SOIL TECHNOLOGY vol. 1, p. 349–364 Cremlingen 1988

THE APPLICATION OF GEOSTATISTICS IN EROSION HAZARD MAPPING

S.A.H.A. **van Beurden** & H.Th. **Riezebos**, Utrecht

Summary

Geostatistical interpolation or kriging of soil and vegetation variables has become an important alternative to other mapping techniques. Although a reconnaissance sampling is necessary and basic requirements of geostatistics have to be met, kriging has the advantage of giving estimates with a minimized error. In erosion hazard assessment this technique may therefore be applied instead of the conventional choropleth mapping, in order to improve the reliability of the input data for an erosion model. In a study, which was performed in the Ardèche district (Southern France) at a detailed scale (1:5000), the two methods are compared directly. It demonstrates a straightforward way to decide which method will give the best overall results at different sampling densities for erodibility controlling variables and the USLE K-factor. To this end a land unit map and several kriging maps of the investigated area (±2,5 km^2) have been prepared. Quotients of kriging estimation variances and within-unit variances of the land unit map were calculated as a relative measure of error reduction. It was found that the kriging estimation variance is 66% (on average) of the within-unit variance for the USLE K-factor at a 150 meter sampling distance. This percentage can be further reduced to about 25% at a 50 meter sampling distance. It is concluded that kriging is an efficient option for mapping erodibility at this scale.

1 Introduction

The assessment of erosion risk in a given area proceeds generally by mapping land units, which provide an extrapolation base for a limited number of observations. This conventional method introduces an error in the maps by assuming a small within-unit variance — representing the scatter of the measurements around the mean — and is often an oversimplification of reality. In this study it was tried to find out whether a geostatistical, optimal interpolation would be a more accurate alternative to that method if exactly the same number of observations was used. In order to do so the variances of kriging maps and of land unit-based maps were compared for different sampling densities and for several erodibility controlling variables. Since this study was confined to erodibility only, the land unit map is in fact a physiographic soil map, in this case based on a classification of the lithology, the slope angle and the vegetation or landuse.

A comparable investigation was car-

ISSN 0933–3630

D–3302 Cremlingen-Destedt, W. Germany
0933–3630/88/5011851/US$ 2.00 + 0.25

ried out in the Netherlands (BREGT et al. 1987) using soil-hydrological data. In the study the purity of conventional soil map classes was compared with the purity of a thematic map based on geostatistical interpolation. It was found that there was no significantly better result when kriging was applied that way. In this study however the kriging maps are not subjected to any form of classification, As a consequence full justice is done to possible error reductions.

2 Location and description of investigated area

The study area in the Ardèche province of France (fig.1) is situated within a sedimentary basin. Jurassic and Cretaceous carbonate rocks such as limestones, calcareous marls and their alternations and mixtures from the geological basement. The area has a humid submediterranean climate (BORNAND et al. 1977) with a mean annual precipitation of 1036 mm, a dry summer and a wet spring and autumn. The mean annual temperature is 13°C. Soils are mainly "sols bruns calcaires" with an A(B)C profile and shallow "rendzines" having an AC or A(B)C profile. Molding of the Tertiary landscape took place during the Pleistocene when periglacial conditions alternated with warmer periods. During the Holocene the human impact on the landscape (deforestation and subsequent agricultural use) led to the initiation of soil erosion and locally to the formation of badlands. The soil erosion became more intense when many cultivated terrains were abandoned after the first world war (BOZON 1978).

The drainage basin considered in this study is shown in fig.2. The watersheds are delineated by the contours of the gridcells just inside the drainage basin to indicate the resolution of the maps (gridcell size = 30 × 30 meter). The 150 meter square sampling grid is also indicated. The reasons for choosing this particular distance are discussed below in the section on kriging.

3 Methods

3.1 Analyzed variables

This study includes the variables that are required for the calculation of the USLE K-factor (WISCHMEIER & SMITH 1978):

1. The organic matter content
2. The saturated hydraulic conductivity
3. The texture

Apart from this the porosity is also considered.

