2)	9
GM-0	18.3 (11.3-23.3)	8	1.51 (1.06-1.85)	10	43.6 (30.9-53.4)	13	48.2 (37.5-57.0)	13
GM-3	16.6 (6.9-22.4)	2	1.44 (0.93-1.80)	6	38.8 (30.7-45.4)	10	44.6 (35.9-51.9)	12
GM-6	16.7 (7.0-22.5)	3	1.41 (0.91-1.77)	1	35.5 (28.0-40.0)	4	38.0 (31.2-43.8)	1
*N*3-0	19.5 (13.1-24.3)	12	1.50 (1.10-1.81)	9	39.4 (29.6-47.2)	12	43.6 (36.0-50.1)	10
*N*3-3	18.6 (8.5-24.8)	9	1.41 (0.92-1.77)	3	36.3 (26.3-44.1)	5	40.7 (32.0-47.8)	6
*N*1-0	20.8 (15.8-24.9)	14	1.70 (1.30-2.03)	13	47.4 (34.1-57.8)	14	55.0 (42.1-65.5)	14
*N*1-3	19.5 (11.0-25.2)	11	1.60 (1.18-1.93)	12	34.2 (27.0-40.1)	2	39.5 (31.4-46.3)	4
ISD-0	17.6 (10.9-22.4)	6	1.48 (1.03-1.82)	8	38.4 (26.2-47.5)	9	40.6 (30.4-48.8)	5
ISD-3	17.1 (6.6-23.3)	5	1.43 (0.97-1.78)	5	34.4 (25.4-41.5)	3	38.2 (30.4-44.8)	2
LSS-0	18.9 (12.2-23.8)	10	1.52 (1.06-1.87)	11	38.8 (28.5-46.9)	11	44.0 (36.7-50.1)	11
LSS-3	16.8 (6.3-22.8)	4	1.41 (0.94-1.75)	2	37.5 (27.7-45.2)	7	42.7 (32.6-50.9)	8
OK-0	17.9 (11.6-22.5)	7	1.46 (1.04-1.79)	7	38.3 (28.1-46.3)	8	42.7 (34.8-49.4)	7
OK-3	16.4 (5.9-22.4)	1	1.42 (0.90-1.79)	4	34.1 (25.1-41.2)	1	39.0 (31.1-45.6)	3

Table 4.4 Estimated spatial mean of absolute error, $\underline{\hat{m}}_A(|\varepsilon|)$, and rank for fourteen estimation methods and four soil properties; between brackets: standard error of estimates

Estimation method	d(A1)		P_{max}		$\bar{W}_{min}$		$\bar{W}_{max}$	
	$\underline{\hat{m}}_A(\|\varepsilon\|)$ (cm)	rank	$\underline{\hat{m}}_A(\|\varepsilon\|)$ (kg m^{-2})	rank	$\underline{\hat{m}}_A(\|\varepsilon\|)$ (cm)	rank	$\underline{\hat{m}}_A(\|\varepsilon\|)$ (cm)	rank
RP	12.2 (1.7)	11	-	-	24.7 (3.3)	7	29.3 (3.7)	6
GM-0	11.4 (1.5)	6	1.05 (0.12)	11	29.5 (3.4)	13	35.8 (3.3)	13
GM-3	9.7 (1.4)	3	0.95 (0.12)	6	25.8 (2.8)	8	30.9 (3.2)	9
GM-6	9.5 (1.5)	2	0.90 (0.12)	2	23.8 (2.2)	4	26.5 (2.5)	2
*N*3-0	13.2 (1.6)	13	1.01 (0.12)	9	28.4 (3.0)	12	33.3 (3.1)	11
*N*3-3	11.7 (1.6)	8	0.90 (0.12)	1	23.7 (2.8)	3	27.6 (3.1)	4
*N*1-0	13.4 (1.8)	14	1.16 (0.14)	13	31.3 (3.9)	14	36.9 (4.5)	14
*N*1-3	12.2 (1.6)	10	1.08 (0.13)	12	24.1 (2.5)	5	27.7 (2.8)	5
ISD-0	11.6 (1.4)	7	0.99 (0.12)	7	26.1 (3.0)	9	29.9 (3.0)	8
ISD-3	10.2 (1.5)	4	0.94 (0.12)	5	23.0 (2.7)	2	26.2 (2.8)	1
LSS-0	12.7 (1.5)	12	1.05 (0.12)	10	27.8 (3.0)	11	33.3 (3.1)	12
LSS-3	10.2 (1.4)	5	0.93 (0.12)	3	24.5 (2.9)	6	29.4 (3.2)	7
OK-0	11.7 (1.5)	9	1.00 (0.12)	8	26.7 (3.0)	10	31.6 (3.1)	10
OK-3	9.5 (1.4)	1	0.94 (0.12)	4	22.2 (2.8)	1	27.0 (2.8)	3

We also calculated the spatial mean of the squared errors for the strata 'close to observation points' and 'distant from observation points' (Fig. 4.3). The results for the stratum 'close to observation points' may reveal the performance of the estimation methods in finer grids. However, situations are not completely identical, because in finer grids there is stronger spatial autocorrelation between the test point and all observation points and between the observation points themselves. The $\sqrt{\hat{m}_u(\varepsilon^2)}$ in the stratum 'close to observation points' (hereafter referred to as *close*) was markedly smaller than 'distant from observation points' (hereafter referred to as *distant*), especially for P_{max}, $\bar{W}_{min}$ and $\bar{W}_{max}$. This suggests that it is justified to assume spatial autocorrelation and to use distance to observation points in estimating. This was underlined by the relatively inaccurate estimates of $\bar{W}_{min}$ and $\bar{W}_{max}$ by *GM*-0. All other methods gave significantly ($\alpha = 0.10$) more accurate estimates of these properties. Figure 4.3 also makes clear that the relatively inaccurate estimates of $\bar{W}_{min}$ and $\bar{W}_{max}$ by *N*1-0 in the total area resulted from the poor performance of this method in the stratum *distant*.

For P_{max} the differences between methods in the stratum *close* were large compared to the stratum *distant* and to the total area. *GM*-3, *GM*-6, *LSS*-3, *OK*-0 and *OK*-3 performed best, with very small mutual differences, and performed significantly ($\alpha = 0.10$) better than *N*1-0, *N*1-3, *ISD*-0 and *ISD*-3. This can be explained by the large nugget of the variogram of P_{max} (Table 4.2), and consequently the spiky surface of this property. Interpolators, such as *ISD* and *N*1, going through the data, give too much credence to these spikes and consequently yielded inaccurate estimates near observation points. *OK* is also an interpolator, but the kriged surface is discontinuous at the observation points because the nugget is non-zero, which explains its relatively good performance.

We will now go into the weighting effect before and after stratification, the stratification effect and the effect of weighting plus stratification. Figures 4.4 and 4.5 show the differences in $\sqrt{\hat{m}_A(\varepsilon^2)}$ and their 90% confidence intervals for pairs of methods.

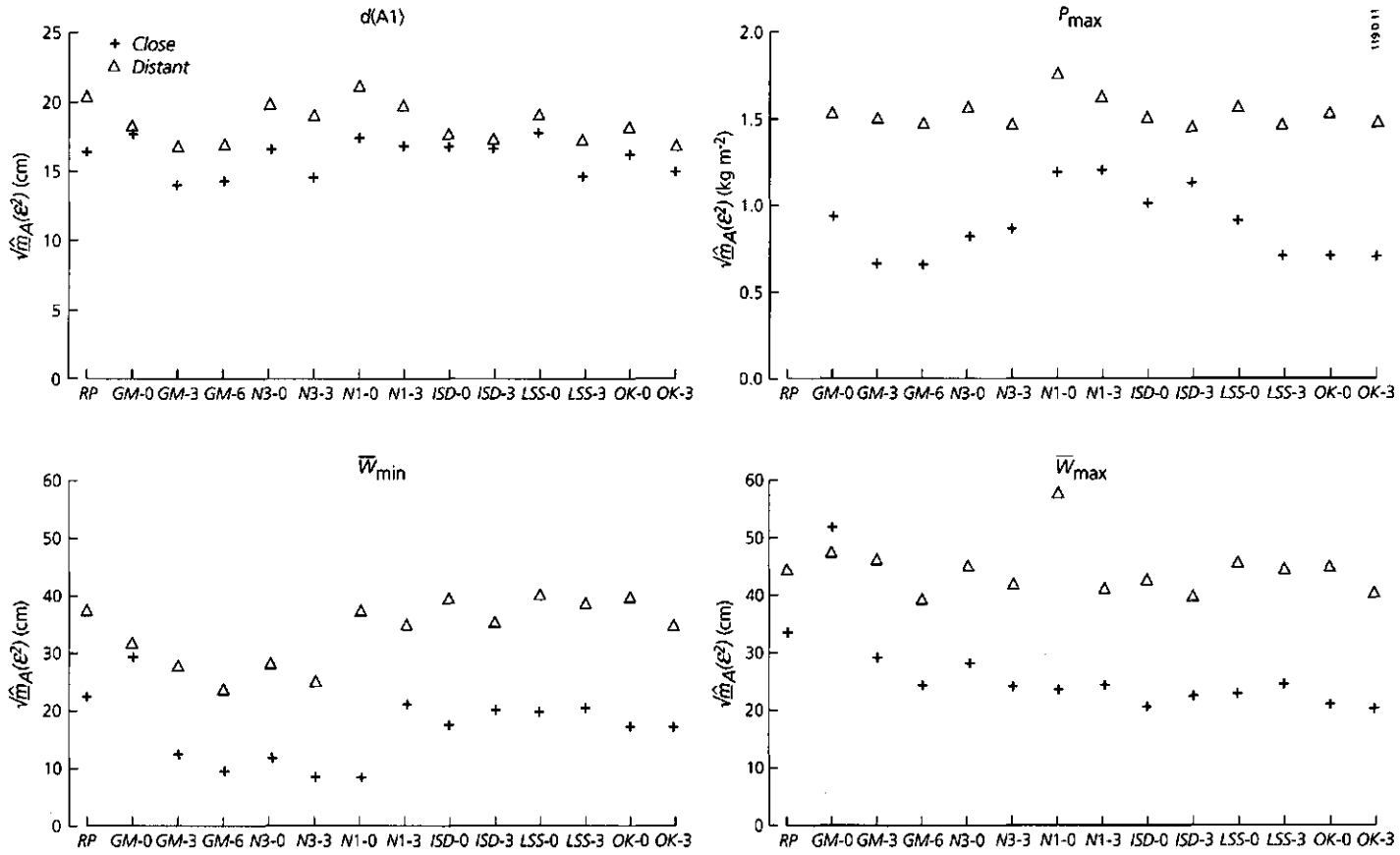


Fig. 4.3 Square root of estimated spatial mean of squared error ($\sqrt{\hat{m}_A(\varepsilon^2)}$) for strata 'close to observation points' and 'distant from observation points' for fourteen estimation methods

Weighting and stratification effect

To find out more about the performance of the methods we assessed the effects of (i) the weighting function, (ii) the stratification, and (iii) the weighting function plus stratification on the spatial mean of squared error. These effects can be traced by comparing pairs of estimation methods. For the weighting effect we compared *N*3-0, *N*1-0, *ISD*-0, *LSS*-0 and *OK*-0 with *GM*-0. Any linear estimator can be written as $\lambda' \mathbf{z}$ where λ' is the transpose of the vector with weights attached to the observation points and $\mathbf{z}$ is the vector with the values of the soil property at the observation points. In all methods except *GM* the weights vary and are a function of the locations of the observation points and the test point. In *GM* all observation points are assigned an equal weight. Comparing the accuracy of the other methods with *GM* therefore yields the effects of weighting.

For the stratification effect we compared *GM*-3 and *GM*-6 with *GM*-0, *N*3-3 with *N*3-0, *N*1-3 with *N*1-0 and so on. The combined effect of stratification and weighting can be assessed by comparing *N*3-3, *N*1-3, *ISD*-3, *LSS*-3 and *OK*-3 with *GM*-0.

The weighting effect may also be assessed by comparing *N*3-3, *N*1-3, *ISD*-3, *LSS*-3 and *OK*-3 with *GM*-3. This yields the weighting effect *after stratification*, which may differ notably from the weighting effect before stratification. It shows whether estimates can be improved (with respect to *GM*-3) by using a weighting function within map units. We quantified these effects by the differences in the estimated roots of the spatial means of squared errors, $\sqrt{\hat{m}_A(\varepsilon^2)}$. The sampling variance of these differences were approximated by Taylor extension (see **Appendix**).

Weighting effect

In general, there was no significant weighting effect[1]: the differences in $\sqrt{\hat{m}_A(\varepsilon^2)}$ were small, whereas the 90% confidence intervals of these differences were wide. Only ISD weighting had a significant, positive effect on estimates of $\bar{W}_{min}$

[1] A significant effect means that the difference in square root of the estimated spatial mean of the squared error differed statistically significantly from zero at $\alpha = 0.10$.

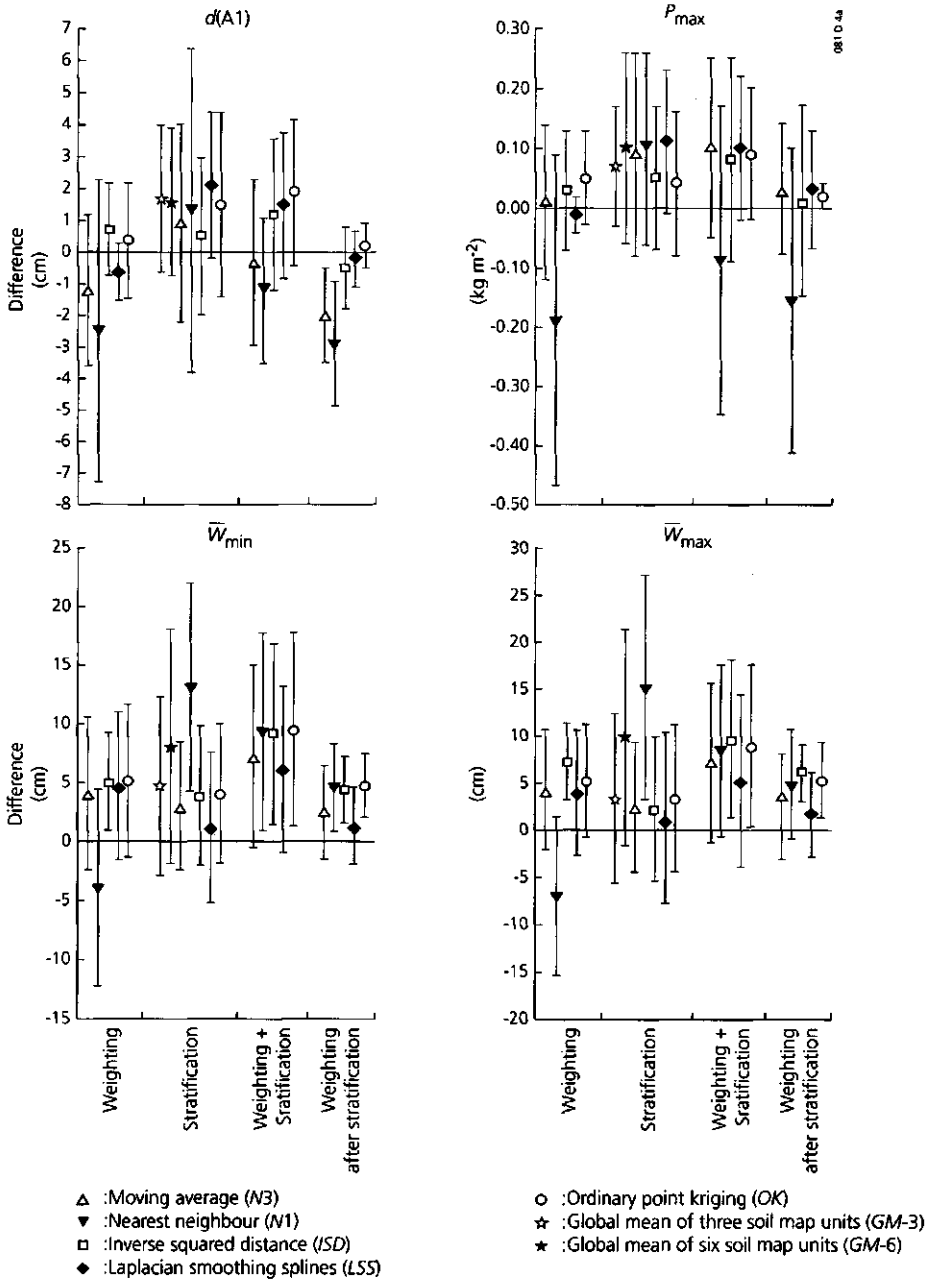


Fig. 4.4 Weighting effect, stratification effect, weighting plus stratification effect and weighting effect after stratification for d(A1), P_{max}, $\bar{W}_{min}$ and $\bar{W}_{max}$ quantified as differences in the root of the spatial mean of squared error for total area. The bars indicate the 90% confidence interval of these differences. How to read this figure: see explanation at the bottom of p. 99

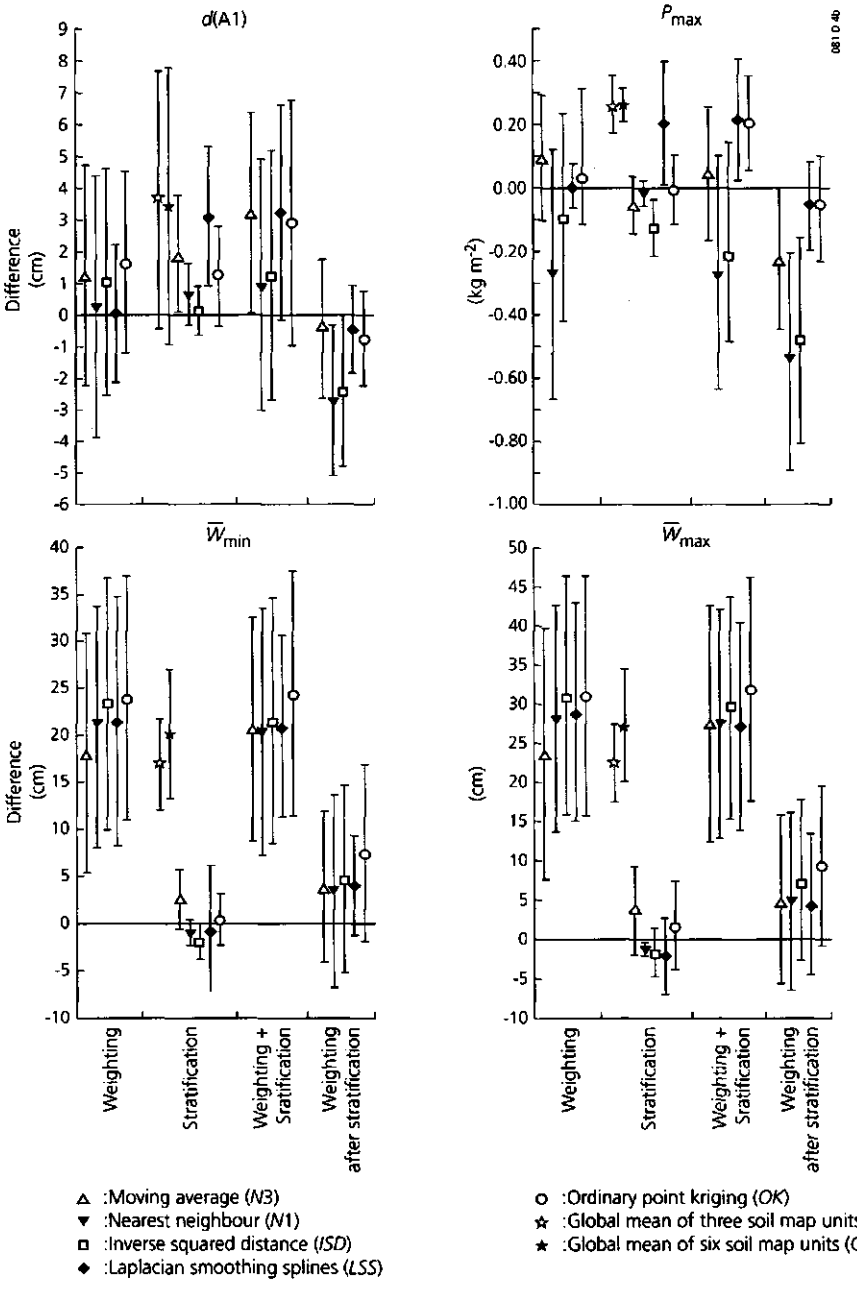


Fig. 4.5 Weighting effect, stratification effect, weighting plus stratification effect and weighting effect after stratification for d(A1), P_{max}, $\bar{W}_{min}$ and $\bar{W}_{max}$ quantified as differences in the root of the spatial mean of squared error for stratum close to observation points. The bars indicate the 90% confidence interval of these differences. How to read this figure: see explanation at the bottom of p. 99

and $\bar{W}_{max}$. Despite this, we can draw some cautious conclusions. *ISD* weighting and *OK* weighting seem to be the most reliable in the sense that these weighting functions slightly decreased the $\sqrt{\underline{\hat{m}}_A(\varepsilon^2)}$ for all properties, whereas the other weighting functions showed a worsening for one or more properties. For all properties the spatial means of squared errors were *in*creased to a greater or lesser degree by *N*1-weighting. Therefore this weighting function seems to be unsuitable.

As opposed to the total area, in the stratum *close* there was a large, significant and relevant effect on the estimates of $\bar{W}_{min}$ and $\bar{W}_{max}$ for all weighting functions. This can be explained by the strong spatial correlation of these properties (small nugget effect; Fig. 4.2). The differences between the weighting functions were small. *OK* weighting also had a significant positive effect on P_{max}, whereas *ISD* weighting had no significant effect on the quality of the estimates of this property. Therefore, *ISD* weighting seems to perform slightly worse than *OK* weighting in this stratum, which underlines the already mentioned poor performance of interpolators near observation points.

Stratification effect

Figure 4.4 shows that generally no significant stratification effect could be traced from the sample of test points. Only for *N*1 estimates of $\bar{W}_{max}$ and $\bar{W}_{min}$ was there a significant, positive effect which can be explained by the poor performance of *N*1 weighting without stratification (see previous section). Nevertheless, for all properties and all methods, soil map stratification slightly reduced $\sqrt{\underline{\hat{m}}_A(\varepsilon^2)}$.

How to read Figs 4.4 and 4.5: for example, the weighting effect of moving average estimation (*N*3) of the thickness of the A1-horizon (*d*(A1)) is -1.2 cm (symbol Δ on the lefthand side at the top of Fig. 4.4) which means that $\sqrt{\underline{\hat{m}}_A(\varepsilon^2)}$ obtained by global mean estimation (*GM*-0) minus $\sqrt{\underline{\hat{m}}_A(\varepsilon^2)}$ obtained by *N*3-0 estimation is -1.2 cm. *N*3-0 performs worse than *GM*-0. The 90% confidence interval of this difference covers the value 0, therefore this weighting effect is not significant at $\alpha = 0.10$.

For the stratum *close* the decrease in $\sqrt{\hat{m}_A(\varepsilon^2)}$ due to stratification was generally smaller than for the total area (Fig. 4.5). For *ISD* estimates of P_{max} and $\bar{W}_{min}$ there was even a significant *negative* effect. Stratification may increase the distance of the test point from the nearest observation point. In situations with gradual transitions between map units, this increase may involve a reduction in correlation if the distance is less than the range of the variogram. As a result, estimation methods assigning large weights to the nearest observation points may yield less accurate estimates. In contrast for *GM* the decrease in $\sqrt{\hat{m}_A(\varepsilon^2)}$ of $\bar{W}_{max}$, $\bar{W}_{min}$ and P_{max} due to stratification in the stratum *close* was much larger than in the total area.

Combined effect of weighting and stratification

Unexpectedly, there was no significant combined effect of weighting and stratification for most properties and estimation methods either (Fig. 4.4). There were significant, positive effects on estimates of $\bar{W}_{max}$ and $\bar{W}_{min}$ by *ISD*-3 and *OK*-3 only. For *d*(A1) and P_{max} the decrease in $\sqrt{\hat{m}_A(\varepsilon^2)}$ obtained by *OK*-3 was nearly significant ($\alpha = 0.10$), whereas *ISD*-3 clearly showed no significant effect for these properties. Again, *OK* appeared to be the most reliable method. An explanation for this is that in *OK* a model of spatial autocorrelation is estimated from the data, whereas, for example, *ISD* implicitly makes use of a postulated, spatial autocorrelation model not estimated from the data. This model may differ markedly from the true model. However, the sample of observation points often fails to meet the requirements for accurate estimation of the variogram (too few samples or inadequate configuration) and consequently *OK* often requires additional data.

For the stratum *close* there was a significant, positive combined effect for all weighting functions on estimates of $\bar{W}_{max}$ and $\bar{W}_{min}$ (Fig. 4.5). The good performance of *LSS*-3 is remarkable: it showed a significant, positive combined effect for P_{max}, $\bar{W}_{max}$ and $\bar{W}_{min}$.

Weighting effect after stratification

There was a relatively small but significant, positive weighting effect after stratification for *OK* and *ISD* on estimates of $\bar{W}_{min}$ and $\bar{W}_{max}$ (Fig. 4.4). For P_{max}, the decrease in $\sqrt{\hat{m}_A(\varepsilon^2)}$ brought about by weighting after stratification was small

for all weighting functions and significant only for *OK*. For d(A1) all weighting functions except *OK* led to an increase of $\sqrt{\hat{m}_A(\varepsilon^2)}$. To sum up: *OK* weighting was the best. It led to a decrease of $\sqrt{\hat{m}_A(\varepsilon^2)}$ for all properties, whereas the other weighting functions increased the $\sqrt{\hat{m}_A(\varepsilon^2)}$ of one or more properties.

For the stratum *close*, application of the interpolators *ISD* and *N*1 after stratification had a significant, negative effect on estimates of d(A1) and P_{max} (Fig. 4.5).

Estimates of expected error derived from theory

For *GM*-0, *GM*-3 and *GM*-6, the estimated spatial mean of p-expected squared error, $\hat{m}_A\{E_p(\varepsilon^2)\}$, was smaller than $\hat{m}_A(\varepsilon^2)$, except for $\bar{W}_{max}$ estimated by *GM*-0 and *GM*-6 (Table 4.5). However, because of the large standard error of $\hat{m}_A(\varepsilon^2)$ these differences were not significant. For *OK*-0 and *OK*-3, the estimated spatial mean of the kriging variance ($\hat{m}_A(\sigma_\xi^2)$) which is equal to the estimated spatial mean of ξ-expected squared errors, $\hat{m}_A\{E_p(\xi^2)\}$, was less than $\hat{m}_A(\varepsilon^2)$ in all cases (Table 4.6). Again, these differences were not significant because of the large standard error of $\hat{m}_A(\varepsilon^2)$. Note the small standard errors of $\hat{m}_A(\sigma_\xi^2)$.

