Spatial Interpolation

A. STEIN

Spatial Interpolation

Ce qui nous plait mieux dans toute la nature,
ce n'est pas ce qu'on voit, c'est ce qu'on se figure.
Jaques Delille, *L'imagination IV*

We have to live with a certain uncertainty.
Karl Popper.

Aan Anita, Judith en Olga
Aan mijn ouders

STELLINGEN

I

Gebruik van statistische interpolatiemethoden leidt tot een optimale strategie om ruimtelijke gegevens te verzamelen, en tot een efficiënt gebruik van beschikbare gegevens bij het vervaardigen van een grafische weergave en de daaraan gekoppelde onzekerheden.
(Dit proefschrift).

II

Het vermijden van Lagrange-multiplicatoren (o.a. [Matheron, 1971]) bij de afleiding en de presentatie van de ruimtelijke interpolatiemethoden kriging en cokriging is verhelderend en ter zake.

> Matheron, G. 1971. The theory of regionalized variables and its applications. Les Cahiers du Centre de Morphologie Mathématique, No. 5. Paris: Ecole de Mines de Paris, 211p.

(Dit proefschrift).

III

Een genesteld bemonsteringsschema (o.a. [Webster, 1985]) is verre van optimaal bij het bepalen van het ruimtelijke semivariogram. Het gebruik ervan wordt ontraden als men ruimtelijke interpolatie wil uitvoeren.

> Webster, R. 1985. Quantitative spatial analysis of soil in the field. In: B.A. Stewart (ed.), Advances in soil science 3. Springer Verlag, New York, pp. 1-70.

(Dit proefschrift)

IV

Het gebruik van een semivariogram om in de aanwezigheid van trends de ruimtelijke structuur van de residuën na aanpassen en aftrekken van deze trend te modelleren kan leiden tot totaal misleidende resultaten. [Webster and Burgess, 1980].

> Webster, R., and T.M. Burgess. 1980. Optimal interpolation and isarithmic mapping of soil properties III. Changing drift and universal kriging. Journal of Soil Science 31, pp. 505-520.

V

Het onderscheid tussen design-based en model-based statistische bemonsteringsprocedures [De Gruijter en Ter Braak, 1990] gaat voorbij aan de formele equivalentie tussen enige semivariogram-waarden en cumulatieve sommen van variantiecomponenten behorende bij een klassiek hiërarchisch bemonsteringsschema.

> De Gruijter, J.J., and C.F.J. ter Braak. 1990. Model-free estimation from spatial samples: a reappraisal of classical sampling theory. Math. Geol. 22, pp. 407-415.

(Dit proefschrift)

VI

Gezien het feit dat een toekomstig stochastisch effect geschat wordt dient het KNMI te spreken over weersvoorspelling i.p.v. weersverwachting.

VII
Verantwoord gebruik van Geografische Informatie Systemen dient uit te stijgen boven het produceren van mooie plaatjes. Pas dan kunnen deze systemen een wezenlijke rol vervullen, zoals bij het zichtbaar maken van beleidsscenario's.

VIII
In veel veldbodemkundige studies is men te snel geneigd op grond van subjectieve ervaringen iets als bewezen te beschouwen voordat een grondige statistische analyse van de verzamelde gegevens heeft plaatsgevonden. Het verdient daarom aanbeveling te streven naar het verzamelen van betrekkelijk eenvoudige, goed te interpreteren kwantitatieve gegevens, dan wel naar goede vertaalfuncties [Bouma, 1989], teneinde een statistische analyse van de gegevens mogelijk te maken.

> Bouma, J. 1989. *Land qualities in space and time*, in *Land qualities in space and time*, edited by J. Bouma and A.K. Bregt, pp. 2-13, Pudoc, Wageningen.

IX
Voor het effectief uitvoeren van wetenschappelijk werk in teamverband is het wezenlijk zo'n team uit verschillende discipline's samen te stellen, waarin de leden verantwoordelijk zijn voor een duidelijk afgebakend en herkenbaar gedeelte van het onderzoek. Goede coördinatie en management is vereist voor de kwaliteit van onderzoeksresultaten, maar dient onderschikt te zijn aan de onderzoeksinspanningen.

X
De term 'Grootschalige kartering' voor kaarten waarvan de schaal relatief groot is, maar die dus betrekking hebben op een relatief klein gebied is weliswaar wiskundig correct maar druist in tegen het algemeen taalgevoel.

XI
Het onderwijs in de statistiek is gebaat met dynamische voorstellingen zoals die met behulp van moderne computerprogramma's mogelijk zijn.

XII
Het doel van de zg. vaderavonden gedurende de zwangerschaps-gymnastiek om de aanstaande vader zich een beetje zwanger te laten voelen mist iedere biologische rechtvaardiging, maar stimuleert het ouderlijke wij-gevoel.

XIII
Omdat Bachs sonaten en partita's voor viool-solo slechts na grondige oefening van 19^e en 20^e eeuwse etudes met enig succes te spelen zijn, en het mede hierom niet erg waarschijnlijk dat Bach deze zelf ooit heeft horen spelen, is een 'authentieke uitvoering' van deze muziek niet mogelijk.

Stellingen behorende bij het proefschrift van A. Stein: 'Spatial Interpolation'.

Wageningen, 12 april 1991.

ABSTRACT

Stein, A., 1991, Spatial interpolation. Doctoral thesis, Agricultural University, Wageningen, The Netherlands.

The theory and practical application of techniques of statistical interpolation are studied in this thesis, and new developments in multivariate spatial interpolation and the design of sampling plans are discussed. Several applications to studies in soil science are presented.

Sampling strategies for collecting spatial data (both the number of observations and their location in the region to be studied) are discussed. It is shown that nested sampling is unsuitable if data are to be collected in an area in order to determine the spatial semivariogram, because semivariogram values for only a few distances are obtained. Furthermore, grid sampling is preferable to nested sampling if spatial interpolation is intended. Sequential sampling is advantegeous if the mean of a variable within an area is to be estimated and there is no spatial correlation. Sequential sampling requires only about 30% of the number of observations required by standard sampling schemes.

In this thesis universal kriging and cokriging (that is kriging and cokriging in the presence of a trend) are formulated in terms of regression procedures. Universal kriging is a special case of universal cokriging. Multivariate increments are extensions of univariate increments and of multivariate stationary variables. Conditions are formulated which permissible polynomial pseudo-covariance and pseudo-cross-covariance functions describing the spatial structure of the variables (or their increments) and their interaction, respectively, have to obey. The coefficients of these functions

5

have been estimated by using the restricted maximum likelihood (REML) method. A practical application of universal cokriging is described.

The application of spatial interpolation in soil science is examined. One of the studies described investigates the problems of scale and the use of soil survey information on moisture deficits caused by groundwater extraction in the Mander area in the Netherlands. In another study the available water and the infiltration rates on terraces of the Allier river in the Limagne area in France are investigated. Cokriging becomes more precise as compared to kriging (and there is a concomitant reduction in costs) if the predictand is strongly correlated with the covariable. This is particularly true if the sampling of the covariable is denser than that of the predictand and the costs of sampling of the covariable are much less than those of sampling the predictand. Stratification of the survey area, e.g. by means of soil map delineations increases the precision of predictions when applying cokriging. An obvious gain in precision is achieved for homogeneous soil units, where the measured values are relatively small, and there is no spatial structure. Also, the use of cokriging permits fewer observations as compared to kriging, if a certain predescribed precision of predictions is defined. When simulation calculation models are used, e.g. to obtain values for moisture deficits, one should first calculate model results for every observation point, and then interpolate, rather than interpolate the input variables, and then calculate model results.

Key words: spatial interpolation, spatial variability, kriging, cokriging, non-stationarity, sampling schemes, soil variables.

6

WOORD VOORAF

Dit proefschrift is ontstaan bij de vakgroep Bodemkunde en Geologie van de Landbouwuniversiteit Wageningen. Binnen deze vakgroep is een regelmatig terugkerende vraag hoe puntwaarnemingen aan bodemkenmerken naar een vlak, bijvoorbeeld een kaart, geïnterpoleerd moeten worden. Hierover gaat dit proefschrift.

Bij het voltooien van dit proefschrift stel ik het op prijs degenen te bedanken die hebben bijgedragen aan de totstandkoming ervan. In de eerste plaats de twee promotoren, die beiden op een eigen, karakteristieke, manier een stempel op het werk hebben gedrukt.

Prof. Corsten ben ik bijzonder dankbaar voor het mede richting geven aan het onderzoek en voor het voeren van telkens terugkerende stimulerende, diepgaande discussies over vrijwel alle aspecten ervan. Zijn vele, vruchtbare ideeën, zijn inspiratie, zijn niet aflatende aandacht voor detail en voor het inbedden van theoretische concepten in gangbare statistische methoden en zijn positief kritische bijdragen aan dit proefschrift zijn behalve opbouwend ook buitengewoon plezierig geweest.

Prof. Bouma ben ik speciale dank verschuldigd voor de mogelijkheden, die hij naast het stimuleren van het onderzoek, het zoeken naar toepassingen en het participeren in afzonderlijke elementen van het onderzoek, binnen de vakgroep Bodemkunde & Geologie heeft gecreëerd. Deze hebben er mede toe bijgedragen dat dit onderzoek een succes is geworden.

Voorts ben ik speciale dank verschuldigd aan Aad van Eijnsbergen en Leo Barendregt voor hun wiskundige en statistische ondersteuning en aan Igor Staritsky voor zijn onder-

7

steuning op het gebied van de computerberekeningen en het gebruik van Geografische Informatie Systemen. Ook allen die hun inbreng hebben gehad in het uitvoeren van het onderzoek, zowel medewerkers aan LUW en het Staringinstituut als studenten, worden op deze plaats bedankt.

Tenminste even waardevol en onmisbaar is de ondersteuning thuis geweest. Ik ben me er van bewust dat de vele avonden, weekenden en vakanties waarin toch weer gewerkt, gestudeerd en geschreven moest worden, maar gedeeltelijk met het verschijnen van dit proefschrift worden gecompenseerd. Zonder de liefdevolle steun van Anita, de vrolijke inbreng van Judith, en zonder de stimulerende belangstelling van mijn ouders, die als regel als eersten inhoudelijk op verschenen publicaties reageerden, was het proefschrift niets, helemaal niets geworden.

Tenslotte wil ik de Programma Commissie voor de Bescherming van de Bodem (PCBB) bedanken voor het beschikbaar stellen van financiële middelen.

Het omslag is ontworpen door E.M. van Cleef van het Grafisch Service Centrum.

Dit proefschrift is opgebouwd uit acht artikelen, die gedurende de afgelopen jaren in verschillende tijdschriften zijn verschenen, dan wel binnenkort zullen verschijnen.

TABLE OF CONTENTS

SYNOPSIS

Introduction

Spatial interpolation is introduced on the basis of the linear regression model with dependent observations. Practical applications to studies on soils are presented. The introduction is concluded with an outline of the thesis and with a selected bibliography of relevant publications.

I. Sampling schemes.

Chapter I.1 considers nested sampling for estimating spatial semivariograms. Proof of the equivalence of the cumulative sum of variance components provided by the classical nested ANOVA and semivariogram values for some distances is given. Nested sampling is compared with other sampling procedures on a simulated random field.

In Chapter I.2 sequential sampling is used to obtain an estimate of the mean value of the infiltration rate in a relatively homogeneous river terrace. Starting originally with 32 observations, it appeared to be sufficient to use only 8 observations to obtain a value for infiltration rate with sufficient precision.

II. Kriging and cokriging

In Chapter II.1 kriging and cokriging are formulated in terms of regression procedures. This formulation turns out to be powerful for understanding the most important equations, because it no longer requires use of Lagrange multipliers.

In Chapter II.2 the constraints that pseudo-covariance and pseudo-cross covariance functions have to obey in order to yield pseudo-positive-definiteness are formulated. The universal cokriging equations are given as well as formulations for block-cokriging and for cokriging of a gradient.

III. Applications to studies in soil science

The financial savings that result when observations on an expensive land quality (the 30-year averaged Moisture Deficit) are used together with observations on a strongly correlated, relatively cheap, covariable (the Mean Highest Water table) are worked out in Chapter III.1.

Chapter III.2 concentrates on the sampling distance necessary to achieve a certain prescribed precision of a predictive map of 'Available Water' in the Limagne area in France. Two covariables are used: 'depth to gravel banks' and 'stoniness of the surface'. Both are relatively easy to observe.

Chapter III.3 considers the advantage of using soil survey information together with spatial interpolation. Cokriging is combined with an existing 1:10000 soil map of the Mander area (in the Netherlands).

In Chapter III.4 universal cokriging is compared with other interpolation procedures when applying simulation calculation models for moisture deficits. As it turns out, one should collect the necessary data at every observation point and then interpolate the simulation model calculations rather than interpolate model input variables first and then use simulation models.

12

INTRODUCTION

1. What is spatial interpolation?

In many environmental studies observations are collected
in space. One may think of observations on soil such as
available water or the concentration of a pollutant like
cyanide, a meteorological characteristic like daily highest
temperature, or a hydrological characteristic like piezometric
heads. Usually such data are collected in regions such as a
polluted area, a river terrace, an aquifer or a country, but
sometimes transects are used as well. There is a common wish
among surveyors and practical researchers to display the
collected data graphically in the form of a map, in particular
if spatial variation of the data in the study area may be
expected. Well-known pictures include the weather map in daily
newspapers and three-dimensional perspectives of elevation
data, displaying the peaks and the valleys of mountainous
regions.

Spatial interpolation is concerned with inference from
point observations to the representation in the form of a map.
Purpose is predicting the value of an environmental character-
istic at any unobserved location, using a linear combination
of available observations. Educated guessing of a random
variable which has not (yet) been observed is difficult, since
the number of measurements is limited. The way to proceed in
spatial interpolation studies is to predict values at many
locations, say at the nodes of a fine-meshed grid, followed by
displaying the resulting values by means of greytones, colours,
isolines or in a three-dimensional presentation.

One of the common properties of environmental data is

13

their spatial dependence: observations close to each other are more similar to each other than observations separated by a larger distance. Lack of dependence is a special case. The range to which this dependence extends may differ for different variables. For a soil variable like 'organic carbon content of the topsoil' this range is likely to be smaller than for 'clay content of the subsoil'. Not only the predictions themselves do matter, but also the prediction errors: how certain can one be about the predictions. Quantification of the prediction error and its variance are main topics to deal with.

When predicting, the spatial dependence is to be taken into account, because the observations close to a prediction location are likely to be more similar to the unobserved value than observation at a larger distance. Predicting therefore will differ for variables with different dependence structures. Further, the interpolated surface will pass through the observations. The prediction procedures studied are termed exact procedures. In this respect interpolated surfaces differ from smoothly fitted trend surfaces which only approximate the observations. If the function describing the spatial structure is continuous, the interpolation surface will be continuous as well.

The measurements and their locations in space play an important role. In the first place, they are used to estimate the dependence structure, and secondly they are used to carry out the predictions. This poses restrictions on the usefulness of any configuration of the observation points in a region. The study of spatial data configurations is of importance.

Predicting is a regularly encountered activity in regression. On the basis of a finite number of observations on

a dependent variable and a number of regressors the parameters of a model are fitted, with which predictions for future regressor values are carried out. In spatial interpolation the regressors are the coordinates of the measurement location, or some simple functions thereof. Because of the dependence among the spatial observations, use of ordinary least squares to estimate the parameters is replaced by general least squares.

Obviously as well, there may exist a relation between the variable to be predicted and other spatial characteristics. If, for example, a rainfall map is to be made for a country which is close to the sea, there likely exists a relation between the amount of precipitation and the distance to the coast. Also, the amount of water available for grassland or maize is likely to be related to both the relative elevation and the presence of gravel at this location. When predicting at unobserved locations this information should be taken into account. It is therefore only natural that the so-called multivariate (that is: using observations on several characteristics) prediction procedures come into view.

When research on the topics of this thesis started, some research was already carried out. Therefore first attention will be paid to the origins of spatial interpolation.

2. The origins of stochastic spatial interpolation.

The results of spatial interpolation are so closely linked with the studies of Matheron and his co-workers, that it seems appropriate to start this section with one of Matheron's statements:

Historically geostatistics are as old as mining itself. As soon as mining men concerned themselves with foreseeing results of future works and, in particular as soon as they started to pick and to analyze samples, and compute mean grade values, weighted by corresponding thicknesses and influence-zones, one may consider that geostatistics were born. In so far as they take into account the space characteristics of mineralization, these traditional methods still keep all their merits (...) Modern developments of the theory have adopted them as their starting point. [Matheron, 1963]

Note that Matheron, like many others working with observations that take their values in a two- or three-dimensional space, uses the term Geostatistics [for example Cressie, 1989; Journel and Huijbregts, 1978]. In this thesis this term is avoided. It seems inappropriate to give spatial interpolation procedures a new label which might suggest that it is a completely new scientific discipline, or some modification of statistics. Existing relations with statistical issues such as the general linear model and time series analysis do not deserve this label, as will be shown below.

The history of spatial variability, spatial statistics and spatial interpolation usually starts with the work of Krige in South African gold mines [Krige, 1951, 1966]. Krige used spatial interpolation to estimate minable block reserves from a number of point observations. Because of the large benefits to be expected from such studies, the gold mines served as a natural basis for spatial interpolation. A publication [Matérn, 1960] in a relatively obscure journal, which, however, contained a good deal of important and still valuable research,

was barely noticed. Matheron's publications [Matheron, 1963, 1965] in the sixties provided a basis of theory, largely resting on measure theory and probability theory. After his work was published in English and American literature [Matheron, 1971, 1973], there was a large increase in applications and a surge of new developments.

The first applications of spatial interpolation were in mining, notably ore grades, especially to estimate *in situ* resources. Later applications concerned a much broader field, such as studies in soil [McBratney and Webster, 1983], hydrology [Delhomme, 1978], atmospheric conditions [Witter, 1984], the design of monitoring networks [Villeneuve et al., 1979], gene frequencies [Piazza et al., 1981] and even the growth characteristics of a single fruit tree [Monestiez et al., 1989]. Journel and Huijbregts [1978] give an overview of many practical applications.

The applications of spatial interpolation procedures in soil studies were primarily concerned with usually rather simple soil survey data such as the clay content, which a trained pedologist estimates by kneading and tasting soil samples [Webster and Cuanalo de la C., 1975; Bouma and Nielsen, 1984]. All these studies assumed that there was no trend. The researchers relied largely on semivariograms, but did not consider it necessary to test whether the conditions were valid to use semivariograms. However, some researchers did recognize the merits of statistical spatial interpolation procedures for using point observations to obtain predictions at unobserved locations, and to quantify the associated uncertainty, [Burgess et al., 1981]. In later studies, there was a shift to the application of spatial interpolation procedures to physical and chemical data on soil, to land qualities and to simulation

17

model output variables [Webster, 1985]. All of these are much more relevant to deal with, because they are more expensive to collect and are not related to landscape features and are often very variable.

In Wageningen, spatial interpolation was taken up seriously after a visit by the French scientist Delfiner in 1976. This resulted in several publications [Witter, 1984; Ten Berge et al., 1983; Bregt et al., 1987, Corsten, 1989] on the development, interpretation and application of statistical interpolation methods. Here it is appropriate to acknowledge the contributions of Prof. Corsten at the Department of Mathematics of the Agricultural University, and of Prof. Burrough (first at the Winand Staring Centre and afterwards at the department of Soil Science and Geology of the Agricultural University) to the promotion and acceptance of spatial interpolation techniques in agricultural research.

The availability of sophisticated hardware and software has enabled spatial interpolation to achieve general acceptance. Nowadays, there exist easy and fast procedures to store spatial observations, investigate their spatial structure, and produce a computerized map with a simple personal computer and matrix printer. This is enhanced by the increasing availability of graphical and interfacing facilities such as Geographical Information Systems, which can be used for spatially interpolating in connection with many different sources of information (for example, digitized soil maps, satellite imagery, height data and other topographic information [Burrough, 1986]).

3. Prediction and regression.

    Spatial interpolation is closely related to predicting in
linear regression theory [Goldberger, 1962; Harville, 1976,
1985; Corsten, 1989]. In this section predictors and their
variances in the context of different stages of generalization,
all based on the linear regression model, will be formulated.
The convention of underlining random variables, and of
indicating the transpose of a vector with a prime will be
followed. The focus will be on the use of multivariate
characteristics, of which the univariate characteristic is a
special case.
    In the most simple situation one considers only one
characteristic, while the observations are independent and
there is no trend. In fact, the prediction problem is reduced
to estimating the mean value within an area. This model is
relevant in studies, where one knows that all observations are
more or less similar, and the assumption of independence may
readily justifiable. An example is given in Chapter I.2 [Stein
et al., 1989b; Finke et al., in press; Wopereis et al., in
prep.].
    Spatially correlated observations provide a more complex
situation even in one-dimensional space (transect). One may
surmise a certain relationship with time series analysis. The
interdependence of observations collected in time can be
defined similarly to the interdependence of observations
collected along a spatial transect. Also, predicting a future
observation in time is similar to predicting a spatial
observation in an unobserved location. This extends also to the
'prediction' of past observations, or of missing observations.
Apart from some analogies, there are important differences.

For example, observations in time are directionally ordered, whereas on a transect they are not. *A fortiori* this holds for two- or three-dimensional space. This prohibits a straight-forward application of auto-regressive or moving average processes to spatial studies. Also, in time series one usually prefers equally spaced observations, whereas in spatial studies one must deal with any pattern of observations. Finally, it is important to deal with monthly or yearly periodicities in time series, which are seldom of importance in spatial studies. The procedures defined and used in this thesis therefore apply to a more general situation: prediction in a single or multi-dimensional observation space, where observations may be collected according to any configuration.

We now turn to a more general situation in which there are observations on p variables. As an example, in a level river terrace one may consider observations on available water, depth to gravel banks, stoniness of the surface, clay content of the B-horizon and depth to mottling. Clearly, $p = 5$ in this example.

Consider as a starting point the linear model, $\underline{y} = X\beta + \underline{e}$. Because each observation is collected in a region, it has besides the measurement itself, also one, two or three coordinates, depending on whether the observations were made in a space of dimension one (transect), two (plane) or three (volume). The vector $\underline{y}$ contains the n measurements on the p characteristics, $\underline{y} = (\underline{y}_1', \ldots, \underline{y}_p')'$, where every $\underline{y}_i$ is a vector containing the n_i observations of the ith characteristic, $n = n_1 + \ldots + n_p$. Different numbers of observations may be collected on these variables possibly in different locations. The matrix X contains values of the coordinates which describe the trend. It may be partitioned into blocks X_{ij}, where every row of X_{ii}

20

contains the powers and the products of the powers of the coordinates of the corresponding observation location for the ith variable up to the degree ν_i of its trend and X_{ij} vanishes for $i \neq j$. For example, in plane sampling, when ν_i equals 2 and the coordinates are given by x_1 and x_2, the values of 1, x_1, x_2, x_1^2, $x_1 x_2$ and x_2^2 are included in every row of X_{ii}. For the distribution of the error vector $\underline{e}$ we assume for the moment that $E[\underline{e}] = 0$ and $Var[\underline{e}] = C$. These assumptions will be relaxed later. The elements of the matrix C are given by covariance functions $c_i(r)$ for the ith variable and cross-covariance functions $c_{ij}(r)$ for the interaction between the ith and the jth variable, all isotropic and depending on the distance r between observations. The relevant question is: how does one optimally predict $\underline{y}_0$, that is: the value of $\underline{y}_1$ in a hitherto unsampled location, using a linear combination of observations on p variables, $\underline{y}_1, \ldots, \underline{y}_p$, such that the variance of the prediction error is minimum? If $p = 1$, prediction is termed kriging, after its original inventor. If $p > 1$ it is termed cokriging. The term 'prediction' is used instead of 'estimation' to indicate that we are attempting to guess the value of a random variable, and not to estimate a parameter.

Let the linear predictor $\underline{t}$ of $\underline{y}_0$ be denoted by

$$\underline{t} = \lambda'\underline{y} = \sum_{i=1}^{p} \lambda_i'\underline{y}_i. \tag{1}$$

Every vector λ_i consists of coefficients (or weights, not necessarily positive) for the observations of the ith variable, and $\lambda = (\lambda_1', \ldots, \lambda_p')'$. We will call $\lambda'\underline{y}$ an unbiased predictor of $\underline{y}_0$ iff. $E[\lambda'\underline{y} - \underline{y}_0] = 0$. In particular, $\lambda'\underline{y}$ is called the best linear unbiased predictor of $\underline{y}_0$ (BLUP) if $Var[\underline{t} - \underline{y}_0]$ is minimal among all such linear unbiased predictors. We want to

21

obtain the BLUP of $y_0 = x_0'\beta + e_0$, where x_0 is analogous to any
of the first n_1 rows of X and $Cov(\underline{e}, \underline{e}_0) = c_0$.

As is shown in Chapter II.1 we obtain the following
expressions for the predictor and the variance of the
prediction error [Stein and Corsten, in press]:

$$t = x_0'\hat{\beta} + c_0'C^{-1}(\underline{y} - X\hat{\beta})$$
$$Var[\underline{t} - \underline{y}_0] = c_{00} - c_0'C^{-1}c_0 + x_a'Vx_a \qquad (2)$$

where $V = (X'C^{-1}X)^{-1}$, $\hat{\beta} = VX'C^{-1}\underline{y}$, the Generalized Least
Squares estimator of β, and $x_a = x_0 - X'C^{-1}c_0$. This procedure
only holds for the situation where the covariance and the
cross-covariance functions are known, or when they can be
estimated from the available data. Estimation of covariance and
cross-covariance functions is only permissible when all the
regression coefficients in β are known, and when the variances
of the variables are finite.

In order to avoid these complications one has introduced
the concept of multivariate increments: attention is restricted
to vectors $\underline{y}$ for which $E[\lambda'\underline{y}]$ exists only for those λ with
$\lambda'(X' ; x_0)' = 0$. Increments must be used in order to overcome
the presence of an unknown trend, including a common unknwon
expectation. In the past, semivariograms were used in the
presence of a trend to describe the spatial structure of the
residuals after estimation and subtraction of the trend
[Webster and Burgess, 1980]. This procedure is known to lead to
completely erroneous results [Matheron, 1971, pag. 196
excercises 1 and 2; Cressie, 1987]. The collection of all λ
defined in this way is termed the permitted collection
[Christakos, 1984]. We do not assume that $E[\underline{y}]$ exists, nor that
$Var[\underline{y}]$ exists. We do assume, however, that $E[\lambda'\underline{y}] = 0$ and that

22

$Var[\lambda'\underline{y}] = \lambda'G\lambda$ exists and is positive for the permitted collection. However, negative values may appear for non-permitted λ. The elements of G are not given by covariance and cross-covariance functions, but by pseudo-covariance functions $g_i(r)$ and pseudo-cross covariance functions $g_{ij}(r)$. Let g_0 contain the values of the pseudo-covariance and the pseudo-cross covariance functions between the observation locations and the prediction location and $g_{00} - g_1(0)$, $g_1(r)$ being the pseudo-covariance function for the predictand. By defining $V = (X'G^{-1}X)^{-1}$, $\hat{\beta} = VX'G^{-1}\underline{y}$ and $x_a = x_0 - X'G^{-1}g_0$, one obtains the following expressions for the predictor $\underline{t}$ and the prediction error variance $Var[\underline{t}-\underline{y}_0]$:

$$\underline{t} = x_0'\hat{\beta} + g_0'G^{-1}(\underline{y}-X\hat{\beta}) \qquad (3)$$
$$Var[\underline{t}-\underline{y}_0] - g_{00} - g_0'G^{-1}g_0 + x_a'Vx_a.$$

which is completely similar to (2). A more detailed account is given in Chapter II.2 [Stein et al., in press].

If one predicts the values at many locations, say at the nodes of a grid with a fine mesh, the predictions may displayed by means of colours or greytones, they may be delineated by means of isolines or they may displayed in a three-dimensional perspective, yielding in any form a map of the area. In this respect, cokriging, used as an automatic contouring procedure may well serve the purpose of spatial interpolation in expert systems.

4. Related interpolation procedures.

The spatial interpolation procedures described in this thesis are not the only interpolation procedures possible. This

section gives a summary of the most common among the latter, and describe their relation to the spatial interpolation procedures described in this thesis. A more detailed presentation is found in [Ripley, 1981].

1. Deterministic interpolation procedures.
    i)   Lagrange interpolation exactly fits polynomials through the observations. The degree of the polynomial is equal to the number of observations minus 1. Therefore, adding or deleting a single observation changes the degree of the polynomial, and hence also the form of the interpolation surface, even at locations far away from the location where this point is added or deleted. This is undesirable.
    ii) Tesselations: neighbouring observations are connected by means of lines, yielding an interpolation surface composed of triangles. Every unobserved location is assigned the value defined by the triangle pertaining to that location.
    iii) Thiessen polygons: every unobserved location is assigned the value of its closest neighbour. This procedure performs well for ordinal data. For quantitative data, the resulting stepwise picture is usually considered to be unrealistic because most quantitative properties change gradually. Besides, outlying observations are overly stressed.
    None of these procedures supplies measures of uncertainty in unobserved locations.
2. Stochastic interpolation procedures.
    i)   Inverse distance interpolation, or nearest neighbour interpolation: observations are assigned weights in

proportion to a function of the inverse distance
(1/r) between the interpolation location and the
observation locations. Commonly used functions are
$1/r$ or $(1/r)^2$. Apart from the diffculty of choosing
any function, this procedure yields sub-optimal
predictions as compared to kriging, because the
spatial structure of the variables is not taken into
account.

ii) Trend surface analysis: this procedure is a special
case of the procedures treated in this thesis, but it
is based on the unrealistic assumption that the
observations are uncorrelated. Predictions coincide
with the expected trend, which is estimated under
Least Squares conditions.

iii) Splines: the thin plate spline is a special case of
kriging. As Dubrule pointed out [Dubrule, 1983], the
interpolating spline is equivalent to kriging with a
trend of degree 1, and a pseudo-covariance function
$g(r) - |r|^2\log(|r|)$. By defining a more general
penalty function the formal equivalence between
splines and kriging is given in [Matheron, 1980].
Among multivariate prediction procedures smoothing
spline functions are equivalent to a particular type
of cokriging; the converse is not true.

5. Random fields.

The probabilistic background to the statistical theory of
spatial interpolation is given by the theory of random fields
[Matheron, 1973; Yaglom, 1986]. Matheron and Yaglom's treatment
of spatial interpolation is largely based on the Gel'fand-

Vilenkin theory of generalized stochastic fields [Gel'fand and Vilenkin, 1964; Gel'fand and Shilov, 1964]. Although one needs topics in generalized function theory, e.g. in order to derive the class of permissible pseudo covariance functions, the theory is not needed for a statistical treatment of the problem of prediction. As is shown in this thesis, prediction equations can be formulated using standard statistical methods.

There are situations in which it is reasonable to simulate random fields [Journel & Huijbregts, 1978, pag. 493], using the turning band method [Mantoglou and Wilson, 1982]. For example, to quantify the influence of spatial variability of soil survey variables on, say, calculations of nitrate leaching, it is better to rely on stochastical simulation of the soil survey variables, with given mean, variance and pseudo-covariance function, because the resulting pattern agrees more with the spatial behaviour of the variables than interpolation maps obtained by kriging and cokriging. These simulated fields do not yield optimal predictions, and are less smooth than the kriged maps. Clearly, however, successively generated fields, all with the same set of parameters, may show apparent differences, as is illustrated in Chapter I.1 [Corsten and Stein, subm.].

6. <u>Practical applications of spatial interpolation</u>.

Spatial interpolation procedures have many practical applications. To mention a few, one may think of a hydrological characteristic such as infiltration depth or piezometric head, a meteorological characteristic such as mean annual rainfall, a geological characteristic such as the content of a certain mineral, but also the regional spread of certain diseases,

26

family income and birth rate. This thesis deals solely with applications in soil science.

In most developed countries soil mapping is complete or is nearly complete, and the data collected for these maps have been stored in national databases. Attention is therefore shifting towards questions about soil quality. Growing awareness of our living environment has stimulated research on how best to use scarce soils of excellent quality, which are sometimes affected by pesticides, soil pollution and the dumping of waste. To be able to make decisions for which purpose to use the soils, the consequences of soil management must be known. Therefore, simulation models are increasingly being used to study variables like nitrate, phosphate and pesticide leaching to groundwater reservoirs. Such models need large amounts of both qualitative and quantitative data, some of which are often already stored in national databases. Both the original observations and the model calculations pertain to points, but the planner and the modeler are usually interested in areas such as parcels, farms, aquifers or even larger tracts of land. Soil studies are traditionally oriented on presentation in the form of a map. Therefore the model input parameters and the results calculated by the model have to be interpolated. Prosecutions for infringements of pollution or groundwater extraction laws lean heavily on figures on the effects of pollution and groundwater extraction; also the quantification of the reliability of model calculations is important. Operational problems such as how to define how many observations are needed and where they should be made so that reliable predictions can be made at unobserved locations are a natural spin-off of this research.

Soil variables are particularly interesting for applying

spatial interpolation procedures. Note the importance of calculating the amount of pollution [Staritsky et al., subm.] and the costs of preservation of the soils and of estimating the 'optimal use' of scarce soil of good quality, especially in developed countries. Furthermore, soil variables may be highly variable, because of the complex structure of the soil, where soil physical, chemical and biological parameters are likely to influence the variables under study. Since soil variables are never observed entirely one has to use finite samples. This distinguishes such observations from, for example, elevation data for which an areal overview can (at least in principle) be obtained.

In most studies, soil observations are unique. There is little reason to suspect that variables such as 'clay content', 'organic matter content' or 'depth to mottling' will change very much over the relevant study period, nor will one be able to collect another observation at the same location. In this regard, studies on soils differ from hydrological studies, where multiple observations in time may be available. However, studies on soil changes, for example under the influence of acid rain, the greenhouse effect, nitrate and phosphate infiltration [Stein and Van Breemen, in press; Bouma, 1989], will require a different approach, maybe supported by the analysis and use of space-time models [Haslett and Raftery, 1989; Rouhani and Wackernagel, 1990; Loader and Switzer, 1989].

7. Outline of this thesis

The first part of this thesis is concerned with sampling strategies in soil studies. Typical questions include: how many observations should one take? which are the most appropriate

28

locations for sampling? Some authors advocate a nested scheme with multiplicative distances, say 1m, 10m, 100m, 1000m, [Miesch, 1975; Pettitt and McBratney, 1990]. However, nested sampling cannot be recommended when the spatial structure is to be determined, and must be avoided if spatial interpolation is intended. Using a grid with fixed distances yields observations which are distributed more evenly over the study area, and hence semivariogram values for more distances; therefore such a grid is preferable, as is shown in Chapter I.1. The use of sequential sampling to collect observations in a relatively homogeneous area is considered in Chapter I.2.

The second part of this thesis is concerned with kriging and cokriging. In Chapter II.1 kriging and cokriging are formulated in terms of regression procedures. In Chapter II.2 the constraints that pseudo-covariance and pseudo-cross covariance functions have to obey in order to yield pseudo-positive-definiteness are formulated. The universal cokriging equations are given as well as formulations for block-cokriging and for cokriging of a gradient.

The third part of this thesis is concerned with two applications of spatial interpolation to studies in soil science. In the Mander study important questions concerning moisture deficits caused by groundwater extraction were considered. The data set was collected during a short period by a single soil surveyor from the Winand Staring Centre. The Limagne study was part of a larger research program carried out by the Department of Soil Science and Geology of Wageningen Agricultural University in the late 1980s. This thesis focuses on the spatial variability of available water and of infiltration rates in sharply delineated river terraces.

In Chapter III.1 [Stein et al., 1988a] savings resulting

from considering observations of a highly correlated, relatively cheap, covariable (the Mean Highest Water table) with observations of a relatively expensive land quality (the 30-year averaged Moisture Deficit) are quantified. In Chapter III.2 [Stein et al., 1989c] attention is focused on the sampling distance necessary to achieve a certain prescribed precision of a predictive map of 'Available Water' in the Limagne area in France. Two covariables 'depth to gravel banks' and 'stoniness of the surface', which are both relatively easy to observe, reduced required observation density considerably. In Chapter III.3 [Stein et al., 1988b] the advantage and practicality of using soil survey information is considered. Use of cokriging is combined with an existing 1:10,000 soil map of the Mander area. In Chapter III.4 [Stein et al., in press] universal cokriging is compared with other interpolation and calculation procedures to obtain values for moisture deficits. Interpolated maps may be used to obtain point values which in turn are used in such procedures. The main problem is whether to interpolate model input variables first, and then use simulation models, or whether to collect the necessary data at every observation point and then interpolate the simulation model calculations.

In every chapter the soil problems that occur have been described as clearly as possible. Because each chapter is a publication in itself, a certain overlap is unavoidable.

30

SELECTED BIBLIOGRAPHY

Barendregt, L.G. 1987. The estimation of generalized covariance when it is a linear combination of two known generalized covariances, *Water Resour. Res. 23*(4), pp. 583-590.

Bouma, J. 1989. Land qualities in space and time, in *Land qualities in space and time,* edited by J. Bouma and A.K. Bregt, pp. 2-13, Pudoc, Wageningen.

Bouma, J., and D.R. Nielsen. 1984. *Soil spatial variability.* Pudoc, Wageningen, 243p.

Bregt, A.K., Bouma, J., and Jellinek, M. 1987. Comparison of thematic maps derived from a soil map and from kriging of point data. *Geoderma 39*, pp. 281-291.

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

Burrough, P.A. 1986. *Principles of geographical information systems for land resources assessment.* Oxford, Clarendon press, 193p.

Carr, J.R. and P.G. McCallister. 1985. An application of cokriging for estimation of tripartite earthquake response spectra. *Math. Geol. 17,* pp. 527-545.

Christakos, G. 1984. On the problem of permissible covariance and variogram models. *Water Resour. Res. 20*(2), pp. 251-265.

Christensen, R. 1990. The equivalence of predictions from universal kriging and intrinsic random-function kriging. *Math. Geol. 22,* pp. 655-664.

Cochran, W.G. 1977. *Sampling techniques.* Wiley, New York, third edition, 428p.

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

Corsten, L.C.A., and A. Stein. Subm. Are nested sampling schemes recommendable for spatial semivariogram estimation? *Submitted to J. Amer. Stat. Ass.*

Cressie, N.A.C. 1986. Kriging nonstationary data. *J. Amer. Stat. Ass. 81*, pp. 625-634.

Cressie, N.A.C. 1987. A nonparametric view of generalized covariances for kriging. *Math. Geol. 19*, pp. 425-449.

Cressie, N.A.C. 1989. Geostatistics. *The American Statistician, 43*, pp. 197-202.

Delfiner, P. 1976. Linear estimation of non-stationary spatial phenomena. In *Advanced geostatistics in the Mining Industry*, edited by M. Guarascio, M. David and C. Huijbregts, pp. 49-68, Dordrecht, Reidel.

Delhomme, J.P. 1978. Kriging in the hydrosciences. *Adv. Water Resour. 1*(5), pp. 251-266.

Dowd, P.A. 1989. Generalized cross-covariances. In *Geostatistics*, edited by M. Armstrong, pp. 151-162, Dordrecht, Kluwer.

Dubrule, O. 1983. Two methods with different objectives: splines and kriging. *Math. Geol. 15*, pp. 245-257.

Finke, P.A., J. Bouma and A. Stein. In press. Measuring field variability of disturbed soils for simulation purposes with geostatistical and morphological techniques. *Soil Sci. Soc. Am. J.*

Gelfand, I.M., and G.E. Shilov. 1964. *Generalized Functions 1*, Academic Press, New York, 423p.

Gelfand, I.M., and N. Vilenkin. 1964. *Generalized Functions 4*, Academic Press, New York, 384p.

Goldberger, A.S. 1962. Best linear unbiased prediction in the generalized linear model. *J. Amer. Stat. Ass. 57*, pp. 369-375.

Harville, D. 1985. Decomposition of prediction error. *J. Amer. Stat. Ass. 80*, pp. 132-138.

32

Harville, D. 1976. Extension of the Gauss-Markov theorem to include the estimation of random effects. *Annals of Statistics* 4(2), pp. 384-395.

Haslett, J., and A.E. Raftery. 1989. Space-time modelling with long-memory dependence: assessing Ireland's wind resource (with discussion). *J. R. Stat. Soc. C, 18*, pp. 1-50.

Journel, A.G. 1983. Nonparametric estimation of spatial distributions. *Math. Geol. 15*, pp. 445-468.

Journel, A.G. and C.J. Huijbregts. 1978. *Mining geostatistics*. Academic Press, London, 600p.

Kitanidis, P.K. 1983. Statistical Estimation of Polynomial Generalized Covariance Functions and Hydrologic Applications. *Water Resour. Res. 19*(4), pp. 901-921.

Krajewski, W.F. 1987. Cokriging Radar-rainfall and Rain Gage Data: *J. Geoph. Res., 92*, pp. 9571-9580.

Krige, D.G. 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. *J. Chem., Metall. & Min. Soc. of S. Africa 52*, pp. 119-138.

Krige, D.G. 1966. Two dimensional weighted moving average trend surfaces for ore evaluation. *J. S. African Soc. Min. Metal.* 66, pp. 13-38.

Loader, C., and P. Switzer. 1989. *Spatial covariance estimation for monitoring data*. SIMS Technical Report 133, Stanford University, Dept. of Statistics, 26p.

Mantoglou, A. and J.L. Wilson. 1982. The turning bands method for simulation of random fields using line generation by a spectral method. *Water Resour. Res. 18*(5), pp. 1379-1394.

Mardia, K.V., and R.J. Marshall. 1984. Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika 71*, pp. 135-146.

Marshall, R.J., and K.V. Mardia. 1985. Minimum norm quadratic estimation of components of spatial covariance. *Math. Geol. 17*, pp. 517-525.

Matérn, B. 1960. Spatial Variation. *Meddelanden från Statens Skogsforskningsinstitut 49*, pp. 1-144.

Matheron, G. 1963. Principles of geostatistics. *Economic Geology 58*, pp. 1246-1266.

Matheron, G. 1965. *Les variables régionalisées et leur estimation*. Paris, Masson, 306p.

Matheron, G. 1971. *The theory of regionalized variables and its applications*. Les Cahiers du Centre de Morphologie Mathématique, No. 5. Paris: Ecole de Mines de Paris, 211p.

Matheron, G. 1973. The intrinsic random functions and their applications. *Adv. Appl. Prob. 5*, pp. 439-468.

Matheron, G. 1980. Splines and kriging: their formal equivalence. Centre de Géostatistique, Fontainebleau, 21p.

McBratney, A.B., and R. Webster. 1983. Optimal interpolation and isarithmic mapping of soil properties. V. Co-regionalization and multiple sampling strategy. *J. Soil Sci. 34*, pp. 137-162.

Meinguet, J. 1979. Multivariate interpolation at arbitrary points made simple. *J. Appl. Mathematics and Physics (ZAMP) 30*, pp. 292-304.

Miesch, A.T. 1975. Variograms and variance components in geochemistry and ore evaluation. *Geological Society of America, Memoir 142*, pp. 333-340.

Monestiez, P., R. Habib and J.M. Audergon. 1989. Estimation de la covariance et du variogramme pour une fonction aléatoire à support arborescent: application à l'étude des arbres fruitiers. In *Geostatistics*, edited by M. Armstrong pp. 39-56, Dordrecht, Kluwer.

Morkoc, F., J.W. Biggar, D.R. Nielsen, and D.E. Myers. 1987. Kriging with generalized covariances, *Soil Sci. Soc. Amer. J. 51*, pp. 1126-1131.

Myers, D.E. 1982. Matrix Formulation of Co-kriging. *Math. Geol. 14*, pp. 249-257.