The support of a soil sample is essential for the application of geostatistical methods. This support refers to the volume, shape and orientation of the sample. A change in the support can have a considerable effect on the presence of short-range variations (STARKS 1986) and thereby on the feasibility of kriging, as will be explained in the kriging section. This was an important consideration in the present study, because it was known from earlier fieldwork that especially the saturated hydraulic conductivity was highly variable at short distances. This problem was tackled by taking 3 samples at each location: 3 points were chosen within a 5 meter radius around the sample location. At each point a metal cylinder (radius 25.0 mm, length

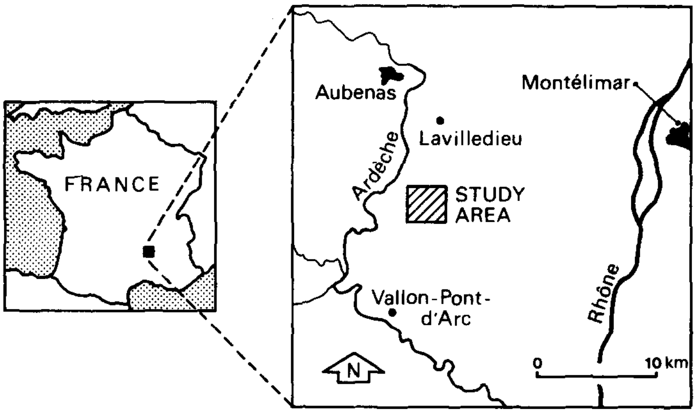

Fig. 1: *Location of the study area.*

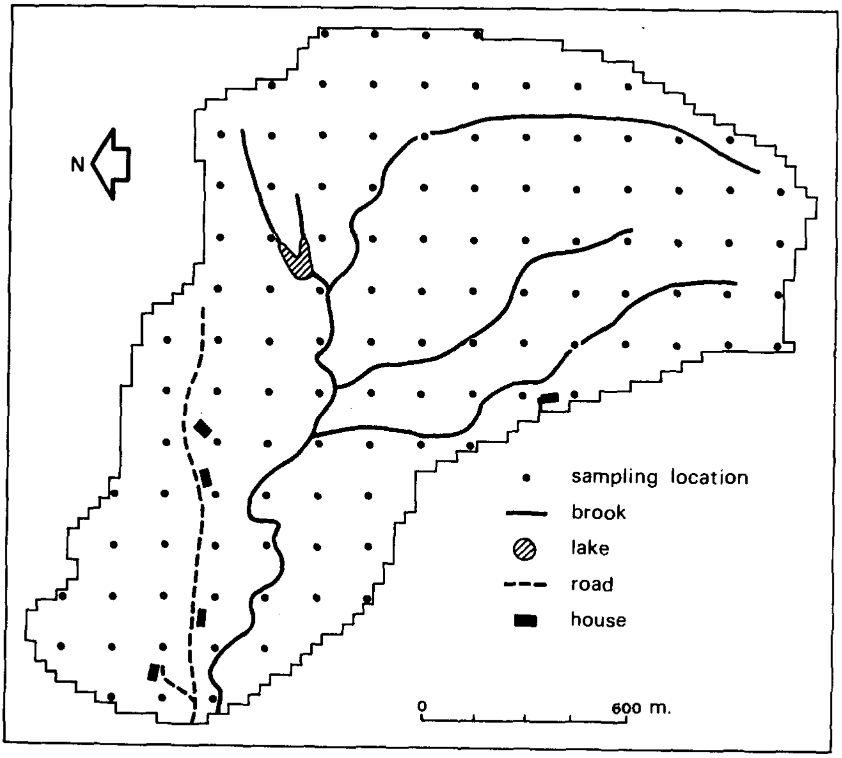

Fig. 2: *Sampling grid of the investigated drainage basin.*

50.5 mm) was driven into the soil vertically after the removal of loose surface matrial, such as gravel pavements or mulch. The cylinders were then dug out and later on saturated in 48 hours. After this preparation the saturated hydraulic conductivity was measured. The saturated samples were then weighed, dried in an oven at 105°C for 24 hours and finally weighed again to calculate the porosity. The complete contents of the 3 cylinders were then mixed and packed for later analysis of the organic matter content (Walkley-Black procedure) and of the texture. The saturated hydraulic conductivity and porosity measurements were combined by means of the "bisquare weight method" (MOSTELLER & TUKEY 1977) which assigns weights to each value depending on the difference with the ordinary mean value. This procedure reduces the effect of outliers which may occur due to the presence of macropores or stones.

3.2 Conventional choropleth map

The physiographic soil map or land unit map is presented in fig.3. The units are defined as indicated in tab.1.