Table 4.5 Estimated spatial mean of squared error, $\hat{m}_A(\varepsilon^2)$, and of p-expected squared error, $\hat{m}_A\{E_p(\xi^2)\}$, for estimation methods global mean (*GM*-0) and means of map units (*GM*-3, *GM*-6); between brackets: standard error of estimates

	d(A1) (cm^2)		P_{max} (kg^2 m^{-4})		$\bar{W}_{min}$ (cm^2)		$\bar{W}_{max}$ (cm^2)	
	$\hat{m}_A(\varepsilon^2)$	$\hat{m}_A\{E_p(\xi^2)\}$	$\hat{m}_A(\varepsilon^2)$	$\hat{m}_A\{E_p(\xi^2)\}$	$\hat{m}_A(\varepsilon^2)$	$\hat{m}_A\{E_p(\xi^2)\}$	$\hat{m}_A(\varepsilon^2)$	$\hat{m}_A\{E_p(\xi^2)\}$
GM-0	334 (123)	334	22.8 (6.0)	16.3	1900 (564)	1896	2325 (549)	2995
GM-3	274 (135)	227	20.6 (7.1)	13.6	1503 (333)	1073	1991 (418)	1599
GM-6	278 (136)	219	19.8 (6.9)	12.0	1193 (244)	1021	1445 (281)	1519

Table 4.6 Estimated spatial mean of squared error, $\underline{\hat{m}}_A(\varepsilon^2)$, and of kriging variance, $\underline{\hat{m}}_A(\sigma_\xi^2)$, for ordinary point-kriging (*OK*-0) and stratified ordinary point-kriging (*OK*-3); between brackets: standard error of estimates

	d(A1) (cm^2)		P_{max} (kg^2 m^{-4})		$\bar{W}_{min}$ (cm^2)		$\bar{W}_{max}$ (cm^2)	
	$\underline{\hat{m}}_A(\varepsilon^2)$	$\underline{\hat{m}}_A(\sigma_\xi^2)$	$\underline{\hat{m}}_A(\varepsilon^2)$	$\underline{\hat{m}}_A(\sigma_\xi^2)$	$\underline{\hat{m}}_A(\varepsilon^2)$	$\underline{\hat{m}}_A(\sigma_\xi^2)$	$\underline{\hat{m}}_A(\varepsilon^2)$	$\underline{\hat{m}}_A(\sigma_\xi^2)$
OK-0	321	250	21.3	15.7	1467	1086	1827	1418
	(111)	(3.7)	(6.5)	(0.04)	(404)	(14)	(367)	(21)
OK-3	268	203	20.0	13.0	1164	876	1524	964
	(139)	(5.9)	(7.1)	(0.5)	(319)	(112)	(330)	(79)

Relative precision of estimated spatial mean of squared error.
Figure 4.6 shows the relative precision of the estimated spatial mean of squared error under stratified simple random sampling using the previously described strata and proportional or optimum allocation. It shows that for proportional allocation the gain in precision was negligible, except for the combinations *GM*-6/$\bar{W}_{min}$ and *GM*-6/$\bar{W}_{max}$. Apparently, only for these combinations were the differences of the mean of squared errors between strata large enough to be exploited in stratified sampling. However, for optimum allocation a considerable gain may be expected in all cases. This means that in particular the variance of the squared error varied considerably between the strata. For the absolute errors the gain for optimum allocation was somewhat smaller, and for proportional allocation there was a small gain too.

For optimum allocation one needs a prior estimate of the standard deviation of the squared error within strata, which might be a problem in practice. However, one may expect relatively large standard deviations further away from observation points and in more heterogeneous map units.

***p*-unbiasedness of estimates at points**
Table 4.7 shows that for all methods the spatial mean of errors differed from zero for most properties; however, this difference was not significant in most cases ($\alpha = 0.05$). The value is related to the sample of observation points. For other systematic samples this value is likely to be different. Therefore the ex-

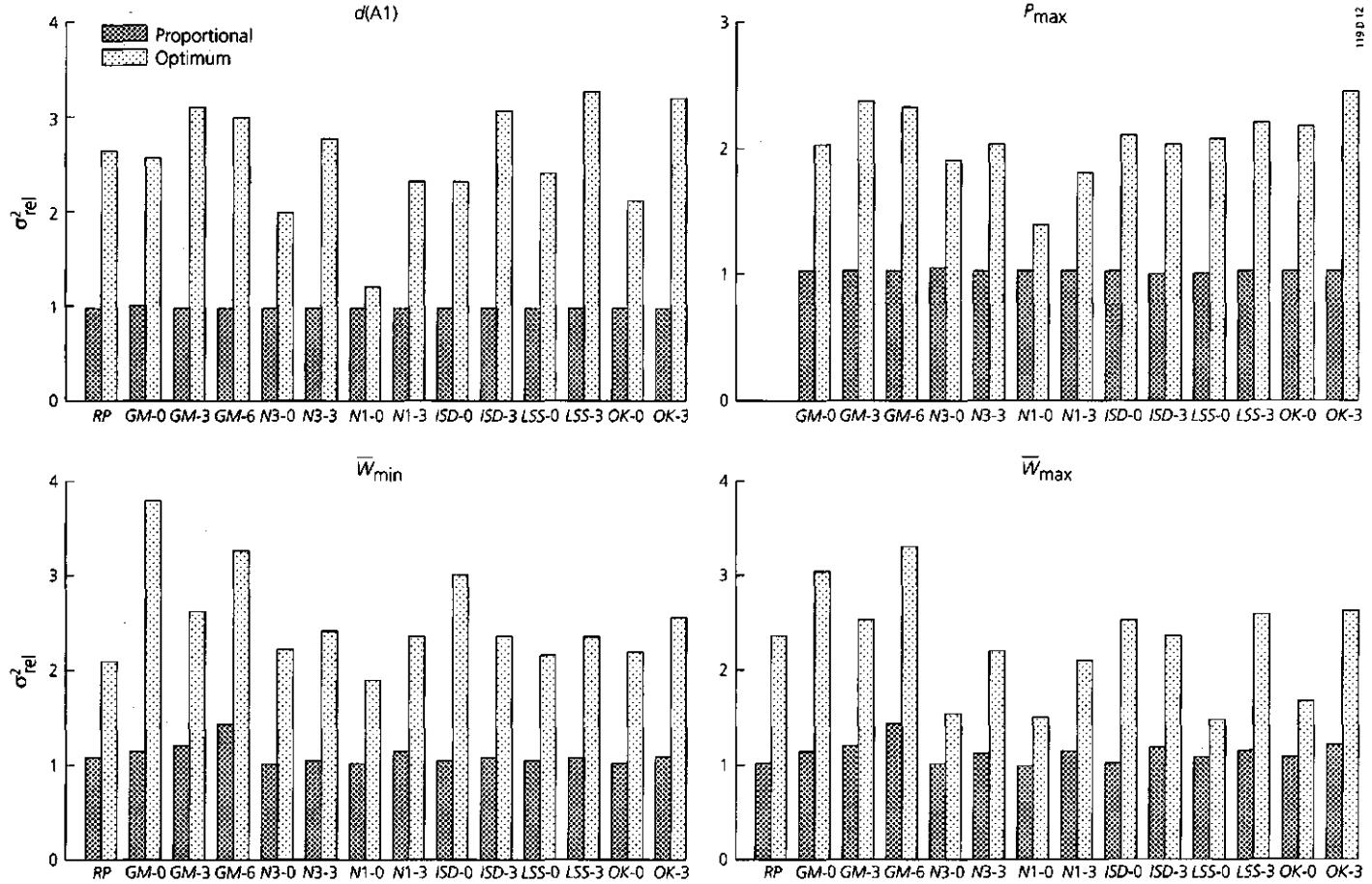


Fig. 4.6 Relative precision of the estimated spatial mean of squared error as estimated by stratified simple random sampling for proportional and optimum allocation

Table 4.7 Estimated spatial mean of error for four soil properties; between brackets: standard error; * value differs significantly from zero ($\alpha = 0.05$)

Estimation method	Estimated spatial mean of error			
	d(A1)	P_{max}	$\bar{W}_{min}$	$\bar{W}_{max}$
RP	-0.7 (1.9)	-	5.5 (4.4)	-7.4 (5.4)
GM-0	2.1 (1.8)	-0.09 (0.15)	-4.8 (3.9)	-9.2 (4.1)*
GM-3	-1.9 (1.8)	-0.09 (0.15)	-4.8 (4.3)	-10.4 (4.8)*
GM-6	-2.0 (1.8)	-0.07 (0.15)	-3.8 (3.9)	-7.7 (4.1)
*N*3-0	-0.6 (2.1)	0.06 (0.16)	-3.4 (4.1)	-7.8 (4.4)
*N*3-3	-1.7 (2.1)	0.00 (0.15)	-3.7 (4.1)	-8.7 (4.4)
*N*1-0	-0.4 (2.3)	-0.02 (0.19)	-4.3 (5.2)	-9.2 (6.1)
*N*1-3	-0.3 (2.2)	-0.09 (0.18)	-3.6 (3.8)	-6.0 (4.3)
ISD-0	-1.2 (1.9)	-0.01 (0.16)	-3.9 (3.9)	-8.2 (3.9)*
ISD-3	-0.5 (1.9)	-0.05 (0.16)	-4.0 (3.9)	-8.8 (4.1)*
LSS-0	-2.5 (2.0)	-0.06 (0.16)	-4.7 (4.1)	-9.3 (4.6)*
LSS-3	-1.8 (1.8)	-0.09 (0.15)	-4.9 (4.2)	-11.5 (4.5)*
OK-0	-1.3 (1.9)	-0.03 (0.16)	-3.2 (4.1)	-8.3 (4.5)
OK-3	-1.2 (1.8)	-0.09 (0.15)	-3.6 (3.9)	-9.7 (4.2)*

pectation of this spatial mean or spatial mean of expected error, over all possible samples under design p is of special importance. We define an estimator of values at points as being p-unbiased if:

$$m_A(E_p[\underline{\varepsilon}]) = E_p[m_A(\underline{\varepsilon})] = 0 \qquad (4.13)$$

It is easy to see that Equation (4.13) holds for *GM*. The other estimation methods make use of a linear estimator, so we have to estimate the p-expectation of the spatial mean of linear combinations. If (i) only points in the area A are used for estimation and (ii) all points have an equal probability of being

included in the sample, then $E_p\{z(\underline{\boldsymbol{x}}_j)\}$ equals the true spatial mean m_A. Moreover, $\Sigma E_p(\underline{\lambda}_{ij}) = E_p(\Sigma\underline{\lambda}_{ij}) = 1$, so:

$$
\begin{aligned}
E_p[m_A(\underline{\varepsilon})] &= E_p[\frac{1}{N}\sum_{j=1}^{N}\{\sum_{i=1}^{n}\underline{\lambda}_{ij}z(\underline{\boldsymbol{x}}_i)\} - z(\boldsymbol{x}_j)] = \\
\frac{1}{N}\sum_{j=1}^{N}\sum_{i=1}^{n}E_p[\underline{\lambda}_{ij}]\ E_p[z(\underline{\boldsymbol{x}}_i)] &+ \sigma_p^2\{\underline{\lambda}_{ij}, z(\underline{\boldsymbol{x}}_i)\} - z(\boldsymbol{x}_j) = \\
\frac{1}{N}\sum_{j=1}^{N}m_A + [\sum_{i=1}^{n}\sigma_p^2\{\underline{\lambda}_{ij}, z(\underline{\boldsymbol{x}}_i)\}] - z(\boldsymbol{x}_j) &= \frac{1}{N}\sum_{j=1}^{N}\sum_{i=1}^{n}\sigma_p^2\{\underline{\lambda}_{ij}, z(\underline{\boldsymbol{x}}_i)\}
\end{aligned}
\qquad (4.14)
$$

where:
$\sigma_p^2\{(\underline{\lambda}_{ij}, z(\underline{\boldsymbol{x}}_i)\}$ = sampling covariance of the weight $\underline{\lambda}_{ij}$ and $z(\underline{\boldsymbol{x}}_i)$.

Note that the weights $\underline{\lambda}_{ij}$ are stochastic because the locations of the observation points $\boldsymbol{x}_i$ are stochastic. So even if requirements (i) and (ii) are met, the linear estimator is *p*-unbiased only if the mean sum of the covariances is 0. If observation points are selected with unequal probability of inclusion, for example by stratified random sampling with non-proportional allocation, *GM* still produces *p*-unbiased estimates at points, because this inequality is taken into account in the estimator. In contrast, all other methods will produce *p*-biased estimates at points, because they ignore the differences in probabilities of inclusion.

Conclusions

How to compare estimation methods

- The spatial mean of estimation error is a more useful criterion for evaluating the performance of estimation methods than the Mean Squared Error of estimates at test points because it is not conditional on the sample of test points.
- Design-based sampling strategies have good potential for calculating confidence intervals of differences in the spatial mean of estimation error and for testing the statistical significance of these differences.
- Stratification of the test sample by distance to observation point and soil

map unit will result in smaller standard errors of the spatial mean of squared error only if sample points are allocated approximately optimally to the strata.

Results of comparison

General results

- For all methods the root of the spatial mean of squared error is very large in proportion to the spatial standard deviation in the total area (at least approximately 80 %). For surveys at scale 1:50 000, a quality criterion based on estimates of means of blocks instead of values at points might be more realistic.
- The differences in the root of the estimated spatial mean of squared error between methods were small, especially when related to their large standard errors. Only few of these differences were statistically significant at $\alpha = 0.10$.
- In circumstances comparable to those under study, the global means of soil map units produce estimates at points as accurate as kriging within groups of soil map units. For the global mean method one needs no variogram, so the costs of sampling will generally be less.

Weighting effect

- No significant weighting effect on the root of the estimated spatial mean of squared errors ($\sqrt{\hat{m}_A(\varepsilon^2)}$) is to be expected if the grid distance exceeds approximately half the range of the property estimated. In finer grids the weighting effect will increase; however, for interpolators like inverse squared distance (*ISD*) this increase will be limited when estimating properties with a noisy spatial variation.
- Weighting according to ordinary point kriging (*OK*) and inverse squared distance (*ISD*) weighting performed best and were generally equivalent; however, the former seemed to be more reliable in the sense that it is less dependent on the spatial variation (variogram) of the property estimated. One advantage of *ISD* over *OK* is that no variogram needs to be estimated and therefore it is a good alternative in the case of coarse sampling grids which make it impossible to estimate the variogram accurately.

Stratification effect

- Soil map stratification will generally have no statistically significant effect on $\sqrt{\hat{m}_A(\varepsilon^2)}$ in circumstances comparable to those in this study. The stratification effect depends on the density of the sampling grid and diminishes with increasing density, except when distance is not applied in weighting, as in the global mean method (*GM*-3, *GM*-6).

Weighting plus stratification effect

- In general, there is unlikely to be a statistically significant effect on $\sqrt{\hat{m}_A(\varepsilon^2)}$ due to weighting plus stratification in circumstances comparable to those we studied. Only stratified *OK* and stratified *ISD* will increase the accuracy compared with the global mean method, for properties with strong spatial autocorrelation such as the mean highest water table and mean lowest water table.
- Stratified *OK* is more reliable than stratified *ISD* and provides more accurate estimates. This superiority increases with increasing density of the sampling grid.

Weighting effect after stratification

- For 1:50 000 soil maps and a grid distance of 500 m, the weighting effect after stratification will generally be small and not significant because no spatial autocorrelation remains within some units, whereas in others the grid distance is generally too long.

Acknowledgements

We thank M.F. Hutchinson for providing a program for estimation by Laplacian smoothing splines, Jan Oude Voshaar for helping with Taylor approximation and Gerard Heuvelink for worthwile comments.

References

Breeuwsma, A., Wösten, J.H.M., Vleeshouwer, J.J., Van Slobbe, A.M. & Bouma, J. 1986. Derivation of land qualities to assess environmental problems from soil surveys. *Soil Science Society of America Journal* 50, 186-190.

Bregt, A.K. 1992. *Processing of soil survey data.* Doctoral thesis. Wageningen Agricultural University Wageningen, The Netherlands.

Brus, D.J., De Gruijter, J.J. & Breeuwsma, A. 1992. Strategies for updating soil survey information: a case study to estimate phosphate sorption characteristics. *Journal of Soil Science* 43, 567-581

Brus, D.J. & De Gruijter, J.J. 1993. Does kriging really give unbiased and minimum variance predictions of spatial means? *Journal of Soil Science* 44, no 4.

Burgess, T.M. & Webster, R. 1980a. Optimal interpolation and isarithmic mapping of soil properties. I. The semivariogram and punctual kriging. *Journal of Soil Science* 31, 315-331.

Burgess, T.M. & Webster, R. 1980b. Optimal interpolation and isarithmic mapping of soil properties. II. Block kriging. *Journal of Soil Science* 31, 333-341.

Cochran, W.G. 1977. *Sampling techniques.* Wiley, New York.

Giltrap, D.J. 1978. *Mathematical techniques for soil survey design.* Doct. Thesis, Univ. Oxford.

Heuvelink, G.B.M. & Bierkens, M.F.P. 1992. Combining soil maps with interpolations from point observations to predict quantitative soil properties. *Geoderma* 55, 1-15.

Hutchinson, M.F. & Gessler, P.E. 1993. Splines - More than just a smooth interpolator. *Geoderma,* in press.

Journel, A.G. & Huijbregts, C.J. 1978. *Mining Geostatistics.* Academic Press, London.

Kendall, M. & Stuart A. 1977. *The advanced theory of statistics.* Volume 1. Distribution theory. Fourth edition. Charles Griffin & Company, London.

Laslett, G.M., McBratney, A.B., Pahl, P.J. & Hutchinson, M.F. 1987. Comparison of several spatial prediction methods for soil pH. *Journal of Soil Science* 38, 325-341.

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.

Särndal, C-E, Swensson, B. & Wretman, J. 1992. *Model assisted survey sampling.* Springer-Verlag, New York.

Soil survey staff. 1975. *Soil Taxonomy.* A basic system of soil classification for making and interpreting soil surveys. Agriculture Handbook 436. Soil Conservation Service. US Department of Agriculture, Washington DC.

Stein, A., Hoogerwerf, M. & Bouma, J. 1988. Use of soil-map delineations to improve (co-)kriging ofpoint data on moisture deficits. *Geoderma* 43, 163-177.

Van der Sluijs, P. & De Gruijter, J.J. 1985. Water table classes: a method to describe

seasonal fluctuation and duration of water tables on Dutch soil maps. *Agricultural Water Management* 10, 109-125.

Van der Zee, S.E.A.T.M., Van Riemsdijk, W.H. & De Haan, F.A.M. 1990. *Het protokol fosfaatverzadigde gronden, deel II (Technische uitwerking).* Agricultural University, Wageningen.

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.

Van Lanen, H.A.J., Hack-Ten Broeke, M.J.D., Bouma, J. & De Groot, W.J.M. 1992. A mixed qualitative/quantitative physical land evaluation methodology. *Geoderma* 55, 37-54.

Van Meirvenne, M., Scheldeman, K., Baert, G. & Hofman, G. 1993. Quantification of soil textural fractions of the Bas-Zaïre using soil map polygons and/or point observations. *Geoderma*, in press.

Visschers, R. & Marsman, B.A. 1991. Meten van de dichtheid van de grond met een gutsboor. *Landinrichting* 31, 33-35.

Voltz, M. & Webster, R. 1990. A comparison of kriging, cubic splines and classification for predicting soil properties from sample information. *Journal of Soil Science* 41, 473-490.

Wahba, G. & Wendelberger, J. 1980. Some new mathematical methods for variational objective analysis using splines and cross validation. *Monthly Weather Review* 108, 1122-1143.

Webster, R. & Oliver, M.A. 1992. Sample adequately to estimate variograms of soil properties, *Journal of Soil Science* 43, 177-192.

Appendix

Taylor approximation of sampling variance of the difference in roots of estimated spatial mean of squared error

The variance of a function f of $\underline{x}$ can be approximated by (Kendall & Stuart, 1977; p. 247):

$$\sigma_p^2\{f(\underline{x})\} \approx [f'\{E_p(\underline{x})\}]^2 \sigma_p^2(\underline{x}) \qquad (4A.1)$$

where:

$f'(E_p(\underline{x}))$ = value of the first derivative of f at $E_p(\underline{x})$.

In our case f($\underline{x}$) is $\sqrt{\underline{x}}$ and $\underline{x}$ is $\underline{\hat{m}}_A(\varepsilon^2)$. From Equation (4A.1) it follows that the sampling variance of $\sqrt{\underline{\hat{m}}_A(\varepsilon^2)}$ can be approximated by:

$$\sigma_p^2\{\sqrt{\underline{\hat{m}}_A(\varepsilon^2)}\} \approx \{\frac{1}{2\sqrt{m_A(\varepsilon^2)}}\}^2 \sigma_p^2\{\underline{\hat{m}}_A(\varepsilon^2)\} = \frac{1}{4}\frac{\sigma_p^2\{\underline{\hat{m}}_A(\varepsilon^2)\}}{m_A(\varepsilon^2)} \qquad (4A.2)$$

The sampling variance of $\{\sqrt{\underline{\hat{m}}_A(\varepsilon^2{}_a)} - \sqrt{\underline{\hat{m}}_A(\varepsilon^2{}_b)}\}$ equals the sampling variance of $\sqrt{\underline{\hat{m}}_A(\varepsilon^2{}_a)}$, plus the sampling variance of $\sqrt{\underline{\hat{m}}_A(\varepsilon^2{}_b)}$, minus twice the sampling covariance of $\sqrt{\underline{\hat{m}}_A(\varepsilon^2{}_a)}$ and $\sqrt{\underline{\hat{m}}_A(\varepsilon^2{}_b)}$. The approximations of the two sampling variances are given by Equation (4A.2); the sampling covariance can be approximated by:

$$\sigma_p^2\{\sqrt{\underline{\hat{m}}_A(\varepsilon_a^2)}\ ,\sqrt{\underline{\hat{m}}_A(\varepsilon_b^2)}\} \approx \frac{1}{4\sqrt{m_A(\varepsilon_a^2)}\sqrt{m_A(\varepsilon_b^2)}}\sigma_p^2\{\underline{\hat{m}}_A(\varepsilon_a^2),\underline{\hat{m}}_A(\varepsilon_b^2)\} \qquad (4A.3)$$

Chapter 5

Strategies for updating soil survey information: a case study to estimate phosphate sorption characteristics

Journal of Soil Science 43, 567-581 (1992)

with Jaap J. De Gruijter
Auke Breeuwsma

Strategies for updating soil survey information: a case study to estimate phosphate sorption characteristics

To support future decisions on alternative strategies for updating soil survey information, the efficiency of four strategies, revision, upgrading, revision plus upgrading and upgrading by two-phase sampling, has been studied. Revision results in a new soil map. Upgrading gives statistical information about means and variances within mapping units of the soil map.

The merits of these strategies were measured in terms of the increase in accuracies of the spatial estimates of the values at unvisited points and of the spatial means of a study area of several soil characteristics. Point values were estimated by assigning an estimate of the mean value to any point in a given mapping unit. In the first strategy, the estimator was derived from the representative profile description of that mapping unit. In the other strategies, the estimator was derived from the statistical sample.

Although the mapping units of the revised map were more homogeneous for some characteristics, the point estimates using the values of the representative profile descriptions, were no more accurate. This was due to the bias of this estimator, which rules out the reduction in spatial variance. If revision was followed by sampling, this bias could be eliminated. As a result, for some characteristics the point estimates became more accurate than those based on the original map.

Estimates of the spatial means of the study area via upgrading by two-phase sampling were more accurate, for all characteristics, than via revision plus upgrading. Using the estimates of the spatial variance within the mapping units of the original map for allocation was apparently more effective than reducing the variance itself.

Introduction

In the 1950s the former Soil Survey Institute of the Netherlands (now part of the Winand Staring Centre) began to map the soil systematically, at a scale of 1:50 000. During 1990, the last hectares were surveyed, and it is expected that all map sheets will be published by 1992. However, some soil characteristics have changed since the survey began, for example the groundwater regime as a result of artificial drainage, and the thickness of peat layers as a result of oxidation. Moreover, recent mapping is more detailed, i.e. there are more delineations. This is a result of surveyors' experience and a changing view on the permitted complexity of the map. Also, mapping was predominantly concerned with delineating soil classes defined in terms of many soil properties relevant to agricultural landuse, whereas nowadays users are also interested in properties related to soil conservation.

Consequently soil scientists in the Netherlands, and also in the USA, began to discuss alternative strategies for updating soil survey information (Marsman & De Gruijter, 1983; Brown, 1985; Arnold, 1988; Bouma, 1988; Wilding, 1988). Some scientists suggest that parts of the soil map assumed by them to be out of date, should be revised, while others regard the benefits of such revisions for practical applications as often not worth the investment. They prefer to leave the map unchanged and to quantify the variation within the units of the existing soil map by statistical sampling. Adding statistical information to an existing soil map in order to enhance its usefulness, will be referred to as *upgrading.*

In this study, four strategies for updating soil maps are compared with regard to costs and merits for spatial estimation of several soil properties, i.e. revision, upgrading, revision plus upgrading, and upgrading by two-phase sampling.

Only one of these strategies was actually carried out in the present case study viz. revision plus upgrading. However, as explained later (in the section **Data collection and analysis**), the results from this strategy have been used to calculate the merits of the other three strategies.

This paper describes the strategies, how to measure the efficiency of these strategies, the case study, and data collection and analysis. Finally, the results are presented and discussed, and some conclusions drawn.

Strategies for updating

As mentioned in the **Introduction**, we compared four strategies for updating. This section describes these strategies.

Revision

The original soil map is only revised. This can be done to a varying degree: we can take all the differentiating characteristics into account (multi-purpose revision) or only those relevant to a specific application (single purpose revision). In doing so, we might change the classification of delineations, change their boundaries, or both. By revising a soil map we are attempting to increase its cartographic purity and the homogeneity of its mapping units. The revision carried out in this case study is described below (see **Case study**).

Upgrading

This strategy leaves the map unchanged, and therefore its purity and homogeneity. A probability sample is taken from the mapping units to give statistical estimates of the means and variances of soil characteristics within these mapping units. Revision does not give such estimates. In this case a stratified simple random sample is taken, using the units of the original map as strata and allocating the sample points proportional to the size of the units.

Revision plus upgrading

Revision is followed by stratified sampling, now using the mapping units of the revised map as strata. This results in statistical estimates of the means and variances of soil characteristics within the mapping units of the revised map.

Upgrading by two-phase sampling

We take a stratified simple random sample from the mapping units of the original map in two phases. In the first phase, a limited number of sample points is allocated proportionally to the mapping units. This preliminary sample is used to estimate the variances within the mapping units. These estimates are then used in the second phase to allocate the remaining sample points in such a way that the total sample is distributed optimally over the strata. By allocating the points proportional to the size and the internal variance of the strata, the sampling variance of the global mean is minimized, assuming equal costs per point (Cochran, 1977, p. 98).

Efficiency

The efficiency of a strategy is determined by its merits and its costs.

Merits

The merits of updating were evaluated in terms of the improvement of spatial estimates. We did so both for estimating values at unvisited points and for spatial means. It is obvious that the accuracy of these estimates depends amongst others on the method to estimate the values at points and means.

Estimation of point values

The method of estimation adopted was to assign an estimate of the mean value of a given mapping unit to any unvisited point in that unit. This procedure worked fairly well for some properties. (Van Kuilenburg *et al.*, 1982). We propose the spatial mean of squared errors, $m_A(\varepsilon^2)$, as the measure of accuracy for a deterministic estimator function $\hat{z}(\boldsymbol{x})$:

$$m_A(\varepsilon^2) \equiv \frac{1}{N}\sum_{i=1}^{N}\{\hat{z}(\boldsymbol{x}_i)-z(\boldsymbol{x}_i)\}^2 \qquad (5.1)$$

where:

N = total number of soil profiles (possible sampling locations) in A (population size);

$\hat{z}(\boldsymbol{x}_i)$ = estimated value of the property at point $\boldsymbol{x}_i$;

$z(\boldsymbol{x}_i)$ = true value of the property at point $\boldsymbol{x}_i$;

A = area to which $\hat{z}(\boldsymbol{x})$ is applied and where the accuracy is to be evaluated.