34

Pettitt, A.N., and A.B. McBratney. In press. Economical spatial surveys to estimate variograms. *Applied Statistics*.

Piazza, A., P. Menozzi, and L. Cavalli-Sforza. 1981. The making and testing of geographic gene-frequency maps. *Biometrics 37*, pp. 635-659.

Rao, C.R. 1989. A Lemma on Optimization of a Matrix Function and a Review of the Unified Theory of Linear Estimation, in *Statistical Data Analysis and Inference*, edited by Y. Dodge, pp. 397-417, Elsevier, Amsterdam.

Ripley, B.D. 1981. *Spatial Statistics*. Wiley, New York, 252p.

Rouhani, S., and H. Wackernagel. 1990. Multivariate geostatistical approach to space-time data analysis. *Water Resour. Res. 26*(4), pp. 585-591.

Seguret, S.A. Filtering periodic noise by using trigonometric kriging. In: In *Geostatistics*, edited by M. Armstrong, pp. 481-491, Dordrecht, Kluwer.

Seo, D.J., W.F. Krajewski and D.S. Bowles. 1990. Stochastic interpolation of rainfall data from rain gages and radar using cokriging. 1. Design of Experiments. *Water Resour. Res. 26*(3), pp. 469-477.

Staritsky, I.G., P. Sloot and A. Stein. Spatial variability of cyanide soil pollution on former galvanic factory premises. Submitted to *Water, Air and Soil Pollution*.

Starks, T.H., and J.H. Fang. 1982. On the Estimation of the Generalized Covariance Function. *Math. Geol. 14*, pp. 57-64.

Stein, M.L. 1987. Minimum Norm Quadratic Estimation of Spatial Covariances. *J. Amer. Stat. Ass. 82*, pp. 765-772.

Stein, A., J. Bouma, W. van Dooremolen and A.K. Bregt. 1988a. Cokriging point data on moisture deficits. *Soil Sci. Soc. Am. J. 52*, pp. 1418-1423.

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

Stein, A., M. Hoogerwerf and J. Bouma. 1989a. Combining soil classification with (co-)kriging of moisture deficits, in *Land qualities in space and time*, edited by J. Bouma and A.K. Bregt, pp. 205-208, Pudoc, Wageningen.

Stein, A., J. Bouma, S.B. Kroonenberg and S. Cobben. 1989b. Sequential sampling to measure the infiltration rate within relatively homogeneous land units, *Catena 16*, pp. 91-100.

Stein, A., J. Bouma, M.A. Mulders and M.H.W. Weterings. 1989c. Using Cokriging in variability studies to predict physical land qualities of a level river terrace. *Soil Technology 2*, pp. 385-403.

Stein, A., J. Bouma, M.A. Mulders and M.H.W. Weterings. 1989d. Predicting physical land qualities of a level river terrace using Cokriging. Contribution to the international conference *Statistics, Earth and Space Sciences*, Leuven, Belgium.

Stein, A. 1990a. Use of Cokriging to predict important land qualities at different scales. *Proc. ISSS Symposium, Kyoto, Japan*, vol. I, pp. 184-189.

Stein, A., and I.G. Staritsky. 1990b. Stratified Cokriging to predict simulated moisture deficits for areas of land as compared to calculated soil unit averages. *Proceedings of EGIS*, pp. 1070-1080, Amsterdam.

Stein, A. 1991. Ruimtelijke variabiliteit van bodem en water. CHO-TNO dag *Ruimtelijke variabilitet van bodem en water*.

Stein, A., and L.C.A. Corsten. In press. Universal kriging and Cokriging as a regression procedure. *Biometrics*.

Stein, A., A.C. van Eijnsbergen and L.G. Barendregt. In press. Cokriging non-stationary data. *Mathematical Geology*.

Stein, A., I.Staritsky, J. Bouma, A.C. van Eijnsbergen and A.K. Bregt. Simulation of moisture deficits and areal interpolation by universal Cokriging. *Accepted for publication in Water Resources Research*.

Stein, A. and N. van Breemen. Time series analysis of changes in the soil solution: evidence for approaching nitrogen saturation in dutch forest soils. *Submitted to Agriculture, Ecosystems and the Environment*.

Ten Berge, H.F.M., L. Stroosnijder, P.A. Burrough, A.K. Bregt and M.J. de Heus. 1983. Spatial variability of physical soil properties influencing the temperature of the soil. *Agricultural Water Management* 6, pp. 213-226.

Van Meirvenne, M., and G. Hofman. 1989. Spatial variability of soil texture in a polder area. II Cokriging. *Pedologie, XXXIX-2*, pp. 209-226.

Vauclin, M., S.R. Vieira, G. Vachaud, and D.R. Nielsen. 1983. The use of cokriging with limited field soil observations. *Soil Sci. Soc. Am. J. 47*, pp. 175-184.

Vecchia, A.V. 1988. Estimation and model identification for continuous spatial processes, *J. R. Statist. Soc. B, 50*(2), pp. 297-312.

Villeneuve, J.P., G. Morin, B. Bobee, D. Leblanc and J.P. Delhomme. 1979. Kriging in the design of streamflow sampling networks. *Water Resour. Res. 15*(6), pp. 1833-1840.

Wahba, G. 1990. Comment on Cressie. *The American Statistician*, 44, 255-258.

Webster, R., and T.M. Burgess. 1980. Optimal interpolation and isarithmic mapping of soil properties III. Changing drift and universal kriging. *Journal of Soil Science 31*, pp. 505-520.

Webster, R. 1985. Quantitative spatial analysis of soil in the field. In: B.A. Stewart (ed.), *Advances in soil science 3*. Springer Verlag, New York, pp. 1-70.

Webster, R., and H.E. Cuanalo de la C. 1975. Soil transect correlograms of north Oxfordshire and their interpretation. *J. Soil Sci. 26*, pp. 176-194.

Witter, J.V. 1984. *Heterogeneity of dutch rainfall*. Thesis, Agricultural University, Wageningen, The Netherlands, 204p.

Wopereis, M.C.S., A. Stein, J. Bouma and T. Woodhead. In prep. Spatial variability and representativity of saturated hydraulic conductivity measurements.

Yaglom, A.M. 1986. *Correlation Theory of Stationary and Related Random Functions. I: Basic Results.* Springer, New York, 526p.

Yates, S.R. 1986. Disjunctive kriging 3. Cokriging. *Water Resour. Res. 22*(10), pp. 1371-1376.

Yates, S.R., and A.W. Warrick. 1987. Estimating soil water content using cokriging. *Soil Sci. Soc. Am. J. 51*, pp. 23-30.

Zimmerman, D.L. 1989. Computationally efficient restricted maximum likelihood estimation of generalized covariance functions, *Math. Geol., 21*, pp. 655-672.

PART I. SAMPLING SCHEMES

CHAPTER I.1. ARE NESTED DESIGNS RECOMMENDABLE FOR SPATIAL SEMIVARIOGRAM ESTIMATION?

by

L.C.A. Corsten and A. Stein

ARE NESTED DESIGNS RECOMMENDABLE FOR SPATIAL SEMIVARIOGRAM ESTIMATION?

ABSTRACT

In spatial studies common use is made of nested survey plans. By applying such plans, one takes observations according to a regular sampling scheme, with distances proportional to each other. It has been observed in the past that variance components provided by the nested sampling plan yield values similar to those obtained by means of semivariogram estimation (Miesch, 1975). From this article it turns out that nested sampling is unsuitable when predictions are to be carried out. Besides, nested sampling can barely be recommended to estimate parameters of a spatial semivariogram.

43

0. Synopsis

In section 1 attention is focused on stationary spatial phenomena in two dimensional space, for which the spatial variogram is to be estimated. In section 2 the layout of a five-stage nested sampling design with three branches per stage is presented. In section 3 the relationship between semivariograms and variance components of a nested design is shown. The relationship is independent of the number of stages, or the degree of balance. In section 4 it is shown how expected mean squares yield estimators for variance components, also holding for unbalanced designs. In section 5 two procedures are outlined for semivariogram estimation: one by using the results of section 4, and one with mean squared pair differences. Equivalence between the two procedures for balanced nested designs is shown in the Appendix. In section 6 the significance of contributions to the estimated semivariogram for values at larger distances as compared to values at smaller distances is tested. The procedure for balanced nested designs is different from that for unbalanced nested designs. In a numerical example, nested sampling is compared with various regular sampling schemes applied to eight simulated random fields.

1. Introductory objectives and assumptions

Consider a quantitative spatial phenomenon, e.g. density of a plant species, a geological characteristic like content of a mineral, a soil characteristic like clay content, annual depth of rainfall, local incidence of human, animal or vegetation disease, heavy metal content in a possibly polluted area, etc. We suppose that this phenomenon is stationary and that it has been measured at n locations, each given by its coordinates. One major problem concerns estimation of its covariance structure, for example in order to predict the value at another, arbitrary point where no such measurement has been made, on the basis of those n observations and the configuration of the n+1 points considered, commonly referred to as Kriging (Goldberger, 1962; Corsten, 1989; Stein and Corsten, 1990).

Let the observations $y_1,\ldots,y_n$ at points $z_1,\ldots,z_n$ be the realisation of a random vector $\underline{y}$. Symbols for random variables will be underlined, in order to distinguish them from fixed values. The assumption of stationarity implies that $\underline{y}_1,\ldots,\underline{y}_n$ have common, but unknown expectation μ and that for each pair z_i and z_j the covariance $cov(\underline{y}_i,\underline{y}_j)$ exists, depending on the difference vector $h = z_i-z_j$, in particular on its length $|h| = r$ only (isotropy). Let $\underline{y}$ be written as $\underline{y} = \mu 1_n + V^{\frac{1}{2}}\underline{u}$, where $\underline{u}$ is a standardized random variable (with expectation vector zero and uncorrelated elements, each with unit variance), μ and V are the expectation and the covariance matrix of $\underline{y}$, respectively, and 1_n is the n-vector consisting of elements 1 only. The function expressing $cov(\underline{y}_i,\underline{y}_j)$ in the distance r between z_i and z_j is $c(r)$, the covariance function, or the covariogram. In general $c(r)$ will not be known, and so it

should be estimated from observations; even if $c(r)$ is assumed to belong to a permissible family of functions, satisfying the condition of being positive definite (i.e. $\text{var}(\sum_{i=1}^{n} \lambda_i y_i)$ – $\text{var}(\lambda'y) = \lambda'V\lambda \geq 0$, for any y_i, any n and all λ, and zero only for $\lambda = 0$) the estimation of parameters would still be necessary (Christakos, 1984).

Estimation of $c(r)$ after subtraction of an estimate of the common unknown expectation, followed perhaps by iterations, may lead to completely erroneous results (Matheron, 1971, p. 196, excercises 1 and 2; Cressie, 1987, pp. 427-428). Therefore attention must be restricted to contrasts, i.e. linear combinations $\lambda'y$ with λ orthogonal to 1_n. Previous assumptions will be replaced with $E(\lambda'y) = 0$ for all y and all permissible λ, while $E(\lambda'y)^2 = \text{var}(\lambda'y)$ will be equal to $\lambda'G\lambda$, where $G = \{g_{ij}\}$ and g_{ij} is the pseudocovariance, or generalized covariance, between y_i and y_j (Starks and Fang, 1982; Corsten, 1989; Stein et al., submitted); $\lambda'G\lambda$ will be positive for all permissible λ, (but will be zero for $\lambda = 0$). This pseudocovariance, or pseudocovariogram, is again dependent on $h = z_i - z_j$, h affects g_{ij} only by its length $|h| = r$ (isotropy) and $g(r)$ is the pseudocovariance function, expressing $\{g_{ij}\}$ in the distance r between z_i and z_j. Note that G is invariant under addition of any constant α to $g(r)$, since $\lambda'(G+\alpha 1_n 1_n')\lambda$ is equivalent to $\lambda'G\lambda$.

The simplest contrast is the difference $y_i - y_j$, with variance $g_{ii} - 2g_{ij} + g_{jj} = 2[g(0) - g(r)]$. If $g(r)$ is chosen such that $g(0) = 0$, this reduces to $-2g(r)$. Hence $\tfrac{1}{2}\text{var}(y_i - y_j)$ $= \tfrac{1}{2}E(y_i - y_j)^2 = -g(r)$, which by definition equals $\gamma(r)$, the semivariogram, always satisfying $\gamma(0) = 0$. An example of a permitted semivariogram is the so-called exponential model with

46

parameters A and b,

$$\gamma(r) = A*(1-\exp(-r/b)). \qquad (1)$$

In many studies, the semivariogram is continuous everywhere but may exhibit a discontinuity at the origin, the nugget-effect. This nugget effect of size C may be added to (1), yielding an extended exponential model:

$$\gamma(r) = C*(1-\delta(r))+A*(1-\exp(-r/b)), \qquad (2)$$

where $\delta(r) = 1$ for $r = 0$, and $\delta(r) = 0$ elsewhere.

2. Nested sampling.

Several authors recommend a nested sampling procedure for exploring the form of the semivariogram in contrast to all points of a regular lattice of squares laid out on a two-dimensional area of interest, or to simple random sampling (e.g. Goss and Garrett, 1985; Burrough, 1986; Webster, 1985; Riezebos, 1989; Oliver and Webster, 1985). The value of the conclusions based on this procedure was questionned by the authors and by J. Bouma, professor in soil science (pers. comm.), which initiated this research.

An example of a five stage balanced nested design in a 750m by 750m area may run as follows (Fig. 1): first choose arbitrarily 3 vertices of an equilateral triangle with sides of 375 meters with its center in the center of the area, that is: with a fixed center, but with random orientation. Next, choose at random 3 new equilateral triangles, each with sides of 75 m and with its center in a vertex of the previous triangle. Next
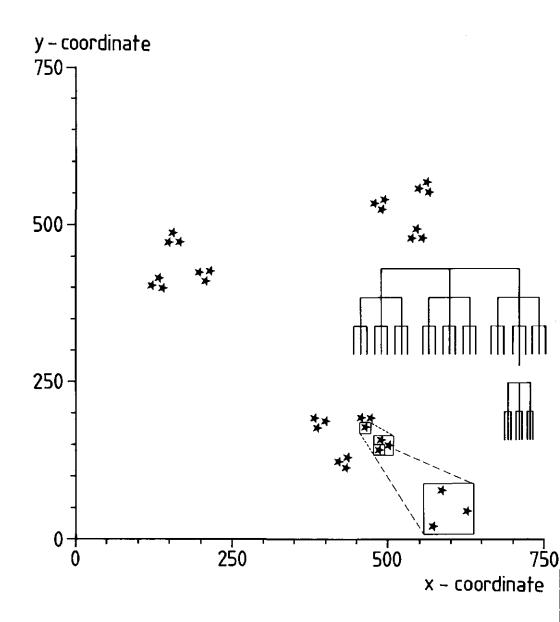
47

Fig. 1. Lay-out of a completely balanced nested sampling scheme
with 5 levels and 3 branches at every stage.

choose at random 9 new equilateral triangles, each with sides of 15m and with its center in a vertex of the last 3 triangles. Next choose at random 27 new equilateral triangles, each with sides of 3m and with its center in a vertex of the last 9 triangles. Next choose at random 81 new equilateral triangles, each with sides of 0.6m and with its center in a vertex of the last 27 triangles. Finally take an observation in each of the 243 vertices or points selected in the last stage (Table 1).

Modifications, for instance by taking one observation in the center of 60 from the last 81 triangles, and three in the vertices of the remaining 21 triangles, implying a reduction to 123 observations, or by unintentional failures, will lead to an unbalanced nested set. Of course, one may choose a different number of stages and two instead of three equidistant points at each stage. Remark that in a two-dimensional space it is impossible to take four equidistant points, in contrast to a three-dimensional space.

The cluster distances are actually approximate distances, as the distances between any element in one cluster and any

Table 1. Lay-out of a completely balanced 5 stage nested sampling scheme with 3 vertices per cluster at every stage.

Stage	Level	# Vertices	Average distance between observations	#clusters
I	4	3	375	1
II	3	9	75	3
III	2	27	15	9
IV	1	81	3	27
V	0	243	0.6	81

element in another cluster are only approximately the same.
Therefore, confounding between different levels may arise when
successive distances in the stages of the design are not
sufficiently different. In practice, ambiguity will be avoided
by taking distances proportionally to each other, with a factor
of, say, at least 3.

3. Estimation of the semivariogram by variance components

It has been claimed that classical analysis of variance
and components of variance procedures in a nested design could
be used for estimation of the semivariogram. The only 'proof'
of this procedure was the demonstration at a balanced numerical
example, given by Miesch (1975).

We shall now show the procedure and its validity for
unbalanced designs as well. Consider a h stage nested design
with fixed distances $r_1 < .. < r_h$ among pairs of observations,
perhaps proportional to each other with a common factor,
generated according to the procedure of section 2. Let the
observations having approximately distance r_i form a class. The
collection of all such classes defines a classification A_i of
the observations. The subscript i decreases as the
classification becomes more refined, in contrast to the
labeling of the stages, which is in reverse order. Let the
number of elements in any class of A_i be denoted by the generic
symbol n_i; in particular, n_0 is always equal to 1, while
$n_h = n$. The set of vectors in R^n which are constant within each
class of any classification A_i will be called A_i as well.
Obviously, the space A_{i+1} is a subspace of A_i, which, by
setting A_0 equal to R^n, holds for i = 0,...,h-1. Note that A_h
is spanned by 1_n.

50

For a classical nested design with independent observations within each stage and with homogeneous variances at the same stage the covariance matrix V of $\underline{y}$ would be $V = \Sigma U_i U_i' \sigma_i^2$, where the jth column of U_i equals the n-vector with elements 1 in the jth class of A_i and 0 elsewhere. For observations collected according to the sampling procedure described above, the variance of contrasts will be governed by a pseudocovariance matrix G of the form

$$G = g_h U_h U_h' + (g_{h-1} - g_h) U_{h-1} U_{h-1}' + \ldots + (g_1 - g_2) U_1 U_1' - g_1 U_0 U_0', \quad (3)$$

where $g_i = g(r_i)$, $i = 1, \ldots, h$. All g_i may be replaced with $-\gamma_i = -\gamma(r_i)$. By adding the constant $\gamma_h U_h U_h'$ and on setting

$$\gamma_i = \sum_{k=0}^{i-1} \sigma_k^2, \quad (4)$$

(3) is seen to be equivalent to

$$G = \sigma_{h-1}^2 U_{h-1} U_{h-1}' + \ldots + \sigma_0^2 U_0 U_0', \quad (5)$$

that is exactly the covariance matrix V for a classical nested design. This shows that unbiased estimation of the variance components σ_i^2 _as if_ one is concerned with a classical multistage sampling procedure can be used for unbiased estimation of the values γ_i of a semivariogram for dependent observations by summing these estimates according to (4). It will be shown in section 5 that estimators for γ_i are non-negative, in contrast to those of the individual components which have a positive probability to be negative, except that for σ_0^2.

51

It will be noted that the previous demonstration is independent of the number of stages, while the requirement of balance is irrelevant. Also for unbalanced designs such an estimation procedure is justified, although the calculation of expected mean squares, necessary for variance component estimation, is more complicated.

4. Expected mean squares

Turning to relevant expected sums of squares under condition (5) for balanced as well as unbalanced nested designs, we define P_i to be the orthogonal projection on the orthogonal complement of A_{i+1} in A_i ($i = 0, \ldots, h-1$). Then $P_i \underline{y}$ is the difference vector between the projection of $\underline{y}$ on A_i and its projection on A_{i+1}, representing variation between the classes of A_i within those of A_{i+1}.

The required expected sum of squares is $E[(P_i\underline{y})'(P_i\underline{y})] = E[\underline{y}'P_i\underline{y}] = tr(P_iV)$, since $E(P_i\underline{y}) = 0$.

To evaluate $tr(P_iV) = \sum\limits_{j=1}^{h-1} tr(P_iU_jU_j')\sigma_j^2$, it is noted first that $P_iU_j = 0$ for $j > i$, since each column of U_j does not alter by projection on A_i or A_{i+1}. Next, observing that P_i may be written as $U_i(U_i'U_i)^{-1}U_i' - U_{i+1}(U_{i+1}'U_{i+1})^{-1}U_{i+1}'$ we find by standard methods that

$$tr(P_iV) = \sum\limits_{j=0}^{i} tr(P_iU_jU_j')\sigma_j^2$$

$$= \sum\limits_{j=0}^{i} \{\sigma_j^2 \sum n_j^2 \, (\frac{1}{n_i} - \frac{1}{n_{i+1}})\}, \tag{6}$$

52

where the second summation extends to all classes of A_j.

The coefficient k_{i0} of σ_0^2 in $tr(P_iV)$ is $\Sigma n_0^2(1/n_i - 1/n_{i+1})$, which equals the number of classes of A_i minus that of A_{i+1}, that is the dimension of P_iR^n, the divisor for obtaining mean squares from sums of squares.

With $k_{ij} = \Sigma n_j^2(1/n_i - 1/n_{i+1})$ for $i \geq j \geq 0$, summation extending to all classes of A_j, we have the following expected sums of squares equivalent to (4):

within A_1

$$E(\underline{y}'P_0\underline{y}) = k_{00}\sigma_0^2$$

between A_i within A_{i+1}

$$E(\underline{y}'P_i\underline{y}) = \sum_{j=0}^{i} k_{ij}\sigma_j^2 \qquad (7)$$

between A_{h-1}

$$E(\underline{y}'P_{h-1}\underline{y}) = \sum_{j=0}^{h-1} k_{h-1,j}\sigma_j^2.$$

Gates and Shiue (1962) at the one hand and Gower (1962) at the other hand gave analogous results, without proof, however. They also provided numerical examples both with four unbalanced stages.

5. Estimation of semivariogram values

Equating $E(\underline{y}'P_i\underline{y})/k_{i0}$ to $MS_i = y'P_iy/k_{i0}$, one can solve the linear equations emerging from (7) to obtain unbiased estimators $\hat{\sigma}_i^2$ of σ_i^2, which by summing according to (4) yield unbiased estimators of h semivariogram values. After some algebraic manipulations, one can deduce with the following successive operators

$$Q_0 = \frac{1}{k_{00}} P_0$$

$$Q_i = \frac{1}{k_{ii}} P_i + \sum_{j=0}^{i-1} \frac{k_{i,j+1} - k_{ij}}{k_{ii}} Q_j \tag{8}$$

that

$$\tilde{\gamma}_{i+1} = y'Q_i y \tag{9}$$

for $i=0,\ldots,h-1$. From (8) and (9) it follows that the estimates $\tilde{\gamma}_i$ are non-negative, as stated in section 4, Q_i being a linear combination of orthogonal projections with positive coefficients, due to the inequality $k_{ij} < k_{i,j+1}$ for all i; this is based on the fact that the sum of squares of a set of positive numbers is smaller than the square of the sum of the same set. With semivariogram values thus estimated one may either guess the graph of the semivariogram or adjust a permissible models

As is shown in the Appendix $\tilde{\gamma}_i$ applied to balanced designs is equivalent to the intuitive unbiased semivariogram estimator $\hat{\gamma}_i$, defined as half the mean of squared differences among all pairs of observations at distance r_i. This is in general not true for unbalanced designs. The latter estimator can also be written as

$$\hat{\gamma}_i = (1/2N_i)y'D_i y, \tag{10}$$

where N_i is the number of pairs with distance r_i and the symmetric matrix D_i equals $U'_{i-1}U_{i-1} - U'_i U_i + H_i$, where H_i is a diagonal matrix with $\{H_i\}_{jj} = n_i - n_{i-1}$, n_i and n_{i-1} being the size of the class of A_i and A_{i-1}, respectively, containing the jth element.

54

Both estimators (9) and (10) can be compared as to their variance under the assumption of normality (or slightly weaker assumptions). It is well known that $\text{var}(\underline{y}'A\underline{y}) = 2\text{tr}(AVAV)$ for any symmetric A (cf. Searle, 1987). Hence $\text{var}(\tilde{\gamma}_i) = 2\text{tr}(Q_{i-1}VQ_{i-1}V)$ and $\text{var}(\hat{\gamma}_i) = (1/2N_i^2)\text{tr}(D_iVD_iV)$, an expression which is more tractable than the one found in (Cressie, 1985). We have established that it depends on the values of γ_i whether one estimator has lower variance than the other. We recall, that an UMVU estimator of the variance components or their combinations only exists for balanced designs (Lehmann, 1983).

6. Testing the significance of contributions to the estimated semivariogram.

Extrapolating the form of the semivariogram from a few points of its graph, one may pose the question if addition of the new variance component estimate $\hat{\sigma}_i^2$ will lead to a significant increase of $\tilde{\gamma}$. As r increases, it is of interest to know if one is close to such an upper bound, the so-called sill, or not at all; observations at a distance r such that $\gamma(r)$ is close to the sill can be considered as uncorrelated, as follows from the corresponding covariance function c(r) which is close to zero at those values of r. Finding a lower bound for the distance at which observations are uncorrelated, the so-called range, may be useful for carrying out statistical tests assuming independence of observations. Within the range more complicated procedures of interpolation are necessary, and thus thoroughly affect the sampling plan.

Testing $\hat{\sigma}_i^2 = 0$ against $\hat{\sigma}_i^2 > 0$ under the assumption of normality of the observations is straightforward with the usual

F-test applied to the test statistic

$$\frac{y'P_i y/k_{i0}}{y'P_{i-1} y/k_{i-1,0}} \tag{11}$$

with k_{i0} and $k_{i-1,0}$ degrees of freedom, respectively, provided the design is balanced. If the design is unbalanced, such a test is justified only for $i = 1$; for higher values of i, the test statistic above is not distributed as the F-statistic under the null hypothesis, deviations becoming larger as the lack of balance increases. In the absence of balance, however, it will remain possible to test $\sigma_1^2 + \ldots + \sigma_i^2 = 0$, equivalent to $\sigma_1^2 = \ldots = \sigma_i^2 = 0$, for $i = 1, \ldots, h-1$ by using $y'P_i y/k_{i0}$ over $y'P_0 y/k_{00}$ as a continuation of testing $\sigma_1^2 = 0$, but this procedure will not be highly informative.

7. Numerical comparison of different sampling procedures

In this numerical excercise one balanced and one unbalanced nested sampling plan are compared with four other plans. As no study is available in which they are applied simultaneously, a random field, with expectation equal to 0 and an extended exponential semivariogram with $C=1$, $A=10$ and $b=125$ (see section 1) was generated using the turning bands method (Mantoglou and Wilson, 1982) to which a nugget error was added. For the one-dimensional realizations along the turning bands the method of Shinozuka and Jan, 1972, has been used. Discrete values were generated with a step size of 7m along 16 turning bands lines, which are evenly spaced along the unit circle of an arbitrary origin. The number of harmonics M was put equal to
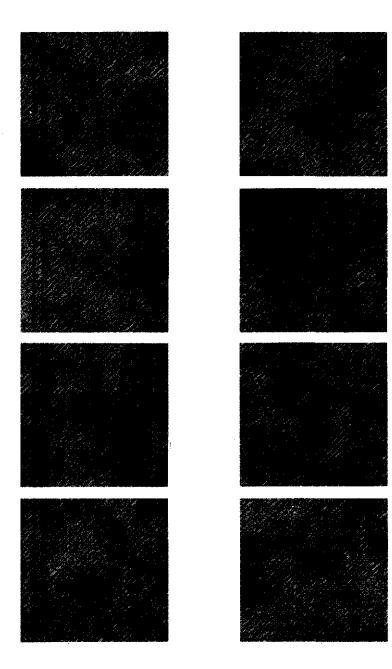
56

Fig. 2. The eight realizations of a stochastic field with mean value 0 and semivariogram $\gamma(r) = 1-\delta(r)+10*(1-\exp(r*0.008))$, $\delta(r) = 1$ for $r = 0$, and 0 elsewhere.

100, the maximum frequency at which the spectrum is truncated, Ω, was set equal to 0.32. We simulated this field eight times, keeping the value for every parameter constant. Visual displays of these fields are given in Fig. 2, black corresponding to the value 16, white to the value -16.

The following plans have been considered:

(a1) a 5 stage nested design with three branches at each distance level, as outlined in section 2. The largest distance equals 375m, next distances decrease proportionally with a factor 1/5, yielding 243 observations;

(a2) a 5 stage nested design like (a1), modified by taking one observation in the center of 60 from the last 81 triangles, and three in the vertices of the remaining 21 triangles, implying a reduction to 123 observations;

(b) a 15 by 15 points grid, with square grid cells of size 50m for which the lower left point was randomly chosen within the 50m by 50m south-west corner of the field, yielding 225 observations (Fig. 3a);

(c) a 14 by 14 points grid, with square grid cells with size 53m for which the lower left point was randomly chosen within the 53m by 53m south-west corner, extended with 50 additional observations at distance 0.6m from randomly chosen grid nodes, yielding 246 observations (Fig. 3b);

(d) a set of 230 randomly located observations (Fig. 3c).

(e) a set of 250 observations on 5 equidistant transects; transects are 150m apart, observations on each transect are 15m apart (Fig. 3d).

In sampling plans (a1), (b), (c), (d) and (e) we aimed at approximately the same costs, positioning 225 to 250
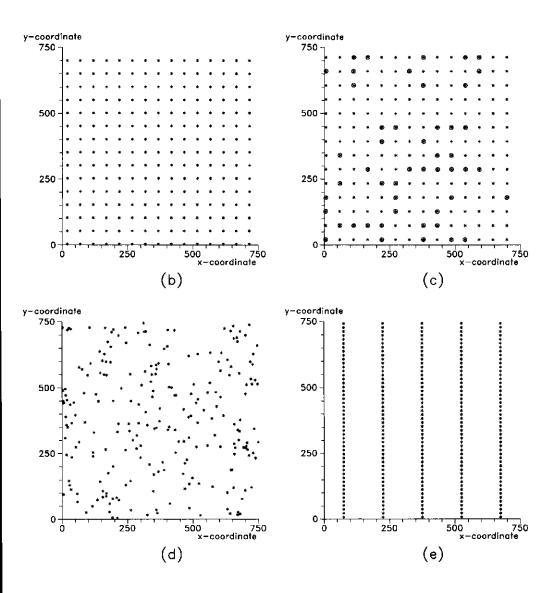
58

Fig. 3. The sampling schemes according to plans (b), (c), (d), and (e). Nodes with neighbours at 0.6m in plan (c) are encircled.

59

observations in a region of 750m by 750m. This observation density is common practice for soil investigations at a scale 1:10,000, for which four observations per squared cm map sheet are recommended. Plan (a2) was included to investigate the effect of passing from a balanced to an unbalanced nested design.

To estimate the semivariogram parameters for the schemes (b) - (e) 25 distance classes of 15m have been used. A weighted least squares regression model was applied to estimate the parameters of the extended exponential model, with weights equal to the number of pairs of observations in a distance class. Estimates for plans (a1) and (a2) are based on the intuitive estimator $\hat{\gamma}_i$, whereas the estimators $\bar{\gamma}_i$, slightly

Table 2. *Estimated variance components and semivariogram values for the nested designs (a1) and (a2), calculated for the second realization.*

Level i	$\hat{\sigma}_i^2$	ANOVA $\bar{\gamma}_i$	Half mean squared pair differences		
			av. distance	# pairs	$\hat{\gamma}_i$
0	1.862	1.862	0.60	243	1.862
1	0.183	2.045	3.02	729	2.045
2	1.135	3.180	15.10	2187	3.180
3	4.744	7.924	75.52	6561	7.924
4	2.864	10.788	377.61	19683	10.788

Balanced

Level i	$\hat{\sigma}_i^2$	ANOVA $\bar{\gamma}_i$	av. distance	# pairs	$\hat{\gamma}_i$
0	2.589	2.589	0.60	63	2.589
1	0.079	2.669	3.01	177	2.450
2	0.943	3.612	15.18	579	3.642
3	5.554	9.166	75.96	1669	9.617
4	2.446	11.612	382.49	5015	11.466

Unbalanced

60

Table 3. *Estimated parameters of semivariogram,* $\gamma(r) =$
$C*(1-\delta(r))+A*(1-exp(r*b))$, $\delta(r) = 1$ *for* $r = 0$ *and* 0
elsewhere, with 6 different sampling schemes for
8 realizations. Input values are $C = 1$, $A = 10$,
$b = 125$.

Sampling plan	a1	a2	b	c	d	e
Realization						
1	1.46	0.92	0	0.49	2.54	0
2	1.39	1.26	1.35	3.18	1.32	0.85
3	0.79	0	0.59	1.17	2.32	0.36
4	0.84	0.74	1.63	0.69	3.57	0.03
5	0.85	0.44	0.19	0.32	1.51	0
6	1.49	1.58	0	0	0	1.19
7	2.90	2.43	2.03	2.47	0.04	0.84
8	2.35	2.45	2.06	0	0	3.07
					Value for C	
1	4.00	3.95	9.80	9.11	8.93	10.20
2	9.43	10.22	9.77	8.88	10.36	8.72
3	7.34	7.51	9.25	8.22	7.52	12.13
4	4.94	6.01	13.63	14.49	42.43	14.22
5	5.94	5.99	8.49	6.88	7.09	7.37
6	10.59	10.87	9.61	8.98	7.71	7.86
7	61.98	15.59	7.53	7.91	10.01	9.01
8	3.79	6.96	7.26	10.40	10.26	5.91
					Value for A	
1	23	20	64	67	102	51
2	64	46	107	166	115	76
3	35	23	106	93	98	102
4	21	22	158	161	137	103
5	19	19	47	43	59	43
6	80	139	60	57	42	71
7	2799	352	100	96	60	57
8	220	488	131	80	67	129
					Value for b	

different for the unbalanced design might have served the same purpose. The latter are presented in Table 2 for the second realization, which yielded a semivariogram closest to the true model.

Comparing the eight realizations we notice apparent differences between the effects of the sampling plans (Table 3). Estimation of the nugget effect C is relatively safe for all sampling plans. Estimation of the sill parameter A exhibits large differences for the nested plans: for example for plan (a1) values of A in the range from 3.79 to 61.98 are obtained, but also the random plan (d) yields one remarkable high value of 42.43. Concerning the parameter b, we notice that this is usually underestimated (in 37 of the 48 cases), which appears to be irrespective of the sampling plan. The values for the nested schemes (a1) and (a2), however, show variation to a far higher extent than the schemes (b) - (e).

In order to examine the effect of the sampling plans on the quality of predictions, the 48 semivariograms in Table 3 were used to obtain best linear predictions in unvisited locations by means of Kriging. To test the quality of the predictions, a test set was created, consisting of 100 ran-domly located points for which values were obtained by means of the same turning band lines which were used to create the eight realizations. Because points at a relatively large distance contribute only little as compared to points at a relatively short distance, neighbourhoods of 2, 4, 8, 12 and 20 points, respectively, were used. Both the mean of the squared errors of the predictions (MSE) and the mean of the prediction error variance (MVP) were calculated and averaged for the eight realizations (Fig. 4). MSE and MVP values do not differ substantially as may be expected from the unbiasedness of the
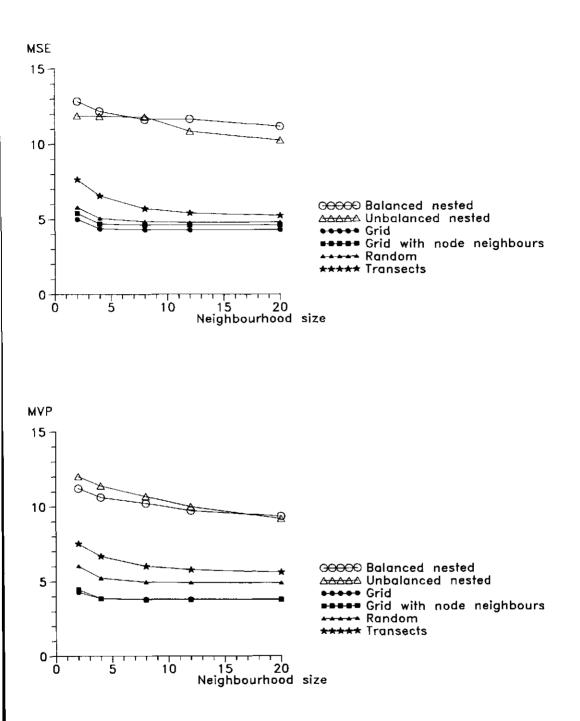
MSE

MVP

Fig. 4. *MSE and MVP values obtained for the six different*
       sampling plans as a function of the neighbourhood size,
       averaged over the eight realizations.

Table 4. *Hierarchical Analysis of Variance for plans (a1) and (a2) at h = 5 levels for the second realization.*

Level	Stage	d.f.	s.s.	Coefficients k_{1j} for σ_j^2				
				σ_0^2	σ_1^2	σ_2^2	σ_3^2	σ_4^2
4	I	2	745.35	2.000	6.000	18.000	54.000	162.000
3	II	6	844.27	6.000	18.000	54.000	162.000	
2	III	18	227.33	18.000	54.000	162.000		
1	IV	54	130.19	54.000	162.000			
0	V	162	301.57	162.000				
Total		242	2248.70				Balanced	
4	I	2	371.23	2.000	3.986	9.672	28.292	81.545
3	II	6	488.84	6.000	11.159	27.515	80.391	
2	III	18	125.53	18.000	32.974	80.911		
1	IV	54	145.61	54.000	72.857			
0	V	42	108.75	42.000				
Total		122	1239.96				Unbalanced	

64

linear predictor. We observe that nested designs lead to prediction error variances which are about twice those obtained with the other designs. Thus, nested designs are inadequate to carry out predictions. It is further noted that plan (e), sampling along transects, provides predictions which are less reliable than those obtained by plans (b), (c) and (d).

Concerning the neighbourhood size we observe an increasing precision with increasing size. The gain is small, however, when 20 instead of 12 observations are used, pointing to an adequate neighbourhood size of 12 observation points for spatial interpolations.

For the second realization a nested analysis of variance has been carried out on the nested data (a1) and (a2) along the lines of section 6 (Table 4). Testing the increase of γ beyond 75m in (a1) gives an F-value equal to 2.65 with 2 and 6 degrees of freedom, which is not even significant at an $\alpha = 0.10$ level (F-value = 3.46), whereas testing the increase of γ beyond 15m gives an F-value equal to 11.14 with 6 and 18 degrees of freedom, significant at an $\alpha = 0.05$ level (F-value = 2.66). This indicates that there is no increase of γ beyond 75m, which is rather strange as compared to the value of 125m for b. We also notice the small number of degrees of freedom available for comparisons at stage I. In plan (a2) the test of $\sigma_1^2 = 0$ yielded an F-value equal to 1.04, with 42 and 54 degrees of freedom, which is not significant at an $\alpha = 0.05$ level (F-value = 1.7). Testing $\sigma_1^2 + \sigma_2^2 = 0$, $\sigma_1^2 + \sigma_2^2 + \sigma_3^2 = 0$ and $\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \sigma_4^2 = 0$, yielded F-values equal to 2.69, 31.47 and 71.69, respectively, all significant at $\alpha = 0.05$ level. Clearly, the increase of γ starts to be significant with respect to stage V at stage III and continues to be so in the next stages as well.

Of course, it is possible, at least in principle, to select a subset of the observations from plan (c) or from plan (d) which more or less exhibits a nested structure, for example with three levels and two observations per cluster, and to carry out the tests along the lines described above for plans (a1) and (a2). But we feel that it is more valuable to have an overall picture of the semivariogram, than to have only the values at a few isolated points.

From this study we conclude that estimation of semivariogram parameters from nested designs performs worse than from the other designs. The 15m by 15m grid performs best. The claim that nested designs are especially appropriate for estimating the nugget effect is not justified. Nested designs are especially unsuitable when they are used to predict values at unvisited locations after estimation of the semivariogram parameters. As nested designs involve very intense sampling of a small number of sub-areas, the observations may be highly influenced by local anomalies.

8. References

Burrough, P.A. (1986), *Principles of geographical information systems for land resources assessment*. Oxford, Clarendon press.

Christakos, G. (1984), "On the problem of permissible covariance and variogram models," *Water Resources Research*, 20, 251-265.

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

Cressie, N.A. (1985), "Fitting variogram models by weighted least squares," *Mathematical Geology* 17, 563-586.

Cressie, N.A. (1987), "A nonparametric view of generalized
    covariances for kriging," *Mathematical Geology* 19, 425-
    449.

Gates, C.E., and Shiue, C. (1962), "The analysis of variance of
    the s-stage hierarchical classification," *Biometrics*, 18,
    529-536.

Goldberger, A.S. (1962), "Best linear unbiased prediction in
    the generalized linear model," *Journal of the American
    Statistical Association*, 57, 369-375.

Goss, T.I., and Garrett, R.G. (1978), "A new unbalanced nested
    ANOVA model in geology - a down to earth design," *American
    Statistical Association, Statistical Computing
    Proceedings*, 360-365.

Gower, J.C. (1962), "Variance component estimation for
    unbalanced hierarchical classifications," *Biometrics*, 18,
    537-542.

Lehmann, E.L. (1983), *Theory of point estimation*. New York,
    Wiley.

Mantoglou, A., and Wilson, J.L. (1982), "The turning bands
    method for simulation of random fields using line
    generation by a spectral method," *Water Resources
    Research*, 18, 1379-1394.

Matheron, G. (1971), "The theory of regionalized variables and
    its application," *Cahiers du Centre de Morphologie
    Mathematique, no. 5*, Ecoles des Mines, Fontainebleau,
    France.

Miesch, A.T. (1975), "Variograms and variance components in
    geochemistry and ore evaluation," *Geological Society of
    America, Memoir* 142, 333 - 340.

Oliver, M.A., and Webster, R. (1986), "Combining nested and
    linear sampling for determining the scale and form of

spatial variation of regionalized variables,"
Geographical analysis, 18 (3), 227-242.

Riezenbos, H.Th. (1989), "Application of nested analysis of
variance in mapping procedures for land evaluation," *Soil
use and management,* 5, 25-30.

Searle, S.R. (1987), *Linear models for unbalanced data.* New
York, Wiley.

Shinozuka, M., and Jan, C.M. (1972), "Digital simulation of
random processes and its applications," *Journal of Sound
Vibrations,* 25, 111-128.

Starks, T.H., and Fang, J.H., (1982), "On the estimation of the
generalized covariance function," *Mathematical Geology*
14, 57-64.

Stein, A., Van Eijnsbergen, A.C., and Barendregt, L.C. (Subm),
"Cokriging non-stationary data," *Mathematical Geology.*

Stein, A., and Corsten, L.C.A., (1990), "Universal Kriging and
Cokriging as a regression procedure," *Biometrics, in
press.*

Webster, R.A. (1985), "Quantitative spatial analysis of soil in
the field," in *Advances in soil science 3,* ed. D.A.
Stewart, Springer Verlag, New York.

68

8. Appendix

In order to evaluate the equivalence of $\hat{\gamma}_i$ and $\tilde{\gamma}_i$ for balanced designs we will make use of the following two identities.

1. Consider the elements y_i, $i=1,..,n$ with mean value $\bar{y}$. Then

$$\sum_{i=1}^{n} (y_i-\bar{y})^2 = \frac{1}{n} \sum_{i<j} (y_i-y_j)^2 \tag{A1}$$

2. Let two classes of A_1 within the same class of A_2 contain the elements y_{1j}, $j=1,..,n_1$ and y_{2j}, $j=1,..,n_1$, respectively. Then

$$(\sum_{j=1}^{n_1} y_{1j} - \sum_{j=1}^{n_1} y_{2j})^2 =$$

$$\sum_{j=1}^{n_1} \sum_{k=1}^{n_1} (y_{1j} - y_{2k})^2 - \sum_{j<k}(y_{1j} - y_{1k})^2 - \sum_{j<k}(y_{2j} - y_{2k})^2. \tag{A2}$$

The first term of the right hand side refers to pairs of elements in different classes of A_1 and the second and the third term to pairs of elements within the same class of A_1.