The lithological distinction between marl and limestone in the map practically coincides with the division of the area in cultivated land and shrubland on flat slopes (usually abandoned fields) on one hand and forest and shrubland on often steeper slopes on the other hand. This division of the area in unit 1 to 3 and unit 4 to 7 is therefore very obvious in the field and seems to be a sound basis for an extrapolation.

3.3 Kriging

Geostatistics constitute a theory about the statistical behaviour of natural phenomena with a spatial variability (DAVIS 1973). Its fundamental concept is the regionalized variable which can always be represented by a continuous surface over the map plane (WEBSTER & BURGESS 1980a). The spatial variation of any variable can be described with three major components (BURROUGH 1986):

1. A deterministic component, for example a constant average or a constant trend.
2. A stochastic component, which can only be characterized by its statistical properties and not by any function (as is possible with the first component).
3. A random, spatially independent component.

The second and third component can be captured in the semivariance τ, the geostatistical measure of spatial dependency for a given sample distance:

$$\tau(h) = 1/2n \cdot \sum_{i=1}^{n} \{Z(Xi) - Z(Xi+h)\}^2$$

with
n = number of sample pair
h = sampling distance interval
$(Xi, Xi+h)$ = pair of sample points at distance h

The semivariance consequently indicates the average similarity of pairs of observations at a given distance. A large semivariance produces a higher probability of a large difference in values at a certain distance. Plotted against the sampling distance a continuous function,

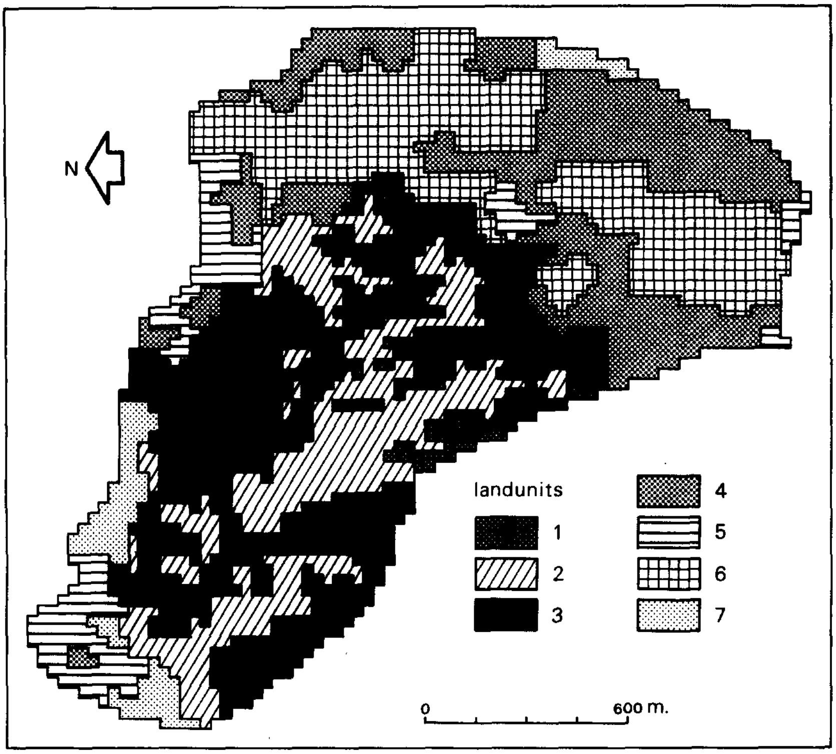

Fig. 3: *Land unit map. The units are described in tab.1.*

unit	lithology	slope	vegetation / landuse
1	marl	> 20%	shrubland
2		≤ 20%	shrubland
3		≤ 10%	cultivated land
4	limestone	> 20%	shrubland
5		≤ 20%	shrubland
6		> 25%	woodland
7		≤ 15%	grassland