In our specific case, since $\hat{z}(\boldsymbol{x})$ is constant within a mapping unit u, with value $\hat{m}_u$, the $m_u(\varepsilon^2)$ in u reduces to:

$$m_u(\varepsilon^2) = (\hat{m}_u - m_u)^2 + \frac{1}{N_u}\sum_{i=1}^{N_u}\{z(\boldsymbol{x}_i)-m_u\}^2 \qquad (5.2)$$

where:

m_u = true spatial mean of z in u;

N_u = total number of soil profiles (possible sampling locations) in u.

The last term in Equation (5.2) is defined as the spatial variance of z in u, which we denote by v_u, so:

$$m_u(\varepsilon^2) = (\hat{m}_u - m_u)^2 + v_u \qquad (5.3)$$

For a given mapping 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:

$$m_u\{E_p(\underline{\varepsilon}^2)\} = E_p\{(\underline{\hat{m}}_u - m_u)^2\} + v_u \quad (5.4)$$

The first term on the right hand side of Equation (5.4) represents the Mean Squared Error (*MSE*) of estimating m_u under design p, and can be broken down into the squared bias and the variance:

$$MSE(\underline{\hat{m}}_u) \equiv E_p\{(\underline{\hat{m}}_u - m_u)^2\} = \{E_p(\underline{\hat{m}}_u - m_u)\}^2 + E_p[\{\underline{\hat{m}}_u - E_p(\underline{\hat{m}}_u)\}^2] \quad (5.5)$$

Denoting the sampling variance under p by σ_p^2 and substituting Equation (5.5) in Equation (5.4), we obtain:

$$m_u\{E_p(\underline{\varepsilon}^2)\} = \{E_p(\underline{\hat{m}}_u) - m_u\}^2 + \sigma_p^2(\underline{\hat{m}}_u) + v_u \quad (5.6)$$

With unbiased estimators the first term cancels and:

$$m_u\{E_p(\underline{\varepsilon}^2)\} = \sigma_p^2(\underline{\hat{m}}_u) + v_u \quad (5.7)$$

For revision, we have no statistical estimates of the spatial means. We have used the representative profiles descriptions and prior information about aluminium and iron content of soil horizons to estimate the spatial means. These estimates might be biased, in which case the first term of Equation (5.6) is not equal to zero. On the other hand, since the estimator is a fixed value, the second term of Equation (5.6) can be dropped.

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

$$m_A\{E_p(\underline{\varepsilon}^2)\} = \sum_{u=1}^{U} W_u \, m_u\{E_p(\underline{\varepsilon}^2)\} \quad (5.8)$$

where:
U = number of mapping units.

Estimation of spatial means

The means of the soil properties are estimated by a stratified simple random sample using the mapping units of the original map (for upgrading) or those of the revised map (for revision plus upgrading) as strata. The means over the entire area are estimated as averages of the sample means in the strata, weighted with the proportions of the total area in the mapping units. The sampling variance of these means is a measure of accuracy of these estimates. For upgrading and revision plus upgrading, sample profiles are allocated proportionally to the strata. The sampling variance of the mean (Cochran, 1977, p. 93) can be estimated by:

$$\hat{\sigma}^2_{STSI(pr)}(\underline{\hat{m}}_A) = \frac{1}{n} \sum_{h=1}^{H} W_h \, \underline{\hat{v}}_h \qquad (5.9)$$

where:

$\hat{\sigma}^2_{STSI(pr)}$ = estimated sampling variance under proportional allocation;
H = number of strata;
n = total sample size;
W_h = weight of stratum h measured as the relative area of stratum h;
$\underline{\hat{v}}_h$ = estimated spatial variance within stratum h.

Upgrading with two-phase sampling implies optimum allocation to the strata. In that case, the sampling variance of the mean (Cochran, 1977, p. 99) can be estimated by:

$$\hat{\sigma}^2_{STSI(op)}(\underline{\hat{m}}_A) = \frac{1}{n} [\sum_{h=1}^{H} W_h \sqrt{\underline{\hat{v}}_h}]^2 \qquad (5.10)$$

where:

$\hat{\sigma}^2_{STSI(op)}$ = estimated sampling variance under optimum allocation.

As with estimation at points, for revision we used the representative profile descriptions of the mapping units to calculate the spatial means of the study area. The accuracy of this biased estimate is calculated in terms of the ab-

solute value of the difference between this estimate and an unbiased estimate of the spatial mean.

Costs

The costs of revision depend on its type and thoroughness. If field work is carried out, the main factor is the density of the augerings. The costs of upgrading depend on sample size and on the costs of measuring the soil characteristics. We assumed that the total costs were equal to the measurement costs per sample profile, times the sample size. For revision plus upgrading, we simply added the costs of revision and of the statistical sample. The costs of upgrading in two phases were calculated in the same way as for upgrading. The costs of revision and upgrading are summarized in Table 5.1.

Table 5.1 The costs of revision and upgrading. Components common to revision and upgrading (e.g. measurement of water table to calculate systematic measurement error of the mean highest water table, $\bar{W}_{min}$) were omitted.

	Days	Cost (Dfl)
Revision		
1 Re-survey of 5020 ha	72	50 400,00
2 Describing profiles	10	7 000,00
Upgrading		
1 Sampling		
a Phosphate sorption char.	6 p.p.d.	116,65 p.p.
b $\bar{W}_{min}$	17 p.p.d.	41,20 p.p.
2 Laboratory analysis	-	167,50 p.s.

p.p.d. = points per day; p.p. = per point; p.s. = per sample

Case study

The use of the soil map and the soil conditions in the study area are important factors governing the merits and costs of the four strategies. This section also describes the way in which we revised the original soil map and the result of this revision.

Use of the soil map and definition of phosphate sorption characteristics

We chose the problem of phosphate leaching from agricultural soils as an example of soil map use. Leaching of phosphate to groundwater and surface water is an environmental problem in large parts of The Netherlands. In areas with intensive animal husbandry, the application rates of manure exceed the phosphorus crop uptake (Van der Zee, 1988), which may lead to saturation of the soil with phosphates (Breeuwsma & Schoumans, 1987). Policy makers need to know how susceptible the soils are to phosphate leaching, the degree of phosphate saturation and where phosphate-saturated soils occur.

The degree of phosphate saturation of a soil profile is determined by the amount of manure applied and by its capacity to adsorb the phosphates. In non-calcareous sandy soils, a strong correlation exists between the maximum mass of P_2O_5 sorbed by soil (P_{max}) and oxalate-extractable aluminium and iron content (Schoumans *et al.*, 1987; Van der Zee *et al.*, 1987, 1988; Van der Zee & Van Riemsdijk, 1988), which, in turn, are correlated with soil type and soil horizon (Breeuwsma *et al.*, 1986, Breeuwsma & Schoumans, 1987; Schoumans *et al.*, 1989). The P_{max} of a column of soil, given its diameter, increases as the depth to which it is calculated, is increased. As we were interested in the leaching of phosphates to the groundwater, it seemed rational to take the depth to the water table as a reference depth. This depth is also indicated on soil maps. Soil maps may therefore be expected to be suitable when estimating P_{max}.

In contrast to P_{max} which is a function of soil type only, other phosphate sorption characteristics such as the actual mass of phosphate sorbed by soil are determined also, or primarily, by human activities such as manuring. Hence, the usefulness of soil maps as a tool for spatial estimation may be different for these characteristics.

The following phosphate sorption characteristics were considered: maximum areic mass of P_2O_5 sorbed by soil above the mean highest water table, P_{max} (kg m^{-2}), maximum volumetric mass of P_2O_5 sorbed by soil above the mean highest water table, P^v_{max} (kg m^{-3}), areic mass of P_2O_5 sorbed by soil above

the mean highest water table, P (kg m^{-2}), the relative mass of phosphate sorbed by soil above the mean highest water table, P_{rel} (-), and the areal fraction saturated with phosphate ($P_{rel} > c$), A_c (-).

The maximum areic mass of P_2O_5 sorbed by soil, P_{max}, was calculated by the following regression equation (Breeuwsma & Silva, 1992):

$$P_{max} = \sum_{l=1}^{L} \{4.6 + 0.39\ M_{ox,l}\}\ \delta_l\ \rho_l\ 0.71 \qquad (5.11)$$

where:

$M_{ox,l}$ = oxalate-extractable metal (aluminium and iron) content of horizon l (mol kg^{-1});

δ_l = thickness of horizon l as far as lying above mean highest water table, $\bar{W}_{min}$ (m) (For a definition of $\bar{W}_{min}$, see Van der Sluijs & De Gruijter, 1985);

ρ_l = volumic mass of horizon l (kg m^{-3});

L = number of horizons;

0.71 = constant to convert the dimension of P_{max} from mol P per m^2 to kg P_2O_5 per m^2 (kg mol^{-1}).

The maximum volumetric mass of P_2O_5 sorbed by soil, P^v_{max}, has been defined as:

$$P^v_{max} = \frac{P_{max}}{\bar{W}_{min}} \qquad (5.12)$$

The areic mass of P_2O_5 sorbed by soil, P, has been calculated by:

$$P = \sum_{l=1}^{L} P_l\ \delta_l\ \rho_l\ 0.71 \qquad (5.13)$$

where:

P_l = oxalate-extractable phosphate content of horizon l (mol kg^{-1}).

The relative mass of phosphate sorbed by soil, P_{rel}, has been defined as:

$$P_{rel} = \frac{P}{P_{max}} \qquad (5.14)$$

The areal fraction saturated with phosphate, A_c, has been defined as the fraction of the area in which P_{rel} of the soil profiles exceeded a given critical value. In this study we used 0.35 as a critical value (Breeuwsma *et al.*, 1989).

Study area

In 1988 the authorities of the National Waterboard and of the province of Gelderland asked the DLO-Winand Staring Centre (SC-DLO) to study the problem of phosphate leaching in an area of about 5020 ha in the centre of the Netherlands. It covers the catchment of the Schuitenbeek which debouches into a narrow lake between the reclaimed IJsselmeerpolders and the mainland. Large concentrations of nutrients in this lake, thought to come from bordering agricultural land, cause rapid growth of algae in summer.

The study area lies on Sheet 32-East of the Soil Map of the Netherlands at a scale of 1:50 000 (Stichting voor Bodemkartering, 1965). It was surveyed in 1959 and 1960. The resulting map was thought to be of little value because of alterations in the groundwater regime since 1960 and the occurrence of soil associations with a large variation in relevant soil characteristics. Paradoxically, these qualities made the map well suited to this case study.

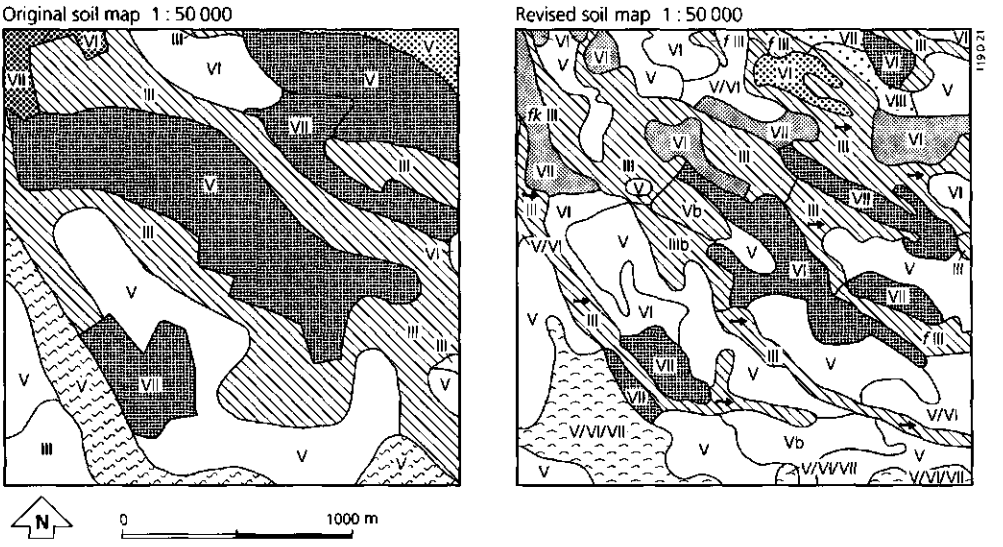
The landscape of the area is intricate comprising ridges of coversand and intervening valleys. Dominant soil types are Typic Haplaquods, Typic Humaquepts, Typic Udipsamments and Plaggepts (Soil Survey Staff, 1975). The dominant landuse on these soils is grassland for permanent pasture. Approximately 14% of the agricultural land is used for growing maize for silage. Twenty-nine percent of the study area is covered by heath and forest.

Revision of soil map of study area

Some 40 new observations per km^2 were made to revise the original map. Both the original and the revised soil maps were made from records ob-

tained by free survey (Steur, 1961), i.e. surveyors chose observation points and delineated soil boundaries in the field by observing landscape features such as geomorphology and vegetation.

The revision resulted in a completely new soil map. Figure 5.1 shows that the revised map is far more detailed than the original one: there are more mapping units (68 versus 22) and many more delineations.



Classification water table classes

Water table class no	III	IIIb	V	Vb	VI	VII	VIII
Mean highest water table (cm below soil surface)	<40	25-40	<40	25-40	40-80	>80	>140
Mean lowest water table (cm below soil surface)	80-120	80-120	>120	>120	>120	>120	>160

Fig. 5.1 Sample areas of the original (1965) and the revised (1990) soil map 1:50 000

Data collection and analysis

We estimated all $m_A\{E_p(\underline{\varepsilon}^2)\}$s and sampling variances of the means for the four strategies of updating by taking only one stratified simple random sample of size 283. This sample was used to estimate the spatial variances within the mapping units of the revised map and within those of the original map. These variances were then used to calculate the $\hat{m}_A\{E_p(\underline{\varepsilon}^2)\}$s (Eqs 5.6 -5.8) and sample variances of the mean (Eqs 5.9 and 5.10), to be expected from the other strategies. (In statistical terms, the mapping units of the original and revised map are combinations of "domains" within strata. Further details of the estimation of means and variances within domains from a stratified simple random sample are given in the **Appendix**).

We used a twofold stratification: geographical and pedological. A square grid with cells of four km^2 was superimposed over the area, and the mapping units of the revised soil map were grouped, by reducing the seven groundwater depth classes into two and grouping map units at the subgroup level (De Bakker and Schelling, 1989). Each group of mapping units within a cell was treated as a stratum for sampling. To each stratum we allocated first two points. The remaining sample points were allocated proportionally to the size of the strata. The points were distributed fairly evenly, but the minimum rule (invoked to guarantee unbiased estimates of the sampling variances) led to substantial variation in density within the strata.

Since phosphate leaching occurs primarily in the agricultural part of the area, we also estimated the variances of the phosphate sorption characteristics within the agricultural parts of the mapping units. To estimate the spatial variance within mapping units, or within the agricultural part of mapping units, their areas within each stratum must be known. These values were obtained from an overlay of the strata-map, the two soil maps and a landuse map derived from the topographical map 1:25 000. The overlay was obtained from ARC/INFO (Environmental Systems Research Institute, 1987).

Results

Phosphate leaches only from agricultural soils, and so we present here the results of the agricultural part of the area. The results of the total area (agricultural land + heath and forest) were similar.

The efficiency of the revision strategy was compared with that of upgrading and the efficiency of revision plus upgrading was compared with that of upgrading in two phases. The first two strategies are the alternatives if only relatively limited funds are available. The last two strategies can be used if more funds are available (Table 5.1). We compared the accuracies of estimates at equal costs (the alternative is to compare costs at equal accuracies). For the first and second strategies this was done for one value, viz. the costs of revision as described above (Table 5.1: Dfl 57,400). This corresponds with a sample size of 202. For the mean highest water table, $\bar{W}_{min}$, the equivalent sample size was 1394 profiles because in this case there are no costs for laboratory analysis.

We compared revision plus upgrading and upgrading in two phases at several cost levels, ranging from the costs of revision plus 50 sample profiles to the costs of revision plus 500 sample profiles.

Revision vs. upgrading

Revision alone does not give estimates of the point values and spatial means of P, P_{rel} and A_c. These characteristics are not only determined by natural conditions, but especially by human activities, so it is pointless to use the descriptions of the representative profiles as an estimator for these characteristics. As such, this a disadvantage of revision.

Table 5.2 presents the results of the estimation at points. It shows that the root of $\underline{\hat{m}}_A\{E_p(\underline{\varepsilon}^2)\}$ for P_{max} and P^v_{max} after upgrading would be almost equal to those after revision. Upgrading would lead to more accurate estimates of $\bar{W}_{min}$ than revision.

Table 5.2 Root of estimated spatial mean of *p*-expected, squared errors of estimates at points of phosphate sorption characteristics and mean highest water table ($\bar{W}_{min}$) for revision and for upgrading (one phase). - = no estimate

	Upgrading (one phase)	Revision
P_{max} (kg m^{-2})	2.105	2.113
P^{v}_{max} (kg m^{-3})	1.865	1.853
P (kg m^{-2})	0.610	-
P_{rel} (%)	20.9	-
A_c (%)	48.2	-
$\bar{W}_{min}$ (cm)	45.2	49.7

Table 5.3 Standard error of mean and error of estimated mean of phosphate sorption characteristics and mean highest water table ($\bar{W}_{min}$) for revision and for upgrading (one phase). - = no estimate

	Upgrading (one phase) (standard error of mean	Revision (error of mean)
P_{max} (kg m^{-2})	0.150	0.152
P^{v}_{max} (kg m^{-3})	0.127	0.164
P (kg m^{-2})	0.042	-
P_{rel} (%)	1.6	-
A_c (%)	3.3	-
$\bar{W}_{min}$ (cm)	0.8	1.8

Table 5.3, presenting the results of the estimation of spatial means, shows that estimates of the spatial means of P^{v}_{max} and $\bar{W}_{min}$ after upgrading would be more accurate than those after revision. For P_{max} these accuracies would be almost equal.

Revision plus upgrading vs. upgrading with two phases of sampling

Figure 5.2, which presents the results of the estimation at points, shows that revision plus upgrading would give more accurate estimates of $\bar{W}_{min}$ than upgrading with two phases of sampling. If more than 60 to 75 profiles are

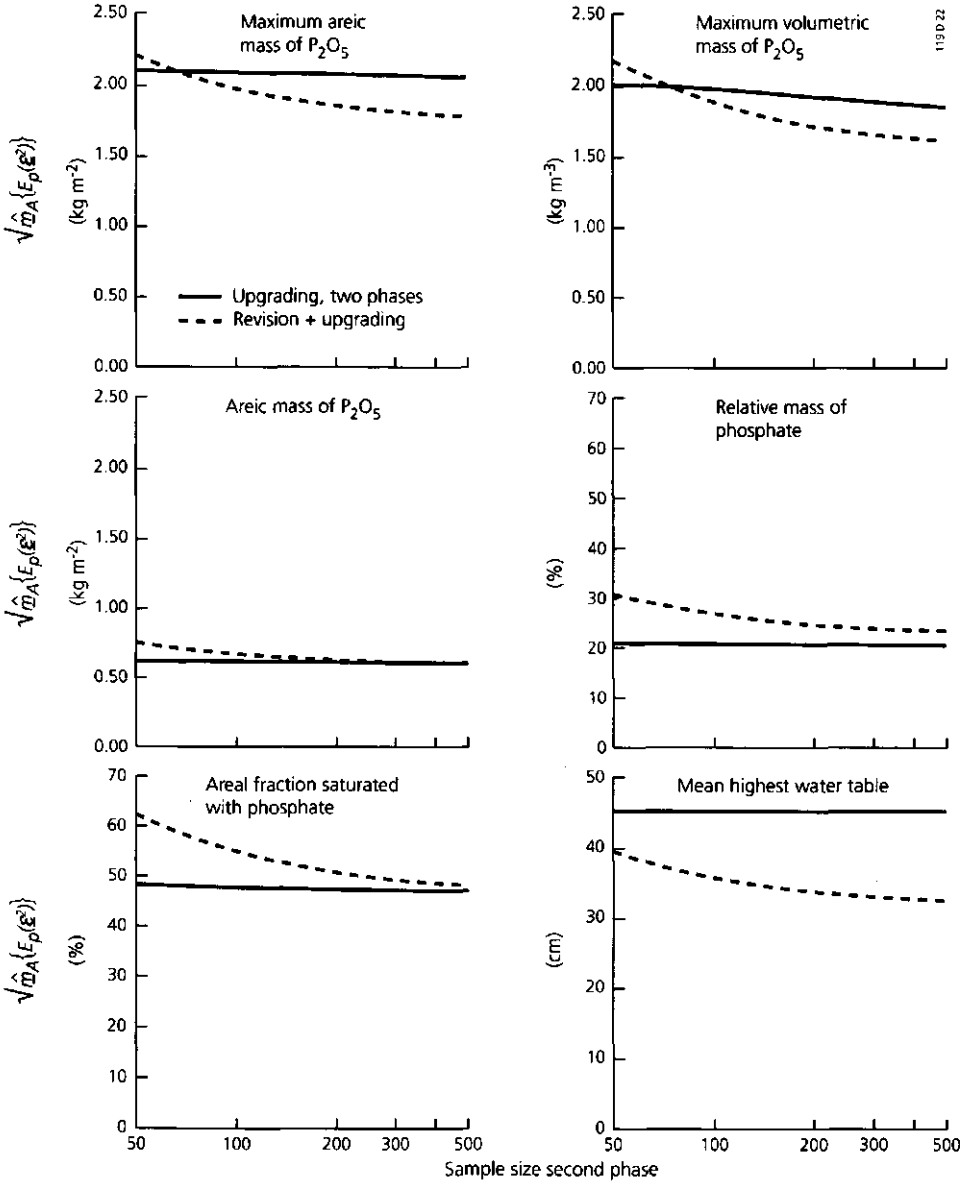


Fig. 5.2 Root of estimated spatial mean of *p*-expected, squared errors of estimates at points of phosphate sorption characteristics and mean highest water table ($\bar{W}_{min}$) for revision plus upgrading and for upgrading with two phases of sampling. Sample size is for second phase of upgrading-two phases, which equals total sample size for revision + upgrading

sampled, which means about one profile per mapping unit, those of P_{max} and P^{v}_{max} would also be more accurate. The estimates of P, P_{rel} and A_c would be less accurate if a small sample is taken after revision, but with larger samples the accuracy would be almost equal to that after upgrading in two phases.

Figure 5.3 shows the results for the estimation of spatial means. It shows that for all characteristics, upgrading with two-phase sampling would give more accurate estimates of the spatial means than revision plus upgrading.

Discussion

The results described above can be explained by the homogeneity of soil properties within the mapping units of the original map and within those of the revised map. From Equations (5.6), (5.8) and (5.9) we can see that the accuracy of the estimates can be increased by reducing the spatial variance. Only if this spatial variance is decreased (the homogeneity is increased), is revision potentially better than upgrading, and revision plus upgrading better than upgrading by two-phase sampling. Figure 5.4 shows the pooled spatial standard deviations within mapping units, calculated as the square root of the pooled spatial variances. The pooled standard deviation of $\bar{W}_{min}$ was markedly reduced, those of P_{max} and P^{v}_{max} were also reduced, but those of P, P_{rel} and A_c were not.

Estimates of point values

Despite greater homogeneity of the mapping units of the revised map with respect to $\bar{W}_{min}$, P_{max} and P^{v}_{max}, the accuracy of estimates at points using the values of the representative profiles (as done in revision), was less for $\bar{W}_{min}$ and almost equal for P_{max} and P^{v}_{max} (Table 5.2). This is due to the bias of this type of estimator, which overrides the reduction of the spatial variance (cf. Eq. 5.6). If, instead, we take the sample means as estimators, as done in revision plus upgrading, we eliminate the bias. As a result the estimates based on the revised map were more accurate than those based on the

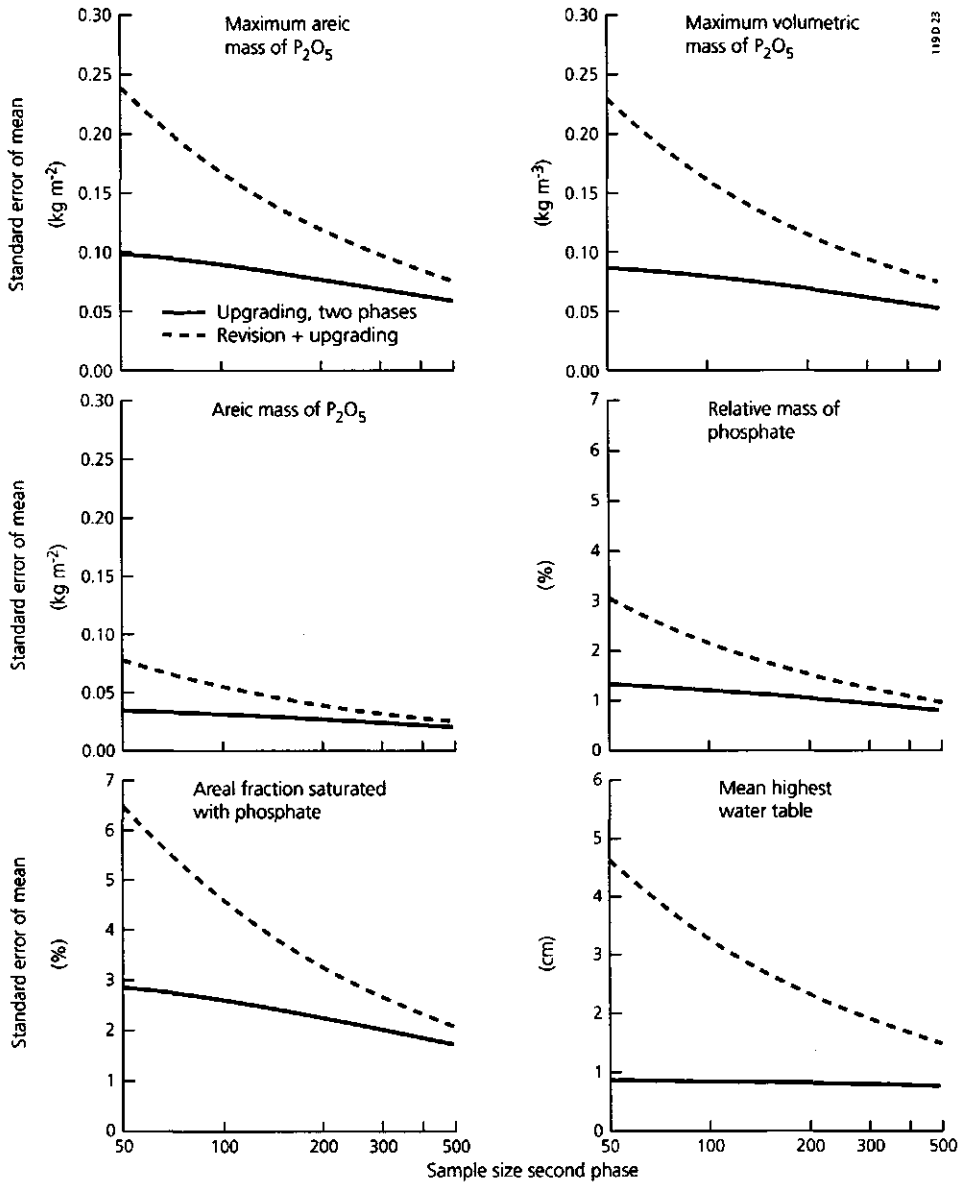


Fig. 5.3 Standard error of mean of phosphate sorption characteristics and mean highest water table ($\bar{W}_{min}$) for revision plus upgrading and for upgrading with two phases of sampling. Sample size is for second phase of upgrading-two phases, which equals total sample size for revision + upgrading

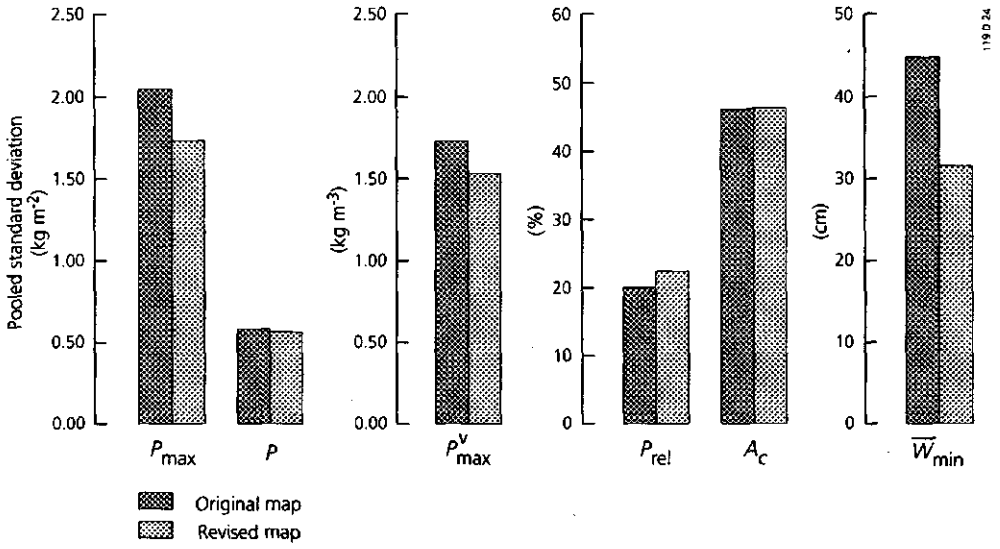


Fig. 5.4 Pooled standard deviation of phosphate sorption characteristics and mean highest water table ($\bar{W}_{min}$) within mapping units of the original and the revised map

original map (Fig. 5.2). Those of P, P_{rel} and A_c (Fig. 5.2) were not more accurate because the spatial variance was not reduced (Fig. 5.4).