In order to establish the equivalence of $\hat{\gamma}_1$ and $\tilde{\gamma}_1$, let the jth element in class i of A_1 be denoted by y_{ij}, and the mean of that class by $\bar{y}_i$. Then as now $k_{00} = (n_1-1)n/n_1$, it follows by use of (A1)

$$\tilde{\gamma}_1 - \frac{1}{k_{00}} y'P_0 y - \frac{1}{k_{00}} \sum_i \sum_{j=1}^{n_1} (y_{ij} - \bar{y}_i)^2$$

$$- \frac{1}{(n_1-1)n} \sum_i \sum_{j<k} (y_{ij} - y_{ik})^2 \tag{A3}$$

Since the total number of different pairs of observations with distance r_1 equals $\frac{1}{2}(n_1-1)n$, (A3) equals $\hat{\gamma}_1$ indeed.

In order to establish the equivalence of $\hat{\gamma}_2$ and $\tilde{\gamma}_2$, let the pth observation in the jth class of A_1 in the ith class of A_2 be denoted by y_{ijp}, the mean value in the jth class of A_1 in the ith class of A_2 by m_{ij} and the mean value of the ith class of A_2 by m_i. Then, as now $k_{11} = (n_2-n_1)n/n_2$, it follows by use of (A1) and (A2) that

$$\frac{1}{k_{11}} y'P_1 y - \frac{n_1}{k_{11}} \sum_i \sum_{j=1}^{n_2/n_1} (m_{ij} - m_i)^2$$

$$= \frac{n_1^2}{(n_2-n_1)n} \sum_i \sum_{j<k} (m_{ij} - m_{ik})^2$$

$$= \frac{1}{(n_2-n_1)n} \sum_i \sum_{j<k} \{ \sum_{p=1}^{n_1} \sum_{q=1}^{n_1} (y_{ijp} - y_{ikq})^2 - \sum_{p<q} (y_{ijp} - y_{ijq})^2 - \sum_{p<q} (y_{ikp} - y_{ikq})^2 \}$$

Since this expression only involves pair differences at distances r_2 and r_1, respectively, it can be written as a linear combination of $\hat{\gamma}_2$ and $\hat{\gamma}_1$. The coefficient of $\hat{\gamma}_2$ will be equal to 1, while that of $\hat{\gamma}_1$ is the opposite of $(k_{11}-k_{10})/k_{11} - (n_1-1)/n_1$.

For higher levels similar arguments apply.

70

CHAPTER I.2. SEQUENTIAL SAMPLING TO MEASURE THE INFILTRATION RATE WITHIN RELATIVELY HOMOGENEOUS SOIL UNITS

by

A. Stein, J. Bouma, S.B. Kroonenberg and S. Cobben.

SEQUENTIAL SAMPLING
TO MEASURE THE INFILTRATION RATE WITHIN
RELATIVELY HOMOGENEOUS SOIL UNITS

A. **Stein**, J. **Bouma**, S.B. **Kroonenberg** & S. **Cobben**, Wageningen

Summary

The statistical prediction techniques Trend Surface Analysis, Kriging and Co-kriging are regularly used to provide predictive single-value soil maps. Kriging and Co-kriging perform well in situations where regionally distributed variables show clear spatial structure. If the amount of variation of the variables under study is relatively small, however, no gain in precision is gained by Kriging and Co-kriging, as compared with simply averaging over an area or by applying Trend Surface Analysis. This study was carried out on an older terrace of the Allier river in Central France where attention was focused on the infiltration rate (Inf). Kriging log(Inf) values did barely improve the predictions as compared with Trend Surface Analysis, whereas deleting part of the observations did not result in serious changes of predictions. A satisfactorial first approach was obtained by regarding the observations as being independent. To determine the sample size which provides sufficient information with respect to the soil characteristic under study use can be made of the sequential t-test as is illustrated.

Application of this test resulted in eight measurements which represented a saving of 70% as compared with the original standard scheme.

1 Introduction

Introduction of environmental laws in various countries is associated with the need for quantitative expressions for environmental land characteristics. Whether or not certain normative levels are exceeded may be a crucial factor in court, where a quantitative statistical analysis is bound to have a higher impact than a comparative qualitative assessment. More emphasis on measurement has, however, major implications for soil characterization programs to be executed in the field by soil survey personnel. The associated costs may be prohibitively high and development of optimal sampling programs within particular land units is therefore important. Recent literature suggests the need for a preliminary sampling program to establish patterns of spatial variability to be followed by the sampling program itself in which the number of samples is optimized while one takes into account the obtained spatial variability structure and the required accuracy. Statistical prediction techniques, such as Kriging and Co-kriging can then be used to efficiently

predict values for any location in the area (e.g. CORSTEN 1985).

However, even a preliminary sampling program can be rather time consuming when complex measurements are involved. This study is concerned with measurement of the infiltration rate, a land characteristic for which measurements are rather complex and time consuming. Rather than to proceed directly with the execution of a standard preliminary sampling program, which may require at least some thirty measurements, we have explored the possibility ot use geological and pedological descriptions of the land unit to be characterized, to possibly reduce the number of measurements. As it turns out, fortunately, some land units have soils or soil horizons with a low spatial variability. Much is gained if this can be established without an extensive preliminary sampling program involving the uncritical application of a standard geostatistical sampling procedure.

2 Study area

The study was carried out on one of a flight of nine fluvial terraces of the Allier river in the Limagne Graben in the French Central Massif. The sediments of the Allier river and its terraces comprise gravels, sands and clay, mainly of granitic-gneissic and basaltic provenance (KROONENBERG et al. 1982). The terraces are numbered according to French usage from Fz (youngest) to Fs (oldest). The bulk of the terraces consists of coarse gravel beds and sand lenses of several metres in thickness, but the uppermost 1 or 2 metres are usually clayey to sandy. Weathering of the volcanic components and progressive clay illuviation lead to a substantial increase in clay content from the soils of the lower, younger terraces to those of the higher, older ones (FEIJTEL et al. 1988). Variability of the soil infiltration rate from one terrace to another, therefore, is mainly dependent upon soil age. Data on spatial variability of infiltration rates within the youngest Fz terraces suggest a strong dependence upon soil texture, which is controlled in turn by sedimentation patterns. However, in the older Fv terrace segment, being studied here, little variability in soil texture could be detected in the field. As impeded drainage is one of the main limiting factors of the soils of this terrace, spatial variability of infiltration rate was investigated in this study, using geostatistical methods.

3 Sampling and statistical methods

The sampling design used here was similar to stratified sampling (STEIN et al. 1988). Different river terraces were distinguished as major land unit by means of geological and pedological descriptions. Within land units a fixed sampling design was followed, which was determined in advance.

Infiltration rates were measured with the standard double-ring infiltrometer (FAO 1979). Measurements were continued until a steady infiltration rate was reached for a period of at least one hour. A total of twenty-eight measurements were made in the top of the Bt-horizon at a depth of 30 cm below surface following a pre-determined grid (fig.1). This number was considered to be a compromise between the minimum number required to apply Kriging and Co-kriging and the maximum number that could reasonably be sampled in the available time. Four-
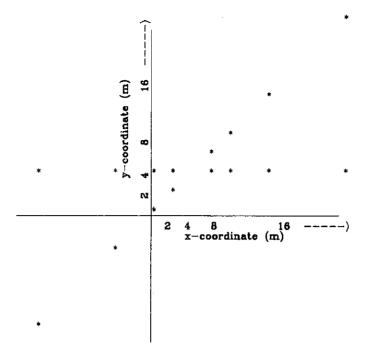
Fig. 1: *Central part of the sampling scheme.*

teen measurements were carried out in two directions on grid nodes with distances that increased exponentially from the centre (2 m, 4 m, 8 m, ..., 128 m). Fourteen additional measurements were carried out perpendicular to these directions at a distance of 2 m from a 2 m node, at a distance of 4 m from a 4 m node, etc. In all, therefore, a total of twentyeight measurements were obtained on two transects crossing with an angle of 45°.

Observations on infiltration rate (Inf) are reduced to their logarithmus (log(Inf)), to obtain a population which is approximately normal (fig.2). Co-kriging was not considered, as the simple soil variables were only slightly correlated with infiltration rate (tab.1).
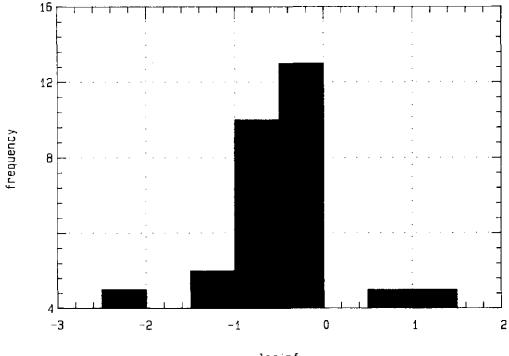
The statistical prediction techniques

Trend Surface Analysis (TSA) and Kriging are well described elsewhere (e.g. WATSON 1972; DELFINER 1976). Two different measures are used to evaluate the quality of predictions of infiltration rates for locations where no measurements were made:

1. The Mean of Squared Errors (MSE). Predictions t_i $(i = 1, ..., n)$ are carried out in each observation point on the basis of the log(Inf) observations y_i at all the remaining $(n - 1)$ points. A measure of the quality of the prediction is given by:

$$MSE = 1/n * \sum_{i=1}^{n} (t_i - y_i)^2$$

2. The Mean Variance of Prediction errors (MVP). As $\underline{t}_i$ $(i = 1, ..., n)$ is

75

Fig. 2: *Histogram of log(Inf) values.*

		Mean	Standard Deviation	Correlation
Log(Inf)	cm/day	-.53	.59	1.00
Thickness Bt	cm	46	9.55	-.05
Bulkdensity Bt		1.86	.13	-.21
Relative height	m	1.02	.35	.32
Shearstress		71	20	-.15
Fraction Mottles	%	16	11	.12

Tab. 1: *Mean and Standard deviation of different variables in the study, as well as the correlation coefficient with log(Inf).*

the (stochastic) predictor in the ith observation point $\underline{y}_{0i}$ ($i = 1, ..., n$) of the log(Inf) value if it would have been measured, the MVP is defined as

$$MVP = 1/n * \sum_{i=1}^{n} \text{var}(\underline{t}_i - \underline{y}_{0i})$$

where $\text{var}(\underline{t}_i - \underline{y}_{0i})$ is given by the formula for the prediction error variance under the prediction model being used, i.e. Kriging or TSA.

Two ways were chosen for investigating the differences in prediction performance by the two methods.

i) Values in the observation points were predicted disregarding the log(Inf) observations in those points and yielding the MVP. Afterwards differences between prediction and measurement were used to calculate the MSE.

ii) Predictions were carried out for 100 randomly located points within the sample area and the standard deviation of prediction error, being the square root of prediction error variance, was calculated.

Observation density was changed as follows. The complete set of 28 observations was reduced by four observations at a time, beginning from the centre, until the four outermost observations remained. A neighbourhood consisted of the observations in the reduced observation set, so its size changed proportional to the changing observation density.

A question we wanted to answer concerned the number of observations necessary to reach a predetermined precision.

This question was reformulated as: if after, say, m measurements some hypothesis is formulated concerning the distribution of the variable under study, is there any need to continue sampling, or can we stop. To investigate the need for additional observations in carrying out a pilot study, use was made of the sequential t-test (WALD 1948, KENDALL & STUART 1973, see Appendix). In this test, a hypothesis concerning the distribution of the variable under study has to be formulated, which after some measurements is accepted or rejected. In this study we hypothesized the mean μ of the log(Inf) values to be equal to some value μ_0. This hypothesis is rejected if $|\mu - \mu_0| > \delta_1 * \sigma$. Here σ is the standard deviation and $\pmb{\delta}_1$ is some value to be determined in advance. The value of δ_1 depends on the allowed tolerance of the soil characteristic under study. In this case, the standard deviation is unknown in advance, so a slight modification of the sequential t-test was used. The method is considered to be an alternative to nested sampling (WEBSTER 1985).

When Kriging was applied, generalized increments were used to determine the degree of the trend and to obtain estimates for the coefficients of the generalized covariance function (STARKS & FANG 1981). Three cases were considered: a 0th order polynomial trend with generalized covariance function $g(h) = \sigma_0 + \sigma_1 * h$, a 1st order polynomial trend with generalized covariance function $g(h) = \alpha_0$ (a pure nugget effect) and a 1st order polynomial trend with generalized covariance function $g(h) = \alpha_1 * h$. Higher order trends were not taken into account because of the supposed homogeneity of the terrace segment, other models for the generalized covariance

Degree trend	Estimated coefficients			Mean rank
	$\alpha-0$	$\alpha-1$	$\alpha-3$	
0	0.2778	-1.244	—	40.54
1	0.4136	0	0	41.75
	0	-5.339	0	45.21

Tab. 2: *Estimated coefficients of the generalized covariance function for a 0th and a 1st order trend and the mean rank of the squared increments.*

nr. obs.	MSE		MVP	
	Kr	TSA	Kr	TSA
28	.67	.36	.31	.47
24	.72	.31	.33	.47
20	.70	.38	.35	.47
16	.77	.47	.39	.48
12	.95	.62	.43	.49
8	1.16	.93	.52	.51
4	.77	.57	.84	.60

Tab. 3: *MSE and MVP values obtained by Kriging (Kr) and Trend Surface Analysis (TSA), respectively, for the 28 observations used as a test set.*

nr. obs.	100 points	
	Kr	TSA
28	.40	.47
24	.40	.47
20	.40	.47
16	.41	.48
12	.41	.49
8	.44	.51
4	.56	.56

Tab. 4: *MVP values obtained by Kriging (Kr) and Trend Surface Analysis (TSA), respectively, for 100 randomly located points in the study area.*

function did not lead to an admitted solution. The trends were compared by average ranks of the squared increments obtained with converged coefficients.

4 Results and discussion

The log(Inf) values are on the average -0.53, with an individual standard deviation of 0.59. This corresponds with an infiltration rate of about 4 mm/day, which is a very low value.

The lack of variability in infiltration rates in this terrace segment may be the result of strong clay illuviation, as well as of a decrease in the textural contrast between sandy and clayey parent materials due to profound weathering of the volcanic components in the sands.

For log(Inf) the squared increments in the observation points were calculated. A 0th order trend fitted slightly better than a 1st order trend, as the mean rank increased from 40.54 to 41.75, when the 0th order trend is compared with a 1st order trend and a pure nugget effect as a generalized covariance function (tab.2).

Using the overall mean of the log(Inf) observations as a prediction for the values in the 28 observation points yields a value for the MSE equalling 0.61. Predictions obained by Kriging are somewhat better, but differences are small (tab.3). The MSE values increase from .36 to .93 when the number of observations decreases from 28 to 8.
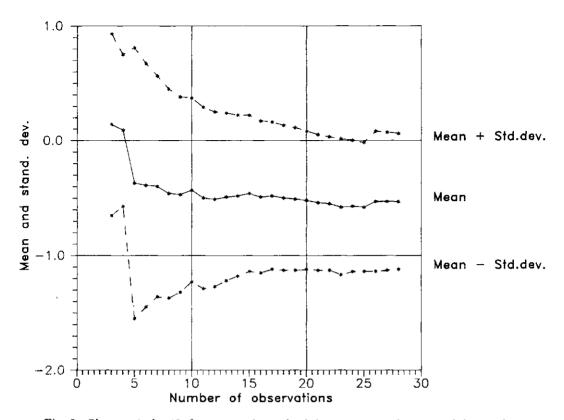
78

Fig. 3: *Changes in log(Inf) mean and standard deviation as a function of the number of observations.*

The change in the variance of the prediction error in the hundred randomly located spots is given in tab.4. No measurements were taken in these spots, so the MSE could not be calculated. The mean variance of the prediction error increases from .63 to 1.06 when the number of observations decreases from 28 to 8.

As these changes in MSE and MVP values are considered small, the practical question must be raised now as to how an unnecessarily expensive exploratory survey of 28 measurements of infiltration rates can be avoided. Firstly, a geological and pedological analysis of the area to be studied is to be made, leading to a stratification of the area (here: the flight of terraces). For every stratum (terrace) a step by step increase in observation density can be made, where the observations are considered to be independent. Next, if a clear spatial structure in the soil variables can be distinguished and modelled, we can turn to the more advanced statistical prediction techniques as Kriging and Co-kriging.

Starting with the four outermost observations and the assumption that the observations were independent, the number and the density of observations was increased. It was noted that estimates for the mean and the standard devia-
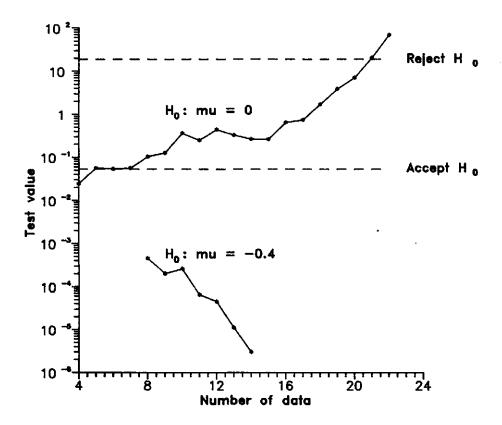
Fig. 4: *Testing the hypotheses H_0 : $\mu = 0$ (after the fourth measurement) and H_0 : $\mu = -0.4$ (after the eight measurement). Between the H_0-acceptance line and the H_0-rejecting line, applying to both hypotheses, is an intermediate zone.*

tion only slightly change (fig.3). For the first four measurements we formulated the hypothesis H_0 that the mean value μ was equal to 0, a hypothesis to be tested with error probability levels 0.05 for incorrectly deciding in favour of H_0 as well as for incorrectly rejecting H_0 in favour of the value δ_1 depending upon a feasible different value of the mean. In this case, a value of $\delta_1 = 1$ seemed to be appropriate, as a soil with an infiltration rate higher than 3 cm per day (log(Inf) ≈ 1) was judged to be different from the

hypothesized 1 cm per day (log(Inf) ≈ 0) from a practical point of view. Although at first the null hypothesis was likely to be accepted (fig.4), as soon as another measurement was taken the test value fell between the H_0-acceptance line and the H_0-rejecting line. In all 22 observations would be needed to finally reject H_0. After eight measurements were taken, H_0 was revised, leading to the hypthesis that μ be equal to -0.4 instead of 0. This hypothesis was confirmed from the start, leading to the conclusion that with only

80

eight observations enough information was obtained to decide in favour of H_0.

The procedure now can be generalized as follows. If, for different reasons, a low degree of variability is expected to occur within a stratum, observations are taken and sequentially followed, working from a low to a high observation density, by reducing the distances between observations step by step. Attention has to be focused on a hypothesis concerning the distribution of the soil characteristic under study. Locations for measurements are to be selected in advance, for instance randomly, or on predetermined transects or grids. The test value of the sequential test is calculated after each new measurement. As soon as the pre-defined limits based on confidence levels are reached, measurements are stopped, otherwise additional measurements have to be taken.

For our study, working from outside to the centre, eight measurements would have sufficed to yield estimates of infiltration rates that would not have been significantly different from results obtained by using the standard, exploratory survey of 28 measurements. The value of the standard deviation reduces from 1.18 to .60. Both values are equivalent from a practical point of view, considering values for infiltration rate. The alternative procedure, discussed in this paper, which includes a geological and a pedological analysis before measurements are made, reduced costs by an estimated 70%.

References

CORSTEN, L.C.A. (1985): Current statistical issues in agricultural research. Statistica Neerlandica 39(2), 156–168.

DELFINER, P. (1976): Linear estimation of non-stationary spatial phenomena. In: M. Guarascio et al. (eds.), Advanced geostatistics in the mining industry. 49–68. D. Reidel, Dordrecht.

FAO (1979): Soil survey investigations for irrigation. FAO Soils Bulletin 42, App. B-2, Rome.

FEIJTEL, T.C., JONGMANS, A.G., BREEMEN, N. VAN & MIEDEMA, R. (1988): Genesis of two planosols in the Massif Central, France. Geoderma, 43, 249–269.

KENDALL, M.G. & STUART, A. (1973): The advanced theory of statistics. Vol. 2 (3rd Edn.). Griffin, London.

KROONENBERG, S.B., MOURA, M.L. & JONKER, A.T.J. (1988): Geochemistry of the sands of the Allier river terraces, France. Geologie en Mijnbouw 67, 75–89.

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.

WALD, A. (1947): Sequential analysis. John Wiley, New York.

WATSON, G.S. (1971): Trend Surface Analysis. Jour. Mathematical Geology, v. 3, no. 3, 215–226.

WEBSTER, R. (1985): Quantitative spatial analysis of soil in the field. Advances in soil science, 3, 1–70.

This paper is a contribution from the Dept. of Soil Science and Geology of the Agricultural University, P.O. Box 37, 6700 AA Wageningen, The Netherlands

Appendix

The sequential probability ratio test

The Sequential Probability Ratio Test (SPRT) is designed to decide between two simple hypotheses. Suppose a random variable z has a distribution $f(z, \mu)$, and we wish to test the hypothesis H_0 that $\mu = \mu_0$ against the alternative hypothesis H_a that $\mu = \mu_1$. The test constructed decides in favour for either μ_0 or μ_1 on the basis of observations $z_1, z_2, ...$; we will suppose that if H_0 is true, we wish to decide for H_0 with probability with at least $(1 - \alpha)$, while if H_a is true, we wish to decide for Ha with probability at least $(1 - \beta)$. We calculate the likelihood ratio L_m as

$$
\begin{aligned}
L_m &= \frac{\text{Probability of observed values given} H_a}{\text{Probability of observed values given} H_0} \\
&= \Pi_{i=1}^{m} \frac{f(z_i, \mu_1)}{f(z_i, \mu_0)}
\end{aligned}
$$

The procedure to follow is: continue sampling as long as $B < L_m < A$. Stop sampling and decide for H_0 as soon as $L_m < B$ and stop sampling and decide for H_a as soon as $L_m > A$. It can be proven that $A \approx (1 - \beta)/\alpha$ and $B \approx \beta/(1 - \alpha)$.

As the standard deviation is unknown, we have to deal with a composite hypothesis, which is based upon weight functions. It is therefore reasonable to impose the following structure. For all μ satisfying $|(\mu - \mu_0)/\sigma| < \sigma_0$ it is preferred to accept the hypothesis H_0, and for all μ satisfying $|(\mu - \mu_0)/\sigma| > \sigma_1$ it is preferred to reject this hypothesis. We therefore specify three regions in the space (μ, σ), two in which H_0 is accepted and rejected, respectively,

Rejection if $|\mu - \mu_0|/\sigma > \delta_1$,

Acceptance if $\mu = \mu_0$

while the remainder is an indifference region.

The likelihood ratio with appropriate weight functions can be formulated as

$$
\frac{1/2 \left[\int_0^\infty \frac{1}{\sigma^m} \exp\{\frac{-1}{\sigma^2} \sum_{i=1}^m (z_i - \mu_0 - \delta_1 \sigma)^2\} d\sigma + \int_0^\infty \frac{1}{\sigma^m} \exp\{\frac{-1}{\sigma^2} \sum_{i=1}^m (z_i - \mu_0 + \delta_1 \sigma)^2\} d\sigma \right]}{\int_0^\infty \frac{1}{\sigma^m} \exp\{\frac{-1}{\sigma^2} \sum_{i=1}^m (z_i - \mu_0)^2\} d\sigma}
$$

where m is the number of observations. The integrals appearing above were solved by means of a numerical integration routine.

Address of authors: Dept. of Soil Science and Geology of the
A. Stein Agricultural University
J. Bouma P.O. Box 37
S.B. Kroonenberg 6700 AA Wageningen
S. Cobben The Netherlands

PART II. PREDICTIONS

CHAPTER II.1. UNIVERSAL KRIGING AND COKRIGING AS A REGRESSION PROCEDURE

by

A. Stein and L.C.A. Corsten

UNIVERSAL KRIGING AND COKRIGING AS A REGRESSION PROCEDURE.

SUMMARY

Prediction of a property on the basis of a set of point measurements in a region is required if a map of this property for the region is to be made. Of the spatial interpolation and prediction techniques Kriging is optimal among all linear procedures, as it is unbiased and has minimal variance of the prediction error. In Cokriging, which has this same attractive property, additional observations of one or more co-variables are used, which may lead to increased precision of the predictions. Both techniques are often applicable in different fields like soil science, meteorology, medicine, agriculture, biology, public health and environmental sciences (e.g. atmospheric or soil pollution). In this study we try to remove the cloud of obscurity covering the notions of Kriging and Cokriging by embedding them into regression procedures. This leads to a straightforward formulation of the two techniques. It turns out that Kriging and Cokriging differ only slightly from each other. The procedures are illustrated by two numerical examples, one to demonstrate the methodology, and one practical problem encountered in a soil study. Cokriging is found to be most valuable when a highly correlated covariable is sampled intensely.

INTRODUCTION

Traditional map production is increasingly being replaced by computer-based statistical prediction. An interesting well known prediction technique has been developed in France since 1963 by Matheron and his co-workers (cf. Matheron, 1963, 1971; Delfiner, 1976) and is generally known as Kriging, named after the South-african D. Krige, the first practitioner of the method in the fifties (cf. Krige, 1951). This technique performs well in predicting the value of a possible but actually not taken observation of a spatially distributed variable such as a mine grade (Krige, 1966), a soil characteristic (Webster, 1985), rainfall (Witter, 1984), or gene-frequency (Piazza, Menozzi and Cavalli-Sforza, 1981). After predicting a spatially distributed variable at many locations in a region of interest (e.g. in the nodes of a grid) from a few point observations, one can produce a map of this variable, e.g. by 'contours' or isopleths, differently shaded areas or graphical techniques for surfaces in three-dimensional space. This activity may be termed interpolation. The predictor in any location will be a linear combination of the observations with positive or negative weights. This technique provides not only a prediction, but also an estimate of the variance of the prediction error. So a general indication of the quality of a map can be given, determined by the data and by the configuration of the observations as a whole. Among all linear predictors Kriging is the one without systematic prediction error and with smallest prediction variance.

Kriging is not the only spatial prediction method, other non-statistical methods include tesselation and splines. Kriging, however, is the only method with a sound statistical

base and is to be applied when uncertainties exist concerning measurement error and when the variation is a function of the distance between mesasurements (e.g. Webster, 1985). Kriging is compared with splines in Dubrule, 1983.

Although Kriging is best in the above sense more information is often available to a scientist or a surveyor than what is contained in just one variable. So quite naturally an extension to Kriging emerged, known as Cokriging (Matheron, 1979; Journel & Huijbregts, 1978). This technique, with attractive properties similar to those of Kriging (e.g. unbiasedness, minimum prediction error variance), utilises observations of more than one variable in predicting one variable. It has found many applications, e.g. in soil science: the prediction of water stored in soil at 1/3 bar pressure (Vauclin et al. 1983), reducing the costs of soil surveys (Stein et al. 1988), mapping top-soil silt content (McBratney and Webster, 1983). It has also been used in other disciplines, e.g. predicting radar measured rainfall data (Krajewski, 1987), or jointly estimating peak values of earthquake response spectra (Carr and McCallister, 1985). In all situations above, the different variables are combined to give optimal predictions of one variable.

Cokriging, however, has up to now two serious disadvantages. Firstly, it is quite difficult to grasp, due to the complexity of notation (cf. Journel and Huijbregts, 1978, p. 171). Secondly, in contrast to Kriging (Cressie, 1987), it could not fully be used in the presence of a trend in the variables due to the fact that the theory of estimation of spatial cross-covariance and its limitations is not completely available. Therefore we shall restrict ourselves to the situation that covariances are known.

89

Often Lagrange multipliers are used to derive the Kriging and Cokriging formulae, and they are used in the presentation of the final formulae as well (see, for example, Journel & Huijbregts, 1978). We believe that the complexity of notation is largely due to their use in the derivation and in the presentation of the results. Further they are not necessary, as we have to minimize quadratic functions only.

It is known how Kriging can be understood in the context of linear regression applied to dependent observations (cf. Corsten, 1985, 1989). The present article shows how Cokriging can be seen in a similar fashion; Kriging emerges as a special case. There are links with other fields in statistics, e.g. query response with best linear unbiased predictors (Gianola and Goffinet, 1982), current topics in experimental designs like Bayesian analysis (Steinberg, 1985) and nearest neighbour models (Gleeson and Cullis, 1987), decomposition of the prediction error (Harville, 1985) and kernel smoothing and splines (Friedman and Silverman, 1989). In all the topics mentioned here, the variable is decomposed into three constituent parts: a fixed effect, locally dependent errors and a purely random effect. As is shown below, this applies to universal Kriging and Cokriging as well.

One of the important aspects to be stressed here is that we aim to estimate a stochastic variable, i.e. to <u>predict</u> a possible future observation, instead of to estimate a parameter, i.e. to <u>estimate the expectation</u> of an observation. In ordinary linear regression theory with uncorrelated disturbances, all with the same variance, the two approaches differ only in their standard deviations. The best prediction and the estimate of its expectation for a new point with regressor values contained in the vector x_0 will, admittedly, take the

90

same value in this case. But the standard deviation of the estimator of the expectation is the residual standard deviation, multiplied by the square root of $x_0'(X'X)^{-1}x_0$, while for the predictor the last factor must be replaced by the square root of $1+x_0'(X'X)^{-1}x_0$. Here X is the matrix where each of the k columns is an n-vector containing the observations of a regressor in the n observation points. The difference between these multiplication factors can be impressively large. Apart from numerical differences prediction of a random variable and estimation of its expectation are fundamentally different. From now on we deal mainly with the estimator of a stochastic effect, i.e. with prediction.

UNIVERSAL COKRIGING AS A REGRESSION PROCEDURE

The term regionalized variable is commonly used to indicate that a variable takes values in a specific 1, 2 or 3-dimensional space. The regionalized variable for which predictions are to be provided is called the target variable, or the *predictand*. It is assumed that the predictand has been observed in a given pattern of m points. In addition, n observations of another covariable in possibly different points are available. We restrict ourselves here to only one covariable, the more general case of k-1 covariables being treated in Appendix B. As far as notation is concerned, we adopt the convention of underlining random variables or random vectors; the transpose of a matrix or a column vector is indicated with a prime.

The prediction problem can be stated in the following way. We require a predictor $\underline{t}$ for the value $\underline{y}_0$ of the predictand at a point z_0 on the basis of the m observations at the points

$z_1, \ldots, z_m$ of the predictand, which are contained in the stochastic m-vector $\underline{y}_1$. In addition, observations for the covariable at n completely or partially different points $z_{m+1}, \ldots, z_{m+n}$ are available and are contained in the stochastic n-vector $\underline{y}_2$. Although in practice there will be shared measurement locations, which are in fact necessary to estimate the joint spatial structure function, there is no need to have those in the present prediction problem.

The expectation of each element of $\underline{y}_1$ and $\underline{y}_2$ is supposed to be a polynomial in the one, two or three coordinates of the observation points z_i $(i=1,\ldots,m)$ and z_j $(j=m+1,\ldots,m+n)$, respectively, with respect to an arbitrary coordinate system, and so the expectations obey the structure:

$$E\underline{y}_1 = X_1\beta_1 \qquad\qquad (2.1)$$
$$E\underline{y}_2 = X_2\beta_2 \qquad\qquad (2.2)$$

Both $E\underline{y}_1$ and $E\underline{y}_2$ are linear combinations of the regressor vectors with as yet unknown parameters β_1 and β_2. The matrix X_1 consists of m rows and p columns, where p is dependent on the degree (d) of the polynomial expectation or *trend* of the variable and the dimension (dim) of the region in which the observations are taken. It is easily shown that p equals

$$p = \frac{(\dim+d)!}{\dim! \, d!} = \binom{\dim+d}{\dim}. \qquad\qquad (2.3)$$

In a two-dimensional region and without trend, X_1 is merely the vector 1_m, consisting of m ones only. If the degree of the trend is equal to 1, each row of X_1 consists of 1, ξ_1 and ξ_2, ξ_1 and ξ_2 representing the coordinates of an observation point of the predictand. If the trend is quadratic, each row of X_1

92

consists of 1, ξ_1, ξ_2, ξ_1^2, $\xi_1\xi_2$, ξ_2^2. Similarly, the matrix X_2 is n by p where each row consists of p monomial values of the coordinates of an observation point of the covariable. The term 'Universal' for Kriging and Cokriging indicates the presence of trend terms.

Obviously now $E\underline{y}_0 = x_0'\beta_1$, in which the vector x_0' consists of one row of p monomial values of the coordinates of the point z_0 where a prediction is required, similarly to the p regressors in X_1.

In contrast to trend surface analysis (Watson, 1971), where observations are supposed to be independent, the observations in Kriging and Cokriging have a special dependence structure. This is modelled by means of the covariance between (actual and hypothetical) observation points or their disturbances with respect to expectated values.

The dependence structure of the vector $(\underline{y}_1', \underline{y}_2', \underline{y}_0)'$ is assumed to be given by the symmetric covariance matrix C^*, of order m+n+1. This matrix may be partitioned as

$$C^* = \begin{pmatrix} C_{11} & C_{12} & c_{01} \\ C_{21} & C_{22} & c_{02} \\ c_{01}' & c_{02}' & c_{00} \end{pmatrix} = \begin{pmatrix} C & c_0 \\ c_0' & c_{00} \end{pmatrix} \tag{2.4}$$

where C_{11} is the covariance matrix of the elements of $\underline{y}_1$, C_{22} of the elements of $\underline{y}_2$, C_{12} ($= C_{21}'$) between the elements of $\underline{y}_1$ and those of $\underline{y}_2$, and c_{01}, c_{02} and c_{00} between $\underline{y}_0$ and the elements of $\underline{y}_1$, $\underline{y}_2$ and $\underline{y}_0$, respectively, while $C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}$. The matrix C^* must be positive definite in order that the variance of any linear predictor be non-negative. Each element of C_{11} and of c_{01} as well as c_{00} is assumed to be known as an isotropic function $c_1(r)$, only of the distance r between the

pair of observation points concerned, the so-called covariance function for the predictand. Likewise, each element of C_{22} is assumed to be known as an isotropic function $c_2(r)$ of the distance r between the pair of observation points concerned, the covariance function for the co-variable, and each element of C_{12} and of c_{02} is assumed to be known as an isotropic function $c_{12}(r)$ of the distance r between the pair of observation points concerned, the so-called cross-covariance function between predictand and co-variable. Further assumptions about the covariance functions are not required. To avoid complications we do not consider here the conditions which $c_1(r)$, $c_2(r)$, $c_{12}(r)$ must satisfy for C^* to be positive-definite. Neither are problems of estimating covariance functions addressed in this paper (see e.g. Ripley, 1981). Nor do we deal with generalized covariance functions, which are only to be applied to linear combinations of possible observations $(\underline{y_1},\underline{y_2})'$ whose coefficient vector must obey certain linear restrictions (see e.g. Delfiner, 1976; Dowd, 1989). The semivariogram, which is a special case of the opposite of such a generalized covariance function in particular in the absence of a trend, is not considered in this expository context, once more in order to keep the material as simple as possible (Corsten, 1989).

We have now the following linear model for actual observations $\underline{y}$ and hypothetical observation of the predictand $\underline{y_0}$:

$$\binom{\underline{y}}{\underline{y_0}} = \binom{X}{x_0^{*'}}\beta + \underline{e}; \quad E\underline{e} = 0; \quad \text{Cov}(\underline{e}) = C^*, \tag{2.5}$$

in which we have introduced the abbreviations:

94

$$X = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix}, \ x_0^* = \begin{pmatrix} x_0 \\ 0 \end{pmatrix}, \ \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}. \tag{2.6}$$

We require the best predictor $\underline{t} = \lambda'\underline{y}$, linear in the observations $\underline{y}$, which satisfies the conditions that

$$E(\lambda'\underline{y} - \underline{y}_0) = 0 \tag{2.7}$$

and

$$Var(\lambda'\underline{y} - \underline{y}_0) \text{ is minimal.} \tag{2.8}$$

With $V = (X'C^{-1}X)^{-1}$ Appendix A proves that

$$\underline{t} = x_0^{*\prime}\hat{\beta} + c_0'C^{-1}(\underline{y} - X\hat{\beta}) \tag{2.9}$$

where $\hat{\beta} = VX'C^{-1}\underline{y}$, the generalized least squares (GLS) estimator of β.

The predictor $\underline{t}$ in (2.9) is the sum of the estimated local expectation of the predictand only, that is $x_0^{*\prime}\hat{\beta} = x_0'\hat{\beta}_1$, and a linear combination of the observed residuals contained in $\underline{y} - X\hat{\beta}$, with the best linear multivariate approximation of the predictand by all m+n observations as coefficients. Because $\underline{t} - x_0^{*\prime}\hat{\beta} = c_0'C^{-1}(\underline{y} - X\hat{\beta})$, the procedure can be interpreted as regression of the residuals of $\underline{t}$ with respect to $x_0^{*\prime}\hat{\beta}$ on the residuals of the $\underline{y}$'s with respect to $X\hat{\beta}$.

The variance of the prediction error equals

$$Var(\underline{t} - \underline{y}_0) = c_{00} - c_0'C^{-1}c_0 + x_a'Vx_a. \tag{2.10}$$

where $x_a = x_0^* - X'C^{-1}c_0$.

This variance of the prediction error (2.10), free of Lagrange multipliers, can be interpreted as follows:
c_{00} is the variance of the variable under study;

$c_0'C^{-1}c_0$ is the reduction of that variance due to the best
linear approximation by the other observations considered
as m+n additional multivariate characteristics;

$x_a'Vx_a - Var(x_a'\hat{\beta})$ is the variance of the sum of estimated
expectations of predictand and covariable, however, not in
the prediction location with regressors contained in $x_0{}^*$,
but in a different but related point with regressors
contained in x_a.

In fact the prediction error variance reflects the
orthogonal decomposition of the prediction error into
$y_0 - c_0'C^{-1}y$ and $x_a'\hat{\beta}$, respectively.

The predictor $\underline{t}$ is an exact predictor: if a prediction is
carried out at an observation point, the vector c_0 is equal to
the corresponding column of C, leading to the observation
itself as the best prediction. Then, it turns out that the
variance of the prediction error will vanish.

Repeated application of $\underline{t}$ at different locations requires
revision of x_0 and c_0 only, whereas the costly inversion of the
matrix C is required only once.

A confidence interval for the predictand can be given if $\underline{y}$
has a multivariate Gaussian distribution. It follows that with
probability 1-α the interval

$$[\underline{t}-u(\alpha/2)*Var(\underline{t} - \underline{y}_0)^{\frac{1}{2}}, \underline{t}+u(\alpha/2)*Var(\underline{t} - \underline{y}_0)^{\frac{1}{2}}]$$

will cover $\underline{y}_0$, where $u(\alpha/2)$ is the value such that the
probability that $\underline{u} < u(\alpha/2)$ equals $1-\frac{1}{2}\alpha$, $\underline{u}$ following the
standard normal distribution.

PARTICULAR CASES

a. **Universal Kriging**.

The Universal Kriging equations are obtained when observations are present for the predictand only, that is with n equal to 0. All sub-matrices with subscript 2 will now vanish. As a result we obtain the same equations for the predictor and the variance of the prediction error as eqs. 2.9 and 2.10, but where y now reduces to y_1, X to X_1, C to C_{11}, c_0 to c_{01}, $\hat{\beta}$ to $\hat{\beta}_1$ and V and x_a to the corresponding sub-matrix and -vector, while x_0^* reduces to x_0.

b. **The nugget effect**.

In many studies, a so-called 'nugget effect', representing measurement error variance plus variation at very small distance, is encountered, often in combination with spatial dependence. We then have for $c_1(r)$ that $c_1(r) = B*\delta(r)+f_1(r)$, where $\delta(r) = 1$ if $r = 0$ and $\delta(r) = 0$ if $r \neq 0$, the coefficient B is the size of nugget effect and $f_1(r)$ models the spatial dependence. Similar structures apply for $c_2(r)$ and $c_{12}(r)$.

A feature sometimes encountered in regionalized studies is the so-called pure nugget effect, i.e. correlation zero for non-zero distance. Then the covariance functions have the structure $c_1(r) = \sigma_1^2\delta(r)$, $c_2(r) = \sigma_2^2\delta(r)$ and $c_{12}(r) = \sigma_{12}\delta(r)$.

Without nugget effect the interpolation surface will be entirely smooth, passing through the observations, while in the presence of a nugget effect the observations will appear as spikes above or below the smooth prediction surface. With pure nugget effect this prediction surface will coincide with the estimated trend surface.

97

Example 1.

To illustrate different aspects of universal Kriging and Cokriging, a simple example was constructed. Along a straight line two observations of the predictand, with values equal to 21 and 23, were situated at locations z_1 and z_2 with coordinates 1 and 3, respectively. Also, three observations of the covariable with values 5, 6 and 6 were situated on locations z_3, z_4 and z_5 with coordinates 1, 2 and 3, respectively. The covariance functions $c_1(r)$ and $c_2(r)$ as well as the cross-covariance function $c_{12}(r)$ were assumed to be known, and to obey an exponential model with nugget effect (cf. 3b): $c(r) = B\delta(r)+A\exp(-r/b)$, where $\delta(r) = 1$ if $r = 0$ and $\delta(r) = 0$ if $r \neq 0$. The values of B, A and b for each of the three covariance functions are presented in Table 1. The values are such that the correlation coefficient between the variables equals +0.73. Attention was focused on the difference between Kriging and Cokriging, with and without a linear trend. Some intermediate steps are shown in Table 2. For Kriging without trend the sole value of β equals the average of the two

Table 1. Values for the coefficients of the covariance functions

	B	A	b
$c_1(r)$	1	3	0.5
$c_2(r)$	0.3	1.7	0.5
$c_{12}(r)$	0.4	1.9	0.5

98

observations. For (Universal) Kriging with a linear trend the values of β are those for a line through the points. However, the values are modified by the introduction of the covariable, and become less clearly interpretable.

Predictions were carried out for all points with coordinates between 0 and 4 (fig. 1), standard deviations of prediction error are given in fig. 2. Numerical attention is focused on the locations P_1, P_2 and P_3 with coordinate 2, 2.5 and 3.5, respectively (Table 3). Point P_1 is of interest as there is an observation of the covariable, but no observation of the predictand, point P_2 is an interior point for which typically predictions are needed, e.g. if a map is to be constructed, and with point P_3 extrapolation can be illustrated.

The prediction in P1 increases by 0.58 if a covariable is used, because of the relatively large value of the observation of the covariable in the prediction location. An increase of the prediction in P_2 by approximately 0.25 is observed, once due to trend and secondly due to the use of a covariable. The prediction in P_3 is barely influenced by the presence of a covariable, but the presence of a trend has major influence. Finally, it is seen from fig.1 how predictions are in fact modifications with respect to the trend surface; also noted are the discontinuities at the observation points of the predictand, due to the nugget effect.

The strongest reduction in the standard deviation of the prediction error when the covariable is included takes place in P_1, due to the large value of $c_0'C^{-1}c_0$ caused by the strong cross-correlation between the two variables. Further, a major part of the variance of the prediction error is due to the fact

Table 2. Intermediate results for example 1.