Tab. 1: *Classification of land units.*

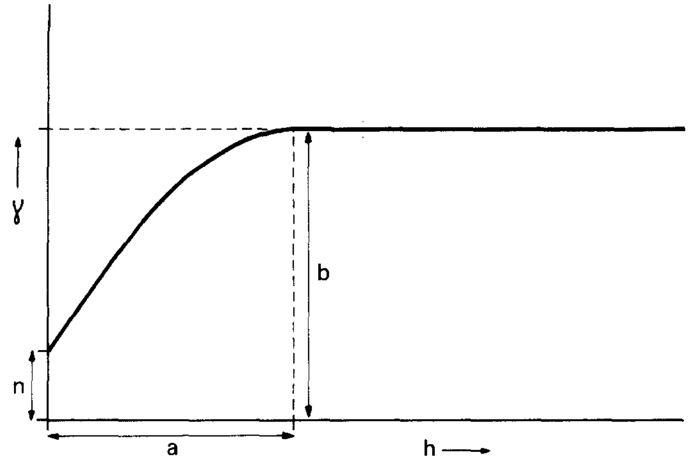

Fig. 4: *Typical, spherical semivariogram model.*
a = range; n = nugget variance; b = sill variance; (see text for explanation)

the semivariogram model, can be fitted through the variances. This is necessary in order to obtain estimates of the semivariance at all possible distances for the interpolation procedure. Fig.4 shows a typical spherical semivariogram model, defined by three parameters:

a the maximal distance of spatial dependency, usually referred to as the range.

n the nugget variance, including variations on very short distances (compared to the sampling distance) and measuring errors.

b the sill variance, which is directly comparable with the total variance in the data.

Although there are several alternatives (linear, exponential, Gaussian), the spherical model is preferred because it is often used in soil surveys, especially when the semivariogram clearly levels off at increasing distance (BURGESS & WEBSTER 1980a). The actual model fitting was performed by applying a weighted least square procedure. This has the advantage that weighing is proportional to the number of observation pairs for each calculated semivariance, thus giving most weight to the values at small distances (CRESSIE 1985).

After the semivariogram model has been calculated, kriging can be applied. Kriging is a weighted average interpolation method, which is based on the unique spatial characteristics of the considered variable. The weights that are assigned to respective observations depends on the semivariances that are associated with the distances of these observations to the point to be interpolated. As a result of this the estimates of the interpolation have a minimal and known error.

Kriging can only be used without the necessity of special adaptions in the procedure if two basic requirements are met (DAVIS 1973):

1. Stationarity of the mean in all parts of the investigated area. This requirement refers to the first component of

the regionalized variable as mentioned above. It means that there should not be any slow change or "drift" of the expected mean value within a certain neighbourhood in order to avoid biased estimation.

2. Stationarity of the variance, which implies that one semivariogram should be valid throughout the area.

If the first requirement is not met, it is possible to subtract the drift of the original values and consecutively treating the residuals as normal variables (WEBSTER & BURGESS 1980). This procedure is called "universal kriging" and comprises three steps:

- The estimation and removal of the drift. In this study first or second order trend surfaces were used to this end. The order depending on an analysis of variance, which compared the variance due to the trend with the variance due to the deviations from the trend (DAVIS 1973).
- The interpolation of the residuals, which have a stationary deterministic component. As stands to reason the semivariogram that is used for this interpolation is calculated from the residuals and not from the original data.
- The addition of the trend surface map and the map of the interpolated residuals.

If the interpolation is intended to give estimates of point values it is called point kriging. Actually these estimates and the associated kriging variances describe a probability distribution for a sample of exactly the same support as the original observations. It is however possible to make estimates over areas, which results in smaller estimation variances and smoother maps. This procedure is called block kriging (BURGESS & WEBSTER 1980b). It is a good option for this study, since erosion models will estimate soil losses of a certain area anyway. Block kriging also has the advantage of being less sensitive to changes in the model parameters than point kriging (BROOKER 1986). The actual block size in this investigation, namely the already mentioned gridcell size 30 × 30 meter, was a compromise between two tendencies:

1. A reduction in size in order to give a precise representation of the slopes. For the application of a deterministic erosion model it is desirable that slopes of average length are divided in at least three segments to allow a minimal differentiation in erosion intensity and/or sedimentation.
2. An increase in size so that smaller block kriging estimation variances can be obtained.