Estimation of spatial means

Upgrading by two-phase sampling was better than revision plus upgrading for all characteristics (Fig. 5.3). We feel that this also holds for many other circumstances (other soil properties, other areas, other maps). Apparently, quantifying the spatial variance of map units is more effective than trying to reduce this variance by remapping.

Increasing homogeneity

In contrast to the phosphate sorption characteristics, $\bar{W}_{min}$ is a differentiating characteristic, so increasing the homogeneity of this property was a direct aim of revision. Generally, differences in the rate of reduction of the pooled standard deviations of soil properties can be explained by the type and strength of the correlation between these properties and the properties used

to differentiate the mapping units. If the relation is linear and the correlation coefficient is large, reduction in the spatial variance within mapping units of the differentiating characteristic would lead to a similar reduction in spatial variance within mapping units of the properties to be estimated.

With a linear relation, and only a weak correlation, spatial variance would not be reduced very much. If the relation is not linear, reduction depends on the type of relation. To see what the relationships are in this case, we plotted scatter diagrams of the phosphate sorption characteristics versus $\bar{W}_{min}$ and calculated correlation coefficients (Fig. 5.5). This figure shows that P_{max} and $\bar{W}_{min}$ were strongly correlated ($r = 0.854$). Revision resulted in a marked reduction of the spatial variance within mapping units of $\bar{W}_{min}$, and so that of P_{max} was also reduced. The relation of P and $\bar{W}_{min}$ was linear, but the correlation was only moderate ($r = 0.414$). Although P was also defined by $\bar{W}_{min}$, the correlation was less strong because P was mainly determined by the amount of manure applied to the soil. This is why the pooled standard deviation of P was not reduced. Figure 5.5 shows that the relation between P_{rel} and $\bar{W}_{min}$ was not linear but more or less rational, i.e. the quotient of two linear functions. The calculated correlation coefficient was only -0.227. With such a relation, decreasing the variance of $\bar{W}_{min}$ within mapping units does not automatically reduce the variance of P_{rel}. Since the pooled standard deviation of P_{rel} was not reduced, neither was that of A_c. Although the correlation between P^v_{max} and $\bar{W}_{min}$ was very small ($r = -0.045$), the pooled standard deviation of P^v_{max} was still reduced. This can probably be explained by the relation between P^v_{max} and the classification unit. At subgroup level the variance ratio is 8.4. Purity with respect to subgroup, was increased from 50% to 63% by revision.

Conclusions

If no funds are available for revision plus upgrading but only for one of them, we recommend upgrading in circumstances comparable to the present case study. Although revision leads to a decrease in variance within mapping units of differentiating characteristics and characteristics strongly correlated

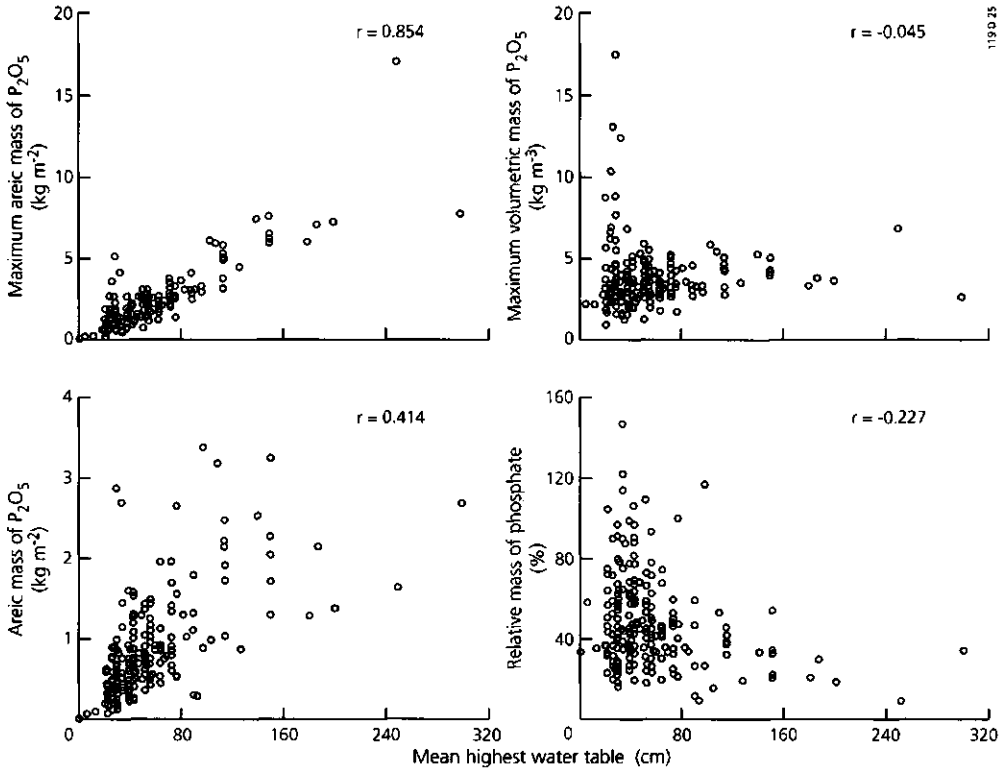


Fig. 5.5 Scatter diagrams of the phosphate sorption characteristics versus the mean highest water table ($\bar{W}_{min}$) and correlation coefficients

with them, estimates at unvisited points, using the values of the representative profiles of the revised mapping units as estimators, are less accurate than those using the estimated means of the original mapping units as estimators. Also, revision gives biased estimates of the means of the study area, and gives no estimates at all of characteristics which need to be measured by laboratory anlysis of samples taken from the study area itself, such as the areic mass of P_2O_5 sorbed by soil.

If funds for revision plus upgrading are available, the alternative is to take a large sample, split it up into two phases, and use the results of the first phase to allocate the sample profiles optimally in the second phase. In doing

this, estimates of the areal means of all characteristics are more accurate than for revision plus upgrading. For the estimates at points, revision plus upgrading is better than upgrading with two phases of sampling for characteristics with increased homogeneity within mapping units.

In 1988, the former Soil Survey Institute of the Netherlands began a nationwide sampling project, in which the Soil Map of the Netherlands at a scale 1:50 000 is used for stratification. This sample gives estimates of the variance within mapping units of several soil properties and characteristics, which can also be used to allocate sample profiles optimally in subsequent samples. This nationwide sampling then takes the place of the first phase of the fourth strategy (upgrading with two-phase sampling). As the relative variances within mapping units in the study area may differ from those in The Netherlands, the resulting allocation may not be optimal.

Acknowledgements

We thank G. Ebbers, H. Van het Loo and P. Chardon for revising the original soil map, J.G.A. Reijerink for processing the results of the chemical analysis, H. De Wijer for his expert work on automated cartography and J. Bouma and R. Webster for helpful comments on an earlier draft.

References

Arnold, R.W. 1988. Classification and correlation for current and future map production. In *Soil resources: their inventory, analysis and interpretation for use in the 1990's.* Proceedings of an international interactive workshop, pp. 54-58. University of Minnesota, Minneapolis.

Bouma, J. 1988. When the mapping is over, then what? In *Soil resources: their inventory, analysis and interpretation for use in the 1990's.* Proceedings of an international interactive workshop, pp. 3-12. University of Minnesota, Minneapolis.

Breeuwsma, A., Wösten, J.H.M., Vleeshouwer, J.J., Van Slobbe, A.M. & Bouma, J. 1986. Derivation of land qualities to assess environmental problems from soil surveys. *Soil Science Society of America Journal* 50, 186-190.

Breeuwsma, A., Reijerink, J.G.A., Schoumans, O.F., Brus, D.J., & Van het Loo, H. 1989. *Fosfaatbelasting van bodem, grond- en oppervlaktewater in het stroomgebied van de Schuitenbeek.* Report 10, DLO The Winand Staring Centre for Integrated Land, Soil and Water Research (SC-DLO), Wageningen.

Breeuwsma, A. & Schoumans, O.F. 1987. Forcasting phosphate leaching from soils on a regional scale. In *Vulnerability of soil and groundwater to pollutants.* Proceedings and Information 38. (eds W.A. Duijvenboden & H.G. van Waegeningh), pp. 973-981. National Institute of Public Health and Environmental Hygiene, The Hague.

Breeuwsma, A. & Silva, S. 1992. *Phosphorus fertilisation and environmental effects in the Netherlands and Italy.* Report 57, DLO The Winand Staring Centre for Integrated Land, Soil and Water Research (SC-DLO), Wageningen.

Brown, R.B. 1985. The need for continuing update of soil surveys. In *Soil and Crop Science Society of Florida.* Proceedings 44. pp. 90-93.

Cochran, W.G. 1977. *Sampling Techniques.* 3rd edn. Wiley, New York.

De Bakker, H. & Schelling, J. 1989. *Systeem van bodemclassificatie voor Nederland.* De hogere niveaus. Tweede, gewijzigde druk. Bewerkt door D.J. Brus en C. Van Wallenburg. DLO The Winand Staring Centre for Integrated Land, Soil and Water Research (SC-DLO), Wageningen.

Environmental Systems Research Institute, INC. 1987. ARC/INFO Users Guide. Redlands, CA.

Marsman, B.A. & De Gruijter, J.J. 1983. Dutch soil survey goes into quality control. In *Soil Information Systems Technology.* Proceedings of the sixth meeting of the ISSS working group on soil information systems. (eds P.A. Burrough & S.W. Bie), pp. 127-134. PUDOC, Wageningen.

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.

Schoumans, O.F., Breeuwsma, A. & De Vries, W. 1987. Use of soil survey information for assessing the phosphate sorption capacity of heavily manured soils. In *Vulnerability of soil and groundwater to pollutants.* Proceedings and Information 38. (eds W.A. Duijvenboden & H.G. van Waegeningh), pp. 1079-1088. National Institute of Public Health and Environmental Hygiene, The Hague.

Schoumans, O.F., Marsman, B.A. & Breeuwsma, A. 1989. Assessment of representative soil data for phosphate leaching. In *Land qualities in space and time.* (eds J. Bouma & A.K. Bregt), pp. 201-204. PUDOC, Wageningen.

Soil Survey Staff. 1975. *Soil Taxonomy.* A basic system of soil classification for making and interpreting soil surveys. Agriculture Handbook 436. Soil Conservation Service. US Department of Agriculture, Washington DC.

Steur, G.G.L. 1961. Methods of soil surveying in use at the Netherlands Soil Survey Institute. *Boor en spade* XI, 59-77.

Stichting voor Bodemkartering. 1965. Bodemkaart van Nederland, Schaal 1:50 000. Toelichting bij kaartblad 32 Oost (Amersfoort). Wageningen.

Van der Sluijs, P. & De Gruijter, J.J. 1985. Water table classes: a method to describe seasonal fluctuation and duration of water tables on Dutch soil maps. *Agricultural Water Management* 10, 109-125.

Van der Zee, S.E.A.T.M. 1988. *Transport of reactive contaminants in heterogeneous soil systems.* Doctoral thesis, Agricultural University, Wageningen.

Van der Zee, S.E.A.T.M., Fokkink, L.G.J. & Van Riemsdijk, W.H. 1987. A new technique for assesment of reversily adsorbed phosphate. *Soil Science Society of America Journal* 51, 599-604.

Van der Zee, S.E.A.T.M., Nederlof, M.M., Van Riemsdijk, W.H. & De Haan, F.A.M. 1988. Spatial variability of phosphate adsorption parameters. *Journal of Environmental Quality* 17, 682-688.

Van der Zee, S.E.A.T.M. & Van Riemsdijk, W.H. 1988. Model for long-term phosphate reaction kinetics in soil. *Journal of Environmental Quality* 17, 35-41.

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.

Wilding, L.P. 1988. Improving our understanding of the composition of the soil-landscape. In *Soil resources: their inventory, analysis and interpretation for use in the 1990's.* Proceedings of an international interactive workshop, pp. 13-35. University of Minnesota, Minneapolis.

Appendix

The mapping units of the original map and those of the revised map were not used as strata themselves. Therefore we needed to estimate spatial variances within domains (subpopulations) cutting across the strata (Cochran, 1977 pp. 142-143). In the present case, the problem is simple because the size of the subpopulations within the strata is known.

We estimated spatial variances within domains (mapping units) using the procedure followed by Marsman and De Gruijter (1986):

$$\underline{\hat{v}}_u(z) = \underline{\hat{m}}_u(z^2) - \{\underline{\hat{m}}_u(z)\}^2 + \hat{\sigma}_p^2\{\underline{\hat{m}}_u(z)\} \qquad (5.A1)$$

where:

$\underline{\hat{v}}_u(z)$ = estimated spatial variance of characteristic z within mapping unit u;

$\underline{\hat{m}}_u(z^2)$ = estimated spatial mean of z^2 in u;

$\underline{\hat{m}}_u(z)$ = estimated spatial mean of z in u;

$\hat{\sigma}_p^2\{\underline{\hat{m}}_u(z)\}$ = estimated sampling variance of the estimator $\underline{\hat{m}}_u$ under sampling design p.

The means of mapping units, cutting across the strata, were estimated according to:

$$\underline{\hat{m}}_u = \sum_{h=1}^{H} W_{hu}\, \underline{\hat{m}}_{hu} \qquad (5.A2)$$

where:

H = number of strata;

W_{hu} = weight of stratum h, measured as the proportion of the total area in mapping unit u in stratum h;

$\underline{\hat{m}}_{hu}$ = estimated spatial mean of z in the part of the area covered by both stratum h and mapping unit u, calculated as the unweighted mean of the values at the sample points in that part.

The mean of z^2 in u was estimated in the same way as the mean of z (Eq. 5.16), except that the values were first squared.

The sampling variance of the mean was estimated by:

$$\hat{\sigma}_p^2(\underline{\hat{m}}_u) = \sum_{h=1}^{H} W_{hu}^2\, \frac{\underline{\hat{v}}_{hu}}{n_{hu}} \qquad (5.A3)$$

where:

$\underline{\hat{v}}_{hu}$ = estimated spatial variance in the intersection of h and u, estimated as the variance among sample points in that intersection;

n_{hu} = number of sample points lying in both h and u.

Measurement error

Estimates of the variances of the phosphate sorption characteristics would be biased by measurement error if calculated directly from the measurements. An important source of error of the measurements of the phosphate sorption characteristics is the measurement of the mean highest water table ($\bar{W}_{min}$). $\bar{W}_{min}$ was estimated mainly from visible profile characteristics. For profiles with an estimated $\bar{W}_{min} < 1.20$ m these estimates were corrected for systematic measurement error by simultaneously measuring water tables at the observation points and at reference points with a known $\bar{W}_{min}$. We used these corrected values of $\bar{W}_{min}$ to calculate the P_{max}, P^{v}_{max}, P and P_{rel} for each sample point with Equations (5.1) to (5.4). To calculate A_c, we introduced a dummy variable: $z_i = 1$ if P_{rel} of profile i > 35%; $z_i = 0$ otherwise.

Estimates of the variances calculated from the values corrected for systematic measurement error, would still be biased by random measurement error. This bias equals the variance of the measurement error plus twice the covariance between measurement error and measured variable. Regressing the estimated $\bar{W}_{min}$ of 32 sampling points on the real $\bar{W}_{min}$ showed that the variance due to measurement error is 290 cm^2, i.e. a standard error of 17 cm. This agrees closely with the standard error of 16.3 cm found by Marsman & De Gruijter (1986). The variance of measurement error of P_{max} was calculated by regressing the P_{max} values calculated from the estimated $\bar{W}_{min}$ on those calculated from the real $\bar{W}_{min}$. The same procedure was followed for P^{v}_{max}, P and P_{rel}.

To compare the usefulness of the original and revised soil map for estimating point values, the variances within mapping units were pooled for each map. The pooled variance within mapping units was calculated by subtracting the estimated bias due to random measurent error from the pooled variance as calculated from the measurements.

To estimate the variance within a given mapping unit, strata with only one sample point in that unit were clustered. Variances within units with less than two sample points in total, could not be estimated. Such units cover 1.4% of the original map, and 1% of the revised one.

Chapter 6

Estimation of non-ergodic variograms and their sampling variance by design-based sampling strategies

Mathematical Geology, accepted

with Jaap J. De Gruijter

Estimation of non-ergodic variograms and their sampling variance by design-based sampling strategies

Design-based sampling strategies based on classical sampling theory offer unprecedented potentials for estimation of non-ergodic variograms. Unbiased and uncorrelated estimates of the semivariance at the selected lags and of its sampling variance can simply be obtained. These estimates are robust against deviations from an assumed spatial autocorrelation model. The same holds for the variogram model parameters and their sampling (co)variances. Moreover, an objective measure for lack of fit of the fitted model can simply be derived. The estimators for two basic sampling designs, simple random sampling and stratified simple random sampling of pairs of points, are presented. The first has been tested in real world for estimating the non-ergodic variograms of three soil properties. The parameters of variogram models and their sampling (co)variances were estimated with 72 pairs of points distributed over 6 lags.

KEY WORDS: statistical dependence, design-unbiasedness, lack of fit, confidence interval, dispersion matrix

Introduction

The variogram plays a central role in geostatistical prediction techniques such as kriging. It describes the spatial correlation within a region. Especially the variance of the prediction error is sensitive to the variogram. Therefore, it is important to develop efficient sampling strategies for estimating variograms, i.e. combinations of sampling designs and estimators that are as accurate as possible, given the budget for sampling and measurement. Besides maximum accuracy, it is of equal importance to get an estimate of this accuracy. In this context Shafer and Varljen (1990, p. 1787) stated: "Unfortunately, because of the presence of correlation, classical statistical theory cannot be applied to make inferences regarding the confidence limits of the variogram estimates". This view seems to be common among geostatisticians and prevented them from developing simple strategies giving direct estimates of these confidence

limits.

Recently De Gruijter and Ter Braak (1990) have shown that the assumption that classical sampling theory is inapplicable in spatial sampling is false. We will show that this assumption is unfortunate as well, because classical sampling theory offers unprecedented potentialities for variogram estimation. The aim of this paper is:

- to draw the attention of geostatisticians to a new set of methods, based on classical sampling theory, for estimating the variogram and its accuracy;
- to show that these new methods have important advantages over existing methods;
- to illustrate the approach by working out two of the possible methods and by applying one of these in a real-world case study.

We will first discuss non-ergodic variograms and their advantages over ergodic variograms, and we will briefly comment on existing methods to estimate variograms. Then we describe and motivate the new approach and illustrate it with examples. Finally we present the application of one of possible methods in a case study.

Sampling variance of non-ergodic variograms

Journel and Huijbregts (1978) made a distinction between theoretical variograms and local variograms. In short the difference between these two is that the first is defined as an average over all realizations of the underlying stochastic model and the second only over the realization actually sampled. The latter is not related to a stochastic model. The importance of this distinction was re-emphasized by Srivastava (1987) and Isaaks and Srivastava (1988). Journel and Huijbregts (1978) also stated that a distinction should be made between 'fluctuation variance' and estimation variance of the variogram. Estimation variance arises from repeated sampling of a single realization. However, even if one has a complete enumeration of an area, and as a result the estimation variance of the local variogram is zero, we still have fluctuation variance of the estimator of the theoretical variogram, that is variance between model-

realizations.

In this paper we will confine ourselves to estimation of the local variogram because we want to use it for inference of parameters of a single realization of the model. According to Isaaks and Srivastava (1988, p. 322) local variograms give more reliable confidence intervals of predictions than theoretical variograms because "it captures the character of spatial variability unique to the domain over which it is defined". Isaaks and Srivastava use the term 'non-ergodic variograms' as a synonym of 'local variograms'. We prefer the adjective 'non-ergodic' to 'local' because it describes its essence better. For non-ergodic variograms the only source of variation is estimation variance, which we will refer to as sampling variance, the usual term for it in classical sampling theory.

Following Journel (1985) and Isaaks and Srivastava (1988) we will use set theory to define non-ergodic variograms. Symbol A refers to a limited area (Fig. 6.1). It is viewed as the set of all possible sample locations. As the number of different possible sample locations is finite, we consider A as a finite population. N denotes the size of A. The subscripts $+\boldsymbol{h}$ and $-\boldsymbol{h}$ are used to denote

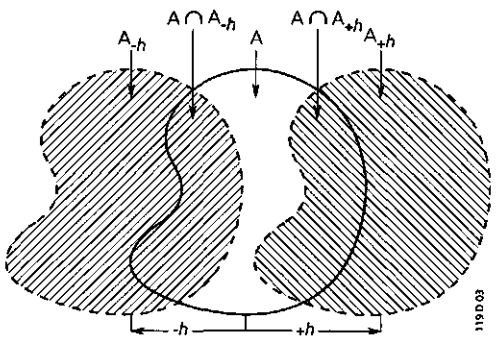


Fig. 6.1 Translation of the area A (After Isaaks and Srivastava, 1988). Points $\boldsymbol{x}$ in $A \cap A_{-\boldsymbol{h}}$ can be paired with points $\boldsymbol{x} + \boldsymbol{h}$ in $A \cap A_{+\boldsymbol{h}}$ and contribute to $\gamma_A(\boldsymbol{h})$

translations of A. Set A_{+h} is obtained by translating the set of locations within A by vector $\boldsymbol{h}$. The set of locations within the intersections of A and translated sets (A_{-h} and A_{+h}) is of special importance. A point $\boldsymbol{x}$ within $A \cap A_{-h}$ is an element of A ($\boldsymbol{x} \in A$), and can be paired with a point $\boldsymbol{x} + \boldsymbol{h}$, which is also an element of A ($\boldsymbol{x} + \boldsymbol{h} \in A$). Point $\boldsymbol{x} + \boldsymbol{h}$ is in the intersection $A \cap A_{+h}$. Now, the non-ergodic variogram, $\gamma_A(\boldsymbol{h})$, is defined as (Isaaks and Srivastava, 1988):

$$\gamma_A(\boldsymbol{h}) = \frac{1}{2N(\boldsymbol{h})} \sum_{A \cap A_{-h}} \{z(\boldsymbol{x}) - z(\boldsymbol{x}+\boldsymbol{h})\}^2 \qquad (6.1)$$

where:

$z(\boldsymbol{x})$ = value of the target variable at location $\boldsymbol{x}$;

$N(\boldsymbol{h})$ = size of $A \cap A_{-h}$.

Note that the locations that contribute to these functions depend on $\boldsymbol{h}$, as with changing $\boldsymbol{h}$, $A \cap A_{-h}$ also changes. Now we will go into the estimation of this function. First we will review shortly the existing methods and comment on their suitability for the purpose.

Existing methods

Much work has already been done in developing efficient sampling designs and methods to compute confidence intervals of variograms. Here, we will review the literature only shortly. For a more extensive review, the reader is referred to Pettitt and McBratney (1993).

Sampling designs

Russo (1984) and Warrick and Myers (1987) proposed algorithms for optimizing configurations of sample points to estimate variograms. The distance classes and the number of pairs for each distance class are decided a priori. The algorithms start with an initial configuration, for example a systematic or random configuration. Next the locations are slightly changed in such a way that the predefined distribution of distance classes is approximated as closely as possible and the variation of $\boldsymbol{h}$ within distance classes is minimal. Russo and Jury

(1988) showed that the optimized sampling configuration produced more accurate estimates of the covariance function than the systematic sampling configuration.

Zimmerman and Homer (1991) proposed a *D*-optimality criterion to optimize selection of additional sampling locations in order to estimate selected attributes of the ergodic variogram. These attributes were the ratio of nugget to sill, "compatibility"-determining attributes and anisotropy.

Corsten and Stein (1993) compared several sampling configurations for variogram estimation, including nested, systematic, simple random and transect sampling configurations. They concluded that:

(i) the nested sampling designs produced relatively inaccurate estimates of variogram-parameters;
(ii) prediction error variances with variograms estimated from nested sampling designs were about twice those obtained with the other designs.

In the next chapter we will further comment on the nested sampling design.

Pettitt and McBratney (1993) proposed a staggered design on linear transects in three orientations as a new and suitable sampling design for estimation of the variogram. They used the *D*-optimality criterion to quantify the efficiency of staggered transects with different order of the distances between points on the transect.

Confidence intervals

Davis and Borgman (1979) derived the probability density function (p.d.f.) of the ergodic variogram estimator. However, this p.d.f. holds only for equally spaced observations from a one-dimensional, stationary Gaussian random function, which is rather specific. Davis and Borgman (1982) showed that for a stationary, second order random function, the quantity $\{\hat{\gamma}(\boldsymbol{h}) - \gamma(\boldsymbol{h})\}/\sigma_p\{\hat{\gamma}(\boldsymbol{h})\}$, where $\sigma_p\{\hat{\gamma}(\boldsymbol{h})\}$ is the sampling standard deviation of the estimated semivariance for lag $\boldsymbol{h}$, is asymptotically standard normal distributed.

McBratney and Webster (1986) and Taylor and Burrough (1986) calculated the accuracy of estimated variograms *indirectly* by simulation of random fields.