	Kriging		Co-kriging	
	No trend	Linear trend	No trend	Linear trend
y	21 23	21 23	21 23 6 6 5	21 23 6 6 5
x	1 1	1 1 1 3	1 0 1 0 0 1 0 1 0 1	1 1 0 0 1 3 0 0 0 0 1 2 0 0 1 3 0 0 1 1
c	4.000 0.055 0.055 4.000		4.000 0.055 0.257 0.035 2.300 0.055 4.000 0.257 2.300 0.035 0.257 0.257 2.000 0.230 0.230 0.035 2.300 0.230 2.000 0.031 2.300 0.035 0.230 0.031 2.000	
v	2.027	9.918 -3.945 -3.945 1.973	1.703 0.888 0.888 0.774	9.593 -3.945 5.418 -2.265 -3.945 1.973 -2.265 1.133 5.418 -2.265 4.712 -1.969 -2.265 1.133 -1.969 0.984
β	22.000	20.000 1.000	22.178 5.654	20.178 1.000 4.654 0.500

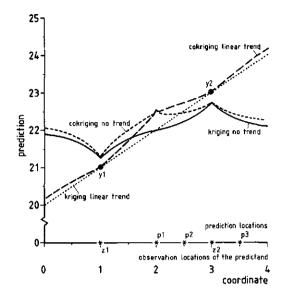|	Kriging		Co-kriging	
	P_1 P_2 P_3	P_1 P_2 P_3	P_1 P_2 P_3	P_1 P_2 P_3
c_0	0.406 1.104 1.104 0.406 0.149 0.020	0.406 1.104 1.104 0.406 0.149 0.020	0.406 1.104 1.104 0.406 0.149 0.020 2.300 0.699 0.699 0.257 0.699 0.095 0.257 0.095 0.013	0.406 1.104 1.104 0.406 0.149 0.020 2.300 0.699 0.699 0.257 0.699 0.095 0.257 0.095 0.013
x_0^*	1.00 1.00 1.00	1.00 1.00 1.00 2.00 2.50 3.50	1.00 1.00 1.00 0.00 0.00 0.00	1.00 1.00 1.00 2.00 2.50 3.50 0.00 0.00 0.00 0.00 0.00 0.00
x_a	0.80 0.69 0.72	0.80 0.69 0.72 1.60 1.64 2.67	0.83 0.75 0.78 -0.94 -0.34 -0.10	0.83 0.75 0.78 1.65 1.80 2.83 -0.94 -0.34 -0.10 -1.89 -0.77 -0.30

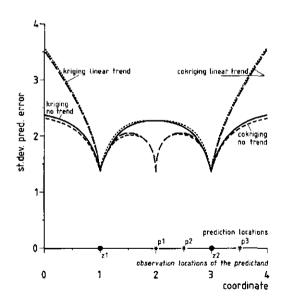*Fig. 1. Predictions carried out with Kriging and with Cokriging, without a trend and with a linear trend.*

Fig. 2. Standard deviation of the prediction error for Kriging and with Cokriging, without a trend and with a linear trend.

101

Table 3. Predictions and negative or positive contributions to the prediction error variance by $c_0'C^{-1}c_0$ and $x_a'Vx_a$ in example 1. The value of c_{00} is equal to 4.0 in all cases.

		Trend	Prediction	$c_0'C^{-1}c_0$	$x_a'Vx_a$	$var(\underline{t} - \underline{y}_0)$
P1	Kriging	None	22.0	0.081	1.297	5.215
		Linear	22.0	0.081	1.297	5.215
	Cokriging	None	22.58	2.666	0.468	1.803
		Linear	22.58	2.666	0.468	1.803
P2	Kriging	None	22.24	0.309	0.968	4.659
		Linear	22.5	0.309	1.099	4.790
	Cokriging	None	22.47	0.507	0.588	4.081
		Linear	22.74	0.507	0.722	4.215
P3	Kriging	None	22.28	0.305	1.059	4.755
		Linear	23.5	0.305	4.021	7.717
	Cokriging	None	22.39	0.311	0.898	4.587
		Linear	23.63	0.311	3.863	7.552

that the parameters in β have to be estimated, implying that $x_a'Vx_a$ is far from negligible (Table 3).

In summary this example shows how Kriging and Cokriging may perform differently, in particular on a small set of data.

Example 2.

The second example stems from a soil study in the Limagne area in Central France (Stein et al. 1989). The central problem of this example is to characterize soil variability within an area without clear physiographic features and to map a crucial soil parameter for crop growth simulation modelling over the area. As a rule physiographic features, like faults, relief, brooks, erosion forms, etc., are useful tools to study

soil spatial variability in an area and to delineate soil units, each with a 'representative profile'.

The study area is located on one of the most recent terraces of the river Allier, some 30 kilometers down-stream of the city of Clermont-Ferrand. The area of 200 ha is naturally bounded at one side by the Allier, and at the other side by a somewhat higher situated and older terrace. Moisture availability is one of the most important land qualities for agriculture as depth to groundwater is greater than 1.2m. On a generalized 1:100.000 soil map, which is the only one available for this region (INRA 1965) no soil delineations are presented, which implies that this region is considered to be fairly homogeneous on this scale. When more detailed investigations are carried out on a larger scale (e.g. 1:20.000) clay, sand and gravel layers exhibit irregular patterns that are difficult to map due to, again, lack of clearly visible physiographic features (Stein, Bouma, Kroonenberg and Cobben, 1989). A regular sampling scheme was therefore used to investigate the spatial variability (Fig. 3).

Four transects were planned with observation points 100m apart. Two transects were located more or less parallel to the Allier river, and two transects were chosen perpendicular to the river, yielding a total of 62 observation points. To investigate variability of soil survey data at short distances, an additional sampling scheme was used (Fig. 3) with observations at distances 1m, 2m, 4m, 8m, 16m and 32m, in two directions from every fourth observation point of the main transect perpendicular to the river. From any of those points one additional observation was taken parallel to the river at a distance 1m, 2m, 4m, 8m, 16m and 32m, respectively. In every observation point a number of basic soil survey data were
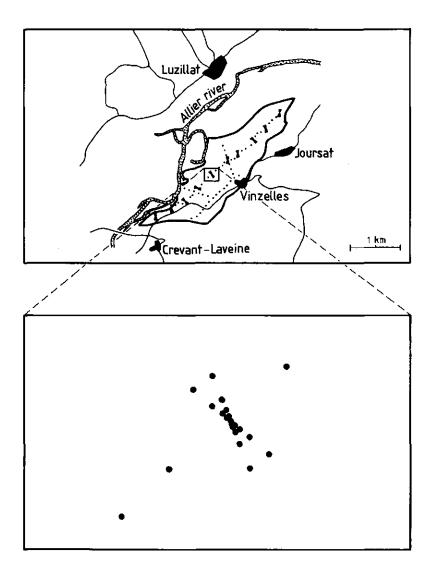
103

Fig. 3. The sampling scheme on the river terrace in the Limagne area in France.
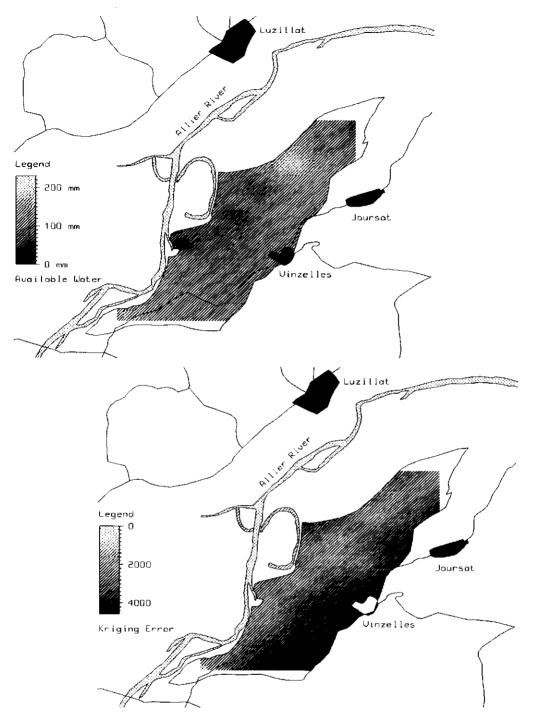
104

Fig. 4. The Co-kriged AVW map (above) and the map with the prediction error variances (below).

measured by means of a boring to 1.2m, or to an impenetrable gravel layer, whichever appeared first. Gravel layers within 1.2m were encountered in 192 of the 304 observation points, sometimes occurring close to the surface. Soil physical properties were measured on the main transects, yielding 62 calculations of 'available water' (AVW), which is defined as the volume of water between field capacity and 15 bar. Available Water is directly related to crop yield (e.g. FAO 1976).

Prediction of the land quality AVW for unvisited locations throughout the terrace, was carried out with Kriging. Since no trend was expected, covariance and cross-covariance functions were estimated by existing standard methods (Journel and Huijbregts, 1978). Because soil physical properties require relatively expensive and time consuming measurement procedures, Co-kriging is attractive to be used as well. The choice for Depth to Gravel (DG) as a co-variable is self-evident, since it is easy to measure and highly correlated with AVW (correlation coefficient 0.88). The values of B, A and b for each of the three covariance functions are presented in Table 4.

The resulting AVW map is presented in Fig. 4. With this map optimal predictions for unvisited locations were obtained.

Table 4. Values for the coefficients of the covariance functions.

		B	A	b
AVW	$c_1(r)$	1310	2580	0.28
DG	$c_2(r)$	306	522	0.27
AVW x DG	$c_{12}(r)$	342	1100	0.25

It is noted how the gravel bank in the center of the area givesrise to restricted water availability.

To compare the gain in precision relative to Kriging, the Prediction Error Variance was determined in 1200 randomly located points within the level river terrace, for both Kriging and Cokriging. Minimum, maximum and average values are presented in Table 5. In order to investigate the relation between the prediction error variance and the distance to measurement locations, the prediction error variances were determined for subset 1, containing the points within 280m from observations of the predictand and for subset 2, containing those points within 100m from observations of the predictand. Finally, the prediction error variance was determined in the 159 data points, all within 64m from an AVW observation where a DG measurement was present, but no AVW observation.

The soil variability patterns were checked by multispectral aerial photographs, made with a Near-Infrared (NIR) filter

Table 5. *Minimum, maximum and mean prediction error variances in example 2, for 1200 random points, for two subsets thereof, and for points with an observation of the covariable but not of the predictand.*

Test set	Number of points	Kriging range	mean	Cokriging range	mean
All random data	1200	1955 - 3985	3099	1591 - 3917	2917
Subset 1	833	1955 - 3907	2909	1591 - 3800	2694
Subset 2	205	1955 - 3198	2541	1591 - 2809	2162
All observations with DG < 1.2m and without observation of AVW.	159	1878 - 3133	2252	1005 - 1206	1080

and with a Green (GR) filter. The (NIR-GR)/(NIR+GR) ratio is used as a vegetation index. For 37 of the 62 locations of the 100m transects containing maize, the greytones were classified according to a standard sensitivity guide. As AVW correlates highly with DG it is likely that differences in greytones were caused by stress situations, due to shortage of available water in the profile. The vegetation index yielded a significant regression equation (at $\alpha = 0.05$ level) with AVW as explanatory variable and the vegetation index as dependent variable. This is considered as a validation of the AVW map obtained by Cokriging.

Conclusions from this example are:

1. The prediction error variance is a markedly increasing function of the distance to observation points. This implies a need for dense sampling of the predictand, which is, however, usually prevented by the associated costs.

2. A gain in precision with respect to Kriging of 50% in terms of the mean variance of the prediction error is observed if a highly correlated covariable is included. This covariable has to be sampled on all grid points where a prediction is carried out so as to produce a predictive soil map. The gain in precision for other points is structural, but small, ranging in the average from 6% to 12%, depending on the test set.

3. Evidence on aerial photographs supports the AVW values obtained in this study.

Acknowledgement

REFERENCES

Carr, J.R. and McCallister, P.G., (1985). An application of co-kriging for estimation of tripartite earthquake response spectra. Mathematical Geology 17, 527-545.

Corsten, L.C.A., (1985). Current statistical issues in agricultural research. Statistica Neerlandica 39, 159-168.

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

Cressie, N.A.C., (1986). Kriging Nonstationary Data. Journal of the American Statistical Association, 81, 625-634.

Delfiner, P., (1976). Linear estimation of non-stationary spatial phenomena. In Advanced geostatistics in the Mining Industry, M. Guarascio, M. David and C. Huijbregts (eds), 49-68. Dordrecht: Reidel.

Dowd, P.A., (1989). Generalized Cross-covariances. In Geostatistics, M. Armstrong (ed.), 151-162. Dordrecht: Kluwer.

Dubrule, O., (1983). Two methods with different objectives: splines and kriging. Mathematical Geology 15, 245-257.

FAO (1976). A framework for land evaluation. FAO Soils Bulletin 32. Rome.

Friedman, J.H. and Silverman, B.W., (1989). Flexible parsimonous smoothing and additive modelling. Technometrics 31, 3-21.

Gianola, D. and Goffinet, B., (1982). Query response: sire evaluation with best linear unbiased predictors. Biometrics 38, 1085-1088.

Gleeson, A.C. and Cullis, B.R., (1987). Residual maximum likelihood estimation of a neighbour model for field experiments. Biometrics 43, 277-288.

Harville, D., (1985). Decomposition of prediction error. J. Amer. Stat. Ass. 80, 132-138.

INRA (1965). Carte Pédologique du Val d'Allier. Feuille sud: Clermont-Ferrand. Service d'Etudes des Sols, Montpellier, France.

Journel, A.G. and Huijbregts, Chr.J., (1978). Mining geostatistics. Academic Press, London.

Krajewski, W.F., (1987). Cokriging radar-rainfall and rain gage data. Journal of Geophysical Research 92, 9571-9580.

Krige, D.G., (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. Journal of the Chemical, Metallurigical & Mining Society of South Africa 52, 119-138.

Krige, D.G., (1966). Two dimensional weighted moving average trend surfaces for ore evaluation. Journal of the South African Society of Mining and Metallurgy, 13-38.

Matheron, G., (1963). Principles of geostatistics. Economic Geology 58, 1246-1266.

Matheron, G., (1971). The theory of regionalized variables and its applications. Les Cahiers du Centre de Morphologie Mathématique, No. 5. Paris: Ecole de Mines de Paris.

Matheron, G., (1979). Recherche de simplification dans un problème de cokrigeage. Centre de Géostatistique, Fontainebleau.

McBratney, A.B., and Webster, R., (1983). Optimal interpolation and isarithmic mapping of soil properties. V. Co-regionalization and multiple sampling strategy. Journal of Soil Science 34, 137-162.

Piazza, A., Menozzi, P. and Cavalli-Sforza, L., (1981). The making and testing of geographic gene-frequency maps. Biometrics 37, 635-659.

Ripley, B., (1981). Spatial Statistics. Wiley, New York.

Stein, A., Bouma, J., Van Dooremolen, W. and Bregt, A.K., (1988). Cokriging point data on moisture deficits. Soil Sci. Soc. Am. J. 52, 1418-1423.

Stein, A., Bouma, J., Mulders, M.A. and Weterings, M.H.W., (1989). Using spatial variability studies to estimate physical land qualities of a level river terrace. Soil Technology 2, 385-402.

Stein, A., Bouma, J., Kroonenberg, S.B., and Cobben, S., (1989). Sequential sampling to measure the infiltration rate within relatively homogeneous land units. Catena, 16, 91-100.

Steinberg, D.M., (1988). Factorial experiments with time trends. Technometrics 30, 259-269.

Vauclin, M., Vieira, S.R., Vachaud, G. and Nielsen, D.R., (1983). The use of cokriging with limited field soil observations. Soil Science Society of America Journal 47, 175-184.

Watson, G.S., (1971). Trend surface analysis. Mathematical Geology 3, 215-226.

Webster, R., (1985). Quantitative spatial analysis of soil in the field. In: B.A. Stewart (ed.), Advances in soil science 3. Springer Verlag, New York, 1-70.

Witter, J.V., (1984). Heterogeneity of dutch rainfall. Thesis, Agricultural University, Wageningen, The Netherlands.

Yates, S.R. and Warrick, A.W., (1987). Estimating soil water content using cokriging. Soil Science Society of America Journal 51, 23-30.

Appendix A - Derivation of predictor and prediction error variance

This appendix gives the derivation of the formulae (2.9) and (2.10). We start with making some preliminary observations and consider a number, say k, of stochastic characteristics y_i, on the basis of which the best linear predictor for y_0 is to be formulated, that is the predictor $\underline{u}$, not necessarily unbiased, with minimal $var(\underline{u} - y_0)$, while C and c_0, the covariances among the y_i and those between the y_i and y_0, are known. Then one has to determine in $\langle y_1, \ldots, y_k \rangle$, the space spanned by the y_i's, the vector $\mu'\underline{y}$ such that $(y_0 - \mu'\underline{y})$ has minimum norm or length, where the inner product is defined as the covariance between two vectors. This is equivalent to finding the orthogonal projection of y_0 on $\langle y_1, \ldots, y_k \rangle$, or $cov(y_0 - \mu'\underline{y}, y_i) = 0$ for $i = 1, \ldots, k$, or $C\mu = c_0$, the normal equations, leading to $\underline{u} = c_0'C^{-1}\underline{y}$. The vector $y_0 - c_0'C^{-1}\underline{y}$ is orthogonal (uncorrelated) to any linear combination of $y_1, \ldots, y_k$ and $var(y_0 - c_0'C^{-1}\underline{y}) = var(y_0) - var(c_0'C^{-1}\underline{y}) = c_{00} - c_0'C^{-1}c_0$.

Now with $E\underline{y} = X\beta$ and $Ey_0 = x_0^{*\prime}\beta$, according to (2.5), we wish to minimize $var(\underline{t} - y_0)$ under the condition that $E(\underline{t} - y_0) = 0$. In order to accomplish this, the prediction error is split into

$$\underline{t} - y_0 = (\lambda'\underline{y} - c_0'C^{-1}\underline{y}) - (y_0 - c_0'C^{-1}\underline{y}) \tag{A.1}$$

the difference of two orthogonal terms, and $var(\underline{t} - y_0) = var(\lambda'\underline{y} - c_0'C^{-1}\underline{y}) + c_{00} - c_0'C^{-1}c_0$.

Now, for $\nu'\underline{y}$, with ν' equal to $\lambda' - c_0'C^{-1}$, we require that $E(\nu'\underline{y}) = x_a'\beta$, with $x_a' = (x_0^{*\prime} - c_0'C^{-1}X)$, because of unbiasedness.

Minimizing $\text{var}(\underline{t} - \underline{y}_0)$ is equivalent to minimizing $\text{var}(\nu'\underline{y})$ while $E(\nu'\underline{y}) = x_a'\beta$. This minimum is attained by the GLS estimator $\hat{\beta} = VX'C^{-1}\underline{y}$, where $V = (X'C^{-1}X)^{-1}$. Hence $\nu'\underline{y} = x_a'\hat{\beta}$ and thus $\underline{t} = \nu'\underline{y} + c_0'C^{-1}\underline{y}$ will be

$$\underline{t} = x_a'\hat{\beta} + c_0'C^{-1}\underline{y}$$
$$= x_0^*{}'\hat{\beta} + c_0'C^{-1}(\underline{y} - X\hat{\beta}) \qquad (A.2)$$

as mentioned in (2.9). Further, the minimum of $\text{var}(\underline{t} - \underline{y}_0)$ is according to decomposition (A.1) equal to $\text{var}(x_a'\hat{\beta}) + c_{00} - c_0'C^{-1}c_0$. Replacement of $\text{var}(x_a'\hat{\beta})$ by $x_a'Vx_a$ yields (2.10).

Appendix B - Extension to k variables.

Extension of the preceding method to k variables, i.e. prediction of a variable with k-1 covariates, is straight-forward. Again, $\underline{y}_1$ is the vector containing observations of the predictand, whereas $\underline{y}_2, \ldots, \underline{y}_k$ are the vectors containing observations of the covariables. In a linear model we have

$$E\begin{pmatrix} \underline{y} \\ \underline{y}_0 \end{pmatrix} = \begin{pmatrix} X \\ x_0^*{}' \end{pmatrix}\beta \qquad (B.1)$$

where $X = \text{diag}(X_1, \ldots, X_k)$, a block-diagonal matrix with the matrix X_i, containing polynomial functions upto the degree of the trend of the coordinates of the observation points of the ith variable, as i^{th} block and the vectors $\underline{y}$ and β are defined as $\underline{y} = (\underline{y}_1', \ldots, \underline{y}_k')'$ and $\beta' = (\beta_1', \ldots, \beta_k')$. With C the covariance matrix of actual observations (of predictand and covariables), c_0 the vector of covariances between the predictand at the unvisited location with all observations

113

contained in $\underline{y}$, and c_{00} the variance of the predictand at the unvisited spot, we obtain with $x_0^{*\prime} = (x_0',0,\ldots,0)$,

$$\underline{t} = x_0^{*\prime}\hat{\underline{\beta}} + c_0' C^{-1}(\underline{y} - X\hat{\underline{\beta}}) \tag{B.2}$$

$$\text{var}(\underline{t} - \underline{y}_0) = c_{00} - c'_0 C^{-1} c_0 + x_a' V x_a \tag{B.3}$$

where now $\hat{\beta}' = (\hat{\beta}_1',\ldots,\hat{\beta}_k')$ and V and x_a are completely analogous to the expressions defined above. The formulae (B.2) and (B.3) are similar to (2.9) and (2.10), respectively.

114

CHAPTER II.2. COKRIGING NON-STATIONARY DATA

by

A. Stein, A.C. van Eijnsbergen and L.C. Barendregt

COKRIGING NON-STATIONARY DATA.

ABSTRACT

Universal cokriging is used to obtain predictions when dealing with multivariate and non-stationary data. In this study attention is focused on multivariate random functions. An important type of non-stationarity is defined in terms of multivariate random functions with increments which are stationary of order k. The covariance between increments of different variables is modeled by means of the pseudo-cross-covariance function. Criteria are formulated to which the parameters of cross-pseudo-covariance functions must comply so as to ensure positive-definiteness. Cokriging equations and the induced cokriging equations are given. The study is illustrated by an example from soil science.

INTRODUCTION

In regionalized variable studies there is a growing emphasis on the measurement and interpretation of characteristics for which relatively advanced measurement procedures are necessary. Land qualities like available water, (Stein et al., 1989), concentrations of heavy metals in the soil (Leenaers et al., 1989) and rain gauge data (Krajewski, 1987) are typical examples. Such variables, as a function of the coordinates of their location are modeled as regionalized variables. Prediction towards unvisited locations is an important activity, e.g. in order to provide a map of the region, or to estimate the total amount of polluted soil or of recoverable reserves.

Sometimes these variables are correlated with other, more easily measured regionalized variables, like depth to gravel, relative altitude or radar-rainfall data. In such situations cokriging (Journel and Huijbregts, 1978; Myers, 1982; Vauclin et al., 1983; Stein and Corsten, 1990) is a regularly used prediction technique which involves observations of several variables at hand. The easily measured variables serve as co-variables when predicting the value of the predictand in an unvisited location.

Upto now, little attention has been given to universal cokriging, i.e. cokriging in the presence of a trend (see, however, Stein and Corsten, 1990). The need for a more profound, statistical treatment of this subject is evident, since the use of cokriging without taking a trend into account may be too gross a simplification.

In this paper attention is focused on the definition and the use of universal cokriging. To do this, cokriging is

118

defined in terms of increments. In section 2 multivariate random functions and multivariate increments are introduced. The question concerning positive-definiteness of permissible covariance functions of increments is treated (Dowd, 1989). In section 3 cokriging is defined in terms of the best linear increment predictors. Finally an example illustrates the estimation of the parameters describing the cross-covariances between increments and the use of universal cokriging in a study on moisture deficits.

As concerns terminology, we have put much effort to describe the concepts as concise and precise as possible. The term universal cokriging is used throughout this paper to indicate that one deals with regionalized variables, which are not stationary themselves, but which have stationary increments of some order. We will use the term 'prediction' instead of 'estimation' to indicate that we estimate a possible future observation and not a parameter. This is in agreement with terminology in time series analysis, where one predicts future observations, but also observations in the past, or missing observations, as well as with terminology in mixed linear models (see also Corsten, 1989 and Christensen, 1990 on universal Kriging). Instead of 'generalized' covariances we prefer the term 'pseudo'-covariances, since we deal with covariances which are not always covariances in the proper sense, since positive-definiteness, an important feature, is not transferred to non-increments. However, one may use as well 'generalized' instead of 'pseudo'.

MULTIVARIATE RANDOM FUNCTIONS

a. Univariate random functions

Attention is focused on the set of observations on a number of variables. We denote the observations on one of these variables, the uth one, by $y_u(x_1), \ldots, y_u(x_{n_u})$; $x_1, \ldots, x_{n_u}$ are the sites in the d-dimensional space, for example R^d, (with respect to an arbitrary origin) where observations are taken and the $y_u(x_i)$ are real numbers. Bold printing indicates a vector. The index u refers to a finite set of variables on which observations are taken. Although in many practical studies like soil surveys, the size of this set may be quite large (Webster, 1985), in this paper a set of size 2 will serve as an illustration.

The underlying model for a single variable is a univariate random function (URF), which will be denoted by $Y_u(x)$, the capital indicating randomness (Ito, 1953; Yaglom, 1986). A URF is seldom, if ever, measured entirely, but in most circumstances at a finite number of observation points only. We therefore turn our attention towards a *finite restriction* of a URF.

A finite restriction of a URF $Y_u(x)$ for any u is that function restricted to a domain consisting of a finite number, n_u say, of observation locations denoted by $x_{u1}, \ldots, x_{un_u}$. The vector of n_u values $Y_u(x_{ui})$ is denoted by Y_u. To be clear, $Y_u(x)$ is the URF and Y_u is the finite restriction.

We will need linear combinations of the observations, as the main direction of this research concerns the formulation of a predictor which is linear in the observations. A *finite linear combination* of $Y_u(x)$ is:

$$\lambda_u' Y_u - \sum_{i=1}^{n_u} \lambda_{ui} Y_u(x_{ui}) \tag{1}$$

where the prime indicates the transpose of a vector. Clearly, the finite linear combination is characterized by the function $\lambda_u(x)$ with the values λ_{ui} in the locations x_{ui} and the value 0 elsewhere. In fact, this finite linear combination is a linear functional acting on the univariate functions on R^d.

In order to deal with non-stationary data, attention has to be paid to a special class of linear combinations, called increments. As is well known (Matheron, 1973; Delfiner, 1976) a *(k+1)th-order increment* of $Y_u(x)$, defined for any non-negative k, is a finite linear combination $\lambda_u' Y_u$ with the property that it annihilates polynomials of degree $\leq k$, i.e. $\lambda_u' p -$

$\sum_{i=1}^{n_u} \lambda_{ui} p(x_{ui}) - 0$ for all polynomials $p(x)$ of degree $\leq k$[1]. The class without any restrictions is the class of random functions where for any given u and any given location x in R^d, $Y_u(x)$ is a 0th order increment. Restrictions increase with increasing k.

We notice that if $\Sigma \lambda_{ui} Y_u(x_{ui})$ is an increment, then $\Sigma \lambda_{ui} Y_u(x_{ui}+h)$ is an increment of the same order, since a polynomial with respect to an origin remains a polynomial of the same order if the origin is translated by a vector with previous coordinates -h. It may be desired to remove trends in the URF $Z_u(x)$ according to a linear filtering operation $Z_u(x) - \Sigma \lambda_{ui} Y_u(x_{ui}+x)$ with input $Y_u(x)$ and output $Z_u(x)$. Therefore, if this operation is applied to a polynomial up to degree k, it will vanish, irrespective of the translation h, which justifies

---

[1]The order k+1 is in agreement with (Gelfand-Vilenkin, 1964). Following (Delfiner, 1976), one should call the order k.

tha name of 'filtering' a polynomial. The Multivariate Random Function (MRF) is introduced as a straightforward extension of the URF.

b. Multivariate random functions

Consider more than one, say p, URF's in a specific region, denoted by $Y_1(x)$ through $Y_p(x)$. Within a level river terrace we may think (cf. Stein et al., 1989) of $Y_1(x)$ as the water limited yield of winter wheat, $Y_2(x)$ as the moisture availability of the top soil, $Y_3(x)$ as the depth to gravel, $Y_4(x)$ as the clay percentage of the B-horizon and $Y_5(x)$ as the stoniness of the surface. Clearly, d is equal to 2 and p is equal to 5 in this example. Of interest is the simultaneous behaviour of the variables with regard to the spatial structure of the variables themselves and of their interrelationship.

An MRF is defined for every $x \in R^d$ as a p-variate random variable $Y(x) = (Y_1(x),...,Y_p(x))'$. Every component $Y_u(x)$, u=1,...,p, of an MRF is a URF. An MRF is used to describe simultaneously variables which may be mutually dependent. This distinguishes an MRF from p univariate functions.

A finite restriction of a MRF $Y(x)$ consists of finite restrictions of each component. The vector of $n=n_1+...+n_p$ values $Y_u(x_{ui})$ is denoted by Y, whence $Y(x)$ is the MRF and Y is the finite restriction. Different sets of points may be used for the different constituting URF's.

A finite linear combination of a MRF $Y(x)$ is the sum of finite linear combinations of the URF's $Y_u(x)$:

$$\lambda'Y = \sum_{u=1}^{p} \lambda_u'Y_u = \sum_{u=1}^{p} \sum_{i=1}^{n_u} \lambda_{ui}Y_u(x_{ui}) \tag{2}$$

122

The finite linear combination $\lambda'Y$ of the MRF $Y(\mathbf{x})$ is called an increment of order $(k_1+1,\ldots,k_p+1)$ iff. for $u = 1,\ldots,p$, the univariate contribution $\lambda_u'Y_u$ is an increment of order k_u+1. We will restrict ourselves to the more common case that all k_u's are equal to the same value k.

An MRF, just like a URF, should have a consistent family of distributions for finite restrictions of $Y(\mathbf{x})$ in the sense of Kolmogorov (Kolmogorov, 1950). If first and second moments of the family of distributions exist, the p-variate expectation vector is defined as $\mu(\mathbf{x})=(\mu_1(\mathbf{x}),\ldots,\mu_p(\mathbf{x}))'$, where $\mu_u(\mathbf{x}) = E[Y_u(\mathbf{x})]$, and for any two locations $\mathbf{x}$ and $\bar{\mathbf{x}}$ the p-variate covariance function $G(\mathbf{x},\bar{\mathbf{x}})$ is defined as the p x p matrix $\{g_{uv}(\mathbf{x},\bar{\mathbf{x}})\}$ with typical element $g_{uv}(\mathbf{x},\bar{\mathbf{x}}) = \mathrm{cov}(Y_u(\mathbf{x}),Y_v(\bar{\mathbf{x}}))$. For an arbitrary finite linear combination we have:

$$E[\lambda'Y] = \lambda'\mu = \sum_{u=1}^{p} \lambda_u'\mu_u = \sum_{u=1}^{p} \sum_{i=1}^{n_u} \lambda_{ui}\mu_u(\mathbf{x}_{ui}) \tag{3}$$

$$\mathrm{Var}[\lambda'Y] = \lambda'G\lambda = \sum_u \sum_v \lambda_u' g_{uv} \lambda_v$$

$$= \sum_{u=1}^{p} \sum_{v=1}^{p} \sum_{i=1}^{n_u} \sum_{j=1}^{n_v} \lambda_{ui}\lambda_{ui} g_{uv}(\mathbf{x}_{ui},\mathbf{x}_{vj}) \tag{4}$$

where μ_u, g_{uv}, μ and G are the finite restrictions of $\mu_u(\mathbf{x})$, $g_{uv}(\mathbf{x},\bar{\mathbf{x}})$, $\mu(\mathbf{x})$ and $G(\mathbf{x},\bar{\mathbf{x}})$, respectively. We now consider more general models for the randomness of a MRF.

c. Pseudo-expectation and pseudo-covariance function

We assume that expectations and covariances of only a restricted set of finite linear combinations, increments of order k, exist. We assume the existence of a multivariate

123

function $\mu(\mathbf{x})$ and a matrix function $G(\mathbf{x},\tilde{\mathbf{x}})$ such that the expectation of an increment $\lambda'Y$ can be calculated with eq. (3) and its variance with eq. (4). The functions $\mu_u(\mathbf{x})$ and $g_{uv}(\mathbf{x},\tilde{\mathbf{x}})$ are not expectations and covariances in the ordinary sense, although eqs. (3) and (4) may suggest this. Therefore the functions $\mu_u(\mathbf{x})$ and $\mu(\mathbf{x})$ are termed *pseudo-expectation functions*. Likewise, the functions $g_{uv}(\mathbf{x},\tilde{\mathbf{x}})$ are termed *pseudo-covariance functions* if u = v, and *pseudo-cross-covariance functions* if u $\neq$ v. $G(\mathbf{x},\tilde{\mathbf{x}})$ is termed a pseudo-covariance function as well. Pseudo-expectations of (permissible) linear combinations are not unique: if $p(\mathbf{x})$ is a polynomial of degree $\leq$ k, then $\mu(\mathbf{x})+p(\mathbf{x})$ is also a pseudo-expectation. Neither is the pseudo-covariance function unique: if $p_i(\mathbf{x})$ are polynomials of degree $\leq$ k for i=1,...,4, and $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ are any functions, then using $G(\mathbf{x},\tilde{\mathbf{x}})+p_1(\mathbf{x})f_1(\tilde{\mathbf{x}})+f_2(\mathbf{x})p_2(\tilde{\mathbf{x}})+p_3(\mathbf{x})p_4(\tilde{\mathbf{x}})$ is equivalent to using $G(\mathbf{x},\tilde{\mathbf{x}})$, as follows from the defining property of increments.

From now on we will deal with MRF's with a pseudo-expectation function and a pseudo-covariance function.

d. Pseudo-stationarity

An MRF $Y(\mathbf{x})$ with existing first and second order moments is called translation invariant or stationary (sometimes called: homogeneous) iff. the following two conditions hold:
1. $E[Y(\mathbf{x}+\mathbf{h})]$ is independent of any translation vector $\mathbf{h}$;
2. $Var[Y(\mathbf{x}+\mathbf{h})]$ is independent of any translation vector $\mathbf{h}$;
An MRF $Y(\mathbf{x})$ is pseudo-stationary of order k if for every increment $\lambda'Y$ of order k the filtered MRF with URF's equal to $Z_u(\mathbf{x}) = \Sigma\lambda_{ui}Y_u(\mathbf{x}_{ui}+\mathbf{x})$ is stationary. The MRF $Y(\mathbf{x})$ is said to have k^{th} order stationary increments. Stationarity of incre-

ments may be interpreted as non-changing expectation and variance with regard to translations[2].

There are several classes of stationarity to be mentioned in order of weakness:

1. For a stationary MRF $Y(x)$ the expectation and the pseudo-covariance are identical to ordinary expectation and covariance, respectively.

2. For a pesudo-stationary MRF $Y(x)$ of order 0 the pseudo-expectation function $\mu(x)$ and the pseudo-covariance function are unique modulo polynomials of degree 0. If $\mu(x)$ is constant, but $\mu(x) \neq 0$ it makes sense to consider increments of order ≥ 1.

3. For a pseudo-stationary MRF $Y(x)$ of order 1 the pseudo-expectation function is a multivariate polynomial of degree ≤ 1, which is not unique modulo a constant. An example of an increment of order 1 is the difference $Y(x) - Y(\tilde{x})$. In this case, the semivariogram H, defined by $H(x,\tilde{x}) = -(1/2)*Var[Y(x) - Y(\tilde{x})]$, is equivalent to $G(x,\tilde{x})$. For first order increments the variances of the constituting MRF's, do not necessarily exist, but if they do exist, and are equal to $G_0(x,\tilde{x})$, say, the well-known relations hold:

$$H(x,\tilde{x}) = -\tfrac{1}{2}G_0(x,x) - \tfrac{1}{2}G_0(\tilde{x},\tilde{x}) + G_0(x,\tilde{x})$$

(Journel and Huijbregts, 1978). If for this case we define $H(h) = H(x,x+h)$, we may use $H(x-\tilde{x})$.

---

[2] For example, the Wiener process as mentioned in [Cressie, 1989] is pseudo-stationary of order 1, which is in agreement with [Gel'fand-Vilenkin, 1964, p. 264-265]. Following [Matheron, 1973] one may call this process an IRF(0).

4. For a pseudo-stationary MRF Y(x) of order k the pseudo-expectation function is a multivariate polynomial of degree $\leq$ k, which is unique modulo a polynomial of degree k-1. If increments are stationary but their expectation is not equal to zero it makes sense to use an increment of an order 1 step higher, for which the pseudo-expectation function will vanish. For such increments, an extended class of covariance functions is available. This is exactly the so-called intrinsic hypothesis (Journel & Huijbregts, 1978, p. 33). By definition, polynomials of degree k-1 are annihilated, because for $Y = Y_1 + p$, where p is the finite restriction of a polynomial of degree < k we have $\lambda'Y = \lambda'Y_1$. Therefore, equivalence classes arise in MRF's, modulo polynomials of degree k-1.

For the sake of completeness, we mention the following property: an MRF Y(x) is called rotation invariant (or: *isotropic*) iff. for any u and any v $g_{uv}(x,\tilde{x})$ depends on the length of the distance vector r between x and $\tilde{x}$ only, and not on its direction. In this case it makes sense to speak about $g_{uv}(r)$ instead of $g_{uv}(x,\tilde{x})$.

e. <u>Pseudo-positive-definiteness and the pseudo-cross-spectrum</u>

If the MRF is stationary, then each of the functions $g_{uv}(h) = \text{cov}(Y_u(x),Y_u(x+h))$ should either have a Fourier transform $f_{uv}(\omega)$ such that

$$g_{uv}(h) = (1/2\pi)^{d/2} \int_{R^d} e^{-i\omega.h} f_{uv}(\omega)\,d\omega \tag{5a}$$

or at least be expressible as

126

$$g_{uv}(h) = (1/2\pi)^{d/2} \int_{R^d} e^{-i\omega.h} dF_{uv}(\omega) \tag{5b}$$

where $dF_{uv}(\omega)$ is a complex valued measure, the so-called spectral density function, and $\omega.h$ denotes the inner product in R^d. Eq. (5b) is the relevant one if there are discontinuities in the $F_{uv}(\omega)$-function, e.g. corresponding to periodicities in the MRF.

The p x p matrix $F(\omega)$, with typical element $F_{uv}(\omega)$ is called the multivariate spectral distribution function. The matrix $f(\omega) = \{f_{uv}(\omega)\}$ is called the multivariate spectral density function. In practice we will deal with $f(\omega)$.

The variance of any linear combination $\lambda'Y$ is

$$\lambda'G\lambda = \sum_{u,v} \lambda_u' G_{uv} \lambda_v = \sum_{u,v} \int_{R^d} \tilde{\lambda}_u(\omega) \tilde{\lambda}_v^*(\omega) dF_{uv}(\omega) \tag{6}$$

with $\tilde{\lambda}_u(\omega) = (1/2\pi)^{d/2} \sum_{k=1}^{n_u} \lambda_{uk} \exp(i\omega.x_{uk})$ and $\tilde{\lambda}_v^*(\omega)$ is the complex conjugate of $\tilde{\lambda}_v(\omega)$. According to Bochner's theorem, the p x p matrix with elements $\int_{\omega \in \Omega} dF_{uv}(\omega)$ must be Hermitian for any bounded, measurable set Ω of R^d. If the Fourier-transforms exist, then this condition is equivalent to the condition that the p x p matrix with elements $f_{uv}(\omega)$ is Hermitian for any $\omega \in R^d$.

If the MRF is non-stationary, but has k^{th}-order stationary increments, then $g_{uv}(h)$ is not necessarily expressible in a form as (5b).[3] For every increment $\lambda'Y$ of $Y(x)$ we have that $Var(\lambda'Y) = \lambda'G\lambda$. The p-variate covariance function $G(h)$ obeys

---

[3] A simple counter example is the URF with 0th-order stationary increments, and with $g(h) = -|h|$.

the requirement that it has to be pseudo-positive-definite, that is, for all x_{ui} and x_{vj} ϵ R^d and all collections of permissible numbers λ_{ui} and λ_{vj} we have:

$$\lambda'G\lambda \geq 0. \tag{7}$$

This is a weaker requirement than positive definiteness: for non-permissible numbers λ_{ui} and λ_{vj} negative values may appear. To judge the permissibility of pseudo-cross-covariance functions we need a criterion to evaluate the positiveness of $\lambda'G\lambda$ for all permissible vectors λ, in fact belonging to a subspace of all vectors λ.

Necessary and sufficient conditions for the non-stationary case are accomplished by a very general theorem in Gelfand-Vilenkin, 1964, page 300. This is in fact an extension of Bochner's theorem to generalized functions with stationary increments.

A sufficient condition for the MRF to have a pseudo-positive-definite covariance function is that there is a matrix $F(\omega)$ with elements $F_{uv}(\omega)$, called the multivariate pseudo-spectral-distribution function, such that (6) holds for any increment $\lambda'Y$ pseudo-stationary of order k. The p x p matrix with elements $\int_{\omega\epsilon\Omega_0} dF_{uv}(\omega)$, which is termed the pseudo-cross-spectrum, must have finite elements and be Hermitian for any compact set Ω_0 of R^d, that does not contain the origin, but may be infinite if Ω contains the origin. There is however the restriction that $\int_{\omega\epsilon\Omega} |\omega|^{2k+2} dF_{uv}(\omega)$ be finite for any bounded

measurable set Ω of R^d.

This does not always lead to a practical evaluation criterion for deciding whether certain cross-covariance functions

128

are permitted, i.e. an expression containing a limited number of parameters. In particular cases, however, when attention is focused on p - 1 (Kriging) or p - 2 (cokriging with only one co-variable), this method can readily be used.

Example 1.

Consider a region of dimension d - 2. We will demonstrate how the previous theory applies for generally used functions $g_{uv}(h)$, linear in the (unknown) parameters $\vartheta_{uv,k}$, without nugget effect. Let $g_{uv}(h)$ be defined as:

$$g_{uv}(h) - \sum_{\gamma=1}^{k} \vartheta_{uv,\gamma} |h|^{2\gamma-1}$$

Again, the value of k corresponds to the degree of non-stationarity. The question we wish to answer is: what requirements the coefficients $\vartheta_{uv,\gamma}$ satisfy in order that the functions $g_{uv}(h)$ guarantee, for $k \geq 1$, a permissible solution?

For the Fourier transform of $g_{uv}(h)$ we find the following expression:

$$f_{uv}(\omega) - \sum_{\gamma=1}^{k} \vartheta_{uv,\gamma} 2^{2\gamma+\frac{1}{2}d-1} \frac{\Gamma(\frac{1}{2}(2\gamma+d-1))}{\Gamma(-\frac{1}{2}(2\gamma-1))} |\omega|^{-2\gamma-d+1} \qquad (9)$$

(e.g. Gelfand and Shilov, 1964). For the cross-spectrum we find a 2 x 2 matrix with elements of the form (9). Investigating the case k - 3, we obtain:

$$f(\omega) - \begin{pmatrix} f_u(\omega) & f_{uv}(\omega) \\ f_{uv}(\omega) & f_v(\omega) \end{pmatrix}$$

$$
-\begin{pmatrix} -\vartheta_{u,1} & -\vartheta_{uv,1} \\ -\vartheta_{uv,1} & -\vartheta_{v,1} \end{pmatrix} |\omega|^{-3} + 9 \begin{pmatrix} \vartheta_{u,2} & \vartheta_{uv,2} \\ \vartheta_{uv,2} & \vartheta_{v,2} \end{pmatrix} |\omega|^{-5}
$$

$$
+225 \begin{pmatrix} -\vartheta_{u,3} & -\vartheta_{uv,3} \\ -\vartheta_{uv,3} & -\vartheta_{v,3} \end{pmatrix} |\omega|^{-7}
$$

$$
= -\theta_1 |\omega|^{-3} + 9 \theta_2 |\omega|^{-5} - 225 \theta_3 |\omega|^{-7}
$$

As is easily verified, the $p \times p$ matrix with elements $\int_{\omega \in \Omega_0} f_{uv}(\omega) d\omega$, has finite elements and is Hermitian for any compact set Ω_0 of R^d, that does not contain the origin, but is infinite if Ω_0 does contain the origin. However $\int_{\omega \in \Omega} |\omega|^{2k+2} f_{uv}(\omega) d\omega$ is finite for any bounded, measurable set Ω of R^d. The same holds for $f_u(\omega)$ and for $f_v(\omega)$.