The feasibility of a geostatistical interpolation, which depends on the size and scale of the spatial variability and the required sampling density, is of course best checked by means of a reconnaissance study preceding a full scale sampling. In this investigation a nested analysis of variance was performed to that end. This technique requires a hierarchical sampling scheme, based on different levels of sampling distance. It is by no means an obligatory step in geostatistical interpolation procedure, but it fits in well because of its objective approach (RIEZEBOS, in press). In this case four levels of sampling distance were chosen: 4, 40, 400 and 2000 meter. An area in the direct vicinity of the study area was then divided into 25 squares of 600 × 600 meter, each being subdivided into 25 subsquares (fig.5). Eight squares were

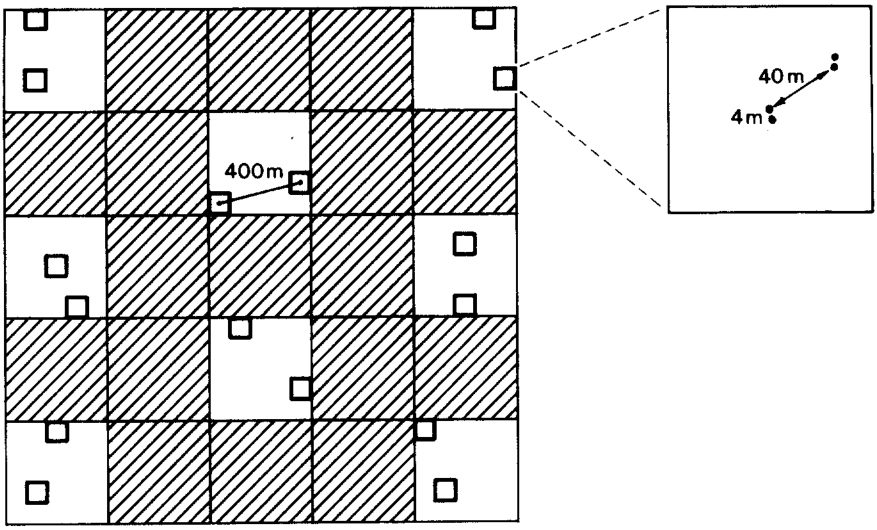

Fig. 5: *Spatial organisation of nested sampling.*

selected at random, having an approximate intersquare distance of 2000 meter. In each of these squares two subsquares were chosen at random, about 400 meter apart. The location of the centres of these 16 subsquares was used to locate the sampling points of the next two levels directly in the field. From the 16 centre points two randomly selected directions served to locate a point at 4 and 40 meter respectively. From the "40 meter point" another sample point was located at 4 meter in a randomly chosen direction. The number of replicates at the first level of 2000 meter sampling distance was 8. The remaining levels were subdivided two at a time. Thus the complete sampling scheme included 8 × 2 × 2 × 2 = 64 samples in all. The resulting information allows rough provisional estimations of the semivariograms, which are then used to establish the optimal sampling distance. the main general considerations for deciding on that distance are:

- the sampling distance should be within the range (BURGESS & WEBSTER 1980a, BURROUGH 1986).
- the cost-effectiveness of the sampling should be in accordance with the overall aims of the investigation.

In addition to that there were the following considerations, which are specific to this study:

- all variables were to be sampled on the same grid, in view of multivariate statistical analyses not discussed in this paper.
- the definitive sampling of an area of at least 2 km^2 was to be performed within a few weeks.

The semivariances of the nested sampling were combined with the semivariances of the final sampling in order to increase the reliability of the semivariogram model. All trend analyses, semivariance analyses and kriging interpolations were performed with the PC-GEOSTAT package (BURROUGH &

the regionalized variable as mentioned above. It means that there should not be any slow change or "drift" of the expected mean value within a certain neighbourhood in order to avoid biased estimation.

2. Stationarity of the variance, which implies that one semivariogram should be valid throughout the area.

If the first requirement is not met, it is possible to subtract the drift of the original values and consecutively treating the residuals as normal variables (WEBSTER & BURGESS 1980). This procedure is called "universal kriging" and comprises three steps:

- The estimation and removal of the drift. In this study first or second order trend surfaces were used to this end. The order depending on an analysis of variance, which compared the variance due to the trend with the variance due to the deviations from the trend (DAVIS 1973).
- The interpolation of the residuals, which have a stationary deterministic component. As stands to reason the semivariogram that is used for this interpolation is calculated from the residuals and not from the original data.
- The addition of the trend surface map and the map of the interpolated residuals.