This procedure was followed also by Russo and Jury (1987) and Corsten and Stein (submitted). However, this procedure produces confidence intervals of the ergodic variogram, showing uncertainty due to fluctuation variance (and sampling variance if one considers more realizations of the sampling design), whereas in this paper we are concerned with sampling variance of non-ergodic variograms. Morris (1991) estimated the minimum variance of the variogram by calculating the 'maximum equivalent uncorrelated pairs', assuming a concave variogram. Again, the calculated minimum variance reflects fluctuation variance plus sampling variance of the theoretical variogram.

To obtain confidence intervals of the non-ergodic variogram, Webster and Oliver (1992) simulated a single random field from an estimated variogram and sampled this realization many times according to several sampling designs. We consider this procedure as less appropriate for calculating confidence intervals in real world situations because the non-ergodic variogram of the simulated realization will generally differ from that of the sampled area for two reasons:

- the quality of the fitted model used in simulation depends strongly on the sample (size, design). Especially with small samples the fitted model may strongly deviate from the 'true' non-ergodic variogram.
- realizations from the same fitted model will have different non-ergodic variograms. There is no guarantee that the non-ergodic variogram of the simulated field equals that of the fitted model.

A second way of indirect estimation of confidence intervals of non-ergodic variograms is subsampling. Chung (1984) and Shafer and Varljen (1990) used the jackknife method. If data are dependent, this procedure becomes rather complicated (Davis, 1987). Gascuel-Odoux and Boivin (1993) used a simple subsampling procedure in which subsamples of points were randomly selected from the sample. However, a given point of the subsample is paired with all other subsample points, so squared differences are certainly correlated. By neglecting this dependency estimates of the sampling variance of the semivariance become biased.

Muñoz-Pardo (1987) derived an expression for calculation of the sampling variance of the non-ergodic variogram. However, in this procedure the non-

ergodic variogram must be known and therefore this procedure is of minor practical importance.

A new approach: design-based sampling strategies

Sampling designs

A common belief among geostatisticians is that there is not a simple way of calculating confidence intervals of variograms because data are dependent but the dependency is unknown (it is this dependency which one wants to model). We agree with this for theoretical (ergodic) variograms with confidence intervals reflecting fluctuation variance plus sampling variance. However, for non-ergodic variograms we believe there is a way out of the problem. To clarify this we first have to stress that the meaning of statistical dependence is unclear without further specification of the source of stochasticity (De Gruijter and Ter Braak, 1990). This source is different for design-based and model-based sampling strategies. In the design-based approach the sampling configuration is random and the model-realization is fixed, whereas in the model-based approach, the other way round, the model-realization is random and the sampling configuration is fixed. From this it follows that dependence has a different meaning in the design-based and model-based approach. Brus and De Gruijter (1993) showed that data which are dependent in the model-based approach still can be independent in the design-based approach and vice versa. For estimation of non-ergodic variograms the only relevant source of stochasticity is sampling variation. Therefore hereafter we use independence in the design-based sense, which we denote by *p*-independence.

In this paper we present a new approach for non-ergodic variogram estimation in which *p*-independence of squared differences is created by the sampling design. This offers the opportunity of simple derivation of the sampling variance of the variogram. This sampling variance can then be used to calculate confidence intervals or to fit models.

The past 50 years have shown large developments of classical sampling theory, leading to a great variety of design-based sampling strategies. Essentially

all these can be used for variogram estimation. There are two main differences, however, with the usual problems of spatial sampling. First, the population elements and sampling units consist of *pairs of points* instead of single points. In other words the population does not consist of all possible locations in area A but of the Cartesian product of A and itself: $A \times A$. Second, a function ($\gamma_A(\boldsymbol{h})$) has to be estimated instead of a single parameter. Efficient strategies for this choose a number of lag vectors $\boldsymbol{h}$ and a sample size for each of them. These choices can be made using theory of experimental design. Then for each $\boldsymbol{h}$ and sample size $n(\boldsymbol{h})$, there is the question how to sample efficiently, which can be dealt with by classical sampling theory. In this paper we focus on the latter part of the problem. Thus the question under study is how to estimate the mean and its sampling variance of squared differences of all pairs of points with a given lag vector $\boldsymbol{h}$. Given a design p, we are looking for estimators with an expected value over all realizations of the sampling process equal to the real value:

$$E_p\{\hat{\gamma}_A(\boldsymbol{h})\} = \gamma_A(\boldsymbol{h}) \qquad (6.2)$$

In other words, the strategy (design + estimator) should be p-unbiased or design-unbiased.

To illustrate the approach we show this for the two designs that many standard books (Cochran, 1977; Särndal *et al.*, 1992) on sampling start with: simple random sampling (*SI*) and stratified simple random sampling (*STSI*). It is not our intention to propagate these two techniques as optimal solutions. They are just simple illustrations of how classical sampling theory can be applied for variogram estimation. After this we will shortly go into cluster sampling and two-stage sampling which still have to be worked out for variogram estimation and finally we will comment on the above mentioned nested sampling design which at first sight may seem to be a suitable design-based strategy for variogram estimation. Finally we will discuss the advantages of the proposed methods in selection and fitting of variogram models and in determining the sampling variances and covariances of the estimated model parameters.

Simple random sampling of pairs of points

In simple random sampling for a lag vector $\boldsymbol{h}$, *pairs* of points separated by $\boldsymbol{h}$ are selected independently from each other and in such a way that the inclusion probability is equal for all such pairs. A practical drawing technique is to take a simple random sample of points, i.e. points are selected at random and independently. These are the first points of the pairs to be selected. In the anisotropic case, i.e. if $\boldsymbol{h}$ is characterized by its length *and* its direction, there is only one possible second point. In other words the second point is selected simultaneously with the first one. If the second point of a pair falls outside the boundaries of the area, the first point of this pair is dropped as well. The whole procedure is done for several, predetermined lag vectors $\boldsymbol{h}$.

In the isotropic case $\boldsymbol{h}$ is characterized by its length only and consequently there are several possible second points. For each selected first point $\boldsymbol{x}_i$ a counterpart is selected with random direction from $\boldsymbol{x}_i$ and at distance $h = |\boldsymbol{h}|$. If the selected second point falls outside the area, the first point is dropped, also if there are other second points (not selected) inside the area. In other words, for all first points, only once a second point is drawn.

As the inclusion probabilty of all pairs of points, given $\boldsymbol{h}$, is equal, the non-ergodic semivariance for lag $\boldsymbol{h}$ can be estimated by taking the unweighted mean:

$$\hat{\gamma}_A(\boldsymbol{h}) = \frac{1}{2n(\boldsymbol{h})} \sum_{i=1}^{n(\boldsymbol{h})} \{z(\underline{\boldsymbol{x}}_i) - z(\underline{\boldsymbol{x}}_i + \boldsymbol{h})\}^2 \qquad (6.3)$$

where:
$n(\boldsymbol{h})$ = number of selected pairs of points.

The estimator of the non-ergodic semivariance is equal to the one given by Isaaks and Srivastava (1988). However, we would like to stress that this estimator is p-unbiased only under the design described above. If a different design is applied such as a stratified random design (see next section), with unequal inclusion probabilities, then the estimator is not p-unbiased.

As the pairs are selected independently, the observed squared differences are

p-independent and the sampling variance of the estimated semivariance, $\hat{\sigma}_p^2\{\hat{\gamma}(\boldsymbol{h})\}$, can be simply estimated by the usual estimator (Cochran, 1977):

$$\hat{\sigma}_p^2\{\hat{\underline{\gamma}}_A(\boldsymbol{h})\} = \frac{\hat{\underline{v}}_A\{0.5\underline{\Delta}^2(\boldsymbol{h})\}}{n(\boldsymbol{h})} = \frac{0.25\sum_{i=1}^{n(\boldsymbol{h})}[\underline{\Delta}_i^2(\boldsymbol{h}) - \hat{\underline{m}}_A\{\underline{\Delta}^2(\boldsymbol{h})\}]^2}{n(\boldsymbol{h})\{n(\boldsymbol{h})-1\}} \qquad (6.4)$$

where:

$\hat{\underline{v}}_A\{\}$ = estimated spatial variance of the quantity between brackets within A;

$\underline{\Delta}^2_i(\boldsymbol{h})$ = squared difference of the ith pair separated by lag $\boldsymbol{h}$;

$\hat{\underline{m}}_A\{\}$ = estimated spatial mean of the squared differences within A.

Stratified simple random sampling of pairs of points

In stratified simple random sampling the population consisting of all pairs of points with a given lag vector $\boldsymbol{h}$ is first divided into disjoint subpopulations called strata. In practice this can be done by dividing the area A into subregions. All pairs of points (with lag vector $\boldsymbol{h}$) within a subregion, form a stratum. But, we also have pairs whose points lie in different subregions. These form additional strata, one for each pair of subregions. From each stratum a simple random sample of pairs of points with lag vector $\boldsymbol{h}$ is taken. This procedure is followed for all selected lag vectors $\boldsymbol{h}$. The stratification need not be equal for all vectors $\boldsymbol{h}$.

Stratification increases the accuracy of the estimated variogram if the strata are more homogeneous for the semivariance than the whole area: that is if the spatial means or variances of squared differences for a given lag $\boldsymbol{h}$ differ between strata. Stein *et al.* (1988) reported differences in ergodic semivariance of several soil properties between soil map units. In general if one has differences in spatial variance between strata there will also be differences in semivariance for lags larger than the range. Stratification produces gain for these lags.

For stratified simple random sampling, the usual p-unbiased estimator of the non-ergodic semivariance is:

$$\hat{\gamma}_A(\boldsymbol{h}) = \sum_l \frac{N_l(\boldsymbol{h})}{N(\boldsymbol{h})} \frac{1}{2n_l(\boldsymbol{h})} \sum_{i=1}^{n_l(\boldsymbol{h})} \{z(\underline{\boldsymbol{x}}_l) - z(\underline{\boldsymbol{x}}_l + \boldsymbol{h})\}^2 \qquad (6.5)$$

where:
$N_l(\boldsymbol{h})$ = size of $A \cap A_{-\boldsymbol{h}}$ in stratum l ($A \cap A_{-\boldsymbol{h}} \cap l$);
$N(\boldsymbol{h})$ = size of $A \cap A_{-\boldsymbol{h}}$;
$n_l(\boldsymbol{h})$ = size of the sample from stratum l.

The ratio of $N_l(\boldsymbol{h})$ to $N(\boldsymbol{h})$ equals the ratio of the area of $A \cap A_{-\boldsymbol{h}} \cap l$ to the area of $A \cap A_{-\boldsymbol{h}}$, and acts as the stratum weight denoted by $w_l(\boldsymbol{h})$.

The sampling variance of the estimated semivariance can be estimated by:

$$\hat{\sigma}_p^2\{\hat{\gamma}_A(\boldsymbol{h})\} = \sum_l w_l^2(\boldsymbol{h}) \frac{\hat{v}_l\{0.5\Delta^2(\boldsymbol{h})\}}{n_l(\boldsymbol{h})} = 0.25 \sum_l w_l^2(\boldsymbol{h}) \frac{\sum_{i=1}^{n_l(\boldsymbol{h})} [\Delta_i^2(\boldsymbol{h}) - \hat{m}_l\{\Delta^2(\boldsymbol{h})\}]^2}{n_l(\boldsymbol{h})\{n_l(\boldsymbol{h}) - 1\}} \qquad (6.6)$$

where:
$\hat{v}_l$ = estimated spatial variance within stratum l;
$\hat{m}_l$ = estimated spatial mean of stratum l.

Other designs
A great variety of other designs is available and in principle all of them may be used for variogram estimation. It is impossible to review them all and we mention only two classes of designs coming into scope if *SI* and *STSI* lead to prohibitive costs because the sampling units are scattered over a wide area. Solutions to this problem are cluster sampling and two-stage sampling.

In cluster sampling the population is partitioned into subpopulations called clusters. A probability sample of clusters is selected and *every* population element in the selected cluster is surveyed. Clusters can be selected according to various designs, for example by simple random sampling, resulting in simple

random cluster sampling (*SIC*). The clusters of pairs can have all kinds of configurations. One example is an equilateral triangle with sides of length h. For each h several clusters are selected independently from each other with random centres, and, for isotropic variograms, with random directions. The points at the vertices of the triangle form three pairs. These three pairs are not selected independently and therefore the squared differences of these three pairs within a cluster are not p-independent. In this case the three squared differences are even logically correlated because each pair has one point in common with the two other pairs. However, the squared differences of pairs from different clusters are p-independent and consequently a robust and p-unbiased estimate of the sampling variance of the estimated mean of squared differences can simply be obtained (Särndal *et al.*, 1992, p.129).

Unless otherwise stated, the word "robust" in this paper is used to describe procedures for inference of semivariances, their sampling variances, variogram model parameters and their sampling variances and covariances that are insensitive to deviations from an assumed spatial autocorrelation model. It is not related to assumptions on the frequency distribution of $z(\boldsymbol{x})$ (robust to outliers) as in Cressie and Hawkins (1980) or to assumptions on the probability distribution of the estimated semivariances.

If we take larger clusters, in each cluster pairs can be formed with different h's and the problem arises whether we should use all pairs or only the pairs separated by a specific h. For example, in a cluster consisting of a square with side h, four pairs exist separated by h and two separated by $h\sqrt{2}$. For each type of pairs, the squared differences of pairs from different clusters are p-independent, which makes it possible to get robust, p-unbiased estimates of the sampling variance of the estimated semivariance for both distances. However, within clusters each pair separated by h has its points in common with the pairs separated by $h\sqrt{2}$, making them logically correlated. Consequently $\hat{\gamma}_A(h)$ is correlated with $\hat{\gamma}_A(h\sqrt{2})$, and this makes it difficult to obtain robust p-unbiased estimates of the variance of the estimated parameters of the fitted model. To avoid this problem we recommend to use only the pairs separated by h.

In general for a given sample size, the sampling variance under *SIC* will be larger than under *SI* due to the tendency for pairs in the same cluster to resemble each other. This is the price to be paid for smaller sampling costs. If one has only a few, large clusters the increase in sampling variance might become inacceptable. In that case, it often pays to select more clusters and, in order to control the cost, to subsample within the selected clusters instead of surveying all pairs within the cluster. Such designs are referred to as two-stage sampling. We may choose any design for the first stage as well as for the second. For example, suppose one wants to estimate the anisotropic semivariance for a given direction. We might then consider all pairs on a line with this direction as one primary sampling unit. In the first stage, a simple random sample of such units is selected with probabilities proportional to their size (i.e. number of pairs). In the second stage, within each selected primary unit a simple random sample of pairs is selected with equal probabilities. As this design is self-weighting, that is all pairs separated by $\boldsymbol{h}$ have equal inclusion probability, the estimator of the sampling variance remains relatively simple.

We already mentioned the nested sampling design (Miesch, 1975; Oliver and Webster, 1986). Also in this design pairs are selected by some probability mechanism which makes some researchers believe that it produces *p*-unbiased estimates of the variance components (semivariance). However, in nested sampling the inclusion probabilities of the pairs generally differ. In the estimator this is not accounted for and as a result the estimates will not be *p*-unbiased. Moreover the inclusion probabilities of certain subsets of pairs can be zero for specific combinations of lag distances at succeeding stages. In that case the sample is not even a probability sample in the strict sense, and the results will be biased whatever estimator is used. Also squared differences are clearly correlated, making it difficult to estimate sampling variances. Finally, we have to stress that nested sampling is designed for estimating ergodic variance components and this is the reason why the differences in inclusion probabilities are not taken into account.

Fitting models

The proposed methods have important advantages when fitting variogram models. These advantages are related to:

(i) selection of a model.
(ii) estimation of the model parameters;
(iii) estimation of the sampling variance and covariance of the estimated model parameters;

Selecting a model

An important advantage of the proposed design-based methods is that they offer the possibility of selecting objectively a model by testing lack of fit by a *F*-test. As we have *p*-independent repeat observations of $\Delta^2(h)$ at given values of $\boldsymbol{h}$, we can compare the variation within $\boldsymbol{h}$-groups with the residual variation of the model under consideration. This is done by comparing the residual deviance of this model ($\underline{D}_{mod}$) to that of the model that contains the $\boldsymbol{h}$-groups as a qualitative predictor ($\underline{D}_{group}$). The test statistic $\underline{F}$ is defined by:

$$\underline{F} = \frac{(\underline{D}_{mod} - \underline{D}_{group})/(df_{mod} - df_{group})}{\underline{D}_{group}/df_{group}} \qquad (6.7)$$

where:

df_{mod} = residual degrees of freedom of the model under consideration;
df_{group} = residual degrees of freedom of the model with the $\boldsymbol{h}$-groups as qualitative predictor.

If the model is correct, $\underline{F}$ is close to 1, if the model is incorrect it will exceed 1. The $\underline{F}$-values can be used to accept or reject a model or to choose between alternative models objectively.

Estimation of model parameters

Generally the means of the squared differences, $\hat{\underline{\gamma}}_A(\boldsymbol{h})$, are used to fit the model. However, to obtain correct estimates of the sampling variance and covariance of the estimated model parameters, the individual squared differences must be used to fit the model. In addition, for selecting a model by testing lack of fit, one needs the individual squared differences, $\underline{\Delta}^2_i(\boldsymbol{h})$ (see next section). Therefore we propose to use the individual squared differences in fitting.

The main advantage of the proposed sampling strategies for estimating variogram model parameters is that *p*-unbiased and robust parameter estimates can simply be obtained because no assumptions need to be made on spatial autocorrelation of squared differences. In the proposed strategies, *p*-independence of squared differences is created by the sampling design. For example, in Maximum Likelihood estimation (*ML*) the loglikelihood can simply be written as a sum of the logarithm of the probabilities of the observations (Cressie, 1991, p. 458). No assumptions need to be made on the spatial autocorrelation model for calculating the joint probabilities of the observations.

The proposed strategies also provide *p*-unbiased estimates of the variance of the squared differences for a given $\boldsymbol{h}$. The reciprocal values of these estimated variances can be used as weights in fitting. For small sample sizes the variance estimates will be rather inaccurate. In this case it is preferable to use these variance estimates to select an appropriate model describing the variance as a function of the expected (fitted) value, and to use these calculated weights in fitting (McCullagh and Nelder, 1989).

Estimating sampling (co)variance of model parameters
The proposed methods also offer good potentials for quantifying the sampling variances and covariances of the estimated parameters of the selected model. These can be calculated as the inverse of the information matrix (Ross, 1990, p. 7). As the squared differences within a group, i.e. for a given $\boldsymbol{h}$, and between the groups are *p*-independent, these estimates of the (co)variances are *p*-unbiased and robust. This dispersion matrix with (co)variances is a more comprehensive measure of uncertainty about the variogram than the confidence intervals of the estimated semivariances at given values of $\boldsymbol{h}$.

Case study

We estimated non-ergodic semivariances by simple random sampling of pairs of points (*SI*) in the Lickebaert polder, which lies in the western part of the Netherlands, near Rotterdam. For $\boldsymbol{h}$, we took 1.5, 4, 12, 32, 90 and 250 m and random directions for all distances. This means that we estimated isotropic

variograms, so that $\gamma_A(\boldsymbol{h}) = \gamma_A(h)$. For each distance h, we selected 15 pairs. Pairs of which one or both points lie in 'non-soil' (e.g. ditch or farmyard) were removed from the dataset. The resulting numbers of pairs are presented in Table 6.1. The configuration is presented in Figure 6.2.

We estimated non-ergodic semivariances for the bulk soil electrical conductivity (C; mS m^{-1}), the mean highest water table ($\bar{W}_{min}$; cm below surface) and the mean lowest water table ($\bar{W}_{max}$; cm below surface). Figure 6.3 shows the estimates and their 90 % confidence intervals assuming a normal distribution of $\hat{\gamma}_A(\boldsymbol{h})$. Except for h = 1.5 and 4 m, the 90 % confidence intervals were rather wide due to a large sampling variance for these lags. Figure 6.3 clearly shows that, in general, the larger the estimated semivariance, the larger its sampling variance. For some lags the lower endpoints of the 90 % confidence intervals were slightly negative. This reflects that the normal distribution assumption for the semivariance is unrealistic, which can be explained by the relatively small values of $n(h)$.

We fitted spherical and exponential models to the estimates by *ML* assuming a gamma distribution and with variance proportional to the fitted value squared. We used Genstat for the computations (Genstat 5 Comittee, 1987). Table 6.2 shows the results. The fitted spherical models are presented in Figure 6.3.

To answer the question whether a spherical or an exponential model fits best, we executed a *F*-test for lack of fit, assuming a gamma distribution again. The results are shown in Table 6.3. For both the spherical and exponential models the *F*-values are below the critical value (for $\alpha = 0.05$), which means that there is no significant lack of fit for both model types, for all variables. For C and $\bar{W}_{min}$ the *F*-values for the spherical models were somewhat smaller than those of the exponential models, for $\bar{W}_{max}$ these were nearly equal. This means that

Table 6.1 Lag distance (h) in m and number of pairs ($n(h)$) of simple random sample of pairs in Lickebaert polder

h	1.5	4	12	32	90	250
$n(h)$	11	14	13	13	11	10

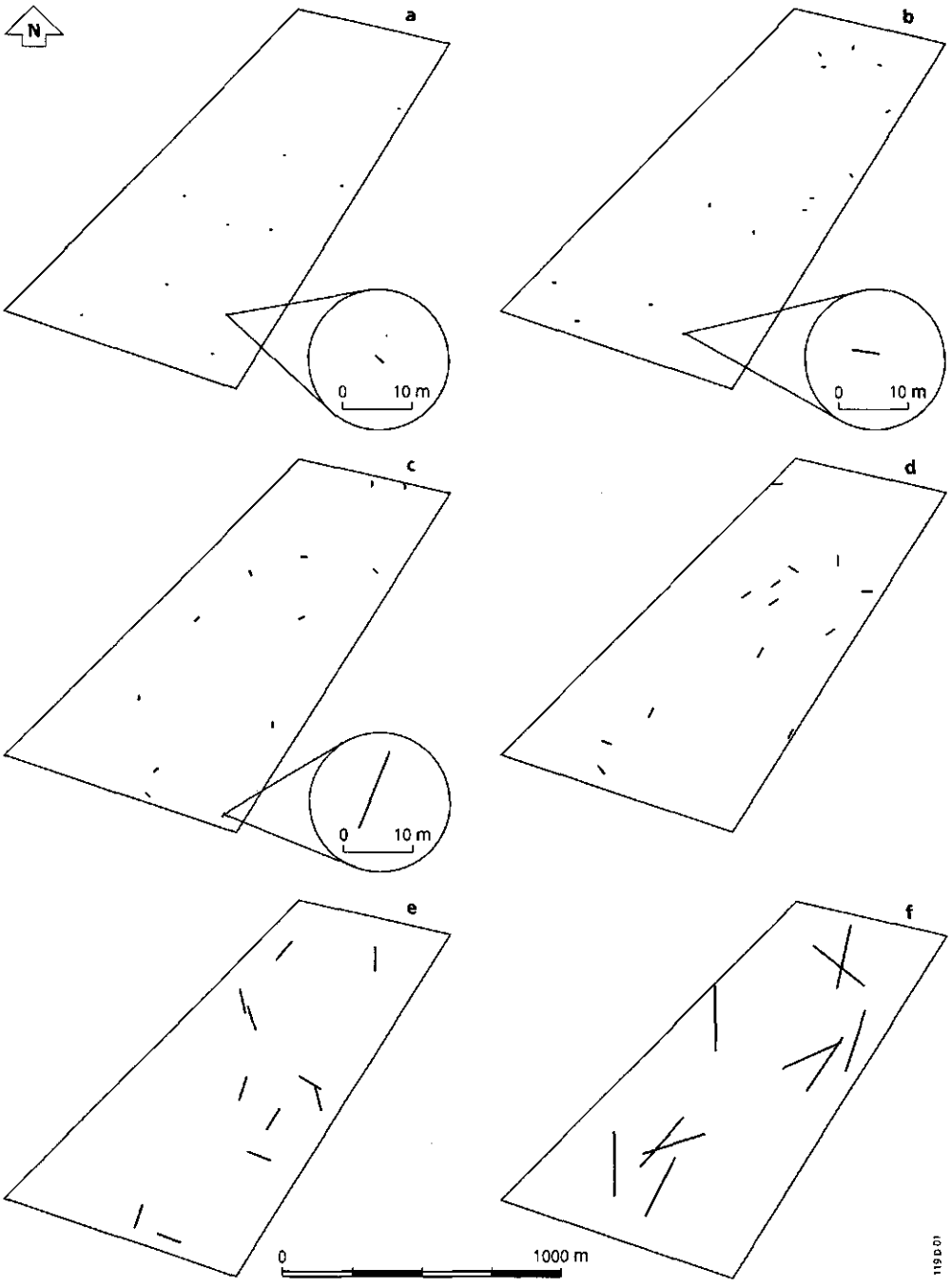


Fig. 6.2 Simple random sample of pairs of points within the Lickebaert polder. Points of a pair are connected by lines; **a**: 1.5 m pairs; **b**: 4 m pairs; **c**: 12 m pairs; **d**: 32 m pairs; **e**: 90 m pairs; **f**: 250 m pairs

Table 6.2 Estimated parameters of spherical and exponential model for non-ergodic variogram of bulk soil electrical conductivity (C), mean highest water table ($\bar{W}_{min}$) and mean lowest water table ($\bar{W}_{max}$); a = range (km), c = sill

	Spherical		Exponential	
	a	c	a	c
C	129	641	122	957
$\bar{W}_{min}$	99.9	125	67.1	142
$\bar{W}_{max}$	108	478	80	525

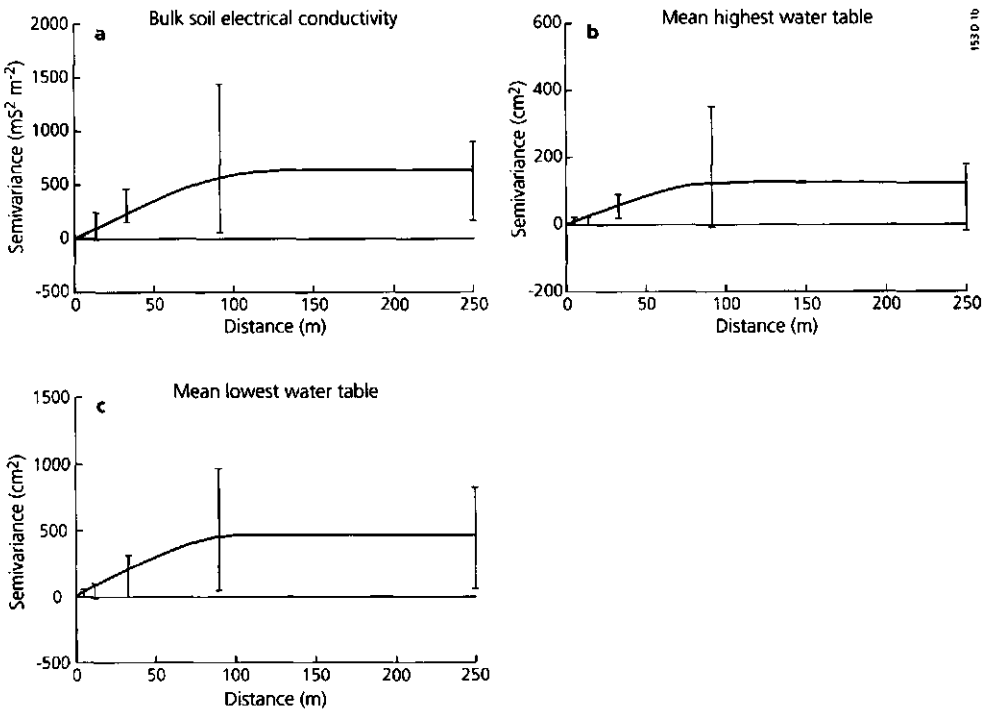


Fig. 6.3 Estimated non-ergodic semivariances. **a**: bulk soil electrical conductivity, **b**: mean highest water table, and **c**: mean lowest water table. The bars show the 90% confidence interval. The fitted models are spherical

Table 6.3 *F*-values of test for lack of fit of spherical and exponential model (for fitted parameters: see Table 6.2)

Model	C	$\bar{W}_{min}$	$\bar{W}_{max}$
Spherical	0.867	0.495	0.240
Exponential	1.235	0.566	0.226

the spherical model fitted somewhat better, although not significantly.