We further notice that for each univariate case the functions $f_u(\omega)$ and $f_v(\omega)$ have to be positive due to Bochners' theorem and hence that $\vartheta_{u,1} \leq 0$, $\vartheta_{u,3} \leq 0$ and $\vartheta_{u,2} \geq -10/3 * \sqrt{\vartheta_{u,1} * \vartheta_{u,3}}$, while similar conditions hold for the coefficients of $f_v(\omega)$, but not for those of $f_{uv}(\omega)$ (Appendix A). For the multivariate case, a sufficient condition for $F(\omega)$ to be positive definite for all values of ω is that $-\theta_1$, θ_2 and $-\theta_3$ are positive definite. This is equivalent to the following conditions, additional to the three before:

1. $\vartheta_{u,1} \vartheta_{v,1} - \{\vartheta_{uv,1}\}^2 > 0$;
2. $\vartheta_{u,2} \vartheta_{v,3} - \{\vartheta_{uv,2}\}^2 > 0$; $\qquad$ (10)
3. $\vartheta_{u,3} \vartheta_{v,3} - \{\vartheta_{uv,3}\}^2 > 0$;

130

MRF's and their increments may be used to carry out predictions of one of the univariate components of the MRF in unsampled locations. This activity is termed cokriging. One of the important aims in cokriging is the improvement of the prediction of one of the univariate components of a MRF by using the information on the other components which are generally better sampled.

We consider an MRF $Y(x)$ with a completely known pseudo-covariance function $G(x,\tilde{x})$, i.e. the function $G(x,\tilde{x})$ is given and can serve as a pseudo-covariance function of a well-defined class of increments of $Y(x)$. We assume that a finite number of multivariate functions $\mu_k(x)$ are given, such that $\mu(x) = \Sigma\ r_k\mu_k(x)$ can serve as a pseudo-expectation vector, linear in the unknown parameters r. The functions $\mu_k(x)$ can be multivariate variables, like moisture content (regarding the values as fixed in terms of the linear expectation model), or indicator variables corresponding to qualitative factors (like vegetation type or treatment). The model matrix for the linear model will be denoted by X. Since $\mu(x)$ is unique up to a polynomial of degree $\leq k$, we assume that X also contains the finite restrictions of the multivariate monomials. In most practical cases a parameter r_k belongs to only one of the components $Y_u(x)$ of $Y(x)$, for example: a treatment variable is relevant only for yield values, and not for organic matter content.

We consider a set of locations constituted of a part where $Y(x)$ is observed and another set where we want to predict one of its components. Throughout this chapter we use the index O to denote the observational part of a vector or matrix, and P

to denote the part for which we wish to make a prediction. The finite restriction of MRF $Y(x)$, arranged in the vector Y, is partitioned into $[Y_0' \ Y_p']'$. Likewise, the weight vector λ is partitioned into $[\lambda_0' \ \lambda_p']'$, as we wish to predict a linear combination $\lambda_p'Y_p$ by a linear predictor $\lambda_0'Y_0$. In most practical cases, the P-part contains one single location, where we wish to predict the value of the main component $Y_1(x)$. Sometimes, however, the P-part consists of a square grid of locations where we wish to predict the mean value or the gradient (in one direction) of the main component $Y_1(x)$. In the example of the point value, $\lambda_p = 1$. In the prediction of the grid mean λ_p will be a vector of equal weights, not necessarily positive, but summing to 1. If we wish to predict the linear gradient along 6 equally spaced points, $\lambda_p = 1/70(-5,-3,-1,1,3,5)'$. Block cokriging, i.e. predicting the average value over an area, is a limiting case.

Now attention is focused on LUI-predictors of $\lambda_p'Y_p$: <u>linear</u> predictors $\lambda_0'Y_0$, such that the prediction error $\epsilon'Y = \lambda_0'Y_0 - \lambda_p'Y_p$ is <u>unbiased</u>, i.e. $E[\epsilon'Y] = 0$ and is an <u>increment</u>. The size of the prediction error can be quantified by its mean squared error $E[\epsilon'Y]^2 = var[\epsilon'Y]$. We are searching for the best LUI-predictor (BLUI-predictor) which is the LUI-predictor with minimum mean squared error. In fact, the prediction error is a zero-expectation increment.

A zero-expectation increment $\epsilon'Y$ satisfies $\epsilon'X=0$ where the matrix X is partitioned into $X=[X_0' \ X_p']'$. The covariance structure is represented by the finite restriction of the pseudo-covariance function, i.e. the matrix G, partitioned as $\begin{pmatrix} G_{00} & G_{0P} \\ G_{P0} & G_{PP} \end{pmatrix}$. The matrix G_{00} consists of the pseudo-covariances between the observations of the predictand, between those of

132

co-variables and those among them. The matrix G_{OP} (= G_{PO}') consists of the pseudo-covariances between observations and actually not taken values in prediction locations and the matrix G_{PP} consists of the pseudo-covariances among the prediction locations.

Any prediction error $\epsilon'Y$ with minimum variance should satisfy the requirement of the lemma in Appendix B, i.e. $\epsilon'G\delta = 0$, for all $\delta = (\delta_O'\delta_P')'$ with $\delta_P = 0$ and $\delta_O'X_O = 0$. This is equivalent to $\delta_O'(G_{OO}\lambda_O - G_{OP}\lambda_P) = 0$ for all δ_O. Therefore $G_{OO}\lambda_O - G_{OP}\lambda_P$ is a vector perpendicular to all δ_O orthogonal to the column space of X_O, i.e. belonging to the column space of X_O. Therefore there is a t such that $G_{OO}\lambda_O - G_{OP}\lambda_P = -X_O t$. This equation and the equation $\epsilon'X = \lambda_O'X_O - \lambda_P'X_P = 0$ are combined into the following partitioned matrix equation:

$$\begin{pmatrix} G_{OO} & X_O \\ X_O' & 0 \end{pmatrix} \begin{pmatrix} \lambda_O \\ t \end{pmatrix} = \begin{pmatrix} G_{OP}\lambda_P \\ X_P'\lambda_P \end{pmatrix} \tag{11}$$

If λ_O satisfies this equation for a certain t, then $\epsilon'Y = \lambda_O'Y_O - \lambda_P'Y_P$ has minimum variance and therefore $\lambda_O'Y_O$ is the BLUI-predictor of $\lambda_P'Y_P$. Therefore equation (11), the so-called 'cokriging'-equation, characterizes the BLUI-predictor of $\lambda_P'Y_P$.

The existence of a BLUI-predictor has been assured, so the cokriging equation is always solvable. However, singularity of the matrix G can cause trouble if a standard way of solving is followed. By adding a large multiple of XX' to G, we will obtain a positive definite symmetric matrix, sharing the properties of the matrix G above (compare Rao, 1989).

Solving λ_O from the first cokriging equation for given t and inserting it into the second equation gives the possibility of solving for t. Inserting the solution of t =

$-(X_0'G_{00}{}^{-1}X_0)^{-1}(X_P-G_{PO}G_{00}{}^{-1}X_0)\lambda_P$ gives the solution of λ_0. Thus we obtain

$$\text{BLUI-predictor of } \lambda_P'Y_P: \quad \lambda_P'(X_PB + G_{PO}F) \qquad (12)$$

where B is defined as $(X_0'G_{00}{}^{-1}X_0)^{-1}X_0'G_{00}{}^{-1}Y_0$ and F as $G_{00}{}^{-1}(Y_0-X_0B)$. Note that B and F are completely determined by the observed values of the MRF and their locations.

Interpretation of (12) is as follows. The first contribution, $\lambda_P'X_PB$, may be defined as the pseudo-trend. Remember that it makes no sense to consider the pseudo-trend XB as an estimation of the pseudo-expectation vector μ, since μ is an artifact and since the estimation error is not even an increment. However, $\lambda_P'X_PB$ itself is a LUI-predictor of $\lambda_P'Y_P$ by invoking generalized least squares for the estimation of the regression coefficients B and it is of polynomial form in the coordinates of the locations concerned. The second contribution $\lambda_P'G_{PO}F$ is free of polynomial effects and of μ-effects. It will improve $\lambda_P'X_PB$ as a predictor of $\lambda_P'Y_P$ by combining knowledge of the interdependence along with that of the pseudo-residuals Y_0-X_0B, again of observed predictand and co-variables with respect to estimated pseudo-expectation.

The mean square error of the BLUI-predictor is obtained by using (11):

$$
\begin{aligned}
\epsilon'G\epsilon &= \lambda_P'G_{PP}\lambda_P - \lambda_P'G_{PO}\lambda_0 - \lambda_0'G_{PP}\lambda_P + \lambda_0'G_{PO}\lambda_0 \\
&= \lambda_P'G_{PP}\lambda_P - \lambda_P'G_{PO}\lambda_0 - t'X_P\lambda_P \\
&= \lambda_P'(G_{PP}-G_{PO}G_{00}{}^{-1}G_{OP})\lambda_P + t'(X_0'G_{00}{}^{-1}X_0)t \\
&= \lambda_P'(G_{PP}-G_{PO}G_{00}{}^{-1}G_{OP} + X_a'VX_a)\lambda_P
\end{aligned}
$$

with $V = (X_0G_{00}{}^{-1}X_0)^{-1}$ and $X_a = X_P-G_{PO}G_{00}{}^{-1}X_0$.

134

Another form of the cokriging equations is obtained by the following consideration. The best predictor for Y_0 is of course Y_0 (exact interpolator). Therefore from (12) we have, replacing the P by the 0, that $X_0B + G_{00}F = Y_0$. Moreover, the columns of F are zero-expectation increments, i.e. $X_0'F = 0$. The two requirements yield the following partitioned matrix equation:

$$\begin{pmatrix} G_{00} & X_0 \\ X_0' & 0 \end{pmatrix} \begin{pmatrix} F \\ B \end{pmatrix} = \begin{pmatrix} Y_0 \\ 0 \end{pmatrix} \tag{13}$$

Solving F from the first cokriging equation for a given B and substitution in the second equation gives the already mentioned expressions for B and F. Equation (13) is the so-called induced equation (Myers, 1988; Rao, 1989).

Example 2

In this example we illustrate the use of universal cokriging in a study on moisture deficits, which we compare with results on universal Kriging (Cressie, 1986). The 404 ha Mander area in the eastern part of the Netherlands has been used for groundwater extraction during the past 30 years (Stein et al., subm.). Due to water shortage caused by this water extraction, water tables in the area were lowered and in dry periods crop yields of some of the local farmers decreased. By means of a 1:10,000 soil survey a total of 499 observations was obtained (Fig. 1). Standard simulation models were used to calculate two different moisture deficits for grassland under the present hydrological situation for every observation point: the moisture deficit for 1976 (MD76, mean value equals 59.5mm), an extremely dry year, and the yearly average moisture deficits for the period 1956 through 1985 (MD30, mean value equals
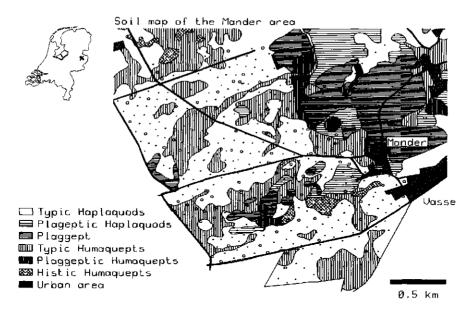
135

Fig. 1. *Soil map of the Mander area in the eastern Netherlands.*
Classification is according to USDA.

11.0mm). In this example we will illustrate the use of universal cokriging to predict moisture deficits, with the Mean Highest Groundwatertable (MHW) as a covariable.

Obviously we have to deal with two variables, both perhaps exhibiting a trend. For bivariate increments the spatial dependence is described by a pseudo-covariance function which is the sum of the functions $g_u(r)$, for the predictand, $g_v(r)$ for the covariable and $g_{uv}(r)$ for their interaction, each isotropic and linearly dependent on a vector ϑ of parameters, according to (8) where the functions $g_k(h)$ for $k > 0$ are equal to $|r|^{2k-1}$.

Simultaneous estimation of the parameters $\vartheta_{u,k}$, $\vartheta_{v,k}$ and $\vartheta_{uv,k}$ is carried out by minimizing the negative log-likelihood function (Kitanidis, 1983). This estimation procedure is commonly called the Restricted Maximum Likelihood (REML) method (Patterson and Thompson, 1975). In order to judge different

136

degrees of trend, Akaikes' information criterion is used (Akaike, 1974; Shibata, 1976).

To estimate the coefficients, the total data set was too large to be handled efficiently. We therefore selected 9 random sets, each of 25 elements. On each of these sets the coefficients of (2) were estimated. Non-permissible coefficients were assigned the value 0 (Delfiner, 1976). The final coefficients are obtained by averaging the coefficients obtained for each of the 9 sets.

Determination of the coefficients ϑ (Table 1a) and evaluation of Akaikes' criterion revealed a second order trend for the three variables. This is not surprising, because low values for the water tables, are associated with soils predominantly occurring in the western part of the area, and pertaining the largest part of the area (appr. 85%), whereas high values are observed for plaggeptic soils which occur in a relatively small area of 15% in the eastern part of the area and moisture deficits are correlated with these.

The coefficients which are estimated for the pseudo-covariance functions for bivariate increments (Table 1b) are in agreement with the corresponding coefficients for the pseudo-covariance functions for univariate increments. For example, the coefficients $\vartheta_{2,1}$ and $\vartheta_{2,2}$ from Table 1b correspond with the coefficients $\vartheta_{1,1}$ and $\vartheta_{1,2}$ for MHW in Table 1a.

Results from universal cokriging are compared with those obtained by means of universal Kriging by means of predicting to a test set of 100 locations which are randomly selected in advance from the complete data set. Whenever universal Kriging was applied, a neighbourhood of 18 observations was used. For universal cokriging the neighbourhood was extended with 19 observations for the covariable. For the 100 test locations the

137

TABLE 1a. *Estimated coefficients for pseudo-covariance functions for different variables.*

Variable		$\vartheta_{1,1}$	$\vartheta_{1,2}$	$\vartheta_{1,3}$	$\vartheta_{1,4}$
MD76	[mm]	885	-82	0	-0.92
MD30	[mm]	48.3	-5.9	0	-0.04
MHW	[cm]	337	-114	0	-1.05

TABLE 1b. *Simultaneous estimated coefficients for pseudo-covariance functions and pseudo-cross-covariance functions for MD30 and MD76 with MHW.*

		MD30	MD76
MD	$\vartheta_{1,1}$	33.4	774.4
	$\vartheta_{1,2}$	-5.2	-36.0
	$\vartheta_{1,3}$	0	0
	$\vartheta_{1,4}$	0	0
MHW	$\vartheta_{2,1}$	411	499
	$\vartheta_{2,2}$	-77	-56
	$\vartheta_{2,3}$	0	0
	$\vartheta_{2,4}$	0	0
MD x MHW	$\vartheta_{12,1}$	36.6	324.2
	$\vartheta_{12,2}$	-15.8	-33.0
	$\vartheta_{12,3}$	0	0
	$\vartheta_{12,4}$	0	0

	MSE		MVP	
	UK	UCK	UK	UCK
MD30	66	66	65	42
MD76	1219	1229	1190	989

Mean Squared Error (MSE), defined as the mean of the squared difference of predicted and observed values, as well as the Mean of the Variances of the Prediction Errors (MVP) is calculated in these points (Table 2).

We notice that MSE values are approximately the same for both universal cokriging and for universal Kriging for the two moisture deficits considered, whereas MVP values are lower for universal cokriging as compared to universal Kriging. Evidently, universal cokriging gives more precise predictions than universal Kriging. Multiple predictions may be combined to yield a predictive map of one of these Moisture Deficits (Fig. 2).
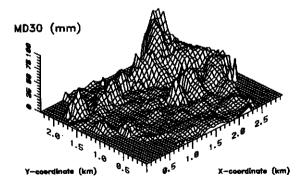

Fig. 2. 30-Year average Moisture Deficit for grassland obtained by means of universal cokriging with the mean highest Groundwater level as a co-variable.

139

CONCLUSIONS

1. Universal cokriging is an extension to current statistical prediction and interpolation procedures; it is now routinely being used in environmental studies like studies in soil science.

2. Universal cokriging allows one to predict values in a configuration, which may constitute a single locations, multiple points or a gradient.

3. For many practical studies when one has to deal with one covariable only, evaluation of the permissibility of pseudo-covariance functions and cross-pseudo-covariance functions is straightforward.

REFERENCES

Akaike, H., 1974, A new look at the statistical model identification, IEEE Trans. Automat. Control, AC-19, p. 716-723.

Christakos, G., 1984, On the problem of permissible covariance and variogram models: Water Resources Research 20, p.251-265.

Christensen, R., 1990, The equivalence of predictions from universal Kriging and intrinsic random-function Kriging. Mathematical Geology 22, 655-665.

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

Cressie, N.A.C., 1986, Kriging Nonstationary Data: J. Amer. Stat. Assoc., v. 81, p. 625-634.

Cressie, N.A.C., 1989, Geostatistics: The American Statistician 43, p.197-202.

Delfiner, P., 1976, Linear Estimation of Nonstationary Spatial Phenomena: in Guarascio, M., et al. (Eds.), Advanced Geostatistics in the Mining Industry, Reidel, Dordrecht, The Netherlands, p. 49-68.

Dowd, P.A., 1989, Generalized Cross-covariances: in Armstrong, M. (Ed.), Geostatistics, Dordrecht, p. 151-162.

Gelfand, I.M., and Shilov, G.E., 1964, Generalized Functions 1, Academic Press, New York, 423p.

Gelfand, I.M., and Vilenkin, N., 1964, Generalized Functions 4, Academic Press, New York, 384p.

Ito, K. Stationary random distributions. Mem. Coll. Sci. Univ. Kyoto 28, 209-233.

Journel, A.G. and C.J. Huijbregts. 1978. Mining geostatistics. Academic Press, London, 600p.

Kitanidis, P.K., 1983, Statistical Estimation of Polynomial Generalized Covariance Functions and Hydrologic Applications: Water Resour. Res., v. 19, p. 901-921.

Kolmogorov, A.N., 1956, Foundations of the theory of probability, second edition. Chelsea, New York, 70pp.

Krajewski, W.F., 1987, Cokriging Radar-rainfall and Rain Gage Data: J. Geoph. Res., v. 92, p. 9571-9580.

Leenaers, H., Ockx, J.P., and Burrough, P.A., 1989, Co-kriging: an Accurate and Inexpensive Means of Mapping Flood Plain Soil Pollution by using Elevation Data, in Armstrong, M. (Ed.), Geostatistics, Dordrecht, p. 371-382.

Matheron, G., 1973, The intrinsic random functions and their applications: Adv. Appl. Prob., v. 5, p. 439-468.

Myers, D.E., 1982, Matrix Formulation of Co-kriging: Math. Geol., v. 14, p. 249-257.

Myers, D.E., 1988. Multivariable geostatistical analysis for environmental monitoring. Sci. de la Terre, v. 27, p. 411-

427.

Patterson, H.D., and Thompson, R., Maximum likelihood estimation of components of variance, in Corsten, L.C.A., and Postelnicu, T. (Eds.) *Proceedings of the 8th International Biometric Conference, Constanta*, 197-207, 1975.

Rao, C.R., 1989, A Lemma on Optimization of a Matrix Function and a Review of the Unified Theory of Linear Estimation, in Dodge, Y. (Ed.), Statistical Data Analysis and Inference, Amsterdam, p. 397-417.

Shibata, R., 1976, Selection of the order of an autoregressive model by Akaike's information criterion. Biometrika 63, 117-126.

Stein, A., Bouma, J., Mulders, M.A., and Weterings, M.H.W., 1989, Using Spatial Variability Studies to Estimate Physical Land Qualities of a Level River Terrace: Soil Technology, v. 2., p. 385-402.

Stein, A., and Corsten, L.C.A., 1990, Universal Kriging and Cokriging as a Regression Procedure: Biometrics, in press.

Stein, A., Staritsky, I., Bouma, J., Van Eijnsbergen, A.C., and Bregt, A.K. Subm. Simulation of moisture deficits and interpolation by universal Cokriging. Submitted to Water Resources Research.

Vauclin, M., Vieira, S.R., Vachaud, G., and Nielsen, D.R., 1983, The Use of Cokriging with Limited Field Soil Observations: Soil Sc. Soc. Amer. J., v. 45, p. 1040-1048.

Webster, R., 1985, Quantitative Spatial Analysis of Soil in the Field. In: R. Stewart (ed.) Advances in Soil Science, Vol. 3, Springer, New York, p. 1-70.

Yaglom, A.M., 1986, Correlation Theory of Stationary and Related Random Functions. I: Basic Results. II:

Supplementary notes and references. Springer, New York, 526p.

APPENDIX A

<u>Two remarks concerning conditional positivity</u>.

1. The conditions to be imposed on the coefficients $\vartheta_{u,k}$ in order to achieve a pseudo-positive-definite pseudo-covariance function. In addition to the parameters introduced in section 2, let $\vartheta_{u,0}$ indicate the nugget effect. Then the requirements that, in a two-dimensional case, $\vartheta_{u,0} \geq 0$, $\vartheta_{u,1} \leq 0$, $\vartheta_{u,3} \leq 0$ and $\vartheta_{u,2} \geq -10/3*\sqrt{\vartheta_{u,1}\vartheta_{u,3}}$ do not apply to pseudo-cross-covariance functions. For example, $\vartheta_{uv,0}$ may be negative. To show this, assume that (Y_u,Y_v) is a bivariate random function where Y_u is a pure nugget effect and $Y_v = -Y_u$ with probability 1. Then $\vartheta_{uv,0} = -\vartheta_{u,0} = -\vartheta_{v,0} < 0$.

2. The above conditions and the conditions (10) that $(\vartheta_{uv,k})^2 \leq \vartheta_{u,k}\vartheta_{v,k}$ are not both necessary in order that the variances of any permissible increment be non-negative. As a counter example we take a MRF with $p = 2$ and $k = 2$, without nugget effect for the two (identical) constituent URF's and with the following parameters for the pseudo-(cross-)covariance functions:

	URF$_1$	URF$_2$	Cross
$\vartheta_{u,0}$	0	0	0
$\vartheta_{u,1}$	1	1	0
$\vartheta_{u,2}$	0	0	$\epsilon/9$
$\vartheta_{u,3}$	1/225	1/225	0

where ϵ is some number > 0. The matrix $F(\omega)$ now becomes with d equal to 2:

$$F(\omega) = \begin{pmatrix} |\omega|^{-3} + |\omega|^{-7} & \epsilon |\omega|^{-5} \\ \epsilon |\omega|^{-5} & |\omega|^{-3} + |\omega|^{-7} \end{pmatrix}$$

This matrix is positive definite for every $\omega \neq 0$ if $|\epsilon| < 2$.

APPENDIX B

In order to find the BLUI predictor of $\lambda_p' Y_p$, we start with a given LUI-predictor and consider the corresponding prediction error $\varphi' Y$. There always exists a LUI-predictor of $\lambda_p' Y_p$, for example $\lambda_p' X_p (X_0' X_0)^{-1} X_0' Y_0$. We assume that the locations at which the observations are made induce a full column rank matrix X_0. The prediction error corresponding to any other LUI-predictor of $\lambda_p' Y_p$ is of the form $\varphi' Y$ minus $\delta' Y$ where $\delta' Y$ is a zero-expectation increment and $\delta = (\delta_0' \delta_p')'$ with $\delta_p = 0$, i.e. δ belongs to a linear space called M.

The following lemma gives the possibility of improving the LUI-estimation error $\varphi' Y$ to the prediction error $\epsilon' Y$ of the BLUI-predictor (Gauss-Markov reduction). However, the lemma holds quite generally. We only assume that the variances and the covariances of the $\varphi' Y$ and the $\delta' Y$ are defined. The key is to determine in the space M the vector δ such that $\varphi - \delta$ has minimum length, where the inner product is governed by a metric equal to the pseudo-covariance matrix. This is equivalent to finding the projection ϵ of φ perpendicular to M in that sense.

144

Lemma Let φ be given and let ϵ be such that $\epsilon-\varphi$ is in M. Then $\epsilon'Y$ has minimum variance iff. $\text{cov}(\epsilon'Y,\delta'Y) = 0$ for all δ in M.

Proof Let φ be given, let ϵ be such that $\epsilon-\varphi$ is in M and let $\epsilon'Y$ satisfy $\text{cov}(\epsilon'Y,\delta'Y) = 0$ for all δ in M. Choose a δ in M. Then $(\varphi-\epsilon)-\delta$ is in M, and

$$\text{Var}((\varphi-\delta)'Y) = \text{var}(\epsilon'Y) + 2\text{cov}(\epsilon'Y,((\varphi-\epsilon)-\delta)'Y)$$
$$+ \text{var}(((\varphi-\epsilon)-\delta)'Y)$$
$$= \text{var}(\epsilon'Y) + \text{var}(((\varphi-\epsilon)-\delta)'Y)$$
$$\geq \text{var}(\epsilon'Y).$$

Because this holds for any δ in M, the if part is proven.

Conversely, let $\epsilon'Y$, with ϵ and φ such that $\epsilon-\varphi$ is in M, have minimum variance and let δ be in M. Then $\omega'Y$ with $\omega = \epsilon - (\text{cov}(\epsilon'Y,\delta'Y)/\text{var}(\delta'Y))\delta$ is in M. Since $\epsilon'Y$ has minimum variance, we have that $\text{var}(\omega'Y) \geq \text{var}(\epsilon'Y)$. On the other hand, working out the variance of $\omega'Y$ gives $\text{var}(\omega'Y) = \text{var}(\epsilon'Y) - (\text{cov}(\epsilon'Y,\delta'Y))^2/\text{var}(\delta'Y) \leq \text{var}(\epsilon'Y)$. Therefore, $\text{cov}(\epsilon'Y,\delta'Y) = 0$, which completes the proof.

PART III. APPLICATIONS TO STUDIES IN SOIL SCIENCE

CHAPTER III.1. COKRIGING POINT DATA ON MOISTURE DEFICIT

by

A. Stein, W. van Dooremolen, J. Bouma and A.K. Bregt.

Cokriging Point Data on Moisture Deficit

A. Stein,* W. van Dooremolen, J. Bouma, and A. K. Bregt

ABSTRACT

Existing computer calculations by simulation of the moisture deficit (MD) were used for 500 point observations in an area of 404 ha with sandy soils in the Netherlands. The statistical prediction techniques kriging and cokriging were used to predict MD-values in 100 points selected at random from the 500 available points. The MD-data at the remaining 400 points were used for the predictions. The mean variance of prediction error (MVPE) and the mean squared error of prediction (MSEP) decreased only slightly when kriging is compared with cokriging using the mean highest water-table level as a covariable. The number of MD-values used in cokriging could, however, be reduced from 400 to 160 with only a small loss of accuracy in using 400 observations on the mean highest water-table (MHW) as a covariable. As the MD-variables being considered are four times as expensive to determine as the covariable, which is routinely estimated during soil survey, this also represents a considerable reduction of costs. Cokriging can thus be a useful technique to more effectively utilize available soil survey information.

COMPUTER SIMULATION techniques for predicting the soil moisture deficit are being used in the Netherlands to estimate yield reductions due to lowering of the water table by water extraction for mu-

A. Stein, Dep. of Soil Statistics, W. van Dooremolen and J. Bouma, Dep. of Soil Science, Agricultural Univ. Box 37, 6700 AA Wageningen, Netherlands; A.K. Bregt, Dep. of Applied Information Science and Statistics, the Netherlands Soil Survey Inst., Wageningen, Netherlands. Received 21 Dec. 1987. *Corresponding author.

nicipal water supply (e.g. Bouma et al., 1980a,b). Soil maps are used to estimate basic hydrological properties such as hydraulic conductivity, moisture retention and water table fluctuations (e.g. Wösten et al., 1985). Simulations are carried out for so-called representative soil profiles for each mapping unit. This is a clear disadvantage, because heterogeneity within units is thus ignored. An alternative, more attractive procedure is to produce simulations for each separate soil boring, followed by predictions for unvisited spots by means of the prediction technique of kriging and averaging to obtain a predictive map (e.g. Bregt et al., 1987; De Wit and Van Keulen, 1987). Costs of borings, data gathering and simulation calculations are, however, relatively high. Procedures have to be developed, therefore, to cut costs to the point that maximum information is provided at minimum cost.

The objective of this study was to explore use of available soil survey information for development of such procedures. A potentially attractive procedure is the statistical prediction technique of cokriging (McBratney and Webster, 1983; Yates and Warrick, 1987), which not only uses observations of a particular variable, such as the moisture deficit, but also observations of covariables such as soil or water table characteristics that are routinely assembled during soil survey. Cokriging could conceivably result in a considerable reduction of costs in achieving a comparable degree of accuracy by using fewer relatively expensive variables and more relatively inexpensive co-variables.

MATERIALS AND METHODS

Study Area and Procedures

In 1985 a detailed soil survey was carried out in the Mander area in the eastern part of the Netherlands by the Dutch Soil Survey Institute (Stiboka) to study the effect of groundwater extraction on the production of grassland (Stoffelsen and Van Holst, 1985; Wösten et al., 1987). The Mander area consists of sandy soils, which are classified as Haplaquods, Humaquepts and Plaggepts (Wösten et al., 1987). Fluctuations of the water table are routinely characterized by Dutch soil survey in terms of mean highest (MHW) and mean lowest (MLW) levels (Van der Sluijs and De Gruijter, 1985; Fig. 1a). They are classified as indicated in Table 1. Lowering of water tables results in change in productions, e.g. of grassland due to lower fluxes of water from the water table to the root zone. Farmers have to be financially compensated for these production losses by municipal water companies. Computer simulation techniques by the simulation model LAMOS are used to obtain the necessary quantitative expressions for the moisture deficit and the associated yields, as has been discussed elsewhere (Bouma et al., 1980a,b). Simulation calculations were made for a 30-yr period. For this study attention was arbitrarily focused on two results, viz. the 30-yr avg. value for the moisture deficit for grassland (MD30) and the value for the dry year 1976 (MD76) (Fig. 1b, c). In the area of 404 ha being studied, 500 soil borings were made. This observation density is generally assumed to be representative for surveys of scale 1:10 000. Hydraulic conductivity and moisture retention data for major soil horizons were used as they were obtained from a previous study (Wösten et al., 1985, 1987). Simulations were carried out for each boring location. To evaluate the predictions by kriging and cokriging, 100 borings were selected at random from the data set to serve as a test set. Predictions were made for the test set with the calculated values for the moisture deficit of the remaining 400 borings (see statistical procedures). Attention was focused on the average precision obtained for the 100 test locations. Concerning the scale of maps a rule of thumb for soil survey recommends that approximately four observations per cm^2 of map area are to be taken. This rule was questioned, however, by investigating the possibility of taking fewer observations. More observations would not be feasible for economic reasons. Data sets with reduced numbers of observations were generated by random numbers corresponding to four map scales (Table 2). Finally an analysis of costs was made to allow a financial comparison between kriging and cokriging.

Statistical Procedures

In carrying out the statistical analyses the statistical prediction techniques kriging and cokriging have been used. Cokriging, closely linked to kriging (Matheron, 1973), allows the use of a second variable (the covariable) in predicting values of the variable of interest (the predictand). The new approach being used here, forms a special version of a more general approach dealing with polynomial trends up to any order. The method is here considered to be a statistical prediction technique (Kendall and Stuart, 1973) rather than a numerical interpolation technique (Vauclin et al., 1983). The

Table 1. Groundwater classes and mean highest and mean lowest groundwater levels.

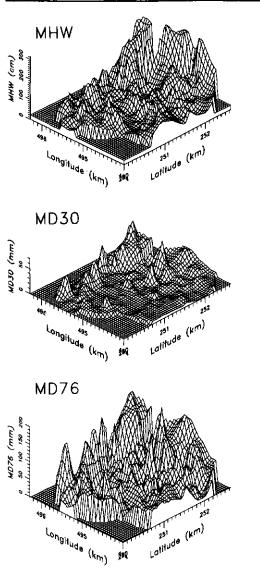| Class	Groundwater level, cm	
	Mean highest	Mean lowest
3	<40	80–120
5	<40	>120
6	40–80	>120
7	>80	>120

Fig. 1. Maps of MHW, MD30, and MD76 in the Mander area, obtained by kriging the original soil survey data.

Table 2. Correspondence between observation density, the approximate mapping scale, and the number of observations in the data sets.

Observ. density. obs./km²	99	44	25	16	11
Approx. map scale	10 000	15 000	20 000	25 000	30 000
No. of observ.	399	177	100	64	44

technique has been developed in the context of the Generalized Linear Model with dependent observations (Rao, 1973), and can be described as follows. Throughout the text we use boldface type to denote that a variable is stochastic, in order to distinguish it from nonstochastic variables. A predictor t is used in a number of points, e.g. the nodes of a grid. Let one such point be denoted by z_0 and let a prediction t of the value of y_0 in z_0 be based on n observations of the predictand in the observation points $z_1, \ldots, z_n$ and m observations of the covariable in the observation points $z_n, \iota, \ldots, z_{n+m}$. Let X be a matrix, $m + n$ by 2, x_0 a two-vector and y a stochastic $m + n$-vector, composed of an n-vector y¹ and an m-vector y² pointing to the predictand and the co-variable, respectively, such that
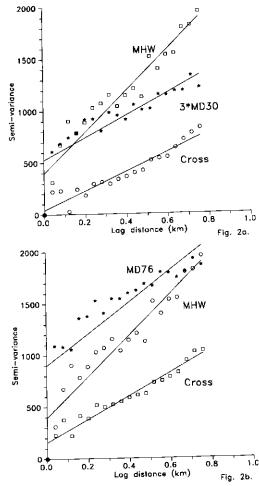

Fig. 2a.

Fig. 2b.

Fig. 2. Semivariograms for MD30, MD76, and MHW, and their cross variograms with MHW. The MD30 values are multiplied by 3 to obtain comparable pictures.

$$X = \begin{bmatrix} 1_n 0 \\ 0 \ 1_m \end{bmatrix}; \ x_0' = (1 \ 0); \ y = \begin{bmatrix} y^1 \\ y^2 \end{bmatrix};$$

1_n is a vector consisting of n ones and 1_m is a vector consisting of m ones. The model can be formulated as

$$y = x\beta + e,$$

with $E(e) = 0$, and var(e) $= C$ and unknown parameter vector β. We have to predict $y_0 = x_0\beta + e_0$ with var(e_0) $= c_{00}$. Under the assumption of stationarity and finite variance the elements of C are given by the covariance functions $c_1(h)$ and $c_2(h)$ for the predictand and the covariable, respectively, and the cross-covariance function $c_{12}(h)$ between both of them, all three depending on the distance h between observation points. The value for c_{00} is given by $c_1(0)$. A generalized least squares estimator for β is $\hat{\beta}$, defined as $\hat{\beta} = (X'C^{-1}X)^{-1}X'C^{-1}y$.

In general, as in this study, the demand of finite variance is too strong. Use can be made of semivariograms $g_1(h)$ and $g_2(h)$ for the predictand and the covariable and of the cross-variogram $g_{12}(h)$ for the spatial interaction between predictand and covariable, respectively, depending on the distance h between observation points. These variograms can be estimated by

$$g_k(h) = \Sigma(Y_i^k - Y_{i+h}^k)^2/2N(h) \text{ for } k = 1,2; \quad [1]$$

$$g_{12}(h) = \Sigma(Y_i^1 - Y_{i+h}^1)(Y_i^2 - Y_{i+h}^2)/2N'(h); \quad [2]$$

Summation in Eq. [1] is for every h taken over all $N(h)$ pairs of observation points of the predictand and the covariable, respectively, being a distance h apart, and in Eq. [2] for every h over all $N'(h)$ pairs of points for which observations of the two variables are available being a distance h apart. Let G be the $n + m$ by $n + m$ matrix with element $g_{ij} = -g(h_{ij})$, being the opposite of the value of the variogram between observation points z_i and z_j and let g_0 be the $(n + m)$ vector with element g_{0i} being the opposite of the value of the variogram between z_0 and the ith observation point. The nugget effect is denoted by g_{00}. We then obtain as a stochastic predictor

$$t = x_0'\hat{\beta} + g_0'G^{-1}(y - X\hat{\beta}). \quad [3]$$

As a measure for the precision of a prediction the variance of the prediction error is used, being equal to

$$\text{var}(t - y_0) = g_{00} - g_0'G^{-1}g_0 + x_a'(X'G^{-1}X)^{-1}x_a[4]$$

where x_a is defined as $x_a = x_0 - g_0'G^{-1}X$.

Since the value of Eq. [4] is zero in an observation point (both kriging and cokriging are exact predictors), a test set of 100 points was randomly selected from the original set of 500 calculated data. For the test set predictions are carried out on the basis of MD-data in a neighborhood of the eight nearest points of the remaining 400 points. The variances of the prediction errors, Eq. [4], were averaged over the test set, yielding the mean variance of prediction error (MVPE). Also, predictions in the test locations were afterwards compared with the corresponding observations, giving a second measure of the performance of the predictor, the mean of squared errors of prediction (MSEP). The two measures are equal to

$$\text{MSEP} = (1/n) * \sum_{i=1}^{n} (t_i - y_i)^2 \quad [5]$$

and

$$\text{MVPE} = (1/n) * \sum_{i=1}^{n} \text{var}(t_i - y_{0i}), \quad [6]$$

respectively. In both Eq. [5] and [6] summation is carried out over the test set, in Eq. [5] expression [4] is used, whereas

153

in Eq. [6] the prediction is compared with the observed value. Although these measures are fundamentally different in that the MVPE is stochastic and the MSEP is not, they can both be used to obtain an indication of the quality of predictions. The relation between the variance of the prediction error and the squared error of prediction is given by

$$E(t - y_0)^2 = \text{var}(t - y_0) + \{E(t - y_0)\}^2 . \quad [7]$$

The term $E(t - y_0)^2$ is estimated by Eq. [5], in other words, the MSEP is an estimator of the mean value of the second moment. Because $\{E(t - y_0)\}^2$ is equal to zero, the prediction model is considered appropriate, if the estimator for $E(t - y_0)^2$ is about equal to the estimator for $\text{var}(t - y_0)$, that is if the MSEP and the MVPE are about equal. In practice, of course, differences may exist between the MVPE and the MSEP, due to deficiencies in the prediction model, i.e. the bias of the model. The MVPE is always smaller than the MSEP.

By means of the relative precision (rp) different prediction techniques are compared with kriging based on 400 observation points. It is defined as

$$\text{rp} = \text{MVPE}^0/\text{MVPE}^1 \quad [8]$$

where MVPE^0 is the MVPE of kriging with 400 observation points and MVPE^1 is the MVPE of the compared prediction technique. A value of the r.p. > 1, for instance, points to a prediction technique that is more precise than kriging with 400 observation points.

Table 3. Coefficients of weighted linear regression models for variograms, $\hat{g}(h) = a + b^*h$ (see also Fig. 2).

	a	b
MD30, mm	74	89
MD76, mm	1030	1560
MHW, cm	495	1780
MD30-MHW	12	295
MD76-MHW	171	1090

Table 4. Correlation coefficients.

		MD30	MD76	MHW	MHL
Moisture deficit					
30-yr avg.	MD30	1.000			
Dry year 1976	MD76	0.934	1.000		
Groundwater level					
Mean highest	MHW	0.587	0.636	1.000	
Mean lowest	MLW	0.551	0.600	0.915	1.000

Table 5.1a. MVPE of MD76 for different observation densities obtained by means of reference variograms.

	Predictand				
Covariable	99	44	25	16	11
99	171	215	269	360	521
44		225	288	384	556
25			294	397	578
16				413	603
11					615

Table 5.2a. MVPE of MD30 for different observation densities obtained by means of reference variograms.

	Predictand				
Covariable	99	44	25	16	11
99	12.4	15.6	19.6	26.1	38.1
44		16.4	21.0	28.0	40.5
25			21.4	28.9	42.1
16				30.0	43.8
11					44.7

Two approaches are followed for the production of the test set. First, it can be assumed that no data are available in the test-set points, except their coordinates, which is the case when a map has to be produced by means of predictions in grid nodes. Second, it can be assumed that in each of the test-set points an observation of the covariable is available. This distinction will be denoted by cokriging 1 and cokriging 2, respectively. To select the most promising covariables, correlation coefficients and crossvariograms were determined between calculated moisture deficit and different soil survey data, such as actual mean highest and mean lowest water table, rooting depth, loam fraction from top- and subsoil, amount of organic matter and clay fraction of the topsoil.

RESULTS

Comparing Kriging with Cokriging

Calculations for the MD30 and the MD76, as made for the 400 boring locations in the study area, were on the average 10.3 and 57.7 mm, while the individual standard deviations were 14.5 and 50.7 mm, respectively. These values could be used for estimation in any unvisited point in the area, if spatial dependency would be absent. Better estimates are, however, obtained when the spatial dependence is taken into account as is inherent to the kriging and the cokriging procedures.
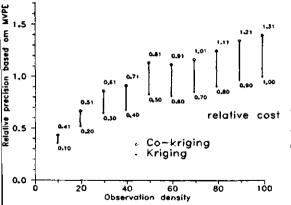
Semivariograms were calculated for the variables being considered. A linear model without sill fits sufficiently well (Fig. 2a, 2b). Fitting was carried out by means of weighted linear regression, weights being based on the number of pairs of points in the distance classes (Table 3). For cokriging the mean highest watertable before extraction (MHW) was used as a covariable as it showed the highest correlation with MD30 and MD76 (Table 4) and a clear spatial dependence. The variables MD30 and MD76 and the covariable MHW are known in every sample location. To simulate differences in scale, the observation density for MD30 and MD76 of the original data set was randomly reduced, such that the number of observations per unit map sheet be equal on every scale. The MSEP and the MVPE values for kriging and cokriging 1 and cokriging 2 were compared for observation densities 11 obs/km^2 for MD30 and MD76 and 99 obs./km^2 for MHW. Semivariograms were calculated for variables and covariables for all data points, as well as for the
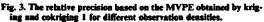
Table 5.1b. MVPE of MD76 for different observation densities obtained by means of scale dependent variograms.

	Predictand				
Covariable	99	44	25	16	11
99	171	242	299	455	539
44		254	325	502	600
25			335	533	654
16				560	699
11					713

Table 5.2b. MVPE of MD30 for different observation densities obtained by means of scale dependent variograms.

	Predictand				
Covariable	99	44	25	16	11
99	12.4	18.0	19.4	33.9	37.9
44		18.9	21.2	36.8	41.8
25			22.0	38.7	45.3
16				40.4	48.1
11					49.1

Fig. 3. The relative precision based on the MVPE obtained by kriging and cokriging 1 for different observation densities.

four sets of reduced numbers of observations. Of particular interest for soil survey is the case of having few observations of the predictand and many of the covariable. The results for the different observation densities are summarized in Tables 5.1a to 5.2b and in Fig. 3. In calculating semivariogram values a minimum of 30 pairs of observations in every distance class is needed. This restricted the number of possible combinations to be tested (Fig. 4a,b). Only semivariograms for MD30 (Fig. 4a) and cross variograms for MD30 and MHW (Fig. 4b) are shown; comparable figures were obtained for MD76.

Results can be summarized as follows:

1. According to the MVPE and the MSEP values obtained, cokriging 2, i.e. cokriging with an observation of the covariable in the prediction point, leads to more precise predictions than cokriging 1. A prediction obtained by cokriging is more precise than a prediction obtained by kriging with the same observations of the predictand. MVPE values indicate that the standard deviations for MD30 and MD76 are 4- to 20-mm lower for cokriging than for kriging (Table 6).