Table 6.4 shows the dispersion matrices of the estimated parameters of the spherical models under the assumption of gamma distributed squared differences. For all variables the parameters a and c are strongly correlated. This means that the 'slope' of the fitted line for $h < a$ remains fairly constant.

Table 6.4 Dispersion matrices (expressed as standard deviations and correlation coefficients) of parameters of spherical model of non-ergodic variogram for bulk soil electrical conductivity (C), mean highest water table ($\bar{W}_{min}$) and mean lowest water table ($\bar{W}_{max}$)

	C		$\bar{W}_{min}$		$\bar{W}_{max}$	
a	76.5		105		114	
c	0.921	325	0.879	106	0.884	417
	a	c	a	c	a	c

Discussion and conclusions

In the presented methods *SI* and *STSI*, each of the $2n$ points is combined with only one other point to form a pair. This results in a set of n pairs. In most existing methods each point is combined with all other points to form a set of $n(2n\text{-}1)$ pairs. At first sight this seems to be more efficient. However, one

should realize that the larger number of pairs is misleading because the pairs are correlated and as a result the decrease in sampling variance will be far less than expected on the basis of the numbers alone. Morris (1991) presented a method to account for this effect, by establishing a maximum equivalent number of uncorrelated pairs. He found considerable differences between the nominal number of pairs and the maximum equivalent number, especially for a triangular configuration of points. Nevertheless, in general the maximum equivalent number will be larger than n. This advantage must be weighted, however, against the drawback that one still needs to postulate a variogram to obtain the sampling variance of the semivariances and model parameters. As this variogram is unknown (we want to estimate it) the estimates of the sampling variances are not robust. In short, the choice is between uncertain estimates of somewhat smaller sampling variances and certain (robust and p-unbiased) estimates of somewhat larger sampling variances. Depending on the circumstances it may be wise to choose the latter.

We conclude that design-based sampling strategies based on classical sampling theory offer strong potentials for variogram estimation which were up till now undiscovered. The p-independent observations of squared differences can be used to obtain p-unbiased and robust estimates of:

- the semivariance and its sampling variance at given values of $\boldsymbol{h}$;
- the variogram model parameters and their sampling variances and covariances;
- the variation within $\boldsymbol{h}$-groups needed in a F-test for lack of fit.

We presented the estimators of the semivariance and its sampling variance for only two elementary designs. Other types of designs such as cluster sampling and two-stage sampling which we dealt with very shortly, should be worked out for variogram estimation and tested in the real world. The efficiency of designs for variogram estimation should be compared. Furthermore theory of experimental design may be used to optimize the choice of h and $n(h)$.

The proposed sampling strategies are applicable only in the pre-sampling situation, i.e. when data still have to be collected, and in the post-sampling situation when the probabilities of inclusion of the sample points are known. In

environmental soil science we often encounter the first situation because available data are insufficient or completely missing. If we already have data but the applied sampling design is unknown, the proposed strategies cannot be applied. This is the price to be paid for losing essential meta-information.

Acknowledgements

We thank J. Oude Voshaar for statistical advice, R. Schuiling for automatic mapping with ARC-INFO and A. Scholten and M. Knotters for collecting the data.

References

Brus, D.J. & De Gruijter, J.J. 1993. Design-based versus model-based estimates of spatial means. Theory and application in environmental soil science. *Environmetrics* 4, 123-152.

Chung, C.F. 1984. Use of the jackknife method to estimate autocorrelation functions (or variograms), in Verly, G., David, M., Journel, A.G. and Marechal, A. (eds): *Geostatistics for natural resources characterization I*, D. Reidel Publishing Company, Dordrecht, pp. 55-69.

Cochran, W.G. 1977. *Sampling techniques*, 3rd edn. Wiley, New York.

Corsten, L.C. & Stein, A. 1993. Nested sampling for estimating spatial semivariograms compared to other designs. *Stochastic Modelling and Data Analysis* (in press).

Cressie, N. A. C. 1991. *Statistics for spatial data*. Wiley, New York.

Cressie, N. & Hawkins, D.M. 1980. Robust estimation of the variogram. *Mathematical Geology* 12, 115-125.

Davis, B.M. 1987. Uses and abuses of cross-validation in geostatistics. *Mathematical Geology* 19, 241-248.

Davis, B.M. & Borgman, L.E. 1979. Some exact sampling distributions for variogram estimators. *Mathematical Geology* 11, 643-653.

Davis, B.M. & Borgman, L.E. 1982. A note on the asymptotic distribution of the sample variogram. *Mathematical Geology* 14, 189-193.

Draper, N.R. & Smith, H. 1981. *Applied regression analysis*, 2nd edn. Wiley, New York.

De Gruijter, J.J. & Ter Braak, C.J.F. 1990. Model-free estimation from spatial samples: a reappraisal of classical sampling theory. *Mathematical Geology* 22, 407-415.

Gascuel-Odoux, C. & Boivin, P. 1993. Variability of sample variograms and spatial estimates according to soil sampling: a case study. *Geoderma* (in press).

Isaaks, E.H. & Srivastava, R.M. 1988. Spatial continuity measures for probabilistic and deterministic geostatistics. *Mathematical Geology* 20, 313-341.

Journel, A.G. 1985. The deterministic side of geostatistics. *Mathematical Geology* 17, 1-15.

Journel, A.G. & Huijbregts, C.J. 1978. *Mining Geostatistics*. Academic Press, London.

McBratney, A.B. & Webster, R. 1986. Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. *Journal of Soil Science* 37, 617-639.

McCullagh, P. & Nelder, J.A. 1989. *Generalized linear models*, 2nd edn. Chapman and Hall, London.

Miesch, A.T. 1975. Variograms and variance components in geochemistry and ore evaluation. *Geological Society of America Memoir*, pp. 333-340.

Morris, M.D. 1991. On counting the number of data pairs for semivariogram estimation: *Mathematical Geology* 23, 929-943.

Muñoz-Pardo, J.F. 1987. *Approche géostatistique de la variabilité spatiale des milieux géophysiques. Application a l'échantillonnage de phénomènes bidimensionnels par simulation d'une foncton aléatoire.* Thèse Docteur-Ingénieur, l'Université Scientifique, Technologique et Médicale de Grenoble et l'Institut National Polytechnique de Grenoble.

Oliver, M.A. & Webster, R. 1986. Combining nested and linear sampling for determing the scale and form of spatial variation of regionalized variables. *Geographical Analysis*, 18, 227-242.

Pettitt, A.N. & McBratney, A.B. 1993. Sampling designs for estimating spatial variance components, *Applied Statistics* 43, no. 4 (in press).

Ross, G.J.S. 1990. *Nonlinear estimation*. Springer-Verlag, New York.

Russo, D. 1984. Design of an optimal sampling network for estimating the variogram: *Soil Science Society of America Journal* 48, 708-716.

Russo, D. & Jury, W.A. 1988. Effect of sampling network on estimates of the covariance function of stationary fields. *Soil Science Society of America Journal* 52, 1228-1234.

Särndal, C-E., Swensson, B. & Wretman, J. 1992. *Model assisted survey sampling*. Springer-Verlag, New York.

Shafer, J.M. & Varljen, M.D. 1990. Approximation of confidence limits on sample semivariograms from single realizations of spatially correlated random fields. *Water Resources Research* 26, 1787-1802.

Srivastava, R.M. 1987. *A non-ergodic framework for variograms and covariance functions*. M.Sc. Thesis, Stanford University, Stanford, California, USA.

Stein, A., Hoogerwerf, M. & Bouma, J. 1988. Use of soil-map delineations to improve (co-)kriging of point data on moisture deficits. *Geoderma* 43, 163-177.

Taylor, C.C. & Burrough, P.A. 1986. Multiscale sources of spatial variation in soil. III.

Improved methods for fitting the nested model to one-dimensional semi-variograms: *Mathematical Geology* 18, 811-821.

Warrick, A.W. & Myers, D.E. 1987. Optimization of sampling locations for variogram calculations. *Water Resources Research* 23, 496-500.

Webster, R. & Oliver, M.A. 1992. Sample adequately to estimate variograms of soil properties. *Journal of Soil Science* 43, 177-192.

Zimmerman, D.L. & Homer, K.E. 1991. A network design criterion for estimating selected attributes of the semivariogram. *Environmetrics* 2, 425-441.

Chapter 7

Incorporating the sampling error of non-ergodic variograms in kriging estimation errors

Mathematical Geology, submitted

with M.J.W. Jansen

Incorporating the sampling error of non-ergodic variograms in kriging estimation errors

The kriging variance of spatial predictions is in practice calculated from sample variograms and does not account for uncertainty about the true variogram, leading to underestimations of the mean squared error. Moreover, ergodic variograms are defined over all model realizations and this too may cause a difference between the mean kriging variance and the mean squared error. The latter cause can be eliminated by using non-ergodic variograms. Moreover, the error component in spatial estimates caused by uncertainty about the variogram can be determined much more simply with non-ergodic variograms than with ergodic ones if pairs of points are selected at random and independently. In a case study, the strong underestimation of the mean squared error by the kriging variance was almost completely eliminated by adding the estimation variance due to sampling error of the non-ergodic variogram.

Keywords: kriging variance, sampling variance, bootstrap, model-based inference, error propagation

Introduction

Spatial estimation techniques using a model of spatial variation (spatial continuity) such as kriging, are potentially more efficient than model-free techniques. Also, an estimate of the precision, the kriging variance, can be obtained directly, without test points with known values. However, the mean kriging variance (*MKV*) often differs considerably from the mean squared error of prediction (*MSE*) at test points. Laslett *et al.* (1987) reported percentages of underestimation (*U*) ranging from 30 - 77%. Bregt *et al.* (1991) found a percentage underestimation of 70%. They adjusted the kriging variances by a factor *MSE*/*MKV* to obtain more realistic kriging variance estimates. However this procedure is not founded on theory and one needs test points for it, eliminating an important advantage of kriging. Overestimations are reported too, see for instance Stein *et al.* (1988).

There are two main causes for the discrepancy between the mean squared error of prediction at test points and the mean kriging variance. The first cause is that in the geostatistical approach, the variogram is defined in terms of expected values over all possible realizations of a probability model (ξ-expectations), i.e. ensemble averages (Isaaks and Srivastava, 1988). Such variograms are referred to as ergodic or theoretical variograms. Suppose that the ergodic variogram is known. In areas where the realization is more erratic than average, the ergodic variogram will underestimate the mean squared error, whereas in areas where the realization is less erratic, the mean squared error will be overestimated.

This problem can be avoided by modelling spatial structure within a deterministic framework (Journel, 1985). Within this framework, the variogram is defined over a single realization, i.e. a single finite and bounded domain. It is referred to as the non-ergodic or local variogram (Journel and Huijbregts, 1978; Srivastava, 1987; Isaaks and Srivastava, 1988). Realizations from the same ergodic variogram will have different non-ergodic variograms. Isaaks and Srivastava (1988) showed for the Wiener-Levy process that if the ergodic and non-ergodic spatial continuity functions are known, the non-ergodic covariance function gives more reliable confidence intervals of predictions than the ergodic variogram because "it captures the character of spatial variability unique to the domain over which it is defined". When we use estimated variograms estimated from data sampled in the study area, this problem may become less serious. In that case the estimated ergodic variogram resembles to a greater or lesser degree the non-ergodic variogram of this area and consequently captures to a certain degree the features peculiar to the study area.

A second cause for the above-mentioned difference is that the kriging variance does not account for the uncertainty about the variogram. Scatter diagrams of semivariance versus lag (variogram clouds) always show a wide spread, nevertheless the function fitted to the sample variogram is assumed to be an errorless estimate of the true variogram in subsequent kriging operations. This is clearly unrealistic.

Gascuel-Odoux and Boivin (1993) analyzed the effect of the sample size (num-

ber of sample points) on the estimation accuracy of:
- the ergodic sample variogram (semivariance at several lags);
- fitted functions (variograms);
- predictions at points by kriging.

They concluded that despite a large variance of the sample variogram, the estimation variance of fitted functions and kriging predictions was relatively small.

Brus and De Gruijter (in prep.) introduced a set of relatively simple methods for design-unbiased (*p*-unbiased) estimation of the sampling variance of the non-ergodic variogram. In these methods, several lags are chosen and for each lag a probability sample (Särndal *et al.*, 1992) of pairs of points is taken, separated by this lag. The question addressed in this paper is how the estimated sampling variance of the variogram can be used to eliminate the second cause of the discrepancy between the mean squared error and the mean kriging variance.

The aim of this paper is:
(i) to show how a probability sample of pairs of points can be used to incorporate the sampling error of the inferred non-ergodic variogram in the estimation error of point kriging;
(ii) to define a new measure of the estimation variance at points that accounts for uncertainty on the variogram, and to test its validity in a real-world case study.

Incorporating sampling error of non-ergodic variogram

Kriging with known non-ergodic variogram

Let us assume first that for a property z the true non-ergodic variogram γ_A in area A is known. The kriging prediction of z in a prediction point $\boldsymbol{x}_0$ given the observations of z at the points $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \boldsymbol{x}_n$ can be written as:

$$\hat{z}(\boldsymbol{x}_0) = f(X, z(\boldsymbol{x}_1)...z(\boldsymbol{x}_n), \gamma_A) = \sum_{i=1}^{n} \lambda_i(X, \gamma_A)\, z(\boldsymbol{x}_i) \qquad (7.1)$$

where:
X denotes the set $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, ...\boldsymbol{x}_n\}$.

The kriging variance is a function of X and γ_A only. The interpretation of the kriging variance in a deterministic setting of kriging is not immediately clear. In probabilistic kriging, using ergodic variograms, the kriging variance of an unbiased predictor is defined as the ξ-expectation of the squared error of prediction over all realizations of the model, given the locations of observation points and prediction point. However, non-ergodic variograms are defined over one realization of the model only, the one actually sampled. Journel (1985) defines a kriging error in a deterministic framework by shifting the configuration $\boldsymbol{X}$ over A (Journel, 1985). All shifts have equal probability subject to the condition that the entire configuration should lie in A. Each allowed shift, say $\boldsymbol{T}$, induces a kriging error, say ε:

$$\underline{\varepsilon}_T = f(\underline{X}_T, z(\boldsymbol{x}_1 + \underline{\boldsymbol{T}})...z(\boldsymbol{x}_n + \underline{\boldsymbol{T}}), \gamma_A) - z(\boldsymbol{x}_0 + \underline{\boldsymbol{T}}) \qquad (7.2)$$

where:
$\underline{X}_T$ denotes the set $\{\boldsymbol{x}_0 + \underline{\boldsymbol{T}}, \boldsymbol{x}_1 + \underline{\boldsymbol{T}}, \boldsymbol{x}_2 + \underline{\boldsymbol{T}}, ...\boldsymbol{x}_n + \underline{\boldsymbol{T}}\}$.

The randomness of the shift produces randomness of the kriging error. Assuming that the configuration X is negligibly small with respect to A, Journel (1985) derives that the expectation of $\underline{\varepsilon}_T = 0$ if f is the usual kriging predictor. Moreover he shows that the variance of $\underline{\varepsilon}_T$ is given by the usual expression for the kriging variance.

Kriging with estimated non-ergodic variogram

In practice, γ_A has to be estimated by sampling. The estimated variogram will differ to some extent from the true variogram (wrong function type or values of parameters). This results in an additional error. Suppose that we have a probability (random) sample of pairs of points for several predetermined lag vectors $\boldsymbol{h}$, say s, and that the location of the sampled pairs is independent of

the location of the configuration X. This sample can be used to obtain unbiased and robust[1] estimates of (i) the semivariances at the lag vectors $\boldsymbol{h}$ and their sampling variances; (ii) the function (variogram) parameters and their sampling (co)variances. The fitted function (variogram) is used in point kriging:

$$\hat{z}(\boldsymbol{x}_0) = f(X, z(\boldsymbol{x}_1)...z(\boldsymbol{x}_n), \hat{\gamma}_{A,s}) \tag{7.3}$$

From Equation (7.2) it follows that the true value $z(\boldsymbol{x}_0)$ can be written as:

$$z(\boldsymbol{x}_0) = f(X, z(\boldsymbol{x}_1)...z(\boldsymbol{x}_n), \gamma_A) - \varepsilon_0 \tag{7.4}$$

The estimation error is given by:

$$\hat{z}(\boldsymbol{x}_0) - z(\boldsymbol{x}_0) = f(X, z(\boldsymbol{x}_1)...z(\boldsymbol{x}_n), \hat{\gamma}_{A,s}) - f(X, z(\boldsymbol{x}_1)...z(\boldsymbol{x}_n), \gamma_A) + \varepsilon_0 = \eta_s + \varepsilon_0 \tag{7.5}$$

where:

η_s = error component in spatial estimation resulting from estimation error of variogram by sample s.

η_s-errors can be generated by repeated sampling of pairs of points under the executed sampling design p for estimation of the variogram.

Not knowing ε_0, Journel (1985) considers ε_0 as a realization of the random variable ε_T, which has, as stated above, approximately, mean 0, the variance being equal to the usual expression of the kriging variance. From that point of view, given $z()$ and the configuration X, the estimation error is the sum of two components:

$$\underline{\hat{z}}(\boldsymbol{x}_0) - z(\boldsymbol{x}_0) = \underline{\eta}_s + \underline{\varepsilon}_T \tag{7.6}$$

Since the sample s and the shift $\boldsymbol{T}$ are independent, the variance of the estimation error (*var*) equals the sum of the variances of $\underline{\eta}_s$ and $\underline{\varepsilon}_T$:

[1] Robust in this paper means insensitive to deviations from an assumed autocorrelation model.

$$\text{var}\{\underline{\hat{z}}(\boldsymbol{x}_0) - z(\boldsymbol{x}_0)\} = \text{var}(\underline{\eta}_s) + \text{var}(\underline{\varepsilon}_T) = \sigma_p^2 + \sigma_K^2 \qquad (7.7)$$

σ_p^2 will be referred to in short as the sampling variance, σ_K^2 as the kriging variance.

The sum of these two variance components, Equation (7.7), can be minimized with respect to:
(i) sampling design for estimation of variogram;
(ii) variogram estimator;
(iii) estimator of value at point.
The kriging variance component is minimal for the kriging estimator (Journel, 1985). Much work has still to be done on sampling designs and estimators minimizing the sampling variance component.

Besides kriging variance and sampling variance caused by the sampling error of the variogram, still a third source of variation of the kriging estimator can be considered, viz. sampling variance due to repeated sampling of observation points. To simplify, we shall not consider this variation. In other words, in this paper inference is conditional on the sample of observation points.

Model-assisted versus model-based inference

The proposed kriging procedure using non-ergodic variograms and the new variance estimator, has some important advantages over the convential kriging procedure with ergodic variograms and the traditional kriging variance estimator:

- The validity of the variance estimates of the proposed kriging procedure does not depend on the correctness of the model. Even if the model used does not hold, the estimated variance (kriging variance plus the sampling variance) equals the variance calculated with the true model. Following Särndal *et al.* (1992, p. 227, 239) we will refer to such estimation procedures as model-assisted. In probabilistic kriging the validity of the variance estimates is dependent on the validity of the model. If the model does not hold, the estimated kriging variance will differ from the kriging variance calculated with the true model. Such prediction procedures are referred

to as model-dependent (Hansen *et al.*, 1983) or model-based (Särndal, 1978; De Gruijter and Ter Braak, 1990).

- In model-based prediction the usefulness of the variance estimates depends on the similarity between the ergodic variogram and the non-ergodic variogram of the area considered. If both variograms differ considerably, the estimated variance will seriously overestimate or underestimate the true variance, even if the true ergodic variogram is used. In model-assisted estimation this problem does not exist because there is only one variogram.

Generating equiprobable sample variograms

The sampling variance can be estimated by generating a set of equiprobable sample variograms from the sample, fitting variograms and then using these variograms one by one in spatial estimation. The sample variograms can be generated in various ways differing with respect to:

(i) data used in simulation: the squared differences of the sampled pairs or the estimated variogram parameters;

(ii) assumptions about the underlying distribution of the semivariances or variogram parameters.

Table 7.1 shows three plausible procedures which we shall describe shortly. With respect to (ii) we distinguish non-parametric procedures (no assumptions about shape of distribution) and parametric procedures. In fact these are the extremes of a continuum with semiparametric procedures in between (Cressie, 1991).

Table 7.1 Procedures for generating equiprobable sample variograms

	Non-parametric	Parametric
Squared difference	classical bootstrap	parametric sample variogram uncertainty analysis
Variogram parameters		variogram parameter uncertainty analysis

In the procedure *classical bootstrap*, the squared differences of the sampled pairs are used in generating equiprobable sample variograms and no assumptions are made about the distribution of the semivariances at a given lag vector $\boldsymbol{h}$. The bootstrap is based on resampling of the data (Cressie, 1991). For vector $\boldsymbol{h}$, it is supposed that the squared differences $\underline{\Delta}^2_1 ... \underline{\Delta}^2_n$ are independent and identically distributed (i.i.d.) random variables with a cumulative distribution function $F(\boldsymbol{h})$. In design-based sampling of pairs of points p-independence is created by the sampling design (Brus and De Gruijter, in prep.) Let us take as an example a simple random sample of pairs. In this design pairs of points separated by $\boldsymbol{h}$ are selected independently and such that the inclusion probability is equal for all such pairs (Brus and De Gruijter, in prep.) The sample size, i.e the numbers of pairs for $\boldsymbol{h}$ will be denoted by $n(\boldsymbol{h})$. For approximation of the sampling distribution of variograms the classical bootstrap then proceeds as follows:

(i) a simple random sample *with replacement* of size $n(\boldsymbol{h})$ is drawn from Δ^2_i, $i = 1$ to $n(\boldsymbol{h})$;
(ii) the sample data are used to estimate the semivariance and its sampling variance for lag $\boldsymbol{h}$;
(iii) steps (i) and (ii) are repeated for the other lags $\boldsymbol{h}$;
(iv) several variograms are fitted to the sample variogram and the best function is selected;
(v) steps (i) to (iv) are repeated B times.

In *parametric sample variogram uncertainty analysis* the squared differences are used too, but type and parameters of the distribution of the squared differences for the lag vectors $\boldsymbol{h}$ are assumed to be known. For example, assuming normality of $\underline{z}$, the squared difference of pairs separated by $\boldsymbol{h}$ follows a $2\gamma(\boldsymbol{h})\chi^2_1$-distribution (Cressie and Hawkins, 1980). In *variogram parameter uncertainty analysis* the estimates of the variogram parameters and of their sampling variance and covariance are used in generating equiprobable variogram functions and these parameters are assumed to be distributed according to a postulated joint density function.

If little can be assumed about the underlying distribution of the variogram parameters, the classical bootstrap is attractive. An advantage of the bootstrap

and the parametric sample variogram uncertainty analysis is that uncertainty about the functional form of the variogram can be taken into account too.

Case study

Study area and data collection

The described procedures to estimate the non-ergodic variogram and the estimation variance at points were tested in a typical coversand area near the village of Wesepe, in the Province of Overijssel, the Netherlands. The area is rectangular, 3 km from east to west and 4 km from north to south and is the southwestern quarter of a 6 x 8 km study area where we compared the performance of several estimation methods (Brus *et al.*, in prep.). Soils were sampled by auger at 415 observation points, 60 pairs of points for variogram estimation and 100 test points (Figs 7.1 and 7.2). The 415 observation points form a square grid of 18 x 24 points, 166.7 m apart, with 17 missing points (build-up area). At each point we measured the maximum areic mass of P_2O_5 which can be sorbed potentially by soil above the mean highest water table (P_{max}; kg m^{-2}). For a definition of P_{max}, see Brus *et al.* (1992). P_{max} is an important property in studies on phosphate leaching which is a considerable environmental problem in the Netherlands.

The soil map shows alternating stream valleys and coversand ridges in a *N* 105° *E* direction. We used the data from the 6 x 8 km area to obtain a prior estimate of the range of an exponential variogram of P_{max} in this direction and the direction perpendicular to it. (Table 7.2). These prior estimates were used to optimize the selection of the lag vector $\boldsymbol{h}_2$ in these two directions. Hereafter we denote the length of $\boldsymbol{h}$ by h. For exponential functions of type:

$$\gamma_A(h) = c_0 + c(1 - e^{-h/r}) \qquad (7.8)$$

an exact locally *D*-optimum design of size n is independent of c_0, c and n and given by (Rasch, 1990):

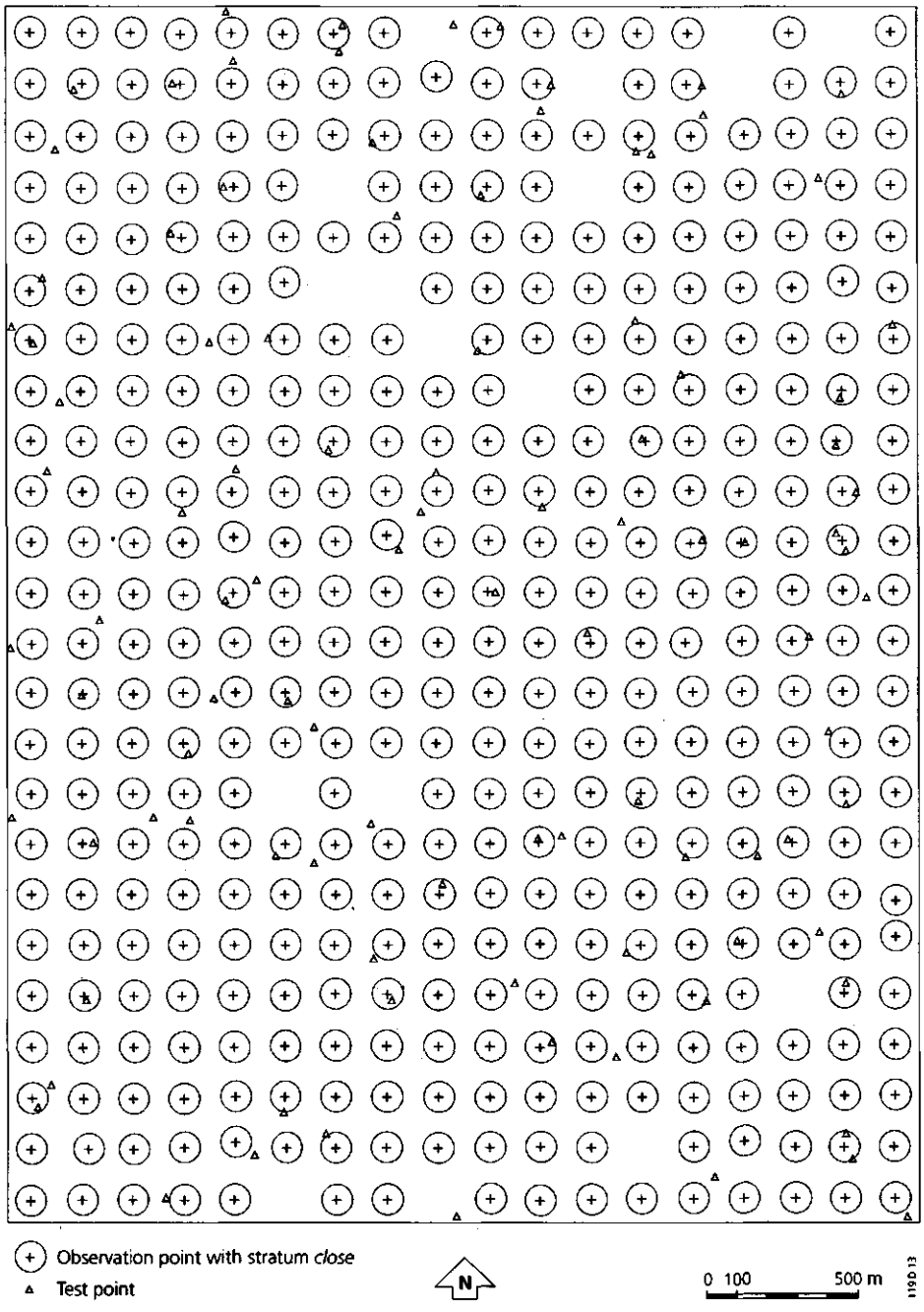


Fig. 7.1 Systematic sample of observation points and stratified simple random sample of test points

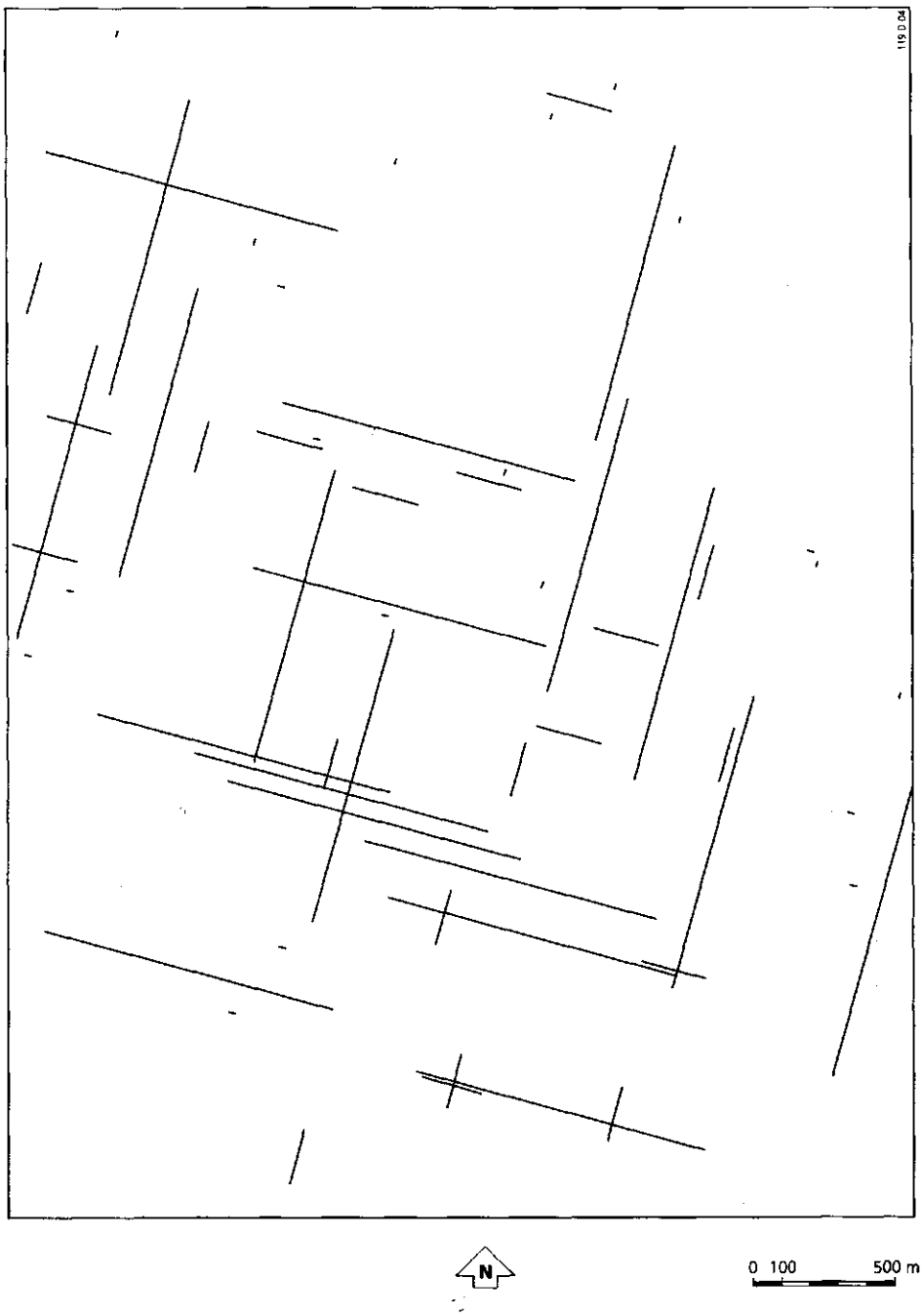


Fig. 7.2 Simple random samples of pairs of points for estimating the non-ergodic variogram. Points of a pair are connected by lines

$$\begin{vmatrix} h_l & h_2 & h_u \\ n/3 & n/3 & n/3 \end{vmatrix} \tag{7.9}$$

where:

h_l = lower boundary of the experimental region;
h_u = upper boundary of the experimental region,

and

$$h_2 = r + \frac{h_l \exp(-h_l/r_0) - h_u \exp(-h_u/r_0)}{\exp(-h_l/r_0) - \exp(-h_u/r_0)} \tag{7.10}$$

where:

r_0 = prior estimate of r.

This design is *D*-optimal under the assumption of a constant residual variance, which is not realistic. However, we believe this design is a useful approximation.

Table 7.2 shows the results. For the lower boundary of the experimental region, h_l, we chose 5 m and for the upper boundary, h_u, 1000 m. These are more or less arbitrary choices. We sampled 60 pairs, which implies (there are two directions and three distances) 10 for each lag. Each set of 10 pairs was selected by simple random sampling.

Table 7.2 Prior estimates of range (r_0) of exponential variogram in two directions and derived optimal lags (h_l, h_2, h_u) in m.

	r_0	h_l	h_2	h_u
N 15° *E*	184	5	185	1000
N 105° *E*	228	5	220	1000

Estimation of non-ergodic variogram

As the inclusion probabilities of all pairs of points, given $\boldsymbol{h}$, are equal, the non-ergodic semivariance for lag $\boldsymbol{h}$ can be estimated by the unweighted mean:

$$\hat{\gamma}_A(\boldsymbol{h}) = \frac{1}{2n(\boldsymbol{h})}\sum_{i=1}^{n(\boldsymbol{h})}\{z(\underline{\boldsymbol{x}}_i) - z(\underline{\boldsymbol{x}}_i+\boldsymbol{h})\}^2 \qquad (7.11)$$

and its sampling variance, $\hat{\sigma}_p^2\{\hat{\gamma}(\boldsymbol{h})\}$, can be simply estimated by:

$$\hat{\sigma}_p^2\{\hat{\gamma}_A(\boldsymbol{h})\} = \frac{\hat{v}_A\{0.5\Delta^2(\boldsymbol{h})\}}{n(\boldsymbol{h})} = \frac{0.25\sum_{i=1}^{n(\boldsymbol{h})}[\Delta_i^2(\boldsymbol{h}) - \hat{m}_A\{\Delta^2(\boldsymbol{h})\}]^2}{n(\boldsymbol{h})\{n(\boldsymbol{h})-1\}} \qquad (7.12)$$

where:

$\hat{v}_A\{\}$ = estimated spatial variance of the quantity between brackets within A;

$\Delta_i^2(\boldsymbol{h})$ = squared difference of the ith pair separated by lag $\boldsymbol{h}$;

$\hat{m}_A\{\}$ = estimated spatial mean of the squared differences within A.

The inverse of these sampling variances were used as weights in fitting various functions for the variogram. Eight functions were used, namely all combinations of - exponential and spherical - with and without nugget - isotropic and isotropic. For the anisotropic functions, the range was described by:

$$r(\theta) = \{r_\phi^2\cos^2(\theta-\phi) + r_\perp^2\sin^2(\theta-\phi)\}^{1/2} \qquad (7.13)$$

where:

ϕ = preferential direction;

r_ϕ = range in direction ϕ;

$r_\perp$ = range in direction perpendicular to ϕ;

θ = direction of the lag vector.

The parameter ϕ was fixed to $N\,105^\circ\,E$ (derived from the soil map). We selected the function with the smallest residual mean sums of squares.

Spatial estimation

Values at points were estimated by local ordinary kriging (Journel and Huijbregts, 1978) with a neighbourhood of 20 points. The variogram fitted on the original sample was used to calculate the kriging variance of the estimated P_{max}.

The classical bootstrap procedure was followed to estimate the sampling variance caused by the sampling error of the non-ergodic variogram at test points: from each of the six samples of pairs, 10 pairs were resampled by simple random sampling with replacement. Next the eight functions, described above were fitted and the best function selected similarly. This procedure was repeated 100 times ($B = 100$). Using these 100 variograms in kriging resulted in 100 estimates at each test point. Given a test point, the variance of these estimates, referred to as the bootstrap variance, was used as an estimate of the sampling variance.

Validation

We tested the validity of (i) the kriging variance and (ii) the kriging variance plus the bootstrap variance, as estimates of the estimation variance at points by the spatial mean of squared deviation ratio, $m_A(r)$:

$$m_A(r) = \frac{1}{N} \sum_{i=1}^{N} \frac{\{\hat{z}(\boldsymbol{x}_i) - z(\boldsymbol{x}_i)\}^2}{\hat{\sigma}_i^2} \qquad (7.14)$$

where:

N = total number of soil profiles in area A (population size);
$\hat{z}(\boldsymbol{x}_i)$ = estimated value of z at location $\boldsymbol{x}_i$;
$z(\boldsymbol{x}_i)$ = true value of z at location $\boldsymbol{x}_i$;
$\hat{\sigma}^2_i$ = estimated estimation variance at location $\boldsymbol{x}_i$.

Ideally, the spatial mean of squared deviation ratio equals 1. We estimated $m_A(r)$ by a stratified simple random sample of test points. At these test points we measured P_{max}, so the estimation error at these points is known. We used the distance to the observation points and soil map unit as stratification criteria. On the basis of the distance we distinguished test points close to the nearest

observation point (≤ 50 m) and further away from the nearest observation point (> 50 m). It is well known that the estimation error is related to the distance to the observation points. The same holds for the soil map units. The spatial variance of soil properties differs considerably between the units of the Soil Map of The Netherlands at a scale of 1:50 000 (Marsman and De Gruijter, 1986, Brus *et. al.* 1992). Given the sample size and the configuration of observation points, the mean estimation error in heterogeneous map units will generally be relatively large. The test points were allocated approximately proportionally to the area of the strata with a minimum of three points per stratum.

For a stratified simple random sample, the spatial mean of squared deviation ratio, $m_A(r)$, can be estimated by (Cochran, 1977 p. 91):

$$\underline{\hat{m}}_A(r) = \sum_{h=1}^{H} W_h \, \underline{\hat{m}}_h(r) = \sum_{h=1}^{H} W_h \frac{1}{n_h} \sum_{i=1}^{n_h} \underline{r}_{hi} \qquad (7.15)$$

where:

H = number of strata;
W_h = weight of stratum h measured as the relative area;
$\underline{\hat{m}}_h(r)$ = estimated spatial mean of squared deviation ratio of stratum h;
n_h = number of test points (sample size) of stratum h;
$\underline{r}_{hi}$ = squared deviation ratio of the ith point in stratum h.

The sampling variance of this estimate can be estimated by (Cochran, 1977 p. 92):

$$\underline{\hat{\sigma}}^2_p(\underline{\hat{m}}_A) = \sum_{h=1}^{H} W_h^2 \, \underline{\hat{\sigma}}^2_p(\underline{\hat{m}}_h) = \sum_{h=1}^{H} W_h^2 \, \frac{\underline{\hat{V}}_h}{n_h} (1 - f_h) \qquad (7.16)$$

where:

$\underline{\hat{\sigma}}^2_p(\underline{\hat{m}}_h)$ = estimated sampling variance of the sample mean of stratum h under sampling design p;
$\underline{\hat{V}}_h$ = estimated spatial variance of stratum h;
f_h = sampling fraction of stratum h.

The sampling fractions are close to zero for all strata so the last term of Equation (7.16) can be dropped.

Results and discussion

Figure 7.3 shows the non-ergodic sample variogram with the estimated semi-variances at the six lags and their 90% confidence intervals. For the 5 m lags, these intervals were rather narrow, however for the larger lags, they were very wide. An isotropic spherical variogram with nugget fitted best to this sample variogram with the following estimated parameter values: nugget = 0.1625 kg^2 m^{-4}, range = 438 m, sill = 2.135 kg^2 m^{-4} (Fig. 7.3). This variogram was used to estimate the kriging variance at the 100 test points.

Figure 7.4 shows a histogram of this kriging variance. It is well known that this kriging variance is completely determined by the variogram and the configuration of observation points; the values at the observation points are of no influence. This explains the relatively large kriging variances of the test points in the stratum 'distant from observation points'.

Figure 7.5 shows the first 25 variograms in direction *N* 105° *E* obtained by the

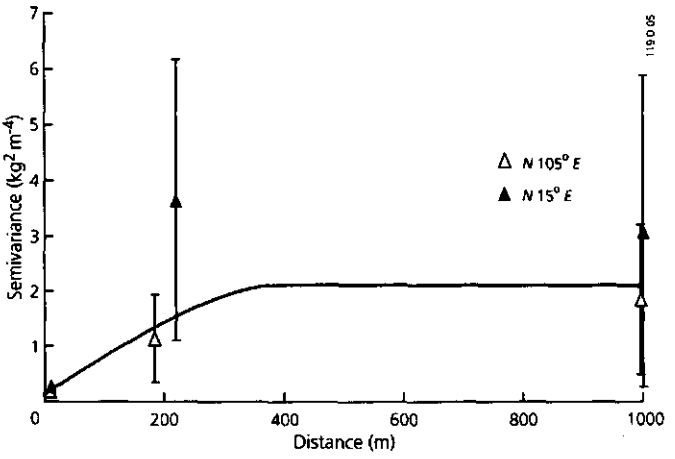


Fig. 7.3 Non-ergodic sample variogram with 90% confidence intervals and fitted model (spherical with nugget, isotropic)

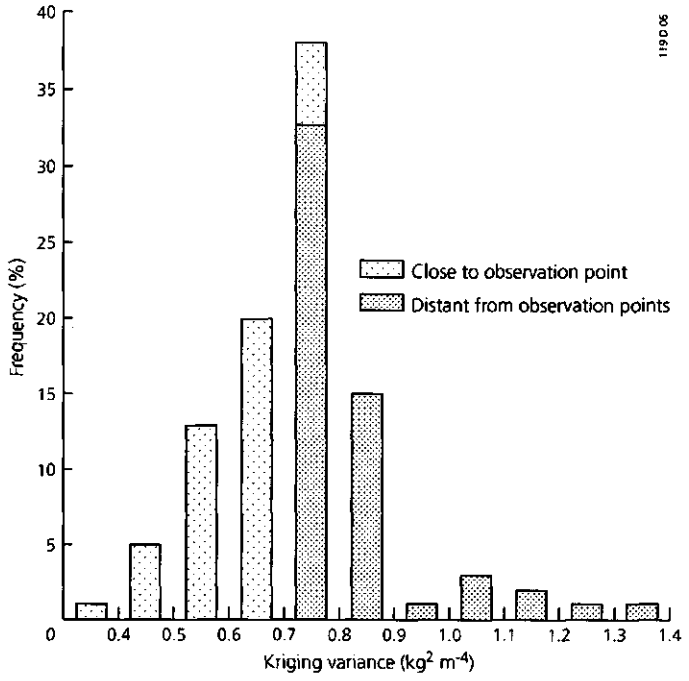


Fig. 7.4 Histogram of kriging variance at 100 testpoints

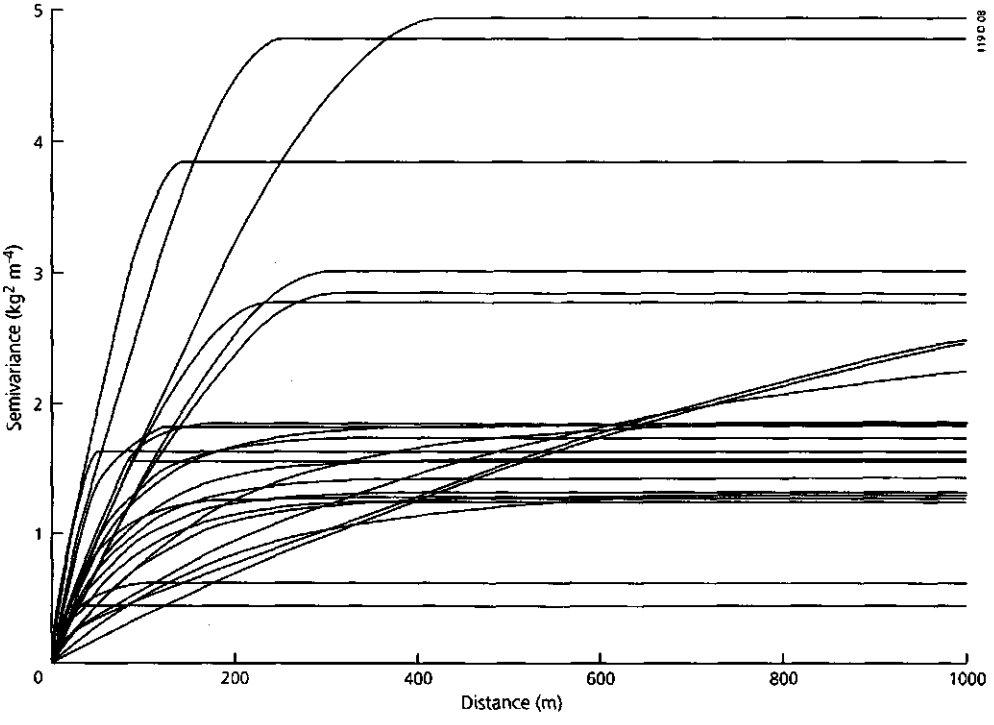


Fig. 7.5 The first 25 variograms obtained by the bootstrap (direction *N* 105° *E*)

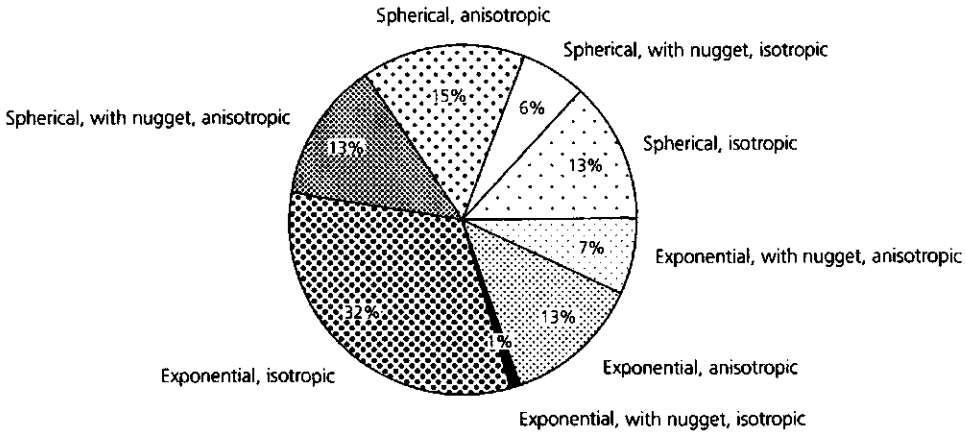


Fig. 7.6 Frequency of eight variogram model types

bootstrap procedure. The variograms differed considerably, especially with respect to the sill and the range. This can be explained by the wide confidence intervals for the largest lags.

Figure 7.6 shows the frequency of the function types. No function type was clearly dominant: there were about as many anisotropic as isotropic variograms and about as many spherical as exponential variograms. Variograms without nugget term slightly outnumbered those with nugget term.

A histogram of the variance of the kriging estimates at the 100 test points, using these variograms is shown in Figure 7.7. On average, the bootstrap variance was somewhat smaller than the kriging variance, but evidently not negligible. The distribution of the bootstrap variance was positively skewed. There was no clear relation between the bootstrap variance and the distance to the nearest observation point: test points in the stratum 'distant from observation points' covered the whole range of the histogram. In contrast to the kriging variance, the bootstrap variance was partly determined by the values at the observation point: large values corresponded with outliers of P_{max} at nearby

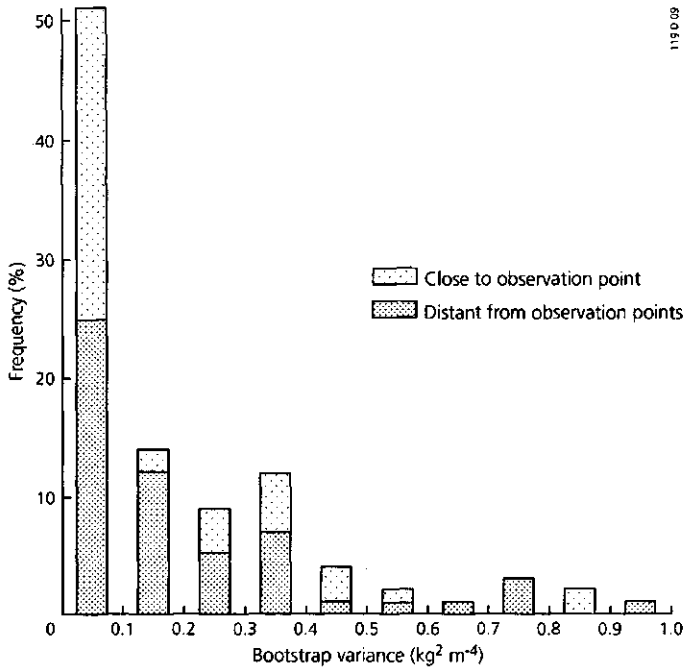


Fig. 7.7 Histogram of bootstrap variance (estimated sampling variance) at 100 testpoints

observation points.

Using the kriging variance as an estimate of the estimation variance at points, the estimated spatial mean of squared deviation ratio was 1.43 (standard error: 0.25). This means that the kriging variance underestimated the estimation variance at points. For the kriging variance plus the bootstrap variance, the estimated spatial mean of squared deviation equaled 1.06 (standard error: 0.19). This means that on average the sampling variance as measured by the bootstrap variance was approximately one third of the kriging variance.

General discussion and conclusions

The role of the model in model-assisted estimation procedures is rather different from that in model-based procedures. In model-based procedures it

describes a process by which the data have been generated, whereas in model-assisted procedures it describes the finite population itself. In model-based procedures it is assumed that the population was really generated by the model ξ and inference is based on this assumption. In model-assisted procedures we hope that the model ξ describes the finite population reasonably well, but inference is not conditioned on this. In variance estimators the sampling error of the estimated model is taken into account. For spatial estimation by kriging this implies that the uncertainty about the variogram is taken into account when calculating the variance of kriging estimates.

Probability samples of pairs of points in which pairs are selected independently, offer strong potentialities for incorporating the sampling error of the non-ergodic variogram in the variance of kriging estimates. For samples like these, the bootstrap is a suitable procedure. In a case study, the sum of the sampling variance and the kriging variance was on average an unbiased estimate of the variance of the error at test points. With 60 pairs (120 points), the sampling variance contributed considerably to the total estimation variance.

References

Bregt, A.K., McBratney, A,B. & Wopereis, M.C.S. 1991. Construction of isolinear maps of soil attributes with empirical confidence limits. *Soil Science Society of America Journal* 55, 14-19.

Brus, D.J. & De Gruijter, J.J. in prep. Estimation of non-ergodic variograms by design-based sampling strategies. *Mathematical Geology* (submitted).

Brus, D.J., De Gruijter, J.J. & Breeuwsma, A. 1992. Strategies for updating soil survey information: a case study to estimate phosphate sorption characteristics. *Journal of Soil Science* 43, 567-581.

Brus, D.J., De Gruijter, J.J., Marsman, B.A., Visschers, R., Bregt, A., Breeuwsma, A. & Bouma, J., in prep. Stratification by soil map units to improve estimates of spatially varying soil properties at points. *Journal of Soil Science* (submitted).

Cochran, W.G. 1977. *Sampling techniques.* Wiley, New York.

Cressie, N.A.C. 1991. *Statistics for spatial data.* Wiley, New York.

Cressie, N. & Hawkins, D.M. 1980. Robust estimation of the variogram: I. *Mathematical Geology* 12, 115-125.

De Gruijter, J.J. & Ter Braak, C.J.F. 1990. Model-free estimation from spatial samples: a reappraisal of classical sampling theory. *Mathematical Geology* 22, 407-415.

Gascuel-Odoux, C. & Boivin, P. 1993. Variability of variograms and spatial estimates due to soil sampling: a case study. *Geoderma* (in press).

Hansen, M.H., Madow, W.G. & Tepping, B.J. 1983. An evaluation of model-dependent and probability-sampling inferences in sample surveys. *Journal of the American Statistical Association* 78, 776-807.

Isaaks, E.H. & Srivastava, R.M. 1988. Spatial continuity measures for probabilistic and deterministic geostatistics. *Mathematical Geology* 20, 313-341.

Journel, A.G. 1985. The deterministic side of geostatistics. *Mathematical Geology* 17, 1-15.

Journel, A.G. & Huijbregts, C.J. 1978. *Mining Geostatistics.* Academic Press, London.

Laslett, G.M., McBratney, A.B., Pahl, P.J. & Hutchinson, M.F. 1987. Comparison of several spatial prediction methods for soil pH. *Journal of Soil Science* 38, 325-341.

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.

Rasch, D. 1990. Optimum experimental design in nonlinear regression. *Commun. Statist. Theory Meth.* 19, 4789-4806.

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.

Srivastava, R.M., 1987. *A non-ergodic framework for variograms and covariance functions.* M. Sc. Thesis, Stanford University, Stanford.

Stein, A., Hoogerwerf, M. & Bouma, J. 1988. Use of soil-map delineations to improve (co-)kriging of point data on moisture deficits. *Geoderma* 43, 163-177.

Chapter 8

Major conclusions

Major conclusions

This chapter describes the main results of this research and draws some major conclusions. First, I will answer questions (i) to (v) in Chapter 1 (p. 4-5). The answer to the question as to what decision rules can be used for choosing between approaches for incorporation of spatial variation models (question vi) is presented in the form of a decision tree. This decision tree is based on all the chapters and therefore can be seen as a synthesis of this thesis. Finally, I will present some ideas for further research.

Main results and conclusions

Question (i): is design-based inference invalid in areas with autocorrelated data?
Design-based inference is always valid, also in areas with autocorrelated data, i.e. dependent data in the model-based sense. These data can be at the same time independent in the design-based sense (Chapter 2).