2. There is a small loss in precision by reducing the number of observations in the predictand up to 40% of the original number. Further reductions result in a larger increase (Fig. 5).

3. A change in the observation density of the covariable had less severe implications for the precision than a change in the observation density of the predictand. Reducing the number of observations by 89% from 399 to 44, for instance, resulted for the predictand in an increase of the MVPE of 400%, and reducing the number of ob-

Table 6. Comparison of cokriging and kriging for 11 obs. per km² for the predictand and 99 obs. per km² for the covariable.

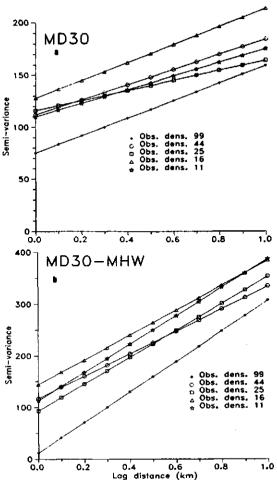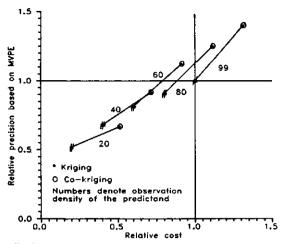|		Kriging	Cokriging 1	Cokriging 2
MSEP	MD30	126	111	102
	MD76	2460	2230	1970
MVPE	MD30	44.7	38.1	31.7
	MD76	615	521	433

Fig. 4. Fitted models for semivariograms for MD30 and cross variograms for MD30 with MHW for different observation densities.
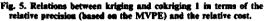
servations of the covariable in an increase of the MVPE of 130%. Incorporation of a covariable reduced the increase of the MVPE from 400 to 300%.

4. Use can be made of semivariograms based on the reduced numbers of observations without seriously affecting the quality of the predictions.

The Associated Costs

The cost for observations and calculations of the predictand is approximately four times as high as the cost for observations of the covariable. On the basis of the relation between relative cost of a survey and relative precision based on MVPE, is graphically presented in Fig. 5. As can clearly be seen from, for example, the 99 observations per km² standard situation, the introduction of observations of a covariable enhances the relative precision (by approximately 40%) and the relative cost (by approximately 30%). Both reduction of costs (by only 10%) and increase of pre-

cision (also by 10%) can, for example, be achieved in the situation of a density of 60 observations per km^2 for the predictand and the original density of 99 observations per km^2 for the covariable. Reduction of observation density from 100 to 20 observations per km^2, reduces relative costs to 0.2 and relative precision to 0.5. Inclusion of cokriging reduces relative costs to 0.5 and relative precision to 0.65. Even larger reductions in costs with increase of precision are obtained when the cost ratio between observations of the predictand and the covariable would be $> 4{:}1$.

DISCUSSION

This study was focused on one important land quality, the moisture deficit, and used only one covariable. The principle being illustrated is, however, valid for other land qualities and covariables still to be selected. Studies are needed to further explore the practical feasibility of using cokriging.

This study was based on the use of variograms. Some considerations should therefore be kept in mind when carrying out cokriging. The necessary number of data to estimate variograms should be more than 30 to 50 per distance class (Journel and Huijbregts, 1978, p. 194). In order to obtain reliable estimates, the number of paired observations (i.e. observations of the predictand and the covariable in one location) must be relatively high (at least some 40 to 60).

Availability of observations of the covariable in a point where a prediction is to be carried out leads to a reduction of the variance of the prediction error by 30%, instead of 15% if no such observation is available and only kriging is applied. Therefore, use of an observation of a covariable, which is highly correlated with a land quality being studied, enhances the precision of the prediction.

There are substantial financial benefits when using cokriging. Sixty percent of observations of expensive variables could be replaced in this study by less expensive variables without serious loss of precision. Observations on these less expensive variables are to

be made in the very locations where a prediction is needed. Up to now it is not clear how to select the most promising covariables. Selection should not only be based upon the correlation coefficient. The spatial relation between the predictand and potential covariables should be taken into account as well. It could be useful to use multivariate techniques with the purpose to combine two or more variables into one new covariable. Studies on this approach are in progress.

In Geographical Information Systems, use of cokriging appears to be attractive, as the performance of a GIS is governed by quality of input data. Kriging offers the opportunity to produce predictive single value maps while cokriging allows incorporation of different soil parameters that are usually available in soil data bases, so as to enhance the precision of output. Computing time will increase when using cokriging, but the availability of high capacity hardware and skillful programming may overcome these problems. Also, in combining two sources of information in terms of variables and covariables, cokriging explicitly uses their spatial relations. Maps of variances of the prediction error can be provided, showing the location of the more or less reliable parts of maps of the predictands.

REFERENCES

Bouma, J., P.J.M. de Laat, R.H.C.M. Awater, H.C. van Heesen, A.F. van Holst, and Th. J. van de Nes. 1980a. Use of soil survey data in a model for simulating regional soil moisture regimes. Soil Sci. Soc. Am. J. 44:808–814.
Bouma, J., P.J.M. de Laat, A.F. van Holst, and Th. J. van de Nes. 1980b. Predicting the effects of changing water-table levels and associated soil moisture regimes for soil survey interpretations. Soil Sci. Soc. Am. J. 44:797–802.
Bregt, A.K., J. Bouma, and M. Jellinek. 1987. Comparison of thematic maps derived from a soil map and from kriging of point data. Geoderma 39:281–291.
de Wit, C.T., and H. van Keulen. 1987. Modelling production of field crops and its requirements. Geoderma 40:253–265.
Journel, A.G., and Chr. J. Huijbregts. 1978. Mining geostatistics. Academic Press, London.
Kendall, M.G., and A. Stuart. 1973. The advanced theory of statistics. Vol. 2. 3rd ed. Griffin, London.
Matheron, G. 1973. The intrinsic random functions and their applications. Adv. Appl. Prob. 5:439–468.
McBratney, A.B., and R. Webster. 1983. Optimal interpolation and isarithmic mapping of soil properties. V. Coregionalization and multiple sampling strategy. J. Soil Sci. 34:137–162.
Rao, C.R. 1973. Linear statistical inference and its applications. John Wiley & Sons, New York.
Stoffelsen, G.H., and A.F. van Holst. 1985. Soil scientific and hydrologic investigations in the Mander watershed area. Internal report. (In Dutch). Soil Survey Inst. Stiboka, Wageningen, Netherlands.
Van der Sluis, P., and J.J. de Gruijter. 1985. Water table classes: A method used to indicate seasonal fluctuation and duration of water tables on Dutch soil maps. Agric. Water Manage. 10(2):109–125.
Vauclin, M., S.R. Vieira, G. Vachaud, and D.R. Nielsen. 1983. The use of cokriging with limited field soil observations. Soil Sci. Soc. Am. J. 47:175–184.
Wösten, J.H.M., J. Bouma, and G.H. Stoffelsen. 1985. Use of soil survey data for regional soil water simulation models. Soil Sci. Soc. Am. J. 49:1238–1244.
Wösten, J.H.M., M.H. Bannink, and J. Bouma. 1987. Land evaluation at different scales: You pay for what you get. Soil Surv. Land Eval. 7:13–24.
Yates, S.R., and A.W. Warrick. 1987. Estimating soil water content using cokriging. Soil Sci. Soc. Am. J. 51:23–30.

156

CHAPTER III.2. USING CO-KRIGING IN VARIABILITY STUDIES TO PREDICT PHYSICAL LAND QUALITIES OF A LEVEL RIVER TERRACE

by

A. Stein, J. Bouma, M.A. Mulders and M.H.W. Weterings

USING CO-KRIGING IN VARIABILITY STUDIES TO PREDICT PHYSICAL LAND QUALITIES OF A LEVEL RIVER TERRACE

A. **Stein**, J. **Bouma**, M.A. **Mulders**, M.H.W. **Weterings**
Wageningen

Summary

On a level river terrace with only limited physiographic features in the Limagne area in France, physical measurements were conducted at 62 sample locations, focusing on the spatial variability of soil-available water. Some 280 observations on a logarithmic sample grid were taken as well to measure basic soil properties. Pedo-transfer-functions were used to estimate soil physical characteristics which are usually difficult to measure in the field. Variability is studied with emphasis on anisotropy and the nugget effect.

As physical soil characteristics are closely related with crop conditions and associated spectral imagery, relations were investigated between multispectral aerial photography and soil physical land characteristics. The vegetation index showed a highly significant relationship (F-value $<.001$) with the available water in the profile.

A predictive soil map was made by Co-kriging. Co-kriging is described in terms of increments. To compare the results of Co-kriging with those obtained with Kriging, use is made of a procedure which defines the necessary number of observations to create a predictive map of a predescribed precision. Application of Co-kriging may either lead to maps which are more precise than maps obtained by Kriging, or to a considerable reduction in costs.

1 Introduction

Spatial variability has traditionally been studied by making soil surveys containing delineated areas that are distinguished on the basis of physiographic features. Each delineated area is represented by one so-called 'representative' profile, the properties of which are being used to make interpretations that are assumed to apply to the delineated area as a whole. This, obviously, can involve gross generalizations (e.g. STEIN et al. 1988a, b). Lack of clearly visible physiographic features at the soil surface hampers construction of a soil map which can, in absence of such features, only be based on a series of systematic soil borings. Questions arise as to how many borings are necessary and which boring pattern should be followed. Recent developments in geostatistics allow a scientific evaluation of this question in contrast to subjective ad-hoc procedures

being used in the past (e.g. WEBSTER 1985). Physiographic features are, however, not the only attributes that can be used to define soil variability patterns. Increasing availability of multispectral aerial photographs, satellite images as well as rapid development of remote sensing provide a solid basis for the application of pattern recognition techniques. However, remote sensing images only reflect conditions either at the soil surface, or those of a crop cover. Moisture stress of a crop can, for example, be registered quite well by Infrared Line Scanning (e.g. MULDERS 1987), but the reasons for the stress situation can usually not be derived and have to be based on on-site investigations. Shortage of water, causing crop moisture stress at a given time may originate from a precipitation deficit or from a variety of possible other causes, such as shallow rooting, occurrence of sandy layers within the soil profile, a deep watertable and management conditions. Remote sensing images are therefore particularly useful for **testing** and **validating** interpretations which are increasingly based on computer simulations, since they offer the possibility to evaluate large areas. Multispectral photography will be applied for validation purposes instead of satellite imagery, which only reveals variation on too small a scale for our purpose.

Returning to the central problem of this study which is the characterization of soil variability within an area without clear physiographic features, it is necessary to specify the object to be studied. Soil survey data and soil physical properties are analyzed with respect to their spatial variability as they are related to crop simulation modelling. The study is carried out on a level river terrace, an area homogeneous on one scale (e.g.

1:100.000), but showing heterogeneity on a larger scale.

A choice can be made to either study variability of various individual soil characteristics, or to focus on interpretation with relevance for practical application. In this study, a choice was made for the latter, and attention is focused on the most important land quality in this area, which is the moisture supply capacity. This land quality is directly related to crop yield (e.g. FAO 1976). Next the level of detail has to be defined. BOUMA (1989) has distinguished six levels of detail, ranging from assembling farmer's experience to detailed computer simulations. The appropriate level is determined by balancing the level of detail required for the type of questions being asked against available data. The common mistake of running a detailed model while very little data are available is to be avoided at all cost. Still, we are faced here with a basic problem. Modern approaches require a quantitative analysis for which computer simulation techniques are increasingly being used. Measurement of all required basic data would, however, be prohibitively expensive. Emphasis is therefore put on the derivation of so-called pedo-transfer-functions which relate available soil data to data that are needed to feed the simulation models. This aspect will be emphasized in this study because application of pedo-transfer-functions allows a quantitative approach using the relatively simple WOFOST simulation model.

In summary, the purpose of this study was to use geostatistical techniques to express and interpret soil spatial variability in a level river terrace, to use pedo-transfer-functions and simulation models to obtain data on spatial patterns of soil moisture availability and associated crop

productions and to investigate the use of spatial variability in the creation of a predictive map of these properties.

2 Materials and methods

2.1 Soils

The study area is located on one of the most recent terraces of the river Allier in the Limagne area in France, near the village of Vinzelles, some 30 kilometers down-stream of the city of Clermont-Ferrand. Soil development is relatively limited, such that alluvial sedimentation is dominant (STEIN et al. 1989). Clay, sand and gravel layers are irregularly distributed in the soil profiles. The area of 200 ha is naturally bounded at one side by the Allier, and at the other side by a somewhat higher situated and older terrace. Land use is predominantly agricultural, the terrace is flooded once every 10–15 years. Moisture availability is one of the most important land qualities for agriculture as depth to groundwater is greater than 1.2 m. On a general 1:100.000 soil map, which is the only one available for this region (INRA 1965) the soil in this terrace is classified as a Eutric Cambisol, but classification as a Fluvisol could have been chosen locally due to the presence of fluvic properties. No soil delineations are presented, which implies that this region is considered to be fairly homogeneous on scale 1:100.000. When more detailed investigations are carried out on a larger scale (e.g. 1:20.000) clay, sand and gravel layers exhibit irregular patterns that are difficult to map due to lack of clearly visible physiographic features. A regular sampling scheme was therefore used instead.

2.2 Sampling methods

Four transects were planned with observation points 100 m apart. Two transects were located more or less parallel to the Allier river, and two transects were chosen perpendicular to the river, yielding a total of 62 measurements. In every observation point a number of basic soil survey data were measured by means of a boring up to 1.2 m, or to an impenetrable gravel layer, whichever appeared first. This included clay percentage of different horizons, visual estimation of the stoniness of the surface, depth to gravel layers, etc. Every location was thoroughly wetted and covered for 24 hours with a plastic sheet to establish moisture content and pressure head at field capacity. Three replicate 100 cc cylindrical samples were taken in the topsoil at each location to measure moisture content at field capacity and bulk density.

In twelve selected points duplicate soil samples were taken to measure moisture retention curves with standard procedures. At the 50 remaining locations moisture content at pressure heads of 0.01, 0.1, 3 and 15 bar, respectively, were predicted by means of a pedo-transfer-function, based on field measured water content at field capacity. 'Available water' was defined as the volume of water between field capacity and 15 bar.

To investigate anisotropy and variability at short distances, a logarithmic rather than an equidistant additional sampling scheme was used (fig.1) with observations at distances 1 m, 2 m, 4 m, 8 m, 16 m, 32 m in two directions from every fourth observation point of the main transect perpendicular to the river, and from any of those points one additional observation was taken at a distance 1 m, 2 m, 4 m, 8 m, 16 m
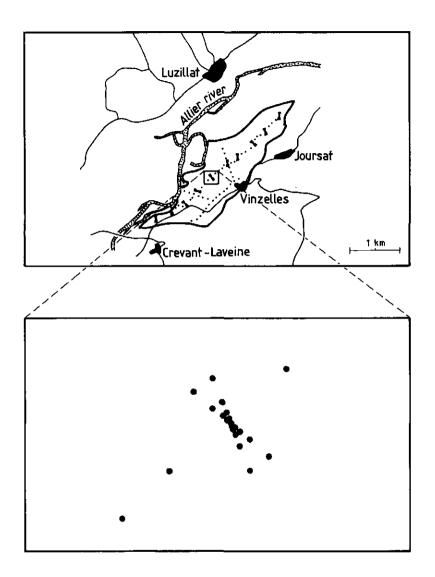
163

Fig. 1: *The sampling scheme on the river terrace in the Limagne area in France (above). One of the ten logarithmic schemes is presented in detail (below).*

and 32 m, respectively, parallel to the river. This comprised a set of 10 pairs of 45°-angled cross-transects. According to this scheme some 280 observations were taken on the basic soil survey data only. The number of measured soil physical properties, therefore, was limited while the number of observations on basic soil properties was abundant.

Multispectral photographs were made during the months July and August, 1988, with a Near-Infrared (NIR, spectral passband 840–900 nm) filter and with a Green (GR, spectral passband 555-580 nm) filter. On these photographs strong variation was visible in both winter wheat and maize plots. The greytones were classified on the negatives in a scale from 1 (light) to 21 (dark) according to a standard sensitivity guide, a 21 step grey scale. In the terrain repeated measurements were performed on 16 maize test plots of 50×50 m size with a hand held radiometer which measures the irradiance and object reflected radiance in Green, Red and Near Infrared spectral passbands (CLEVERS 1986, MULDERS 1987). The reflectance, which is defined as the ratio between irradiance and the object reflected radiance shows a high correlation with photographic density as it is measured on the same data at which aerial photographs were taken in August. Thus the greytones of the multispectral aerial photographs each with their specific distribution, as influenced by filmtype, exposure time and photographic processing can be calibrated by expressing them in the percentage of reflection, which is directly related to terrain properties.

The degree of vegetation coverage has a strong influence on the spectral properties. For instance, a high vegetation coverage produces a high NIR reflectance.

The difference between NIR and visible reflectance is typically high for vegetation in contrast to soil which shows a relatively low difference of all possible ratios. The (NIR-GR)/(NIR+GR) ratio which shows a good linear relation with vegetation coverage, is therefore a good predictor for plant growth, and is further used as a vegetation index. For 37 of the 62 locations of the 100 m transects containing maize, the greytones were estimated on the available negatives.

Based upon soil data, particular crop characteristics and climatic conditions the crop simulation model WOFOST (VAN DIEPEN et al. 1988) was used. Concerning the crop, attention was focused on water limited wheat production, as wheat is commonly cultivated on the Fz terrace being studied. Climatic data were obtained from the weather station in the nearby village of Puy-Guillaume. Data from the aerial photographs were used to check the effects of soil physical properties (governing infiltration, water movement and water uptake by crop) on crop growth, as it is evident by reflection properties that healthy mature crop will show a higher difference between NIR reflectance (e.g. 50%) and GR reflectance (e.g. 15%), than soil with low vegetation coverage and bare soil.

Prediction of land qualities for unvisited locations throughout the terrace, was a major objective of this study. As soil physical properties require relatively expensive and time consuming measurement procedures, the statistical prediction technique Co-kriging (JOURNEL & HUIBREGTS 1978, STEIN & CORSTEN, subm) is attractive to be used. Promising covariables were selected among basic soil characteristics.

Most projects are confronted with lack of data and every project with the ques-

tion as to how available soil survey data can be used most efficiently. The case study to be reported here has therefore a relevance that goes considerably beyond the Limagne area.

2.3 Spatial variability and Co-kriging

In this study interest is focused on the use of more than one variable, say $\{z_u(x), x \epsilon R^2\}$ for $u=1,2,...,p$ describing the spatial behaviour of characteristics of interest in a specific two-dimensional region. We follow the convention that bold printing indicates that a variable is considered stochastic. Examples in this study are the water limited yield of winter wheat, the moisture availability of the topsoil, the depth to gravel, the clay percentage of the B-horizon and the stoniness of the surface. To describe the spatial stucture, use can be made of the covariance function (JOURNEL & HUIJBREGTS 1978) defined as a function of the separation distance h between measurement locations:

$$c_u(h) = E[z_u(x + h)z_u(x)] - \mu_u^2 \qquad (1)$$

where $\mu_u = E[z_u(x)]$, the expectation of $z_u(x)$. The covariance function can be defined if the variable $z_u(x)$ is stationary, that is if μ_u is constant and if $Var[z_u(x)]$ exists. In anisotropic circumstances $c_u(h)$ may differ from $c_u(-h)$. Anisotropy may be encountered for instance in transect data, or in areas where the variation parallel to a river differs from variation perpendicular to this river. For isotropic spatial charactristics, however, $c_u(h)$ is an even function.

The spatial interaction between two variables can be modelled by means of the cross-covariance function, $c_{uv}(h)$, defined as a function of the distance h be-

tween the measurement locations $x + h$ for z_u and x for z_v:

$$c_{uv}(h) = E[z_u(x + h)z_v(x)] - \mu_u\mu_v \qquad (2)$$

where μ_u and μ_v are the expectations for $z_u(x)$ and $z_v(x)$, respectively. The cross-covariance function describes the change of the covariance, if the distance between a measurement location of the first variable and a measurement location of the second variable changes.

For $h = 0$, it follows that $c_u(0) = Var[z_u(x)]$ and that $c_{uv}(0) = Cov[z_u(x), z_v(x)]$. In anisotropic circumstances $c_{uv}(h)$ may differ from $c_{uv}(-h)$, and hence from $c_{vu}(h)$. For isotropic spatial characteristics, however, $c_{uv}(h)$ is an even function.

If $Var[z_u(x)]$ or $Var[z_v(x)]$ do not exist, that is if the variables are not stationary, then often use can be made of expressions of the form $t_u(x, h) := (z_u(x+h)-z_u(x))$ and $t_v(x, h) := (z_v(x+h)-z_v(x))$, the simplest form of increments of order 1, which are also 1st order differences (DELFINER 1976). Attention is focussed on first order differences, being the special case of first order increments, where those increments are comparisons between the variable on location x and the variable at location $x + h$. If a first order difference is stationary, i.e. $E[t_u(x, h)] = 0$, and $Var[t_u(x, h)]$ is independent of the location x, semi- and cross-variograms are used as generalizations of the covariance function and the cross-covariance functions, respectively. They are defined as:

$$\gamma_u(h) = (1/2) \cdot E[t_u(x, h)^2] \qquad (3)$$

$$\gamma_{uv}(h) = (12) \cdot [t_u(x, h)t_v(x, h)] \qquad (4)$$

Therefore, $g_u(h)$ is equal to half the variance of first order differences for distance h (CHRISTAKOS 1984). Similarly,

166

$\gamma_{uv}(h)$ is equal to half the covariance between first order differences both for distance h.

An advantage of the use of the variograms over covariance functions is the fact that the expectations μ_u and μ_v do not occur in the definition of the semi- and cross-variograms, so that they need not to be estimated or guessed. Estimating the expectations μ_u and μ_v is a source of serious errors, unless the covariance functions are known beforehand. In most studies, use is made of variograms, as they can easily be estimated from collected data. Through the mean values in distance classes estimation is carried out by means of least squares estimation of the parameters of, for instance, an exponential model (CRESSIE 1985),

$$\hat{\gamma}(h) = C_O + A \cdot (1 - exp(-h/R)) \qquad (5)$$

Weights are given by the reciprocal of the variance within a distance class or by the number of pairs of observation points. The sill is given by $C_O + A$, the range by R and the nugget effect by C_O.

The spatial prediction technique Kriging (DELFINER 1976) may be described by means of increments. Let the variable $z_u(x)$ be observed in n_u locations $\{x_{ui}\}$. The value of $z_u(x)$ in the location x_O is predicted by means of a predictor $\mathbf{p}$ linear in the observations of $z_u(x)$, given by

$$\mathbf{p} = \sum_{i=1}^{n_u} \lambda_i z_u(x_{ui}) \qquad (6)$$

where the λ_i's are coefficients, or weights, to be assigned to the observations. The predictor $\mathbf{p}$ is unbiased, that is $E[\mathbf{p} - z_u(x_O)] = 0$. In the stationary case this holds if and only if $\Sigma \lambda_i = 1$. Therefore, the prediction error $\mathbf{p} - z_u(x_O)$ is a stationary increment. The predictor $\mathbf{p}$ is best among all unbiased predictors which are linear in the observations of $z_u(x)$, as it has minimal variance of the prediction error, that is: $Var[\mathbf{p} - z_u(x_O)]$ is minimal. This variance of the prediction error can be formulated as the variance of an increment of first order, and therefore as a linear combination of expressions of the form (3).

Co-kriging (VAUCLIN et al. 1983, YATES & WARRICK 1987) can also be described in terms of increments. Let in addition to $z_u(x)$ the variable $z_v(x)$ be observed, in, say, n_v locations $\{x_{vj}\}$. Now, the value of $z_u(x)$ in the location x_O is predicted by means of a predictor $\mathbf{p}$ linear in the observations of both variables, which is given by

$$\mathbf{p} = \mathbf{p}_u + \mathbf{p}_v = \sum_{i=1}^{n_u} \lambda_i z_u(x_{ui}) + \sum_{j=1}^{n_v} \eta_j z_v(x_{vj}) \qquad (7)$$

where the λ_i's and the η_j's are coefficients to be assigned to the observations. Again, the predictor $\mathbf{p}$ is unbiased, which implies that $E[\mathbf{p}_u - z_u(x_O)] = 0$ and $E[\mathbf{p}_v] = 0$. In the stationary case this holds if and only if $\Sigma \lambda_i = 1$ and $\Sigma \eta_j = 0$. Therefore the prediction error $\mathbf{p} - z_u(x_O)$, being a stationary increment, is composed of two stationary increments, $\mathbf{p}_u - z_u(x_O)$ and $\mathbf{p}_v$, respectively. The predictor $\mathbf{p}$ is best among all unbiased predictors, which are linear in the observations of the two variables, $z_u(x)$ and $z_v(x)$, as it has minimal variance of the prediction error. This variance of the prediction error can be decomposed as:

$$Var[\mathbf{p} - z_u(x_O)] = Var[\mathbf{p}_u - z_u(x_O)]$$
$$+ Var[\mathbf{p}_v] + 2 \cdot Cov[\mathbf{p}_u - z_u(x_O), \mathbf{p}_v] \qquad (8)$$

As (8) is the sum of two variances of stationary increments and the covariance

167

between them, it can be formulated in terms of (3) and (4).

To investigate anisotropy, a fairly general model for the spatial structure must be formulated, for example one in which (3) and (4) are described as a function of the length $| h |$ of the distance vector and an anisotropy factor. For example, an exponential model now looks like:

$$\hat{\gamma}(h, \theta) = $$
$$C_O + A \cdot (1 - \exp(- | h | \cdot R(\theta))) \qquad (9)$$

where $R(\theta) = \{B_1^2 \cos^2(\theta - \varphi) + B_2^2 \sin^2(\theta - \varphi)\}^{1/2}$, where the parameters C_O, A, B_1, B_2 and φ are to be estimated (compare BURGESS et al. 1981).

An interesting point concerns the question whether a covariable should be included, and if so: which one, and how inclusion of a covariable affects prediction error variances as a function of the distance from the observation points. For this purpose different test sets may be created, each with observation points at a specific distance from those observation points with which a prediction is carried out.

The gain in precision achieved by Co-kriging as compared with Kriging, can be quantified by means of the Mean of the Variances of the Prediction Errors (MVPE). For all observation points of a test set T containing n points the variances of the prediction error are calculated. The MVPE is defined as the mean value for this test set:

$$MVPE = (1/n) \cdot \sum_{x_o \epsilon T} var[\mathbf{p} - \mathbf{z}_{x_o}] \qquad (10)$$

Another question concerns the number of observations necessary to achieve a certain predescribed precision of a predictive map. This question is faced in the context that the maximum variance of the prediction error for interior points should be as low as possible. If observations were independent this number could easily be determined (COCHRAN 1977). For spatially dependent observations, this question ca: only be answered if semi- and c ss-variograms are available and if also normality it assumed. If a specific precision is prescribed for the predictions, that is that the predictions lay within a, say 95%, confidence region it is possible to arrive at the distance between observation points to achieve this precision, and hence the observation density and the number of observations.

In most practical studies short distance variability of soil properties is of crucial importance for spatial interpolation. Often one uses only a limited number of observations located at relatively short distances from the prediction locations (a so-called neighbourhood) because those points have preponderant coefficients if a larger neighbourhood would be used for prediction. In setting up the sampling grid, attention was focused on this short-versus long-range variability. Semi-variograms were estimated on the basis of observation points with separation distance of 100 m. Exponential models were fitted by means of non-linear regression methods, from which the spatial variability for distances shorter than 100 cm could be estimated by interpolation. The results were compared with the available spatial variability for the same distances as obtained from the logarithmically distributed short-distance observation points.

Symbol	Description	Unit	Mean	Median	S.d.	Coeff. Var.	Min	Max	Nr.
Soil survey data									
THICKA	thickness a-horizon	cm	39.1	40	11.9	.30	10	70	304
DGRAVEL	depth to gravel	cm	63.4	60	28.0	.44	10	120	192
CB	clay% b-horizon	%	22.4	20	4.9	.22	15	35	181
STONI	stoniness	%	4.4	1	7.1	1.61	0	40	291
Soil physical properties									
BD	bulk density	g/cm^3	1.33	1.37	.15	.11	.9	1.5	62
FC	field capacity	cm^3/cm^3	.203	.206	.033	.16	.114	.263	58
AVWPR	avw* profile	mm	125	125	60	.48	28.6	226.0	58
mc_{sat}	mc* at saturation	cm^3/cm^3	.50	.48	.06	.12	.42	.65	62
mc_{fc}	mc* at field capacity	cm^3/cm^3	.30	.28	.06	.20	.17	.43	62
Yield simulation data									
YIELD	water limited yield winter wheat	ton/ha	6.15	6.46	1.15	.19	3.45	7.85	58

* avw = available water; mc = moisture content

Tab. 1: *Descriptive statistics.*

	Semi-Variograms			Cross-variograms with AVWPR		
	C_O	A	R	C_O	A	R
DGRAVEL	306 cm^2	522 cm^2	0.27 km	342 mm*cm	1100 mm*cm	0.25 km
STONI	19%^{2}	14%^{2}	0.15 km	0 mm%	-150 mm%	0.40 km
AVWPR	1310 mm^2	2580 mm^2	0.28 km			

Tab. 2: *Coefficients of fitted exponential model,* $\gamma(h) = C_O + A * (1 - exp(-h/R))$ *for semi-variograms and cross-variograms estimated from the 100 m transects.*

Symbol	Description	Mean	Median	S.d.	Min	Max	Nr.
mc_{sat}	mc at saturation	.51	.50	.06	.42	.63	12
$mc_{.01}$	mc at .01 bar	.42	.42	.05	.36	.51	12
mc_{fc}	mc at field cap.	.34	.34	.05	.27	.45	12
mc_3	mc at 3 bar	.20	.18	.05	.12	.29	12
mc_{15}	mc at 15 bar	.13	.11	.05	.07	.21	12

Tab. 3: *Moisture retention curves.*

3 Results and discussion

3.1 Variability measurements

An overview of descriptive statistics of some selected variables is presented in tab.1. The symbols in the leftmost column are used throughout this section. Gravel layers within 1.2 m depth are encountered in 192 of the 304 observations points, sometimes close to the surface. If a B-horizon is encountered (in 181 locations), a relatively high clay% was measured, which is thought to result from weathering, because micromorphological observations did not reveal clay coatings. Stoniness of the surface ranges from zero to 40%, with a median value of 1%, suggesting that there is a relative increase of gravel at the surface, which may be due to ploughing activities and erosion of fine particles by rainwash and occasional flooding. For AVWPR, DGRAVEL and STONI (see list of symbols) the parameters of an exponential model of the semi-variograms are estimated (tab.2). The estimated variance of AVWPR and DGRAVEL is slightly higher than the sill value ($C_O + A$) as it theoretically has to be. The larger difference between them for STONI must be attributed to its skew distribution, as is also illustrated by the difference between mean and median value.

Spatial variation was estimated in two specific directions: parallel to the Allier and perpendicular to it (fig.3). Semi-variograms for STONI and DGRAVEL showed apparently slight differences. The presence of anisotropy was tested by comparing predictions in 100 randomly located points in the area using equation (9) with those obtained with equation (5). A paired t-test showed no significant differences between isotropic and anisotropic predictions nor between prediction error variances, assuming independence. To estimate the nugget effect, the semi-variograms for the 100 m transects were compared with the semi-variograms for the more refined log-schemes. The nugget-effect as obtained from the interpolated 100 m transect data equals 306 for DGRAVEL and 19 for STONI, respectively, whereas the values for the log-schemes are equal to 470 and 12, respectively. It is noticed that differences up to 50% are quite natural, which emphasizes the fact that extrapolation remains a dangerous activity, although in many circumstances the only possibility.

3.2 Physical characteristics

Remarkably homogeneous variables 'Field capacity' and 'Bulk density' were encountered, with a coefficient of variation of only .11 and .16, respectively, due to the homogeneous topsoil in the measurement locations. The variation of 'available water' for the soil profile, however, is larger, because it reflects depth of rooting.

To predict the moisture content at saturation and at pressure head values of 0.01, 3 and 15 bar, linear models were fitted with measured mc_{fc} as explanatory variable. To do this, 12 samples on selected spots were analyzed in terms of their mc-values at five pressure heads (tab.3). This yielded the equations with the individual standard deviations of the coefficients:

$$mc_{sat} = .23 \pm .10 + .81 \pm 29 mc_{fc} \qquad R^2 = .43 \quad (11)$$

$$mc_{.01} = .12 \pm .05 + .90 \pm 15 mc_{fc} \qquad R^2 = .78 \quad (12)$$

$$mc_3 = -.12 \pm .06 + .92 \pm 17 mc_{fc} \qquad R^2 = .74 \quad (13)$$

$$mc_{15} = -.15 \pm .06 + .81 \pm 16 mc_{fc} \qquad R^2 = .72 \quad (14)$$
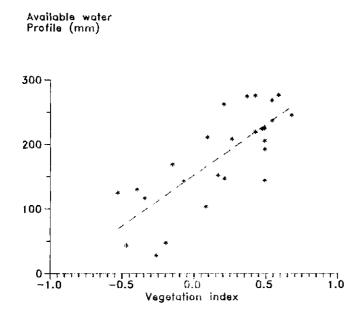
170

Available water
Profile (mm)

Fig. 2: *The relation between the vegetation index (NIR-GR)/(NIR+GR) as observed for maize and the available water in the profile.*

By means of the relations (11) to (14) the $mc_{sat}, mc_{.01}, mc_3$ and mc_{15} values could be predicted for the 62 locations where a mc_{fc} measurement was available. The standard error of the prediction equals 0.050 for mc_{sat} and 0.028 for $mc_{.01}, mc_3$ and mc_{15}. To calculate the available water, only the mc_{15} prediction was actually needed.

The example demonstrates use of regression analysis to predict moisture contents at different pressure heads. Use of this procedure implied an estimated reduction in analysis costs of 80% as compared with measurement of all values at all sites.

3.3 Available water and remote sensing

As YIELD (tab.1) correlates highly with DGRAVEL (corr. coeff. .94) and AVWPR (corr. coeff. .88), it is likely that

Greytone	Relation	R^2
GR	Y = 1.29X - 0.51	.88
NIR	Y = 1.60X - 0.22	.81

Tab. 4: *The relations between the density (X) and the log reflection percentage (Y) for maize during the August flight.*

differences in greytones for one crop type on aerial photographs of one acquisition date were caused by stress situations, due to shortage of available water in the profile. To further evaluate these stress situations, the relation between greytones and AVWPR was investigated. As validation was based on observation in maize plots, attention was focused in the available set of observation points on the observations containing maize, observed during the August flight. The relations
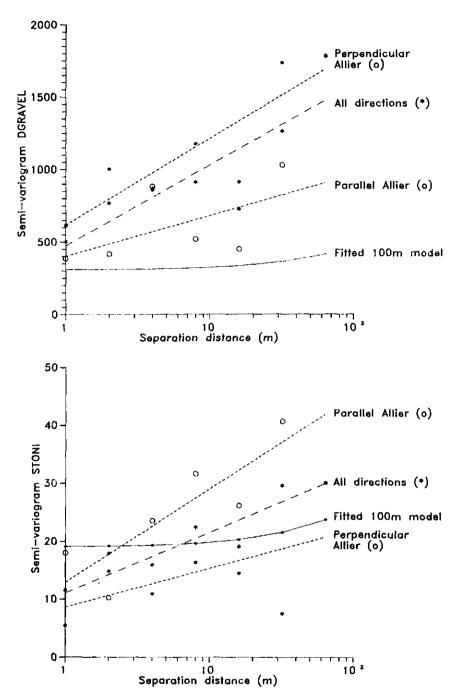
171

Fig. 3: *Semi-variograms for DGRAVEL and STONI for both the logarithmic sampling scheme and the 100 m distant observation points. Because of the nugget effect, the exponential models are slightly convex.*

172

	THICKA	DGRAVEL	CB	STONI	BD	YIELD
corr. coeff.	.41	.96	-.64	-.51	-.06	.88
number of paired observations	31	31	13	13	31	31

Tab. 5: *Correlation coefficients and number of paired observations of AVWPR with easy measureable soil characteristics and with YIELD.*

between the density (X), obtained from the grey tones of GR and NIR and the log reflection percentage (Y) were estimated (tab.4), showing high coefficients of determination (.81 and .88 for NIR and GR, respectively). No clear relationship with AVWPR was discovered for the individual reflectance values. The vegetation index (NIR-GR)/(NIR+GR), however, yielded a significant regression equation (at $\alpha = .05$ level) with AVWPR as explanatory variable and the vegetation index as dependent variable (fig.2). Thus evidence of aerial photographic origin supports the findings on the AVWPR values used in this study regarding their effect on YIELD as derived from simulation modelling.

3.4 Use of Co-kriging

One of the main objectives of this study was the transition from point observations towards statements concering the area as a whole. To do this, predictions of AVWPR values can be carried out in a number of comparative steps. In a first approach the so-called 'representative profile' (see Appendix) was used to summarize the value of AVWPR throughout the region, giving a value of 164 mm. No precision can be assigned to this value.

As a next step the average value for AVWPR throughout the terrace was determined by means of the mean and the median of the 62 observations, giving a value of 125 mm for both mean and median. The mean value has a standard deviation of 60 mm. The confidence interval under the (dubious) assumption that AVWPR is normally distributed and the observations are independent replicates, contains the value from the 'representative profile'. The value of the 'representative profile' as well as the mean of a large number of observations, however, neglect the fact that variability varies spatially due to sedimentological properties of the soil.

Estimated semi-variograms were used to apply Kriging. In order to apply Co-kriging, a co-variable had to be found. STONI and DGRAVEL were chosen as the most promising co-variables, a selection based on the relatively high values of the correlation coefficients (tab.5). The variable CB was left out, despite its high correlation with AVWPR, because the number of paired observations (12) is too restrictive to estimate properly the cross-variogram. The negative value of the correlation coefficient is due to the ability of the soil to more strongly bind the water as the clay percentage increases and water availability decreases.

The spatial correlation thus observed was used to carry out mapping of AVWPR throughout the terrace, yielding a predictive AVWPR map, which was actually created with the help of the Geographic Information System ArcInfo
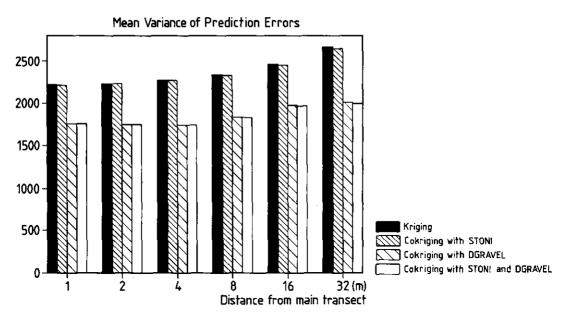
Fig. 4: *The MVPE as a function of the distance between observation points, estimated from the log-scheme, for Kriging (Kr), Co-kriging with STONI (ST), Co-kriging with DGRAVEL (DG) and Co-kriging with DGRAVEL and STONI.*
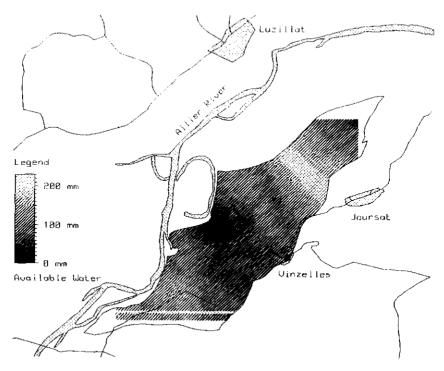

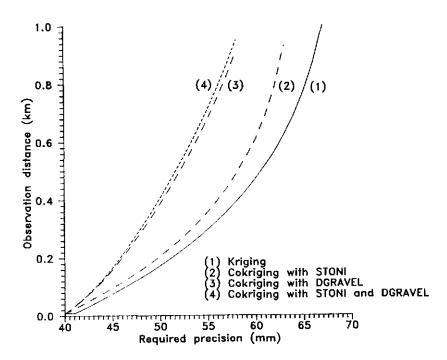Fig. 5: *The Co-kriged AVWPR map.*

174

Fig. 6: *The required precision as a function of the distance between observation points on a square grid for Kriging, Co-kriging with DGRAVEL, Co-kriging with STONI and Co-kriging with both STONI and DGRAVEL.*

(fig.5). With this map optimal predictions for unvisited locations can be obtained, using the spatial structures of AVWPR and of DGRAVEL as well as the spatial correlation between AVWPR and the covariable DGRAVEL.

By means of the MVPE the precision of the predictions in terms of standard deviations of the prediction error are compared for Kriging (Kr), and Co-kriging with STONI (St) and DGRAVEL (DG) as a covariable (fig.4). The MVPE was calculated for six different test sets, each consisting of 40 points at distances of 1 m, 2 m, 4 m, 8 m, 16 m and 32 m, respectively, from the centre points on the main transect with observation points at 100 m distance. The

MVPE is both a function of the distance to observation points, and of the spatial correlation between predictand and co-variable. The choice for the covariable DGRAVEL is self-evident, since it is easy to measure and highly correlated.

Another way to compare Kriging and Co-kriging, is to determine the necessary number of observations to be taken on a future grid with equidistant observation points, so as to arrive at a specific precision of a predictive map. Different simulated observation densities on a square grid were compared, ranging from 1 m to 1 km distant observation points (fig.6). For Kriging a neighbourhood of 4 observation points was used (where 16 would have served as well) and for Co-kriging

175

a twice as dense grid was applied where the 9 neighbouring observation points for the covariable were used. Doubling the observation density of the covariable a second time lead to completely similar results. A prediction was carried out in the centre of the grid cells, where an observation of the covariable was supposed to be present, but not of the predictand, as this point contains the highest uncertainty among interior points when a prediction by Kriging is carried out.

To arrive at a map with a required prescribed precision the following procedure must be followed

- If the precision ranges from 40 mm to 60 mm application of Kriging requires an observation density of the predictand which is twice as high as the required observation density when Co-kriging is applied.

- If the desired precision ranges from 60 mm to 70 mm Kriging may be applied.

- When software is at hand to carry out interpolation with two covariables, a further — minor — reduction in the observation density, and hence in the total costs can be achieved.

The procedure outlined above is based on existing semi- and cross-variograms. If they have to be estimated, a minimum of 40–50 observations are necessary to acquire a required minimum of 30 pairs of observation points per distance class, a number which may far exceed the necessary number of observations needed to create the predictive map. In many existing databases in various countries, however, they can easily be calculated for a large number of characteristics for every soil unit that is considered to be homogeneous on the required scale. On the basis of classification and soil genetic considerations, they might be applied also to areas where no measurements were done so far.

Acknowledgement

References

BOUMA, J. (1989): Land qualities in space and time. In: J. Bouma & A.K. Bregt (Eds.), Land qualities in space and time, 3–13. Pudoc, Wageningen.

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.

CHRISTAKOS, G. (1984): On the problem of permissible covariance and variogram models. Water Res. Res. 20(2), 251–265.

COGHRAN, W.G. (1977): Sampling techniques. Wiley, New York, third edition.

CLEVERS, J.G.P.W. (1986): Application of remote sensing to agricultural field trials. PhD thesis, Agricultural University Wageningen.

CRESSIE, N. (1985): Fitting variogram models by weighted least squares. Mathematical Geology, Vol. 17, No. 5, 563–586.

DELFINER, P. (1976): Linear estimation of non-stationary spatial phenomena. In: M. Guarascio et al. (Eds.), Advanced geostatistics in the mining industry. Reidel, Dordrecht.

FAO (1976): A framework for landevaluation. FAO Soils bulletin 32, Rome, Italy.

INRA (1965): Carte Pédologique du Val d'Allier. Feuille sud: Clermont-Ferrand. Service d'Etudes des Sols, Montpellier, France.