Question (ii): are kriging estimates also optimal in the sense of the classical sampling theory, and what is the worth of the calculated kriging variance?
Optimal estimates in the model-based sense are not necessarily optimal in the design-based sense. Model-unbiased (ξ-unbiased) estimates can simultaneously be design-biased (p-biased). Kriging estimates of spatial means and values at points will generally be p-biased. Moreover, a small model variance (ξ-variance) of the kriging estimator does not necessarily imply a small sampling variance (p-variance) of this estimator in the area sampled. For reasonably large samples a small ξ-variance of the estimated spatial mean implies that the sampling variance of this estimator averaged over all model realizations, will be small too: $E_p\{\sigma_\xi^2(\hat{\underline{m}}_A)\} = E_\xi\{\sigma_p^2(\hat{\underline{m}}_A)\}$ (Chapter 2). The same holds for estimates at points (Chapter 4). If the realization in the area sampled is more erratic than average, the kriging variance will underestimate the true estimation variance; if it is more erratic, it will be an overestimate.

The kriging variance generally is an underestimate of the true estimation vari-

ance because the uncertainty of the model (ergodic variogram) is not incorporated in it (Chapter 7).

The uncertainty of ergodic variogram estimators is difficult to assess because the uncertainty is caused by fluctuation variance (variance between model realizations) besides sampling variance. This fluctuation variance is unknown because the model is unknown. Therefore, I prefer non-ergodic variograms: the uncertainty of their estimators can be assessed relatively simply because the only source of variation is sampling. Design-based sampling strategies, selecting pairs of points at random and independently from each other, are most suitable for obtaining robust and p-unbiased estimates of the sampling variances and covariances of the variogram parameters (Chapter 6).

Probability samples of independently selected pairs of points are also particularly apt for assessing the error component in kriging estimates accruing from uncertainty about the non-ergodic variogram. The bootstrap is a convenient procedure for obtaining an estimate of the variance of this error component, because it is a model-free procedure in which uncertainty about variogram parameters *and* functional type of variogram can be taken into account (Chapter 7).

The underestimation of the mean squared estimation error by the kriging variance was eliminated by adding the bootstrap variance accounting for uncertainty about the non-ergodic variogram, to the kriging variance (Chapter 7).

Model-assisted spatial estimation by kriging with non-ergodic variograms is a promising alternative to model-based spatial estimation by kriging with ergodic variograms because:

- the estimation variance estimates are always valid, i.e. they do not depend on the correctness of the model;
- inference is conditioned on the realization of the ergodic model actually sampled; the only source of stochasticity is sampling (Chapter 7).

Question (iii): can estimates of global means (Chapter 3) and of values at points (Chapter 4) be improved by soil map stratification, and how strong is this effect? How strong is the effect of estimation with variograms on the accuracy of estimates at points? (Chapter 4) How strong is the effect of soil map stratification plus estimation with variograms on the accuracy of estimates at points? (Chapter 4)

Stratification of an area according to the clustered units of the 1:50 000 Soil Map of the Netherlands led to a considerable increase in precision of the design-based estimates of the *global mean* of the maximum areic mass of P_2O_5 sorbed by soil (P_{max}) and of the areic mass of P_2O_5 sorbed by soil (P). For the mean relative mass of phosphate sorbed by soil (P_{rel}) and the areal fraction saturated with phosphate (A_c), there was only gain in the area with small historical phosphate load. For A_c the gain was strongly dependent on the critical value of P_{rel} (Chapter 3).

Estimates *at points* provided by the estimation methods global mean, moving average, nearest neighbour, inverse squared distance, Laplacian smoothing splines and ordinary point kriging, were generally not significantly ($\alpha = 0.10$) improved by 1:50 000 soil map stratification in a case study with a low density of the observations points (1 observation per 25 ha). The only exceptions to this were the estimates of the mean highest water table ($\bar{W}_{min}$) and of the mean lowest water table ($\bar{W}_{max}$) provided by nearest neighbour estimation. This can be explained by the bad performance of this estimation technique without stratification.

For this observation density, kriging estimates of values at points were not significantly more accurate than estimates with the estimated global mean. In other words, in this case there was no significant effect of estimation with variograms on the accuracy of estimates at points. On the other hand, estimates of $\bar{W}_{min}$ and $\bar{W}_{max}$ by stratified kriging with different variograms for the soil map strata, were significantly more accurate than those with the global mean estimator. In other words, for these properties the combined effect of soil map stratification and estimation with variograms was significant ($\alpha = 0.10$). For the thickness of the A1 horizon and the maximum areic mass of P_2O_5 this combined effect was nearly significant. (Chapter 4).

Question (iv): which approach gives the most accurate estimates of values at points: estimation by random sampling stratified by soil map, or kriging? (Chapter 4)
In a regional survey with a low observation density, estimates at points with the means of map units as estimators were slightly more accurate than those provided by unstratified kriging. However, this difference was not significant ($\alpha = 0.10$). In other words, the effect of soil map stratification in design-based estimation was as strong as the effect of estimation with variograms. Soil map stratification in design-based inference has the advantage of *p*-unbiased and robust estimates of the spatial means or the values at points and of their estimation variances (Chapters 4 and 7).

Question (v): does it pay to revise a soil map before using it in spatial estimation?
It generally does not pay to revise a soil map in order to obtain an improved model of spatial variation before using the model in spatial estimation. If revision is followed by random sampling, then the revised model may become profitable for estimation of values at points (Chapter 5).

Question (vi): what decision rules can be used for choosing between the two approaches?
The answer to this question will be presented as a decision tree (Fig. 8.1).

Choice of sampling strategy

In Chapters 2-7 several factors were reviewed, that determine the efficiency of an approach to incorporate a spatial variation model. There are many factors involved and therefore the decision process can become rather complicated. This section deals with these factors more systematically and presents a decision tree to support these decisions. Although the model-assisted approach is very new, the potentials of this procedure are sufficiently clear to include it in the decision tree.

The choice of a sampling strategy (sampling design plus estimator) is governed by (Domburg *et al.*, 1993):

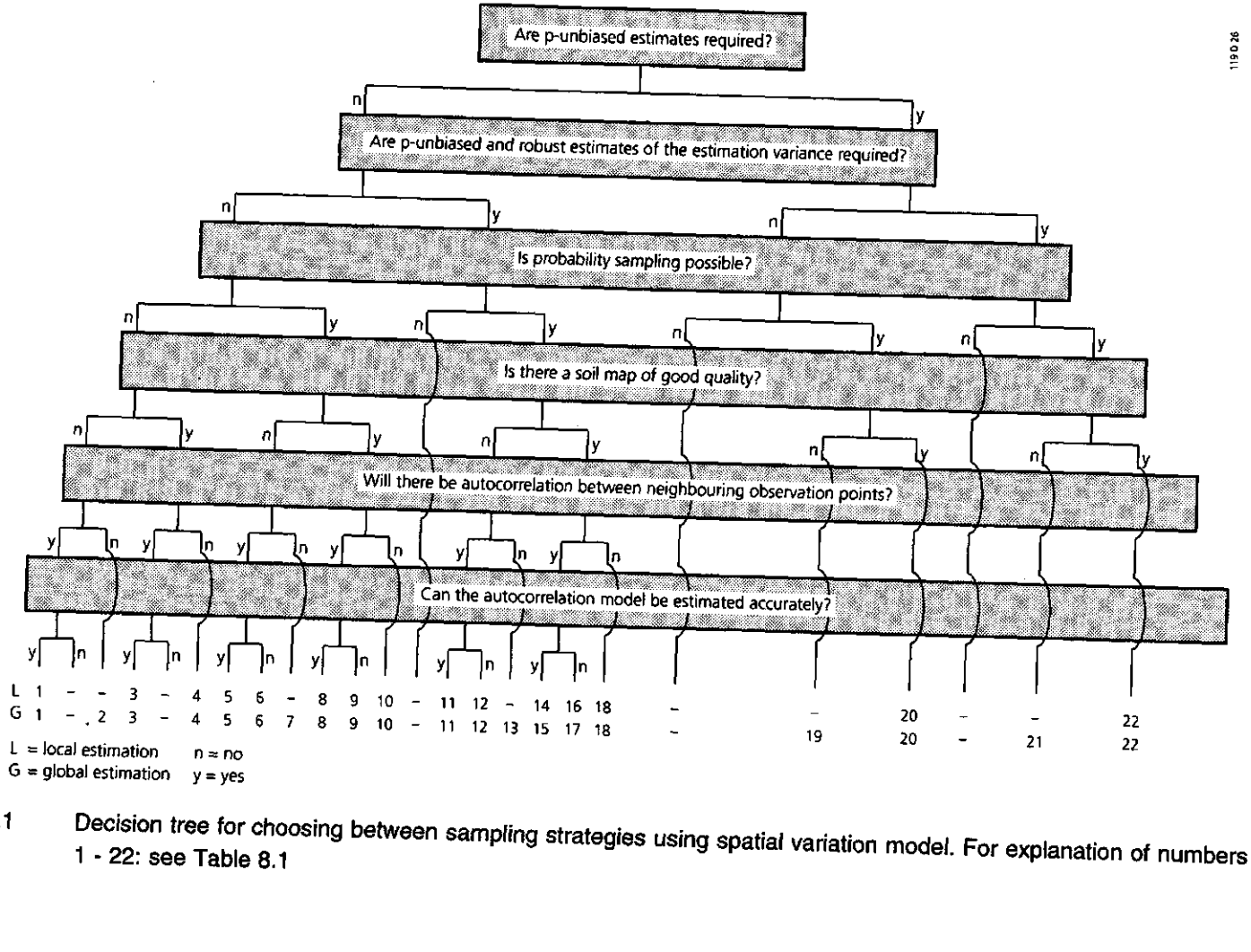


Fig. 8.1 Decision tree for choosing between sampling strategies using spatial variation model. For explanation of numbers 1 - 22: see Table 8.1

Table 8.1 Explanation of sampling strategies (see Fig. 8.1)

No	Sampling	Inference
1	purposive	model-based
-		
2	purposive	sample mean
3	as 1, stratified	model-based
-		
4	as 2, stratified	sample mean
5	purposive: centred, regular grid	model-based
6	observation points: centred, regular grid pairs of points: probability	model-assisted
7	purposive: centred regular grid	sample mean
8	as 5, stratified	model-based
9	as 6, stratified	model-assisted
10	as 7, stratified	sample mean
-		
11	probability	model-assisted
12	probability	model-assisted
13	probability	design-based
14	as 11, stratified	model-assisted
15	as 11, stratified	design-based or model-assisted
16	as 12, stratified	model-assisted
17	as 12, stratified	design-based or model-assisted
18	stratified probability	design-based
19	probability: regular grid with random direction and starting point	design-based
20	as 19, stratified	design-based
21	probability, not systematic	design-based
22	as 21, stratified	design-based

- aim of survey (local or global estimation; criterion for assessment of quality of estimates);
- constraints (accessibility; available budget and sampling costs);
- prior information (soil map).

Figure 8.1 shows a decision tree consisting of seven questions related to the three factors mentioned above. Obviously more questions are relevant in this context, however I included only the questions dealt with in this thesis. I will briefly comment on these questions.

Is it the aim to estimate locally?
The aim of the survey strongly influences the choice of the sampling strategy. If the aim is to estimate the values at points or the means of many blocks, i.e. local estimation, model-based or model-assisted inference is preferable to design-based inference because:
- estimates will be more accurate in certain circumstances;
- for estimating at points, model-based and model-assisted estimates of the estimation variance vary continuously in space, which is generally more realistic than the discontinuous variation of the design-based estimation variance.

To keep the decision tree as simple as possible, this question is incorporated in the base of the tree.

Are p-unbiased estimates required?
The answers to this and following question further specify the aim of the survey. The meaning of *p*-unbiasedness in local estimation differs from that in global estimation (see page 104): for estimation of values at points these estimates are *p*-unbiased if the spatial mean of the *p*-expected errors equals 0. For global estimation each estimate of a block-mean should be *p*-unbiased separately. If this question is answered with yes, design-based strategies are the only option. For model-based and model-assisted inference, estimates will not be *p*-unbiased.

Are p-unbiased and robust estimates of the estimation variance required?
If *p*-unbiased and robust estimates of the accuracy of the estimates are required, probability sampling combined with design-based or model-assisted inference is preferable because:

- the design-based approach is model-free;
- uncertainty about the model is taken into account in model-assisted inference of the estimation variance.

If *p*-unbiased and robust estimates of the estimation variance are not required, purposive sampling, for example by a centred, regular grid generally is superior to probability sampling because its estimation variance is generally smaller than that of a non-centred grid of equal configuration (strategies 5 - 10). In addition estimation variance using model-based or model-assisted inference is generally smaller than with design-based inference if there is autocorrelation between neighbouring observation points (strategies 5, 6, 8, 9).

Is probability sampling possible?
If probability sampling is impossible, for example due to inaccessibility of large parts of the survey region, purposive sampling combined with model-based inference is the only alternative. In this case there is no suitable strategy if *p*-unbiased estimates are required or if *p*-unbiased and robust estimates of the estimation variance are required.

Is there a soil map of good quality?
If there is a soil map of the survey region and the map units are more homogeneous for the property under study than the region as a whole, the efficiency of sampling strategies can be increased by incorporating the map in the sampling design, for example by stratified random sampling or stratified purposive sampling. This stratification should be taken into account in the inference of estimates, also in model-based and model-assisted inference (stratified kriging).

Will there be autocorrelation between neighbouring observation points?
If the available budget is small and the sampling costs are high, the sampling density will be low. This density should be related to the range of the variogram. For relatively low densities there will be no or only little autocorrelation between neighbouring observation points and no gain in accuracy can be ex-

pected by using an autocorrelation model or variogram in spatial estimation. In this case, design-based inference is preferable (strategies 7, 13, 18). If there is no autocorrelation and no soil map of good quality local estimates will be very inaccurate. In this case the aim of the survey and the constraints (available budget) do not match.

Can the autocorrelation model be estimated accurately?
The presence of autocorrelation is one thing, an accurate model describing this autocorrelation is another. If the available budget is small and the sampling costs per observation point are high, the sampling design will be unsuitable for estimating the model (ergodic or non-ergodic variogram) accurately. Consequently, the estimated model may strongly deviate from the true underlying model, probably leading to a large estimation error. Also, the sampling error of the model (variogram) will be large. In model-based inference, the sampling error of the ergodic variogram is not incorporated in the kriging error and as a result the kriging error will strongly underestimate the estimation variance. Therefore, in this case I recommend model-assisted inference because it incorporates the uncertainty about the non-ergodic variogram in the kriging error (strategies 6, 9, 12, 16, 17).

Further research

This thesis compares the efficiency of estimation by random sampling, stratified by soil map with that of kriging for surveys with low observation density. This comparison has to be extended to higher densities. To compare the efficiency of the two methods for estimating spatial means, sampling from simulated fields is a suitable procedure, because the true means are then known.

Furthermore, this thesis presents only the first steps towards a new approach for incorporating a spatial variation model in spatial estimation: the model-assisted approach. It avoids the main drawback of the model-based approach, namely the unreliability of the estimation variance estimates. Much work has still to be done on this:

- the variance of model-assisted estimates was obtained by the bootstrap,

which is a rather laborious procedure. Variance estimators similar to those of the regression estimator with estimated regression coefficients, need to be derived;

- model-assisted spatial estimation was conditioned on the sample of observation points used in kriging. The variance estimate only accounts for the randomness of spatial variation and for the uncertainty about the variogram. Procedures and estimators also accounting for sampling variance of observation points have to be developed;
- the efficiency of sampling designs for estimation of the non-ergodic variogram have to be compared;
- I used separate and independent samples for estimation of the non-ergodic variogram and for kriging. This might not be very efficient and it may be advantageous to use the sample for estimation of the variogram in kriging too. However, robust estimation of the estimation variance might then become troublesome. In short, efficient and robust sampling strategies for simultaneous estimation of variogram and for kriging need to be developed in future.

References

Domburg, P., De Gruijter, J.J. & Brus, D.J. 1993. A structured approach to designing soil survey schemes with prediction of sampling error from variograms. *Geoderma* (in press).

Samenvatting

Ruimtelijke-variatiemodellen in bemonsteringsstrategieën voor de bodem

De efficiëntie van bemonsteringsstrategieën voor de bodem kan worden vergroot door gebruik te maken van voorinformatie over de ruimtelijke variatie van de bodemkenmerken. Deze ruimtelijke variatie kan worden beschreven met een model. Op het hoogste niveau kunnen twee benaderingen worden onderscheiden die van elkaar verschillen in de wijze waarop dergelijke modellen worden gebruikt:

(i) bij het selecteren van bemonsteringslocaties volgens een bepaalde steekproefopzet;
(ii) bij het berekenen van schattingen.

Het hoofddoel van dit proefschrift is het analyseren van de voor- en nadelen van beide benaderingen vanuit theoretisch en praktisch oogpunt. Schatting door middel van een aselecte steekproef, waarbij de bodemkaart wordt gebruikt voor stratificatie van het onderzoeksgebied, diende als voorbeeld voor de eerste benadering. In deze benadering is de bodemkaart het model van de ruimtelijke variatie. De bemonsteringspunten worden binnen de kaarteenheden aselect gekozen. Wanneer een bodemkenmerk binnen de kaarteenheden minder sterk varieert dan in het totale gebied, dan zal over het algemeen een dergelijke gestratificeerde steekproefopzet een nauwkeurigere schatting van het gebiedsgemiddelde opleveren dan een enkelvoudige aselecte steekproefopzet.

De interpolatietechniek kriging diende als voorbeeld voor de tweede benadering. Sinds 1980 staat deze interpolatietechniek bij bodemkundigen sterk in de belangstelling. In tegenstelling tot de vorige benadering wordt in deze benadering bij het selecteren van de bemonsteringspunten geen gebruik gemaakt van een bodemkaart. Vaak wordt bemonsterd volgens een regelmatig grid. Het model van de ruimtelijke variatie wordt gebruikt bij het berekenen van de schatting. Dit model lijkt in geen enkel opzicht op een bodemkaart. Het beschrijft de ruimtelijke variatie van het bodemkenmerk door middel van vergelijkingen. Een belangrijk onderdeel van dit stochastische model is de autocorrelatiefunctie of semivariogram. Een autocorrelatiefunctie is een functie die de correlatie van

een kenmerk, gemeten op twee punten, beschrijft als functie van de afstand en richting tussen deze twee punten.

De keuze tussen de twee benaderingen is in de eerste plaats een keuze tussen de klassieke-steekproefbenadering en de geostatistische benadering. In de geostatistische literatuur komt men echter veelvuldig de opvatting tegen dat de klassieke-steekproefbenadering niet geldig is wanneer de gegevens autocorrelatie vertonen. Ook wordt verondersteld dat de geostatistische benadering optimaal is. Als deze veronderstellingen juist zijn, dan bestaat er geen keuzeprobleem. In dit proefschrift laat ik zien dat de opvatting over het niet-valide zijn van de klassieke-steekproefbenadering op een misverstand berust. Ook maak ik duidelijk dat optimaliteit in de geostatistische benadering een andere betekenis heeft dan in de klassieke-steekproefbenadering. De twee benaderingen gebruiken namelijk verschillende criteria om de kwaliteit van de schattingen te bepalen. Een optimale schatting in de ene betekenis hoeft niet optimaal te zijn in de andere betekenis. Conclusie: er is wèl een keuzeprobleem.

De efficiëntie van de bodemkaart en het semivariogram als modellen van de ruimtelijke variatie zijn onderzocht voor het schatten van waarden op niet-bezochte punten en voor het schatten van gebiedsgemiddelden. Dit is gedaan in vier dekzandgebieden, variërend in grootte van 1233 ha tot 4814 ha, met een steekproefdichtheid van één boring per 20 à 25 ha. In deze gebieden zijn met name bodemkenmerken gemeten die relevant zijn voor het onderzoek naar fosfaatuitspoeling op landbouwgronden.

De nauwkeurigheid van de geschatte *gebiedsgemiddelden* van het fosfaatbindend vermogen (*FBV*) en het huidige fosfaatgehalte van de bodem kon aanzienlijk worden vergroot door de Bodemkaart van Nederland, schaal 1:50 000 te gebruiken voor stratificatie in een aselecte steekproef. Voor de gemiddelde fosfaatverzadigingsgraad en de fosfaatverzadigde oppervlakte leverde deze stratificatie alleen winst op in het gebied met een lage fosfaatbelasting. Voor de fosfaatverzadigde oppervlakte was de winst bovendien sterk afhankelijk van de kritische waarde van de fosfaatverzadigingsgraad.

In een studie met één boring per 25 ha werd de nauwkeurigheid van de ge-

schatte dikte van de A1-horizont (*d*(A1)), gemiddeld hoogste grondwaterstand (*GHG*), gemiddeld laagste grondwaterstand (*GLG*) en het fosfaatbindend vermogen (*FBV*), *op niet-bezochte punten* niet significant groter door het gebied eerst te stratificateren met de 1:50 000 Bodemkaart van Nederland. Dit gold voor alle onderzochte interpolatietechnieken: gebiedsgemiddelde, lokaal gemiddelde, Thiessen polygonen, inverse van gekwadrateerde afstand, Laplacian smoothing splines en kriging. De enige uitzonderingen hierop waren schattingen van de *GHG* en *GLG* door middel van Thiessen polygonen, wat verklaard kan worden door de zeer slechte schattingen met deze techniek zonder stratificatie.

Voor deze waarnemingsdichtheid waren de schattingen verkregen met kriging niet nauwkeuriger dan die met het gebiedsgemiddelde. Daarentegen leverde kriging binnen drie groepen van kaarteenheden wel nauwkeurigere schattingen op dan het gebiedsgemiddelde. Schattingen van *d*(A1), *GHG*, *GLG* en *FBV* op punten door een aselecte steekproef, gestratificeerd met behulp van een bodemkaart waren iets nauwkeuriger (maar niet significant bij $\alpha = 0.10$) dan die met kriging.

Bodemkaarten zijn soms verouderd, bijvoorbeeld door een daling van de grondwaterstand. Sommigen stellen daarom voor om de bodemkaart eerst te reviseren en daarna de gereviseerde bodemkaart te gebruiken als model van de ruimtelijke variatie bij ruimtelijke interpolatie. Uit dit proefschrift blijkt dat deze aanpak over het algemeen niet efficiënt is. Het is efficiënter om de voor revisie benodigde middelen te gebruiken voor het bemonsteren van aselect gekozen punten. De metingen op deze punten kunnen dan worden gebruikt voor het schatten van waarden op niet-bezochte punten of van gebiedsgemiddelden, waarbij de bestaande bodemkaart wordt gebruikt als model van de ruimtelijke variatie. Revisie van de bodemkaart is alleen efficiënt wanneer ook een aselecte steekproef wordt genomen en deze gegevens worden gebruikt voor het schatten van waarden op niet-bezochte punten, gebruikmakend van de gereviseerde bodemkaart. Dit geldt alleen voor bodemkenmerken die sterk gerelateerd zijn aan de bodemkaarteenheden zoals bijvoorbeeld de gemiddeld hoogste grondwaterstand.

Kriging levert behalve geschatte waarden op punten of geschatte gebiedsgemiddelden ook een schatting van de nauwkeurigheid van deze schattingen op: de kriging-variantie. Deze kriging-variantie is echter over het algemeen een onderschatting van de werkelijke variantie van de schattingsfout. Een belangrijke oorzaak hiervan is de verwaarlozing van de onzekerheid over het variogram bij de bepaling van de schattingsfout. In dit proefschrift laat ik zien dat dit probleem opgelost kan worden door te interpoleren met een niet-ergodisch variogram in plaats van met een ergodisch variogram. Een niet-ergodisch variogram is gedefiniëerd voor één realisatie van het stochastische model van de ruimtelijke variatie, namelijk die in het onderzochte gebied. Ergodische variogrammen zijn gedefiniëerd voor alle realisaties van dit model.

In dit proefschrift wordt een set van nieuwe methoden gepresenteerd die zeer goede mogelijkheden biedt voor zuivere en robuuste schatting van de steekproeffout van het niet-ergodische variogram en voor het opkrikken van de kriging-variantie met de variantie van deze schattingsfout. Op deze wijze kon de sterke onderschatting van de schattingsvariantie geheel worden geëlimineerd. De resultaten zijn de eerste stappen naar een geheel nieuwe benadering in de ruimtelijke statistiek, de model-ondersteunde benadering. In deze benadering speelt het model een wezenlijk andere rol als in de geostatistische benadering. In de geostatistische benadering beschrijft het model een proces dat de gegevens in het studiegebied heeft gegenereerd, terwijl in de model-ondersteunde benadering het model het studiegebied zelf beschrijft. In de geostatistische benadering wordt verondersteld dat de populatie werkelijk door dit proces is gegenereerd en de schattingen worden gebaseerd (geconditioneerd) op deze veronderstelling. In de model-ondersteunde benadering echter wordt verondersteld dat het model de populatie redelijk goed beschrijft, maar de schattingen worden hier niet op geconditioneerd. In variantie-schatters wordt de onzekerheid over het model immers meegenomen.

De efficiëntie van de klassieke-steekproefbenadering, de geostatistische benadering en de model-ondersteunde benadering wordt bepaald door een groot aantal factoren. De keuze tussen de drie benaderingen kan worden ondersteund met de beslisboom van Hoofdstuk 8. In deze beslisboom zijn acht vragen opgenomen die gaan over het doel van de inventarisatie, de criteria voor

de bepaling van de kwaliteit van de schattingen, de mogelijkheden voor aselecte bemonstering, de aanwezigheid van voorinformatie in de vorm van een bodemkaart, verwachte autocorrelatie tussen dichtbij elkaar gelegen bemonsteringspunten en de kwaliteit van het autocorrelatie-model.

Curriculum vitae

Derk Jan Brus werd geboren op 2 mei 1956 te Breedenbroek. In 1974 behaalde hij het diploma Atheneum B aan de Christelijke Scholengemeenschap te Aalten. Daarna studeerde hij Geologie aan de Vrije Universiteit te Amsterdam. In 1977, na het behalen van het kandidaatsdiploma Geologie, begon hij de studie Fysische Geografie. Hierin studeerde hij in 1982 af met als hoofdvak laaglandgenese en als bijvak bodemkunde (Landbouwhogeschool Wageningen). In 1983 kwam hij in dienst bij De Stichting voor Bodemkartering (Ministerie van Landbouw en Visserij) op de afdeling Geomorfologie, Geologie en Pollenanalyse. Hier heeft hij gewerkt aan het deel Bodem van de Atlas van Nederland (1987), en aan de Geomorfologische kaart van Nederland 1:50 000. Op zijn initiatief werd begonnen met het schrijven van toelichtingen bij de 1:50 000 kaarten. De eerste drie toelichtingen zijn door hem geschreven.

In 1986 kreeg hij de functie van onderzoeker bodemclassificatie en bodemgeografie bij het Stafbureau van de Hoofdafdeling Karteringen. Tijdens deze periode werd bij hem door Ben Marsman en Jaap de Gruijter de belangstelling gewekt voor toepassing van kwantitatieve (statistische) methoden in de bodemgeografie en bodemclassificatie.

Na het opgaan van de Stichting voor Bodemkartering in DLO-Staring Centrum, Instituut voor Onderzoek van het Landelijk Gebied in 1989, wordt hij senior wetenschappelijk onderzoeker bij de nieuwe afdeling Landinventarisatiemethoden. In zijn onderzoek ligt de nadruk op de ontwikkeling en toepassing van bemonsteringsstrategieën en interpolatiemethoden.