JOURNEL, A.G. & HUIJBREGTS, Chr. J. (1978): Mining Geostatistics. Academic Press, New York.

MULDERS, M.A. (1987): Remote sensing in soil science. Developments in Soil Science 15, Elsevier, Amsterdam.

STEIN, A. & CORSTEN, L.C.A.: Universal Kriging and Co-kriging as regression procedures. Submitted to biometrics.

STEIN, A., VAN DOOREMOLEN, W., BOUMA, J. & BREGT, A.K. (1988a): (Co-)Kriging point data on moisture deficit. Soil Sci. Soc. Am. J. 52, 1418–1423.

STEIN, A., HOOGERWERF, M. & BOUMA, J. (1988b): Use of soil map delineations to improve (Co-)kriging of moisture deficits. Geoderma 43, 163–177.

STEIN, A., BOUMA, J., KROONENBERG, S.B. & COBBEN, S. (1989): Sequential sampling to measure the infiltration rate within relatively homogeneous land units. CATENA 16, 91–100.

VAN DIEPEN, C.A., RAPPOLDT, C., WOLF, J. & VAN KEULEN, H. (1988): CWIS Crop growth simulation model WOFOST. Documentation, vs. 4.1. Centre for World Food Studies, Wageningen, The Netherlands.

VAUCLIN, M., VIEIRA, S.R., VACHAUD, G. & NIELSEN, D.R. (1983): The use of Co-kriging with limited field soil observations. Soil Sci. Soc. Am. J. 47, 175–184.

WEBSTER, R. (1985): Quantitative spatial analysis of soil in the field. In: B.A. Stewart (ed.), Advances in Soil Sciences 3, 1–70.

YATES, S.R. & WARRICK, A.W. (1987): Estimating soil water content using Co-kriging. Soil Sci. Soc. Am. J. 51, 23–30.

Appendix. The Representative profile.

Profile description of Vinzelles, Fz.

I Information on the site sampled

Profile number: VIN.
Classification FAO: Eutric cambisol.
Date of examination: 13/8/87.
Authors description: A.G. Jongmans, A. Stein.
Elevation: 280 m.
Location: Top. map 2630 Ouest Maringues, coord. N 2104.250 — E 682.100.

Landform:
1. Physiographic position of the site: Fz terrace
2. Topography of the surrounding country: flat
3. Microtopography: slightly undulating with gullies.
Slope: class 1, flat.
Vegetation: Arable land, wheat.

II General information on the soil

Parent material: Alluvial deposits of the Allier, Fz level.
Drainage class: well drained.
Moisture condition of the soil: moist throughout.
Depth of the groundwater table:
1. Actual: >150 cm
2. Presumed highest: >150 cm
3. Presumed lowest: >150 cm
4. Flooding: irregularly flooded (10/15 years).
Presence of surface stones or rock outcrops: class 0.
Evidence of erosion.
Present of salt or alkali: class 0.
Human influence: ploughed till 25 cm.

III Description of soil horizons

0–25 cm, Ap: Sandy clay loam, slightly gravelly; 10YR3/2, moist; very coarse angular and subangular cloths; partly friable, partly slightly firm; very few fine and fine discontinuous random tubular pores; very few subrounded and rounded fresh weathered quartz granite basalt gravel; few brick fragments; common fine roots; locally dead plant fragments (maize) in clusters; abrupt and smooth on:

25–70 cm, Bw: Sandy clay loam, slightly gravelly decreasing in clay content with increasing depth; 10YR4/2, moist; moderate fine and medium subangular blocky tending to sponge structure; friable; many very fine and fine continuous random transped tubular simple pores; common (5–10) medium and coarse continuous vertical and random transped tubular simple pores; very fine subrounded and rounded gravel as in Ap; many wormcasts; locally krotovinas; common fine roots; clear and wavy on:

70–103 cm, C1: Alternation of bands consisting of coarse sand (Md50 500–1000μm), loamy fine sand and gravel; 10YR7/3 in sand bands, moist; disturbed stratification; sand bands: very friable; gravel bands: loose; in loamy sand bands: many very fine discontinuous random transped tubular simple pores; few (1–2) medium and

177

coarse continuous vertical tubular simple pores; in sand and loamy sand: many burrows filled up with material of the overlying horizon, few fine roots concentrated in the filled burrows, abrupt and smooth on:

103–120 cm, C2: Alternation of bands consisting of coarse sand (Md50 500–1000μm), and gravel; undisturbed stratification; loose; the sand and gravel consist mainly of a mixture of granite, quartz and basalt fragments fresh and weathered, locally rounded pieces of material with sandy clay loam texture.

List of Symbols

E	: Expectation;
Var	: Variance;
Cov	: Covariance;
$\sum$	: Summation.
$\{z_u(x), x \epsilon R^2\}$	: regionalized variables, as a function of the location x;
$\{z_v(x), x \epsilon R^2\}$	
$c_u(h), c_v(h)$	: covariance function of $z_u(x)$ and $z_v(x)$ as a function of the distance h;
μ_u, μ_v	: expectation of $z_u(x)$ and $z_v(x)$
$c_{uv}(h)$	: cross-covariance function between $z_u(x)$ and $z_v(x)$
$t_u(x,h), t_v(x,h)$	: first order differences.
$\gamma_u(h), \gamma_v(h)$	: semi-variograms;
$\gamma_{uv}(h)$	: cross-variogram;
$\hat{\gamma}(h)$	: estimated variogram;
C_O, A, R	: sill, range and nugget of estimated variogram;
n_u, n_v, n_{uv}	: number of observation locations;
$\{x_{ui}\}, \{x_{vi}\}$	: set of observation locations;
p, p_u, p_v	: predictors;
λ_i, η_j	: weights;
x_0	: prediction location;
MVPE	: Mean Variance of Prediction Errors;
ϵ	: element of;
THICKA	: thickness a-horizon;
DGRAVEL	: depth to gravel;
CB	: clay% of the b-horizon;
STONI	: stoniness;
BD	: bulk density;
FC	: field capacity;
AVWPR	: available water of the profile;
mc_{sat}	: moisture content at saturation;
mc_{fc}	: moisture content at field capacity;
$mc_{.01}$	: moisture content at .01 bar;
mc_3	: moisture content at 3 bar;
mc_{15}	: moisture content at 15 bar;
YIELD	: water limited yield winter wheat;
NIR, GR	: Near-Infrared and Green;

Addresses of authors:
A. Stein, Soil statistician
J. Bouma, Professor of soil science
M.A. Mulders, Soil scientist
M.H.W. Weterings, Student
Department of Soil Science and Geology
Agricultural University
P.O. Box 37
6700 AA Wageningen, The Netherlands

178

CHAPTER III.3. USE OF SOIL-MAP DELINEATIONS TO IMPROVE (CO-)KRIGING OF POINT DATA ON MOISTURE DEFICITS

by

A. Stein, M. Hoogerwerf and J. Bouma

Use of Soil-Map Delineations to Improve (Co-)Kriging of Point Data on Moisture Deficits*

A. STEIN, M. HOOGERWERF and J. BOUMA

Department of Soil Science and Geology, Agricultural University, P.O. Box 37, 6700 AA Wageningen (The Netherlands)

ABSTRACT

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

Predictions of 30-year average moisture deficits (MD30) were carried out by means of kriging and co-kriging, using simulations for 500 point observations in an area of 404 ha of sandy soils in The Netherlands. From the above point observations 100 points were selected at random to function as an independent test set. Attention was focused on improving the precision of kriged and co-kriged MD30-maps, as characterized by two error measures, the mean variance of the prediction error and the mean squared error of predictions. To do so the survey area was stratified by means of soil-map delineations according to soil type and water-table classes based on the ground-water table. In unstratified maps the standard deviation of the prediction error largely depends on the observation pattern. Stratification resulted in an increase of precision of predictions in strata with low MD30 variability and an apparent decrease in strata with high MD30 variability. Major soil-map delineations, as distinguished by a soil survey, had significantly different internal variability. Use of co-kriging resulted in an average increase of precision of MD30-maps of about 10%. This study illustrates the use of available soil-survey information for stratifying a survey area so as to enhance precision of predictions when using kriging and co-kriging of point data.

INTRODUCTION

Kriging and co-kriging are becoming well established statistical prediction techniques to produce predictive soil maps (McBratney and Webster, 1983; Yates and Warrick, 1987). The ready availability of user-friendly software and the fact that the techniques are well founded and the results can be clearly interpreted have made it easy and attractive to apply them. Obviously, the quality of the results obtained is directly related to the quality and the varia-

---

181

bility of the basic data being used and to the selected patterns of observations. These can be regular grids with various configurations or patterns with random components. Preselected grids usually do not take into account geological or pedological differences which form the basis for different delineations on the soil map.

A basic question may be raised as to whether soil-map delineations are of value in creating predictive soil maps. Much variation is likely to be found when a sampling grid is placed over an area with contrasting soil types using (co-)kriging. Webster (1985) recognizes this problem by stating: "The sampling scheme was deliberately designed to distinguish major geological formations in the first stage, since it would have been foolish to have ignored such obvious features." Implicitly it may be postulated therefore that sampling within major soil delineations would result in less overall variability as compared with a situation in which the sampling grid covers several delineations. So far, no soil-survey research has been reported covering this particular aspect. This is surprising because if indeed predictions would be more precise when applied within major soil-map delineations as a form of stratified sampling, a new use of existing soil maps would be introduced. The purpose of this study was to test the feasibility of using sample stratification in creating predictive soil maps on the basis of soil-map delineations.

MATERIALS AND METHODS

The survey area

In 1985 a detailed soil survey was made in the Mander area in the eastern Netherlands by the Dutch Soil Survey Institute (Stiboka) to study the influence of groundwater extraction on the production of grassland (Wösten et al., 1987). The Mander area consists of sandy soils, which are classified as Haplaquods, Humaquepts and Plaggepts (Wösten et al., 1987). Fluctuations of the groundwater table (Gt) are routinely characterized by Dutch soil surveyors in terms of mean highest (MHW) and mean lowest (MLW) levels (Van der Sluys and De Gruijter, 1985; Fig. 1b, Table I). In the Mander area water-table levels were recorded from 1953 to 1955 and in 1985. Simulated values with a 10-day interval were obtained by means of other data in the intermediate period. The groundwater classes in the area are based on the MHW and the MLW as indicated in Table I. Lowering of water tables results in a change in production, e.g. of grassland, due to lower fluxes of water from the water table to the root zone. Farmers are financially compensated for production losses by municipal water companies. In the 404 ha study area, 500 soil borings were made. Hydraulic conductivity and moisture retention data for major soil horizons were measured and averaged as described by Wösten et al. (1985, 1987). The computer simulation model LAMOS (Bouma et al., 1980a, b) was used to obtain
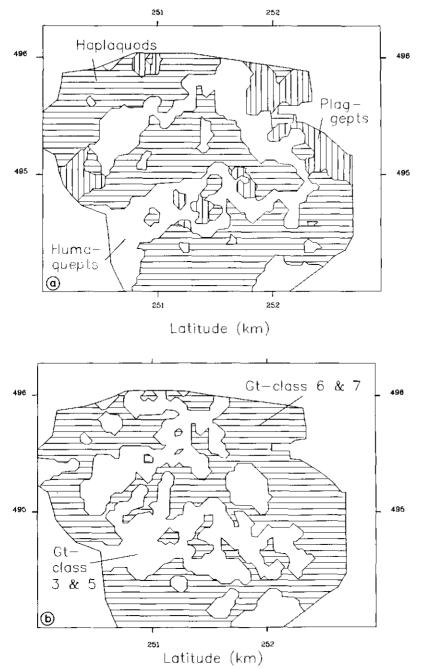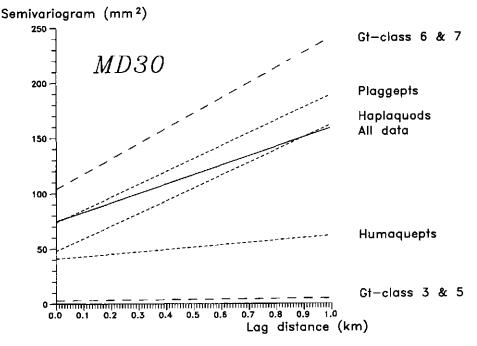
Fig. 1. Stratification of the Mander area according to Humaquepts, Haplaquods and Plaggepts
(a) and according to Gt-classes $3+5$ and $6+7$ (b).

183

TABLE I

Groundwater classes and mean highest and mean lowest groundwater levels, occurring in the Mander area

Class	Groundwater level (cm)	
	mean highest	mean lowest
3	< 40	80–120
5	< 40	> 120
6	40–80	> 120
7	> 80	> 120

the necessary quantitative expressions for the moisture deficits and the associated yields. Simulation calculations were made for each boring location for a 30-year period. For this study attention was focused on the 30-year average value for the moisture deficit (MD30). To evaluate the predictions by kriging and co-kriging, 100 borings were selected at random from the data set to serve as a test set. Predictions were made for the test set using the calculated values for the moisture deficits of the remaining 400 borings (see: Statistical procedures). Attention was focused on the average precision obtained for the 100 test locations. The two Gt-class strata contained 141 and 258 borings and the three soil strata contained 121 borings in Humaquepts, 227 in Haplaquods and
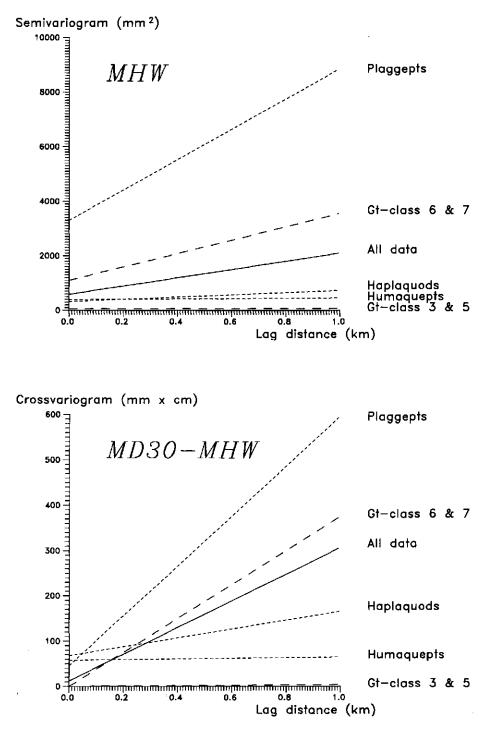

Semivariogram (mm^2)

Fig. 2. Semi-variograms for MD30 (2a) and MHW (2b) and cross-variograms for MD30 with
MHW (2c) according to different strata.

TABLE II

Number of observations of MD30 and the correlation coefficient of MD30 with MHW in different strata

	Number	Corr. coeff.	
All data	399	.587	
Humaquepts	121	.387	
Haplaquods	227	.453	
Plaggepts	51	.604	
Gt-class 3 + 5	141	.158	
Gt-class 6 + 7	258	.405	

51 in Plaggepts, respectively (Table II). The two Gt-class strata contained 47 and 53 test borings and the three soil strata contained 42 test borings in Humaquepts, 47 in Haplaquods and 11 in Plaggepts, respectively.

Statistical procedures

In this study the statistical prediction techniques kriging and co-kriging are used to provide predictors for the values of the so-called predictand (MD30) in the test points. These predictors have certain favourable properties: they are unbiased, they depend linearly on the observations and they have minimal variance of the prediction error (Vauclin et al., 1983). Apart from the prediction itself, the variance of the prediction error is calculated, sometimes referred to as the 'kriging error' (see, for instance, Journel and Huijbregts, 1978), which gives an indication of the performance of the predictor. In this study we assumed that soil variables are isotropic and obey the intrinsic hypothesis within soil strata. Thus use was made of semi- and cross-variograms. In co-kriging, the variable MHW was used as a co-variable, because of its high correlation with MD30 (correlation coefficient .587) and its clear spatial correlation (Fig. 2c). For the prediction of MD30 in any point of the test set by kriging eight neighbouring observations of MD30 were used. In co-kriging the observations of the co-variable MHW in the prediction points concerned was used as well, to give more precise results than when it is left out (Stein et al., 1988). The predictions were afterwards compared with the observed values in the test set.

Throughout this study stochastic variables are denoted boldface to distinguish between a variable and the value which might be assigned to this variable. Two measures were used to obtain insight in the performance of the predictors, the Mean Variance of the Prediction Error (MVPE) and the Mean of Squared Errors of Predictions (MSEP). They are defined as follows:

$$\text{MVPE} = (1/n) \sum_{i=1}^{n} \text{Var}(t_i - y_{0i}) \tag{1}$$

$$\text{MSEP} = (1/n) \sum_{i=1}^{n} (t_i - y_i)^2 \tag{2}$$

where $t_i =$ the (stochastic) predictor provided by kriging or co-kriging in the ith test point; $y_{0i} =$ the variable y when a prediction in the ith test point is being carried out; $t_i =$ the prediction in the ith test point; $y_i =$ the observation in the ith test point; $n =$ the number of observations in the test set.

As we have for a single prediction point the well-known relation

$$E(t - t_0)^2 = \text{Var}(t - y_0) + \{E(t - y_0)\}^2$$

and $E(t - y_0)^2$ is estimated in the mean by:

$$(1/n) \sum_{i=1}^{n} (t_i - y_i)^2$$

averaging over the test set yields the equation:

$$\text{MSEP} = \text{MVPE} + \text{bias}$$

where the bias is due to model deficiency. This gives the connection between the MSEP and the MVPE. Taking square roots yields a comparison between the measures in units of the predictand.

Stratification can be expressed in terms of the partitioning of an area. After stratification the total area A is partitioned into, say, p disjoint non empty sub-areas or strata, $A_1, A_2, \ldots, A_p$, with $A_1 \cup A_2 \cup \ldots \cup A_p = A$, $A_i \cap A_j = \phi$ for $i \neq j$ and $A_i \neq \phi$ for all i. On each stratum A_i the k soil properties are being described by k regionalized variables $y_i^{(j)}(x)$, $j = 1, \ldots, k$, depending on the location vector x, here in two dimensions. Within each stratum every variable is considered to follow the intrinsic hypothesis, that is for all i and for all j:

$$E\{y_i^{(j)}(x) | x \in A_i\} = m_i^{(j)} \text{ (constant)}$$

$$\text{Var}\{y_i^{(j)}(x + h) - y_i^{(j)}(x) | x + h \in A_i \text{ and } x \in A_i\} < \infty$$

The isotropy of the variables implies that the variance depends solely on $|h|$, the length of the vector h, and not on its direction. Predictions in the entire region have to be carried out with respect to the stratified area; that is for every point in the test set the predictor has to be entirely based on observations belonging to the same stratum.

Stratification of the region was based upon groundwater classes (Fig. 1a) and upon soil types (Fig. 1b), which are routinely determined during soil survey. Every observation was assigned to one particular stratum. In stratification according to groundwater classes, classes 3 and 5 were combined as well as

187

classes 6 and 7 to have sufficient observations per stratum to yield reliable estimates for semi- and cross-variograms.

In order to investigate the significance of the encountered differences, a test was developed, which incorporates the spatial dependence structure of the variables within the distinguished strata (see Appendix). As the sample sizes for the individual strata were too large to be handled by personal computers, random selections of 50 observations per stratum were used,

For every stratum the appropriate variograms were estimated and used in the predictions. In general, the MVPE and the MSEP were expected to change after stratification had been carried out. As the variance of the prediction error depends on the pattern of the observations, the MVPE is likely to increase when stratification is being carried out, because the observation points to be used in the predictions in general are located further away from the point where an MD30 prediction is being made. On the other hand, as the determination of the variograms is also based upon individual strata, a general decrease of the MVPE may also be expected as well as a decrease within areas with low variability. Due to the fact that observations from outside the predefined strata no longer were used, the MSEP was expected to decrease after stratification.

For the semi- and the cross-variograms a linear model without sill, $g(h) = a + bh$, was fitted by means of weighted least squares; weights were based upon the number of pairs of points in a distance class (Table III, Fig. 2). Predictions are based on the eight nearest observation points of MD30, and in the case of co-kriging also on the eight nearest observation points of MHW.

Statistical calculations were carried out with the computer package LAND-STAT on personal computers at the Agricultural University Wageningen; results were displayed with the computer package SURFER, which is a release of Golden Software, Inc., Golden, Colorado.

TABLE III

Estimated coefficients of a linear model without sill for semi- and cross-variograms for MD30 and MHW in different strata (see also Fig. 2)

	MD30		MHW		MD30-MHW	
	a	b	a	b	a	b
All data	75	84	587	1520	12	295
Humaquepts	41	21	396	60	58	7
Haplaquods	48	114	323	411	68	98
Plaggepts	74	115	3300	5550	46	549
Gt-class 3 + 5	3	2	53	20	0	4
Gt-class 6 + 7	104	137	1100	2460	0	375

RESULTS

Stratification

The strata differ with respect to moisture deficits as estimations for the mean values and standard deviations around the mean within the strata are different (Table IV). The calculated significance at the 0.001 level of all differences obtained (according to chi-squared testing) should however be interpreted with care, because the sampling scheme is based on well-known soil-landscape relations, which implies that the estimators for means and standard deviations were necessarily biased.

The strata also differ according to the fitted variograms (Fig. 2) as intercept (the nugget effect) and slope differ between different strata. For instance, when comparing the stratum defined by Gt-class 3 and 5 with the stratum defined by Gt-class 6 and 7, the semi-variogram for MD30 has a lower intercept (3 and 104, respectively) and a less strong slope (2 and 137, respectively) (Fig. 2a). Differences among MD30 values within areas occupied by Humaquepts are relatively small. Increasing the sampling density in Humaquept areas has a minor effect on decreasing the standard deviation of the prediction error of MD30 predictions. Increasing the sampling density in Haplaquod and Plaggept areas, however, more strongly decreases the standard deviation of the prediction error of MD30 predictions (Fig. 2a).

The MVPE and the MSEP resulting from the predictions of MD30 for the test set are presented in Table V as averages for the entire test set and for the different strata. The MVPE decreases from 9.30 to 72.8 after stratification according to Gt-class and from 93.0 to 79.8 after stratification according to soil type. The MSEP decreases from 68.1 to 56.4 after stratification according to Gt-class and increases to 75.6 after stratification according to soil type. This clearly indicates that an overall increase of precision is obtained. Table V also

TABLE IV

Estimations for mean and standard deviation of the mean based on spatial structure of MD30 in the different strata

	Mean	S.d. of the mean
All data	10.1	3.2
Humaquepts	9.9	3.4
Haplaquods	4.9	1.3
Plaggepts	25.1	5.5
Gt-class 3 + 5	1.8	0.1
Gt-class 6 + 7	16.3	4.5

TABLE V

Results of the test that no significant differences exist between the strata

	T	Degrees of freedom	Significance
Between Gt-classes	22.4	1	< 0.001
Between soil types	716.0	2	< 0.001
Haplaquods–Humaquepts	24.2	1	< 0.001
Haplaquods–Plaggepts	540.8	1	< 0.001
Humaquepts–Plaggepts	410.1	1	< 0.001

Fig. 3. Co-kriged MD30 map of the Mander area in a three-dimensional presentation. This map is of uniform precision, with decreasing precision only at the borders, as indicated by Fig. 4.

indicates that separation of the two Gt-classes has been quite meaningful, because MVPE and MSEP values were very low for the relatively homogeneous Gt-class 3 and 5 and very high for the relatively heterogeneous Gt-class 6 and 7. This means that the two strata contain two different MD30 populations. Differences are also observed among the three soil strata. The MSEP value for Plaggepts is relatively high, due to the high mean value of MD30 within the Plaggept stratum.

By creating strata from a standard soil map, variability can be allocated to areas where it naturally belongs. When strata are ignored, average MVPE and MSEP values are obtained that mask the fact that variability is characteristically lower in certain strata and higher in others.

Fig. 4. Map of the standard deviation of the prediction error of the co-kriged MD30 map in a three-dimensional presentation.

Fig. 5. Co-kriged MD30 map of the Mander area based on stratification according to soil type in a three-dimensional presentation. This map is not of uniform precision, as indicated by Fig. 6.

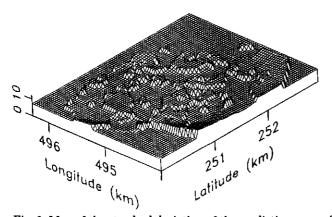Fig. 6. Map of the standard deviation of the prediction error of the co-kriged MD30 map after stratification according to Soil type. Sharp boundaries exist between the different strata. The pattern is comparable with the pattern of soil types (Fig. 1b).

191

TABLE VI

The MVPE and MSEP values according to kriging and co-kriging for the region stratified according to soil type (soil-strat.) and groundwater classes (Gt-strat.); the values for the MVPE and the MSEP for the unstratified region as well as within different strata are included

	MVPE		MSEP	
	kriging	co-krig.	kriging	co-krig.
Unstrat.	93.0	90.7	68.1	61.2
Soil-strat.	79.8	68.3	75.6	70.0
Humaquept	101.3	97.3	48.6	48.2
Haplaquods	67.3	48.5	53.6	57.3
Plaggepts	51.5	41.9	273.0	207.6
Gt-strat.	72.8	72.0	56.4	50.2
Gt-class 3 + 5	3.7	3.7	4.1	4.2
Gt-class 6 + 7	134.2	132.6	102.8	91.0

Co-kriging

Co-kriging generally resulted in more precise predictions than did kriging (Table V). The increases in precision ranged from 0 to 25%. Some MSEP values were larger for co-kriging than for kriging. This may be due to numerical instability (rounding), caused by the fairly small size of the test set for the individual strata. The combination of stratification and co-kriging was nearly always succesful, in that the MSEP decreased from 61.2 to 50.2 after stratification according to Gt-class but increased to 70.0 after stratification according to soil type, probably due to the fact that observation points used in the predictions are located further away. The MVPE value decreased from 90.7 to 72.0 after stratification according to Gt-class and to 68.3 after stratification according to soil type.

When one creates a predictive MD30 map, the average standard deviation of the prediction error will be 9.5 without stratification and will range from 7.2 to 10.1 in the case of stratification according to soil type. The predictive MD30 soil map obtained without stratification is of more or less uniform precision (Figs. 3 and 4). As is well known, precision only decreases in the parts of the area which are less densely sampled, such as near the border of the area. A predictive soil map obtained with stratification, however, does not show uniform precision (Figs. 5 and 6). Strata showing high spatial variability are the less precise parts of predictive soil maps as compared with the strata which show low spatial variability.

DISCUSSION

There are interesting consequences for soil survey by analyzing the above results. When a specific degree of accuracy of predictions is required it is feasible to decrease the number of observations in strata showing relatively low variability and to increase the number of observations in strata showing relatively high variability. The variograms (Fig. 2) allow a quantitative analysis which relates the accuracy of predictions obtained to distance between observation points. In this study, an existing soil map was used for stratification purposes. Thus, available soil survey expertise was used in an innovative manner. In many countries systematic soil surveys are completed. One possible future activity would be to determine the internal variability of existing major land units so as to allow statistically founded quantitative predictions of relevant land qualities rather than qualitative estimates based on the properties of "representative" profiles. This study has demonstrated that existing mapping criteria can be relevant to define effective strata for (geo-)statistical analysis.

When no soil maps are available, it would be advisable to distinguish major land units by means of common soil survey techniques, including remote sensing, and to focus further variability studies on these units rather than on the landscape as a whole with random observations. Continued research is needed to determine which characteristics are most promising for use in stratification and by which criteria they are to be selected. One of the criteria is that the number of data points available to estimate the variogram must be large enough to give reliable estimates. Stratification therefore has to be carried out on the basis of units containing at least 30 to 40 data points. As indicated in Table VI, stratification according to the two Gt-classes was slightly more succesful than stratification according to the three soil types. Different results are likely to be obtained if different land units and land qualities are considered.

Application of co-kriging gives more precise results than ordinary kriging. In this study use could be made of a co-variable (MHW) which was rather highly correlated with the predictand (MD30). Search for such variables is of crucial importance in succesful application of co-kriging.

Both stratification and co-kriging are valuable in geographic information systems, becoming common tools in soil science. Maximizing the accuracy of predictions and, at minimum cost, dealing with variability aspects as a function of different land units is an important task for soil scientists in the future.

University and of A.K. Bregt and J.J. de Gruijter from the Netherlands Soil Survey Institute.

APPENDIX

A test was developed to investigate the significance of the differences in mean values between strata when the observations of a regionalized variable are (spatially) related. Suppose p strata are investigated, from every stratum it is known that the spatial dependency structure is given by the correlation function $c_i(r)$ for $r \geq 0$, $i = 1, \ldots, p$. As an estimator for the mean and the variance within the ith stratum we have:

$$\hat{\mu}_i = \frac{1_n' C_i^{-1} y}{1_n' C_i^{-1} 1_n} = \frac{1_n' G_i^{-1} y}{1_n' G_i^{-1} 1_n} = \frac{1_n' \Gamma_i^{-1} y}{1_n' \Gamma_i^{-1} 1_n}$$

where the matrix C_i contains values of the covariance function within the ith stratum, the matrix G_i contains values of the semi-variogram in the ith stratum and the matrix Γ_i contains values of the generalized covariance function between observations in the ith stratum. Of course, C_i, G_i and Γ_i depend also on the variable under study. The variance of the mean is equal to:

$$\text{Var}\,(\hat{\mu}_i) = (1_n' C_i^{-1} 1_n)^{-1} = -(1_n' G_i^{-1} 1_n)^{-1} = -(1_n' \Gamma_i^{-1} 1_n)^{-1} =: \frac{1}{g_i}$$

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

H_0: $\mu_1 = \mu_2 = \ldots = \mu_p$
H_1: at least one μ_j differs from the other μ_i's, $i \neq j$

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

$$T = \sum_{i=1}^{p} g_i \hat{\mu}_i^{\,2} - \left(\sum_{i=1}^{p} g_i \hat{\mu}_i \right)^2 \Big/ \left(\sum_{i=1}^{p} g_i \right)$$

which has under H_0 a chi-squared distribution with $p-1$ degrees of freedom. Of course, in practical studies the spatial structure has to be estimated from the data. As the test value will only slightly change, the same chi-square distribution was being used.

REFERENCES

Bouma, J., De Laat, P.J.M., Awater, R.H.C.M., Van Heesen, H.C., Van Holst, A.F. and Van de Nes, Th.J., 1980a. Use of soil survey data in a model for simulating regional soil moisture regimes. Soil Sci. Soc. Am. J., 44(4): 808–814.

Bouma, J., De Laat, P.J.M., Van Holst, A.F. and Van de Nes, Th.J., 1980b. Predicting the effects of changing water-table levels and associated soil moisture regimes for soil survey interpretations. Soil Sci. Soc. Am. J., 44(4): 797–802.

Journel, A.G. and Huijbregts, Chr.J., 1978. Mining Geostatistics. Academic Press, New York, N.Y.

McBratney, A.B. and Webster, R., 1983. Optimal interpolation and isarithmic mapping of soil properties, V. Coregionalization and multiple sampling strategy. J. Soil Sci., 34: 137–162.

Stein, A., Van Dooremolen, W. and Bouma, J., 1988. (Co-)kriging point data on moisture deficit. Soil Sci. Soc. Am. J. In press.

Van der Sluys, P. and De Gruijter, J.J., 1985. Water-table classes: a method used to indicate seasonal fluctuation and duration of water tables on Dutch soil maps. Agric. Water Manage., 10(2): 109–125.

Vauclin, M., Vieira, S.R., Vachaud, G. and Nielsen, D.R., 1983. The use of cokriging with limited field soil observations. Soil Sci. Soc. Am. J., 47: 175–184.

Webster, R., 1985. Quantitative spatial analysis of soil in the field. In: Advances in Soil Science, Vol. 3. Springer-Verlag, New York, N.Y.

Wösten, J.H.M., Bouma, J. and Stoffelsen, G.H., 1985. Use of soil survey data for regional soil water simulation models. Soil Sci. Soc. Am. J., 49: 1238–1244.

Wösten, J.H.M., Bannink, M.H. and Bouma, J., 1987. Land evaluation at different scales: you pay for what you get. Soil Survey and Land Evaluation, 7: 13–24.

Yates, S.R. and Warrick, A.W., 1987. Estimating soil water content using Cokriging. Soil Sci. Soc. Am. J., 51: 23–30.

CHAPTER III.4. SIMULATION OF MOISTURE DEFICITS AND AREAL INTERPOLATION BY UNIVERSAL COKRIGING.

by

A. Stein, I.G. Staritsky, J. Bouma, A.C. van Eijnsbergen
and A.K. Bregt.

SIMULATION OF MOISTURE DEFICITS AND AREAL INTERPOLATION BY UNIVERSAL COKRIGING.

ABSTRACT

Simulation models for calculating moisture deficits for areas of land require interpolation procedures to arrive from point observations towards area covering statements. We distinguish: calculate first, interpolate later (CI-procedures), which interpolate calculated model results for test locations, and procedures which interpolate basic soil data towards test locations, followed by model calculations: interpolate first, calculate later (IC-procedures). In this study several CI- and IC-procedures are compared, which simulate moisture deficits, by means of a test set of 100 observations, yielding the Mean Squared Error (MSE). CI-procedures consistently produced lower MSE values than IC-procedures. Parameters of the pseudocovariance function (PCF), which models the spatial structure of bivariate increments in universal cokriging, were estimated by means of the Restricted Maximum Likelihood procedure. Compared to universal kriging, universal cokriging yielded comparable MSE values, but a lower Mean Variance of the Prediction Error. Best results in this study were obtained by pointwise simulation model calculations, followed by statistical interpolations.

KEY WORDS Spatial Variability, Cokriging, Moisture Deficits, Interpolation, Simulation modelling

INTRODUCTION

Computer simulation models for calculation of moisture deficits form an integral part of modern soil survey interpretation methods. These models are usually based on soil and water information for specific points in space (Feddes *et al.*, 1978; Bouma, 1989a). A problem is encountered, when calculations for points are to be extrapolated towards areas of land.

Two different procedures can be distinguished to obtain data for areas of land. In the first procedure, simulation calculations are carried out for all points where observations of all necessary variables are available. By means of a statistical interpolation method model output is predicted for other points in the area. This procedure may be described as 'calculate first, interpolate later' (e.g. De Wit and Van Keulen, 1985; Bouma, 1989b; Stein *et al.*, 1988a,b).

In the second procedure, variables to be used in a simulation model are mapped in the area of interest by means of pedological and functional clustering (e.g. Wösten *et al.*, 1985). The simulation model is applied, based on predicted variables for the representative profile for the mapping unit in which the observations are located. This method may be described as 'interpolate first, calculate later' (De Wit and Van Keulen, 1985) and has traditionally been used in soil survey interpretations where 'representative soil profiles' are being considered.

Each procedure has clear implications for soil sampling and the associated costs. The first procedure emphasizes the use of statistical sampling schemes, e.g. grid sampling or nested sampling. The second procedure emphasizes use of socalled judgement sampling, in which observations are collected

200

and interpreted according to physiographic features. The selection of any procedure will also be determined by the accuracy of the simulation model involved.

In this study different versions of the two procedures will be compared. Attention is focused on use of one simulation model which is routinely being used in the Netherlands for calculations of the moisture supply capacity, the LAMOS model (Bouma *et al.*, 1980a,b). This model is applied to calculate moisture deficits in the Mander area in the Netherlands. Concerning statistical interpolation methods, use is made of Kriging and Cokriging using the spatial structure of the different variables in the region (see, for example, Delfiner, 1978; Vauclin *et al.*, 1983; Cressie, 1986; Yates and Warrick, 1987; Stein *et al.*, 1989). As no preliminary insight exists as to which degree non-stationarity is encountered, the relatively advanced Restricted Maximum Likelihood (REML) estimation procedure is used to estimate coefficients of the functions describing the spatial dependence structure (Kitanidis, 1983; Mardia and Marshall, 1984; Zimmerman, 1989). These functions are used in non-stationary Kriging and Cokriging procedures (Stein and Corsten, in press; Stein *et al.*, in press).

Although this study focuses on one simulation model applied in one particular area, the presented methodology is general and applicable in any study concerning modelling of water movement, followed by spatial interpolation.

a. Study area

The 404 ha Mander area in the eastern part of the Netherlands has been used for groundwater extraction (Bregt and Beemster, 1989). Due to water shortage caused by this water extraction, water tables in the area were lowered and in dry periods crop yields of some of the local farmers decreased. By means of a study carried out by the Winand Staring Centre, individual losses of farmers were calculated. A total of 499 observations on basic soil properties, like rooting depth, organic matter content, clay percentage and former and present mean highest and mean lowest water table were obtained by means of a 1:10,000 soil survey, carried out in 1985. In the study area six different soil types were found that are classified according to Soil Survey Staff, 1975 (Figure 1). All soils

Figure 1. Soil map of the Mander area with observation (.) and test (o) points.

202

depth (cm)

Figure 2. Schematic diagram summarizing representative soil profiles for the six soil types being distinguished.

developed in cover sands and are relatively wet, as indicated by soil mottles, except for the Plaggepts. The schematic diagram in Figure 2 summarizes representative soil profiles for the six soil types being distinguished. Current water tables were measured. Levels that occurred before the period of water extraction were estimated by the soil surveyors using soil mottling criteria. A general lowering of 50 cm of the groundwater table was obtained during the 30 year period of groundwater extraction.

b. The simulation model and physical data

The simulation model LAMOS (De Laat, 1980) was used to calculate moisture deficits caused by lowering of groundwater tables in the Mander area. Calculations of this model are based

203

TABLE 1. Soil input parameters for LAMOS.

- moisture retention curve of the topsoil (MRC)
- moisture retention (MRC) and hydraulic conductivity (HC)
    curves for all subsoil layers.
- composition of subsoil (type and number of layers and their
    depth of occurrence).
- rooting depth for grassland (RD)
- current mean highest and mean lowest groundwater table (MHW
    and MLW, respectively).
- previous mean highest and mean lowest groundwater table (PMHW
    and PMLW, respectively).

on observations of soil variables listed in Table 1, as discussed by Bouma *et al.* (1980a,b), on meteorological data and on data on groundwater extraction.

Two different moisture deficits for grassland were calculated under the present hydrological situation, the moisture deficit for 1976 (MD76), an extremely dry year, and the yearly average moisture deficits for the period 1956 through 1985 (MD30).

The LAMOS model makes a distinction between topsoil and subsoil. The latter can be composed of different layers. Moisture retention curves are needed for both top- and subsoil. Hydraulic conductivity curves are needed for the subsoil only. Measurements were made in the context of the original study (Bouma *et al.*, 1980a,b). Hydraulic conductivity and moisture retention curves for all 499 points being considered in this study are based on measured data in well defined soil layers as discussed by Wösten *et al.*, (1985). Thirteen moisture retention curves were distinguished for the topsoil and fifteen curves for layers of the subsoil. In addition, fifteen hydraulic conductivity curves were distinguished for layers of the subsoil.

204

TABLE 2. Van Genuchten/Mualem parameters for values of α, n and l for soil layers occurring in the Mander area. Included are values for K_S and ϑ_S, the number of occurrences and a description of the layers in terms of organic matter content (OM) and texture (% loam).

Code	#Occur-rences	In Soil-type[3]	Percentage OM	Loam	Van Genuchten parms α	n	l	K_S	ϑ_S
Topsoil[1]									
I-1-2	4	1,4,6	10.8	13.5	0.026	1.27			0.590
I-1-5	4	4,6	18.1	13.5	0.020	1.25			0.610
II-1-3	12	1,3,4	1.1	22.5	0.028	1.45			0.420
II-1-5	16	1,4	2.1	14.5	0.041	1.39			0.430
II-1-9	68	1-4,6	3.5	22.5	0.036	1.34			0.460
III-1-5	59	1,4-6	3.8	13.5	0.027	1.36			0.460
III-1-8	26	1,4-6	6.2	13.5	0.026	1.32			0.500
III-1-12	8	4,6	10.8	23.5	0.019	1.28			0.590
V-1-2	36	1,4	4.0	13.5	0.045	1.39			0.480
V-1-5	73	1-5	6.0	13.5	0.042	1.33			0.500
V-1-8	61	1-6	9.5	13.5	0.030	1.29			0.530
VII-1-5	14	1,4	6.2	13.5	0.021	1.31			0.500
VII-1-6	18	1,3,4	6.2	23.5	0.025	1.25			0.500
Subsoil[2]									
B1	2	4	0-15	0-10	0.022	1.58	-0.319	47.8	0.375
B2	105	1-6	0-15	10-18	0.026	1.37	-0.825	59.3	0.434
B3	50	1-6	0-15	18-33	0.022	1.40	0.001	28.5	0.453
B4	1	***	0-15	33-50	0.017	1.52	-0.180	67.8	0.419
B16	13	5	23-100		0.015	1.31	1.740	105.1	0.738
O1	163	1,4-6	0-3	0-10	0.027	1.82	0.153	169.9	0.356
O2	420	1-6	0-3	10-18	0.023	1.60	0.260	94.2	0.385
O3	75	2-4	0-3	18-33	0.024	1.58	-0.118	87.8	0.352
O4	12	4	0-3	33-50	0.018	1.53	0.137	185.5	0.359
O5	6	***	0-3		0.049	1.87	0.605	396.3	0.334
O16	27	6	35-100		0.019	1.27	-0.364	22.6	0.879
M	16	***			0.013	1.12	0.000	7.2	0.778
H8	22	6			0.020	1.33	-1.578	8.0	0.625
S	6	***			0.006	1.28	-4.241	0.4	0.375

[1] Topsoil codes refer to Krabbenborg et al., 1983.
[2] Subsoil codes refer to Wösten and Van Genuchten, 1988.
[3] Soil types refer to Table 3. For the subsoil the occurrence of a layer as the upper layer of the subsoil next to the topsoil is considered only.
*** Not occurring in the upper layer of the subsoil.

Soil layers were defined in terms of soil texture (i.e. loam content) and organic matter content as indicated in Table 2. Calculation of moisture retention curves was based on regression analysis using organic matter and loam content as explanatory variables (Krabbenborg *et al.*, 1983). The associated hydraulic characteristics, presented in Table 2, were expressed in terms of the Van Genuchten/Mualem coefficients α, n and *l*, the saturated water content ϑ_S and the measured saturated hydraulic conductivity K_S (Van Genuchten and Nielsen, 1985).

c. Statistical procedures

Use is made of kriging and cokriging to interpolate spatially dependent observations and Thiessen polygons for ordinal data (Ripley, 1981; Corsten, 1989). In the current study area there was some conjecture, based on the soil map, that the different variables were non-stationary, although a stationary approach was proven to be succesful in the past (Stein *et al.*, 1988a,b). This is important, because in situations of severe non-stationarity sample auto- and cross-semivariograms cannot be used when kriging and cokriging are applied, but use has to be made of pseudo covariance functions (PCF), also termed: generalized covariance functions (Kitanidis, 1983; Morkoc *et al.*, 1987; Stein *et al.*, submitted). We applied the REML estimation method to estimate coefficients of these functions (Patterson and Thompson, 1975; Zimmerman, 1989). (Co-)kriging, with use of PCF's is termed universal (Co-)kriging, which is outlined in another paper (Stein and Corsten, in press).

In this study, commonly used polynomial trends of degree 1

206

and 2 are used to model the non-stationarity of the variables as compared to the stationary situation of a trend of degree 0 (Matheron, 1971). One way to deal with this type of non-stationarity is to use increments of the observations, instead of the observations themselves. The increments are assumed to be stationary. For such increments, the spatial structure is modeled by means of PCF's. In this place we will investigate the use of bi-variate increments, for p-variate increments a straightforward extension is formulated in the Appendix.

Several steps are necessary to deal with bivariate non-stationary data. The first step is to form bi-variate increments of the observations. The second step concerns the choice for a model to describe their spatial structure and the estimation of its parameters.

To obtain bivariate increments, consider n_1 observations of the variable Y_1 and n_2 observations of the variable Y_2, which are summarized in a vector Y, of length $n = n_1 + n_2$. The degree of the trend of the ith variable is denoted by ν_i. The monomials of a degree upto the degree of the trend of the coordinates of the observation locations of Y_i are contained in the matrix X_{ii}. On Y the transformation $Z = \binom{Z_1}{Z_2} = \binom{C_1 Y_1}{C_2 Y_2} = CY$ is applied in which $C = \begin{pmatrix} C_1 & 0 \\ 0 & C_2 \end{pmatrix}$ of size $N = N_1 + N_2$ by n, is such

that both Z_1 and Z_2 are vectors of univariate increments, i.e. $C_i X_{ii} = 0$ for $i = 1, 2$. For example, a matrix consisting of any N_i rows of $C_i = (I - X_{ii}(X'_{ii} X_{ii})^{-1} X_{ii})$, one of the possibilities to subtract a polynomial trend from the observations, is a valid choice. For this particular choice, the PCF's describe the spatial structure of the corresponding residuals. In principle, the values of N_i, $i = 1, 2$, need not be the same for the two

variables. They are chosen such that a polynomial trend of degree ν_i is filtered. In order to estimate the parameters of the PCF's, however, they need to be the same.

The second step concerns the estimation of the parameters of the PCF's. For bivariate increments, the PCF is the sum of the functions $g_{11}(r)$, $g_{22}(r)$ and $g_{12}(r)$, each isotropic and linearly dependent on a vector γ of parameters:

$$g_{ij}(r) = \gamma_{ij,0}\delta(r) + \gamma_{ij,1}r + \gamma_{ij,2}r^3 + \gamma_{ij,3}r^5 \qquad (1)$$

for $i,j=1,2$, where r is the distance between observation points, $\delta(r) = 1$ if $r = 0$ and $\delta(r) = 0$ if $r \neq 0$. Here, $g_{11}(r)$ describes the spatial dependence of the predictand, $g_{22}(r)$ of the covariable, and $g_{12}(r)$ describes the spatial interaction among predictand and covariable. As is well known (Delfiner, 1976), the number of terms included in (1) depends on the degrees ν_i of the polynomial trends, i.e. $\gamma_{ii,2}r^3$ can be included only if ν_i exceeds 0, and $\gamma_{ii,3}r^5$ only if ν_i exceeds 1, whereas for $i \neq j$ $\gamma_{ij,2}r^3$ can be included only if ν_{ij}, the minimum of ν_i and ν_j, exceeds 0, and $\gamma_{ij,3}r^5$ only if ν_{ij} exceeds 1, due to the requirement of positive-definiteness (Stein *et al.*, in press).

The variance of increments of the observations, $Q = var(Z) = E[ZZ'|\gamma]$, has a structure similar to (1), i.e.

$$Q = \sum_{k=1}^{\nu_1+2} \gamma_{11,k}Q_{11,k} + \sum_{k=1}^{\nu_2+2} \gamma_{22,k}Q_{22,k} + \sum_{k=1}^{\nu_{12}+2} \gamma_{12,k}Q_{12,k} \qquad (2)$$

where the matrices $Q_{ij,k} = C_iG_{ij,k}C_j'$ are of size $N_i \times N_j$. Since summation is only possible for matrices of equal size we notice that N_i and N_j need be the same when applying this procedure. If the degrees of the trends or the numbers of observations are

208

different for the differenet variables one has to define neighbourhoods of appropriate size to achieve this. Elements of the matrix $G_{ij,k}$ are obtained from each of the constituting parts of (1).

We assume that the N-dimensional Gaussian field followed by the N increments of the n observations is non-degenerate. Restrictions are imposed by the use of increments of the observations, in stead of the observations themselves. Estimation of the $\gamma_{ij,k}$ is provided by minimizing the negative log-likelihood function. This estimation procedure is commonly called the Restricted Maximum Likelihood (REML) method (Patterson and Thompson, 1975). In order to judge different degrees of trend, Akaikes' information criterion is used (Vecchia, 1988).

To estimate the coefficients, the total data set was too large to be handled efficiently. We therefore selected 9 random sets, each of 25 elements. On each of these sets the coefficients of (2) were estimated. Non-permissible coefficients were assigned the value 0 (Delfiner, 1976). The final coefficients are obtained by averaging the coefficients obtained for each of the 9 sets. Whenever universal kriging was applied, a neighbourhood of 18 observations was used. For universal cokriging the neighbourhood was extended with 19 observations for the covariable.

Results from all procedures are compared by means of a test set of 100 locations which are randomly selected in advance from the complete set of the 499 borings. For the 100 test locations, the Mean Squared Error (MSE), defined as the squared difference of predicted and observed values, is calculated:

$$\text{MSE} = 1/100 * \sum_{i=1}^{100} (\text{MD}_{ci} - \text{MD}_{pi})^2 \qquad (3)$$

where MD_{ci} is the calculated moisture deficit on the basis of measured profile characteristics in the ith test point, and MD_{pi} is the predicted moisture deficit in the ith test point. Besides, to compare the kriging and the cokriging results, the Mean Variance of the Prediction Error (MVP) is calculated as well in these points.

To investigate whether predicted values differ signifi-cantly in the mean from the original data, an adapted t-test is used, in which the spatial structure of both the directly calculated data and predicted data is included (Stein *et al.*, 1988a).

To distinguish among different procedures in terms of dispersion, Pitman's test is applied (Pitman, 1939). This test is appropriate to compare two dependent estimates, s_1^2 and s_2^2, of the same population variance σ^2, assuming normality. Every MSE value is such an estimate, and every pair of MSE-values is compared in this way. One calculates $\text{Cov}[(X_1-X_2),(X_1+X_2)]$, where X_1 denotes a prediction error of one procedure, i.e. $X_{1i} = \text{MD}_{ci} - \text{MD}_{pi}$, and X_2 denotes a prediction error of another procedure. Assuming that X_{1i} and X_{1j} are independent for $i \neq j$, as well as X_{2i} and X_{2j}, this results in a test stochastic, which follows a t-distribution.

Finally, bi-plots are constructed (Gabriel, 1971). Bi-plots graphically display the scatter of the units (rows) and the configure of the variables (columns) in a way which allows recovery of the observations. In this study, the different procedures and the observations are established, in order to show the relation between the quality of the prediction

210

procedures and the different soil types.

d. Calculation and interpolation procedures.

Different procedures are distinguished to obtain moisture deficits which are representative for areas of land (Figure 3). We will make a distinction between CI procedures (= Calculate first, Interpolate later) and IC (= Interpolate first, Calcu-- late later) procedures. Both procedures are based on use of the soil map.

In the CI procedures, MD30 and MD76 were calculated for every location where observations were available (e.g. Stein *et al.*, 1988a,b). This yielded a total of 399 values for both variables. Interpolation to each of the 100 test locations was carried out by:

(i) the CI_0-procedure, where the average value of all point calculations within each of the six mapping unit was taken. This is equivalent to ignoring the spatial structure of the variables and applying a zeroth-order trend surface in a stratified area.

(ii) the CI_1-procedure, which applied Kriging;

(iii) the CI_2-procedure, which applied Kriging in the area stratified according to the soil map (Stein *et al.*, 1988a); some soil units were combined in order to have strata of at least 30 observation points: soil types 1 and 2 were combined as Haplaquods, and soil types 4, 5 and 6 were combined as Humaquepts. Stratification was justified considering minor differences among soil types (Table 3).

(iv) the CI_3-procedure, which applied Cokriging with the current Mean Highest Watertable as the highest correlated

211

499 observations

100 TEST | 399

IC_0, IC_1, IC_2

Basic soil data for 100 test locations

Simulations
Moisture Deficits (MD)

MD for 399 locations

CI_0, CI_1, CI_2, CI_3

MD for 100 test locations
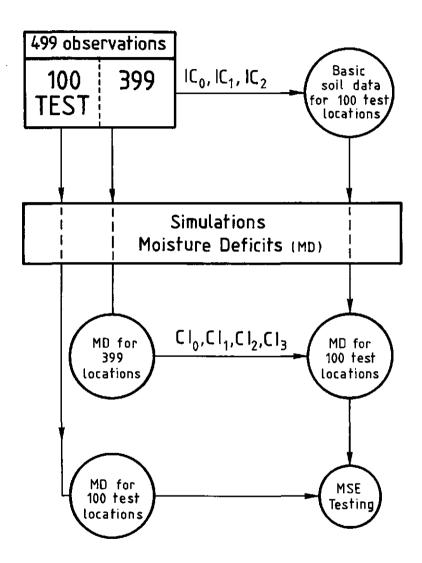
MD for 100 test locations

MSE Testing

Figure 3. *Different procedures to obtain point data on moisture deficits.*

TABLE 3. *Mean, standard deviation (SD) and coefficient of variation (CV) for calculated moisture deficits [mm], and continuous soil characteristics for the different soil types.*

		Total	st1	st2	st3	st4	st5	st6
MD76	mean	59.5	59.6	90.1	120.6	26.9	42.8	42.9
[mm]	SD	52.1	45.1	56.4	51.1	36.1	37.1	39.5
	CV	0.88	0.76	0.63	0.42	1.34	0.87	0.92
MD30	mean	11.0	10.1	17.9	27.2	4.5	5.6	6.4
[mm]	SD	15.3	15.1	18.3	15.0	9.6	5.4	6.6
	CV	1.40	1.50	1.02	0.55	2.16	0.96	1.03
MHW	mean	88.8	74.8	130.3	210.1	51.0	93.3	68.9
[cm]	SD	74.1	41.3	96.7	103.7	35.7	54.2	39.9
	CV	0.83	0.55	0.74	0.49	0.70	0.58	0.58
MLW	mean	171.5	163.2	199.1	270.2	136.8	165.4	153.3
[cm]	SD	66.1	47.3	89.5	70.1	43.7	62.7	42.9
	CV	0.39	0.29	0.45	0.26	0.32	0.38	0.28
PMHW	mean	42.9	30.2	61.2	140.3	19.7	36.1	22.2
[cm]	SD	55.3	21.8	66.4	90.3	20.7	35.8	16.2
	CV	1.29	0.72	1.08	0.64	1.05	0.99	0.73
PMLW	mean	131.2	121.7	147.4	227.1	103.2	115.0	111.7
[cm]	SD	59.7	37.2	62.2	80.8	34.8	58.4	30.8
	CV	0.46	0.31	0.42	0.36	0.34	0.51	0.28
RD	mean	30.8	29.5	34.7	39.5	28.0	35.0	28.9
[cm]	SD	4.7	2.9	2.1	2.1	3.5	0.0	2.2
	CV	0.15	0.10	0.06	0.05	0.12	0.00	0.08
OM	mean	5.8	5.0	5.5	6.9	6.3	6.2	11.9
[%]	SD	3.8	1.9	1.2	1.1	6.1	1.3	9.3
	CV	0.66	0.39	0.22	0.15	0.96	0.20	0.78
LT	mean	15.5	14.0	19.0	19.0	15.8	19.0	17.3
[%]	SD	3.7	2.2	3.1	2.6	4.5	3.9	4.7
	CV	0.24	0.16	0.17	0.14	0.29	0.21	0.27
nr. obs.		399	209	17	51	101	12	9

Soil types	Variables	
st1 – Typic Haplaquod;	RD	Rooting depth for grassland
st2 – Plaggeptic Haplaquod;	PMHW	Previous MHW
st3 – Plaggept;	PMLW	Previous MLW
st4 – Typic Humaquept;	MHW	Current MHW
st5 – Plaggeptic Humaquept;	MLW	Current MLW
st6 – Histic Humaquept;	OM	Organic matter content
	LT	Loam content of the top soil

213

covariable among the basic soil characteristics (correlation coefficient with MD30: 0.69, with MD76: 0.74).

For the IC procedures, observations concerning the variables listed in Table 1 were interpolated towards the 100 independent test points. The interpolation procedure reflects the effects of points surrounding each test point. Thus, the area is represented. Next, calculations were made for the 100 test points.

(i)   In the IC_0-procedure, calculated MD30 and MD76 values were based on the moisture retention curve of the topsoil which most often occurred within the mapping unit in which the test point was located. For the subsoil the moisture retention curve and the hydraulic conductivity curves were taken which most often occurred in the upper layer of the subsoil of the mapping unit in which the test point was located. For all other variables, like rooting depth, the average value within the mapping unit was taken.

(ii)   In the IC_1-procedure Kriging was used to predict values for the continuous soil characteristics for each of the 100 test points (RD, MHW, PMHW, MLW and PMLW). For the remaining soil characteristics (i.e. the moisture retention curve of both the top and the subsoil and the hydraulic conductivity curve for the sub soil) the characteristics of the nearest neighbour were taken for each test point. This procedure, commonly known as Thiessen polygons (Ripley, 1981), is readily available within GIS-packages.

(iii) In the IC_2 procedure, Kriging was used to predict texture and organic matter contents for each test point. Selection of a moisture retention curve was made next, using crite-

ria of Table 2. Concerning composition of the subsoil, some identical layers occurred twice or thrice in one profile. When all these layers were taken into account, depths and thicknesses of an unrealistic total of more than 40 different layers should have been predicted for each point. In this study we use therefore a simplified procedure by considering for each test point only depths and thicknesses of those layers that occurred in the nearest observation point. Further, layers occurring in fewer than 30 observations points were ignored. The depth to subsoil layers in each test point was then predicted by Kriging (e.g. Vanmarcke, 1977).

In summary, the IC_0 procedure is simple, because direct use of available soil-map data is being made. The IC_1-procedure is closely connected with direct availability of soil data and a relatively simple interpolation method in a Geographic Information System. The IC_2-procedure is a refinement of the IC_1-procedure by predicting values for variables at test points which, in turn, allow estimation of moisture retention, hydraulic conductivity and composition of the subsoil. In the CI-procedures, MD30 and MD76 calculations are made for all 399 locations. Values in test locations are derived by means of simple averaging (CI_0), or by more advanced prediction methods, like Kriging (CI_1), stratified Kriging (CI_2) or Cokriging (CI_3).

215

RESULTS AND DISCUSSION

a. Summary statistics.

Summary statistics for the total survey area and for the several soil types are presented in Table 3. Lowering of MHW and MLW by 50cm is illustrated by comparing the values of PMHW and PMLW with current MHW and MLW. As is clear from high values for the coefficient of variation, MHW observations are difficult to obtain. MD30 and MD76 values are easily obtained in soil type 3 (Plaggepts) as is illustrated by the relatively low CV value of 0.42 and 0.56, in contrast with, for example, Typic Humaquepts (CV equals 1.34 and 2.16, respectively). We further note relatively high values (120.6 mm and 27.2 mm) for MD30 and MD76 and for all water tables in Plaggepts. Mean values for the other soil types equal 59.5 mm for MD76 and 11.0 mm for MD30. A high mean value for the organic matter content is observed for the peaty soil st6 (Histic Humaquepts), with a high standard deviation. The fact that 1976 was a dry year is illustrated by the high value of the moisture deficit for this year (MD76 = 59.5 mm) as compared to the 30-year average (MD30 = 11.0 mm).

b. Spatial variability

Determination of the coefficients γ (Table 4) and evaluation of Akaikes' criterion revealed a second order trend for every variable in the area. This is not surprising, because low values for the water tables, both previous and current, loam content of the top soil and rooting depth are associated with soils predominantly occurring in the western part of the area, and pertaining the largest part of the area (appr. 85%),

216

whereas high values are observed for plaggeptic soils which occur in a relatively small area of 15% in the eastern part of the area (Table 3, Figure 1).

Looking at the spatial variability of MD76 and MD30 within the different strata we observe that the covariance function pf MD30 only slightly changes for increasing distance: the observations are more or less independent. This points to the fact that the scale of changes within these units corresponds with the second degree polynomial, and that the actual observations are more or less random deviations from this trend surface. As concerns MD76 we observe a similar pattern within the Haplaquods and within the Humaquepts. Within the Plaggepts, however,a clear spatial structure is still present. This is due to the specific structure of the Plaggepts, which were formed by man over periods of hundreds of years by adding organic material to the soil surface.

A comparison of the spatial structure of PMHW with MHW and of PMLW with MLW shows only a slight change, although mean values are clearly different. Both results support the idea that the change of the waterlevel is homogeneous throughout the study area.

The coefficients which are estimated for the PCF's for bivariate increments (Table 4b) are in agreement with the corresponding coefficients for the PCF's for univariate increments. For example, the coefficients $\gamma_{22,1}$ and $\gamma_{22,2}$ from Table 4b correspond with the coefficients $\gamma_{11,1}$ and $\gamma_{11,2}$ for MHW in Table 4a.

We finally observe, that the coefficient $\gamma_{11,1}$, which is an estimation of the nugget effect, is for every variable smaller than the estimation of the variance, which is in agreement with theoretical considerations. Only when we look to

TABLE 4a. *Estimated coefficients for pseudo-covariance functions for different variables.*

Variable		$\gamma_{11,0}$	$\gamma_{11,1}$	$\gamma_{11,2}$	$\gamma_{11,3}$
MD76	[mm]	885	-82	0	-0.92
Haplaquods		1446	-6.1	0	-0.12
Plaggepts		539	-133	0	-1.24
Humaquepts		905	-3.7	0	-0.05
MD30	[mm]	48.3	-5.9	0	-0.04
Haplaquods		51.6	-5.1	0	-0.05
Plaggepts		127	-4.0	0	-0.04
Humaquepts		74.6	-3.3	0	-0.02
RD	[cm]	4.9	-1.2	0	0
MHW	[cm]	337	-114	0	-1.05
MLW	[cm]	1679	-85	0	-0.39
PMHW	[cm]	197	-71	0	-0.38
PMLW	[cm]	1675	-85	0	-0.39
OM	[%]	10.2	-0.9	0	0
LT	[%]	8.2	-0.6	0	0

TABLE 4b. *Simultaneous estimated coefficients for pseudo covariance functions and pseudo cross covariance functions for MD30 and MD76 with MHW.*

		MD30	MD76
MD	$\gamma_{11,0}$	33.4	774.4
	$\gamma_{11,1}$	-5.2	-36.0
	$\gamma_{11,2}$	0	0
	$\gamma_{11,3}$	0	0
MHW	$\gamma_{22,0}$	411	499
	$\gamma_{22,1}$	-77	-56
	$\gamma_{22,2}$	0	0
	$\gamma_{22,3}$	0	0
MD x MHW	$\gamma_{12,0}$	36.6	324.2
	$\gamma_{12,1}$	-15.8	-33.0
	$\gamma_{12,2}$	0	0
	$\gamma_{12,3}$	0	0

PMLW and MLW, we notice that the two values are relatively close to each other.

c. Comparing different procedures.

Table 5 lists average values of the moisture deficits (MD30 and MD76) in the test points, obtained by the IC and the CI procedures. Mean values of MD30 and MD76, calculated for the test data, are included as well. All procedures yield moisture deficits which are *in the mean* in agreement with the calculated data for the test set. A t-test with dependent observations yielded no significant differences among predicted and observedvalues. Procedure CI_0, however, yield predictions with a different, lower, standard deviation, due to the fact that variation within soil types is not accounted for.

TABLE 5. Mean moisture deficits [mm], standard deviations [mm] and MSE values [mm^2] for MD76 and MD30 for the 100 testpoints, as obtained by seven procedures.

	MD76			MD30		
	Mean	St. Dev.	MSE	Mean	St. Dev.	MSE
Measured	50	44		8	10	
CI_0	50	18	1586	8	3	86
CI_1	51	37	1219	9	9	66
CI_2	54	42	1423	10	11	104
Haplaquods	61	24	1214	9	6	57
Plaggepts	124	53	4349	29	15	442
Humaquepts	27	27	890	5	10	69
CI_3	52	30	1229	8	7	66
IC_0	56	34	2007	8	7	107
IC_1	49	45	2299	8	10	109
IC_2	42	41	1857	7	9	102

The quality of the predictions was tested by means of the MSE values. The MSE values show apparent differences for the various methods. MSE values obtained for the CI-procedures are lowest. The IC-procedures all yield relatively high values, although the IC_2 procedure is slightly better than the relatively simple IC_1 procedure. But even an advanced and complex procedure, such as IC_2, did not result in very reliable predictions. In fact, results were worse than those of the relatively simple CI_0 procedure. A possible explanation is that a very detailed description of the subsoil resulted in values of limited confidence. In short, CI-procedures are more attractive to be used than IC-procedures. However, under practical circumstances high costs are associated with obtaining moisture retention curves and hydraulic conductivity curves for both top- and subsoil in the 30-40 observation points which are needed to apply a CI-procedure. Measurements in well defined soil horizons are necessary to reduce the total number of measurements (Wösten *et al.*, 1985). Especially if a simulation model yields relatively general results, one will most likely turn towards an appropriate IC-procedure.

Stratified Kriging did not result in an overall improvement of predictions (compare CI_1 with CI_2-procedure), due to the relative large MSE-value of 4349 obtained for st3, the Plaggepts. Remarkable low MSE-values (890 for MD76 and 69 for MD30) are, however, obtained for Humaquepts, due to their low internal variability. Stratification by distinction of soil types was therefore quite effective. Higher costs are also associated with stratified Kriging, as compared to unstratified Kriging, as sufficient observations must be available in every stratum, whereas in unstratified Kriging sufficient observa-

220

tions are needed only in the region as a whole.

With respect to MSE-values, universal Cokriging did not result in an improvement as compared with universal Kriging. This may be due to the fact that 12 parameters in the bivariate PCF were estimated, instead of only 4 in the univariate PCF. However, MVP-values for MD30 were 65.0 and 41.7, for Kriging and Cokriging, respectively, and 1190 and 989 for MD76. We therefore conclude, that universal Cokriging results in more precise predictions than universal Kriging.

The significance of the differences in dispersion, tested by means of Pitman's test, is shown in Table 6. Clearly, differences observed between the CI- and the IC-procedures are mainly significant, whereas there is no significant difference among the CI-procedures, nor among the IC-procedures, with some exceptions, notably that the CI_2-procedure (stratified Kriging) more resembles the IC-procedures than the IC-procedures when predicting MD76.

Considering the bi-plots (Figure 4) we observe that the Plaggepts (symbol b, with centroid B) form a different set as

TABLE 6. *Significance of differences in MSE values between the different procedures.*

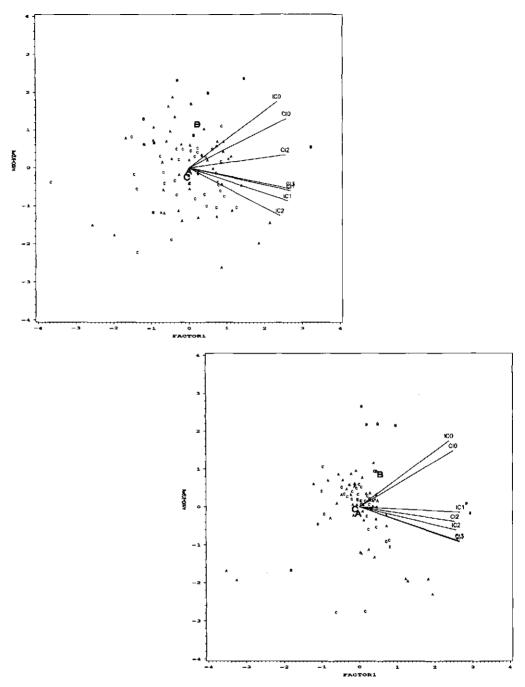|	MD30							MD76					
	CI_1	CI_2	CI_3	IC_0	IC_1	IC_2		CI_1	CI_2	CI_3	IC_0	IC_1	IC_2
CI_0				**		**					***		*
CI_1				**	***	***			**		***	***	***
CI_2				*	*	***				**			
CI_3				**	***	***					***	***	***
IC_0													
IC_1						**							

221

Figure 4. Bi-plot for MD30 (left) and for MD76 (right). Labels denote the soil types: a = Haplaquods, b = Plaggepts and c = Humaquepts. Capitals denote the mean values.

compared to the nearly similar Humaquepts (symbol a, with centroid A) and Haplaquods (symbol c, with centroid C), and that both the CI_0- and the IC_0-procedure are sensitive for prediction errors within the Plaggepts, notably for MD76, and hence give too general results to be succesfully applied for this soil unit. Besides we notice that both the CI_1-and the CI_3- procedure are overlapping, which points to an analogous behaviour in predicting the moisture deficits.

CONCLUSIONS

1. When values for moisture deficits are to be obtained for areas of land, it appears from this study that the best way to proceed is to simulate first for every location where observations are available, and to then apply statistical interpolation techniques for the area, although higher costs for data collection may be prohibitive.

2. Stratification according to soil type is quite effective to improve the quality of predictions within the strata.

3. Universal cokriging performs better than universal kriging, as evidenced by lower prediction error variances for the test set.

ACKNOWLEDGMENTS

REFERENCES

Barendregt, L.G., The estimation of generalized covariance when it is a linear combination of two known generalized covariances, *Water Resour. Res.*, *23*, 583-590, 1987.

Bouma, J., De Laat, P.J.M., Van Holst, A.F., and Van de Nes, T.J., Predicting the effects of changing water-table levels and associated soil moisture regimes for soil survey interpretations, *Soil Sci. Soc. Am. J.*, 44, 797-802, 1980a.

Bouma, J., De Laat, P.J.M., Awater, R.H.C.M., Van Heesen, H.C., Van Holst, A.F., and Van de Nes, T.J., Use of soil survey data in a model for simulating regional soil moisture regimes, *Soil Sci. Soc. Am. J.*, 44, 808-814, 1980b.

Bouma, J., Land qualities in space and time, in Bouma, J. and Bregt, A.K. (Eds.) *Land qualities in space and time*, Pudoc, Wageningen, 2-13, 1989a.

Bouma, J., Using soil survey data for quantitative land evaluation, in Stewart, B.A. (Ed.) *Advances in soil science IX*, Springer Verlag, Berlin, 177-213, 1980b.

Bregt, A.K., and Beemster, J.G.R., Accuracy in predicting moisture deficits and changes in yield from soil maps, *Geoderma*, *43*, 301-310, 1989.

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

Cressie, N., Kriging non-stationary data, *J. Amer. Stat. Assoc.*, *81*, 625-634, 1986.

De Laat, P.J.M., Model for unsaturated flow above a shallow water table, applied to a regional subsurface problem, *Agr. Res. Rept. 895*, Pudoc, Wageningen, 1980.

De Wit, C.T., and Van Keulen, H., Modelling production of field

crops and its requirements, *Geoderma, 40,* 253-267, 1985.

Delfiner, P., Linear Estimation of Nonstationary Spatial Phenomena, in Guarascio *et al.* (Ed.), *Advanced Geostatistics in the Mining Industry,* Reidel, Dordrecht, 49-68, 1976.

Feddes, R.A., Kowalik, P.J., and Zaradny, H., *Simulation of field water use and crop yield,* Pudoc, Wageningen, 1987.

Gabriel, K.R., The biplot-graphic display of matrices with application to principal component analysis. *Biometrika, 58,* 453-467, 1971.

Kitanidis, P., Statistical estimation of polynomial generalized covariance functions and hydrologic applications, *Water Resourc. Res., 19,* 909-921, 1983.

Krabbenborg, A.J., Poelman, J.N.B., and Van Zuilen, E.J, *Standaard-vochtkarakteristieken van zandgronden en veenkoloniale gronden, Deel I & II,* Stichting voor Bodemkartering, Wageningen, 1983.

Mardia, K.V., and Marshall, R.J., Maximum likelihood estimation of models for residual covariance in spatial regression, *Biometrika,* 71, 135-146, 1984.

Matheron, G., The theory of regionalized variables and its applications, *Cahiers du Centre de Morphologie Mathématique, vol. 5,* Ecole des Mines de Paris, Fontainebleau, France, 1971.

Morkoc, F., Biggar, J.W., Nielsen, D.R., and Myers, D.E., Kriging with generalized covariances, *Soil Sci. Soc. Amer. J.,* 51, 1126-1131, 1987.

Patterson, H.D., and Thompson, R., Maximum likelihood estimation of components of variance, in Corsten, L.C.A., and Postelnicu, T. (Eds.) *Proceedings of the 8th International Biometric Conference, Constanta,* 197-207,

1975.

Pitman, E.G., A note on normal correlation, *Biometrika, 31*, 9-12, 1939.

Ripley, B.D., *Spatial Statistics*, Wiley, New York, 1981.

Silvey, S.D., *Statistical Inference*, Chapman and Hall, London, 1975.

Soil Survey Staff. Soil Taxonomy, A basic system of soil classification for making and interpreting soil surveys. *USDA-SCS Agric. Handb. 436*, Washington, D.C., 1975.

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

Stein, A., Bouma, J., Van Dooremolen, W. and Bregt, A.K., Cokriging point data on moisture deficit, *Soil Sci. Soc. Amer. J.*, 52, 1418-1423, 1988b.

Stein, A., Bouma, J., Mulders, M.A., and Weterings, M.H.W., Using spatial variability studies to estimate physical land qualities of a level river terrace, *Soil Technology*, 2, 385-402, 1989.

Stein, A., and Corsten, L.C.A., Universal kriging and Cokriging as a regression procedure, *Biometrics* (in press), 1990.

Stein, A., Van Eijnsbergen, A.C., and Barendregt, L.G., Cokriging non-stationary data. *Math. Geology. in press*.

Vanmarcke, E.H., Probabilistic modeling of soil profiles, *Journal of the geotechnical engineering division, ASCE, Vol. 103, No. GT11, Proc. Paper 13365*, 1227-1246, 1977.

Van Genuchten, M. Th., and Nielsen, D.R., On describing and predicting the hydraulic properties of unsaturated soils, *Ann. Geophys.*, 3, 615-628, 1985.

Vauclin, M., Vieira, S.R., Vachaud, G., and Nielsen, D.R. The use of cokriging with limited field soil observations,

     Soil Sci. Soc. Am. J., 47, 175-184, 1983.

Vecchia, A.V., Estimation and model identification for
     continuous spatial processes, *J. R. Statist. Soc. B*, 50,
     297-312, 1988.

Wösten, J.H.M., Bouma, J., and Stoffelsen, G.H., Use of soil
     survey data for regional soil water simulation models,
     Soil Sci. Soc. Am. J., 47, 175-184, 1985.

Wösten, J.H.M., and Van Genuchten, M.Th., Using texture and
     other soil properties to predict the unsaturated soil
     hydraulic functions, *Soil Sci. Soc. Am. J.*, 52, 1762-1770,
     1988.

Yates, S.R., and Warrick, A.W. Estimating soil water content
     using cokriging, *Soil Sci. Soc. Am. J.*, 51, 23-30, 1987.

Zimmerman, D.L., Computationally efficient restricted maximum
     likelihood estimation of generalized covariance functions,
     Math. Geol., 21, 655-672, 1989.

APPENDIX: Estimating the parameters of a linear, pseudo-covariance function with the REML estimation method.

Let there be n observations on p variables, collected in the vector Y. The upper n_1 part contains the n_1 observations of the predictand, the following n_2 part the n_2 observations of the first covariable, etc., $n = \Sigma n_i$. The degree of polynomial trend which accounts for the non-stationarity of the ith variable is denoted by ν_i. An increment is any linear combination $\lambda'Y$ of the observations for which:

1. $E[\lambda'Y] = 0$;
2. $Var[\lambda'Y] = \lambda'G\lambda$, where the matrix G is derived from pseudo-covariance functions $g_{uv}:R^d x R^d \rightarrow R$, which depend on a m-vector γ of parameters; G has to be such that $\lambda'G\lambda \geq 0$ for all $\lambda'Y$ with $E[\lambda'Y] = 0$.

The collection of λ's satisfying 1 and 2 is called the collection of permissible λ's. It forms a finite-dimensional linear space, for which the elements of any basis of the space be arranged as rows of the matrix C. Any linear combination of the elements of the vector $Z = CY$ is an increment. Conversely, any increment of the observations can be expressed as a linear combination of the elements of the vector $Z = CY$ (Kitanidis, 1983; Barendregt, 1987). For example, the matrix

$$C = \begin{pmatrix} C_1 & & 0 \\ & C_2 & \\ & & \cdot \\ & & \quad \cdot \\ 0 & & C_p \end{pmatrix} = (I - X(X'X)^{-1}X') \quad (A1)$$

can be used, where X is a partitioned matrix with blocks X_{ij}, with $X_{ij} = 0$ for $i \neq j$ and X_{ii} containing k_i multivariate

monomials of the coordinates of the observation locations of the ith variable upto ν_i. In this case, the last k_i rows of X_{ii} may be skipped, as the rank of C_i equals N_i, yielding a matrix C of size $N = \Sigma N_i$ by n. Incidentally, we notice that the value of N_i is equal to n_i-k_i, where $k_i = 1$ if $\nu_i = 0$, $k_1 = 3$ if $\nu_1 = 1$, and $k_i = 6$ if $\nu_i = 2$ and hence that

$$N = \sum_{i=1}^{p} N_i = \sum_{i=1}^{p} n_i - \sum_{i=1}^{p} (\nu_i+1)(\nu_i+2)/2 \qquad (A2)$$

Concerning the distribution we assume that the vector Z has a Gaussian distribution with zero mean and variance matrix $Q = CGC'$, linear in the vector γ:

$$Q = \gamma_1 Q_1 + \gamma_2 Q_2 + \ldots + \gamma_m Q_m \qquad (A3)$$

where $Q_1,\ldots,Q_m$ are known matrices and $\gamma_1,\ldots,\gamma_m$ are unknown parameters. As is shown by standard methods, the likelihood equations are $\text{tr}(Q^{-1}Q_iQ^{-1}(Z.Z'-Q)) = 0$ for $i=1,\ldots,m$, which is equivalent to the following system of linear equations:

$$\text{tr}(Q^{-1}Q_1Q^{-1}Z.Z') = \text{tr}(Q^{-1}Q_1Q^{-1}Q_1)\gamma_1 + \ldots + \text{tr}(Q^{-1}Q_1Q^{-1}Q_m)\gamma_m$$
$$\ldots\ldots \qquad (A4)$$
$$\text{tr}(Q^{-1}Q_mQ^{-1}Z.Z') = \text{tr}(Q^{-1}Q_mQ^{-1}Q_1)\gamma_1 + \ldots + \text{tr}(Q^{-1}Q_mQ^{-1}Q_m)\gamma_m$$

A procedure to solve these equations is, after a prior estimate of Q, to solve the γ's and to calculate a new Q, which is equivalent to Newton's method (Silvey, 1975; Kitanidis, 1983). The calculation is repeated until convergence is achieved. This procdure, commonly known as the restricted maximum likelihood (REML) approach, maximizes the likelihood of the parameter γ for the observed increments. It is noted that the REML-estimate

does not depend on the choice of the basis of increments.

A key statistic to distinguish between various degrees of trends is Akaikes's information criterium. This is defined as:

$$A = l(Z|\gamma) + 2m. \tag{A5}$$

The degree of the trend which minimizes (A5) is considered the appropriate degree of non-stationarity.

SAMENVATTING

A. Stein: Ruimtelijke interpolatie

Bij het uitvoeren van ruimtelijke interpolatie, tracht men op grond van een beperkt aantal waarnemingen aan één of meer kenmerken die in een gebied verzameld zijn voorspellingen te doen voor locaties waar geen waarnemingen gedaan zijn. Hierbij kunnen we denken aan een bodemkenmerk, zoals 'beschikbaar vocht' of 'organisch stof gehalte', aan geologische gegevens, zoals het gehalte aan een mineraal en aan meteorologische gegevens, zoals de hoeveelheid neerslag, maar ook aan sociografische kenmerken als gezinsinkomen of aantal ziektedagen. Als dit gebied een regio betreft, dan kunnen we voorspellingen uitvoeren op de knooppunten van een fijnmazig rooster en dit vervolgens weergeven in de vorm van een kaart in een twee- of driedimensionaal perspectief. In dit proefschrift komen verschillende aspecten van de ruimtelijke interpolatie aan de orde.

I. Bemonsteringsstrategieën

In hoofdstuk I.1 wordt aandacht besteed aan genestelde bemonsteringsschema's. Voor alle genestelde (hiërarchische) schema's is de equivalentie bewezen tussen semivariogram waarden die verkregen worden met variantiecomponenten uit een klassiek genesteld ANOVA model en variantiecomponenten die verkregen worden met behulp van het gemiddelde van paarverschillen. Bovendien leveren de schattingsmethoden identieke resultaten voor gebalanceerde schema's, hetgeen zich overigens niet uitstrekt tot ongebalanceerde schema's. De praktische

aspecten van genesteld bemonsteren zijn onderzocht t.o.v.
verschillende andere regelmatige bemonsteringsschema's in een
studie waarin gebruik gemaakt is van acht gesimuleerde toevals-
velden. Genesteld bemonsteren kan niet aanbevolen worden voor
het bepalen van het semivariogram en moet afgeraden worden
indien met de aldus verzamelde waarnemingen ruimtelijke inter-
polatie wordt uitgevoerd. Belangrijkste reden is dat er m.b.v.
genestelde bemonsteringsstrategieën voor te weinig afstanden
een semivariogram verkregen wordt.

In hoofdstuk I.2 wordt een sequentiële bemonsteringsproce-
dure gebruikt om waarnemingen aan de infiltratiesnelheid van
water in een rivierterras te verzamelen. Deze procedure is
geschikt om de gemiddelde waarde van een kenmerk te bepalen en
deze als een vlakdekkende uitspraak te hanteren indien het te
interpoleren kenmerk weinig variatie kent. Bovendien is deze
procedure efficiënt uit het oogpunt van een reductie van het
benodigde aantal waarnemingen, en de daarmee geassocieerde
kosten. Er wordt een beslissingscriterium gehanteerd om na
iedere nieuwe waarneming vast te stellen of de bemonstering
moet doorgaan of kan stoppen.

Deze twee hoofdstukken gaan over enkelvoudige kenmerken.

II. Voorspellingen

Bij een statistische studie is het in eerste instantie van
belang een zuivere voorspeller te vinden met minimale variantie
van de voorspelfout voor de waarde die in een niet bezocht punt
gemeten zou kunnen worden. Deze voorspeller is een lineaire
combinatie van de waarnemingen die gemeten zijn aan het kenmerk
dat men wil voorspellen of aan andere, hoog gecorreleerde, va-
riabelen die bij voorkeur met een grotere dichtheid bemonsterd

232

zijn. Deze waarnemingen kunnen via een willekeurige configuratie over het studiegebied verspreid zijn. Als regel is er sprake van (ruimtelijke) afhankelijkheid tussen de waarnemingen: waarnemingen dicht bij elkaar lijken sterker op elkaar dan waarnemingen die verder van elkaar verwijderd zijn. De ruimtelijke afhankelijkheidsfunctie, d.w.z. de pseudo-covariantie functie, legt deze afhankelijkheid vast.

In hoofdstuk II.1 worden de enkel- en meervoudige voorspellingsmethoden kriging en cokriging geïnterpreteerd in het licht van regressie-procedures bij gegeven covariantiestructuur. De resultaten zijn van toepassing op situaties waarin sprake is van een trend. Een belangrijk resultaat uit deze studie is, dat niet langer Lagrange multiplicatoren nodig zijn in de afleiding of de presentatie van de belangrijkste formules. Bovendien worden essentiële formules beter interpreteerbaar worden. De studie wordt geïllustreerd aan de hand van het Limagne-onderzoek.

In hoofdstuk II.2 worden voorwaarden geformuleerd waaraan de pseudo-crosscovariantiefuncties die de ruimtelijke structuur beschrijven van meervoudige kenmerken moeten voldoen om toegelaten te zijn. De parameters van deze functies worden bepaald op grond van de verzamelde waarnemingen met behulp van de Restricted Maximum Likelihood methode. In dit hoofdstuk wordt expliciet aandacht besteed aan meervoudige toevalsvelden waarvan de incrementen een polynomiale trend uitfilteren. Bovendien is een generalisatie van cokriging geformuleerd: ook meervoudige lineaire combinaties van waarnemingen, zoals een blokgemiddelde of een gradiënt, kunnen voorspeld worden.

III. Toepassingen in bodemkundige studies.

De toepassingen stammen uit twee bodemkundige studies: een studie waarin onderzoek is gedaan naar vochttekorten in het Mandergebied in Overijssel en een tweede studie naar beschikbaar vocht en infiltratiesnelheid van water in de terrassen van de Allier in de Limagne-slenk in Frankrijk.

In hoofdstuk III.1 is de reductie van gemaakte kosten bij het uitvoeren van ruimtelijke interpolatie berekend. Verschillende kenmerken zijn op een verschillende schaal geïnventariseerd. Het bleek mogelijk te zijn om voor een lagere prijs een hogere precisie te bereiken. Deze studie had betrekking op het Mandergebied.

In hoofdstuk III.2 is de afstand tussen bemonsteringspunten vastgesteld om een bepaalde van te voren vastgelegde precisie te bereiken. Het bleek mogelijk te zijn om een reductie in de totale kosten te realiseren door een covariabele bij het uitvoeren van ruimtelijke interpolatie te betrekken. Deze studie had betrekking op beschikbaar vocht, dat in de Limagne slenk in Frankrijk hoog gecorreleerd is met het eenvoudig te bepalen 'diepte tot grindbanken'.

In hoofdstuk III.3 is cokriging gecombineerd met een stratificatie van het studiegebied o.a. in de vorm van bodemeenheden die voorkomen op beschikbare bodemkaarten. Enerzijds zien we een toename van de totale precisie, anderzijds zien we zowel een toename van de precisie in homogene kaarteenheden als een afname in heterogene kaarteenheden. Ruimtelijke interpolatie zonder gebruik van dergelijke informatie heeft derhalve een vrij globaal karakter, en kan verfijnd worden door gebruik te maken van beschikbare bodemkaarten.

In Hoofdstuk III.4 worden verschillende interpolatie-

234

procedures met elkaar vergeleken bij het gebruik van een specifiek model om vochttekorten te bepalen. Belangrijkste vraag is of modelberekeningen op veel waarnemingslocaties moeten worden uitgevoerd, gevolgd door ruimtelijke interpolatie, of dat de basisgegevens voor het model moeten worden geïnterpoleerd, gevolgd door modelberekeningen plaatsen waar niet gemeten is. De eerste methode, 'reken eerst, interpoleer later' leverde structureel betere resultaten op dan de tweede methode 'interpoleer eerst, reken later'. In deze studie is universeel cokriging operationeel toegepast.

CURRICULUM VITAE

Alfred Stein is geboren op 1 april 1958 te Rijswijk (Z.H.). In 1976 behaalde hij het eindexamen gymnasium B aan het Lorentz Lyceum te Eindhoven. In 1983 studeerde hij af in de Wiskunde en Informatica met specialisatie toegepaste statistiek aan de Technische Universiteit Eindhoven. Het afstudeeronderzoek bij prof. Doornbos betrof een statistische procesanalyse bij Forbo Krommenie, B.V. Aansluitend was hij werkzaam bij de stralingsbeschermingsdienst van de TU. Vanaf 1985 is hij als universitair docent in dienst bij de vakgroep Bodemkunde en Geologie van de LUW. Hij verzorgt onderwijs en onderzoek in de ruimtelijke statistiek en statistische ondersteuning van bodemkundig en geomorfologisch onderzoek.

Het promotie-onderzoek is gestart in 1986. Het sluit aan bij een algemene vraag om te komen tot grafische weergave van bodemvariabelen, inclusief de daaraan gekoppelde onzekerheden. Op uitnodiging van Prof. Fiering verbleef hij gedurende 1 maand aan Harvard University, Division of Applied Sciences, USA, waar hij samenwerkte met Dr. Christakos. Op uitnodiging van Dr. Wackernagel verbleef hij gedurende enige tijd aan het Centre de Géostatistique in Fontainebleau, Frankrijk. Hij is editor van Soil Technology.

236