If the interpolation is intended to give estimates of point values it is called point kriging. Actually these estimates and the associated kriging variances describe a probability distribution for a sample of exactly the same support as the original observations. It is however possible to make estimates over areas, which results in smaller estimation variances and smoother maps. This procedure is called block kriging (BURGESS & WEBSTER 1980b). It is a good option for this study, since erosion models will estimate soil losses of a certain area anyway. Block kriging also has the advantage of being less sensitive to changes in the model parameters than point kriging (BROOKER 1986). The actual block size in this investigation, namely the already mentioned gridcell size 30 × 30 meter, was a compromise between two tendencies:

1. A reduction in size in order to give a precise representation of the slopes. For the application of a deterministic erosion model it is desirable that slopes of average length are divided in at least three segments to allow a minimal differentiation in erosion intensity and/or sedimentation.
2. An increase in size so that smaller block kriging estimation variances can be obtained.

The feasibility of a geostatistical interpolation, which depends on the size and scale of the spatial variability and the required sampling density, is of course best checked by means of a reconnaissance study preceding a full scale sampling. In this investigation a nested analysis of variance was performed to that end. This technique requires a hierarchical sampling scheme, based on different levels of sampling distance. It is by no means an obligatory step in geostatistical interpolation procedure, but it fits in well because of its objective approach (RIEZEBOS, in press). In this case four levels of sampling distance were chosen: 4, 40, 400 and 2000 meter. An area in the direct vicinity of the study area was then divided into 25 squares of 600 × 600 meter, each being subdivided into 25 subsquares (fig.5). Eight squares were

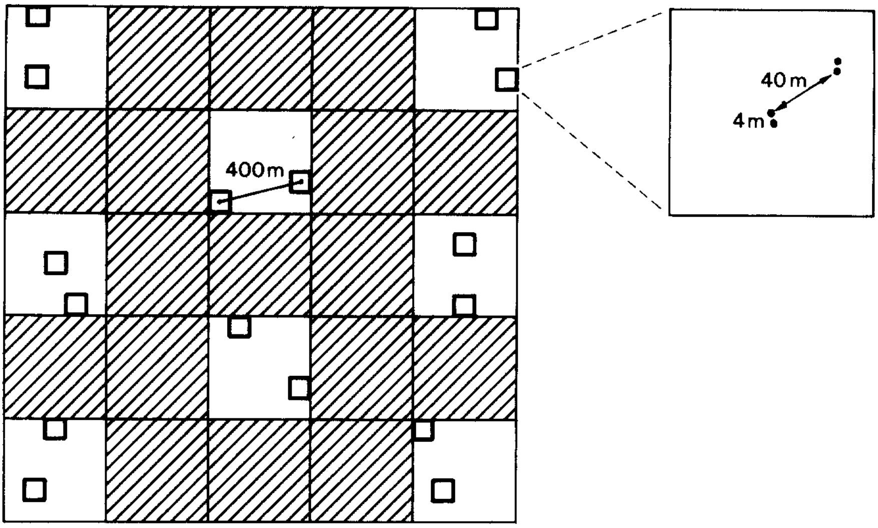

Fig. 5: *Spatial organisation of nested sampling.*

selected at random, having an approximate intersquare distance of 2000 meter. In each of these squares two subsquares were chosen at random, about 400 meter apart. The location of the centres of these 16 subsquares was used to locate the sampling points of the next two levels directly in the field. From the 16 centre points two randomly selected directions served to locate a point at 4 and 40 meter respectively. From the "40 meter point" another sample point was located at 4 meter in a randomly chosen direction. The number of replicates at the first level of 2000 meter sampling distance was 8. The remaining levels were subdivided two at a time. Thus the complete sampling scheme included 8 × 2 × 2 × 2 = 64 samples in all. The resulting information allows rough provisional estimations of the semivariograms, which are then used to establish the optimal sampling distance. the main general considerations for deciding on that distance are:

- the sampling distance should be within the range (BURGESS & WEBSTER 1980a, BURROUGH 1986).
- the cost-effectiveness of the sampling should be in accordance with the overall aims of the investigation.

In addition to that there were the following considerations, which are specific to this study:

- all variables were to be sampled on the same grid, in view of multivariate statistical analyses not discussed in this paper.
- the definitive sampling of an area of at least 2 km^2 was to be performed within a few weeks.

The semivariances of the nested sampling were combined with the semivariances of the final sampling in order to increase the reliability of the semivariogram model. All trend analyses, semivariance analyses and kriging interpolations were performed with the PC-GEOSTAT package (BURROUGH &

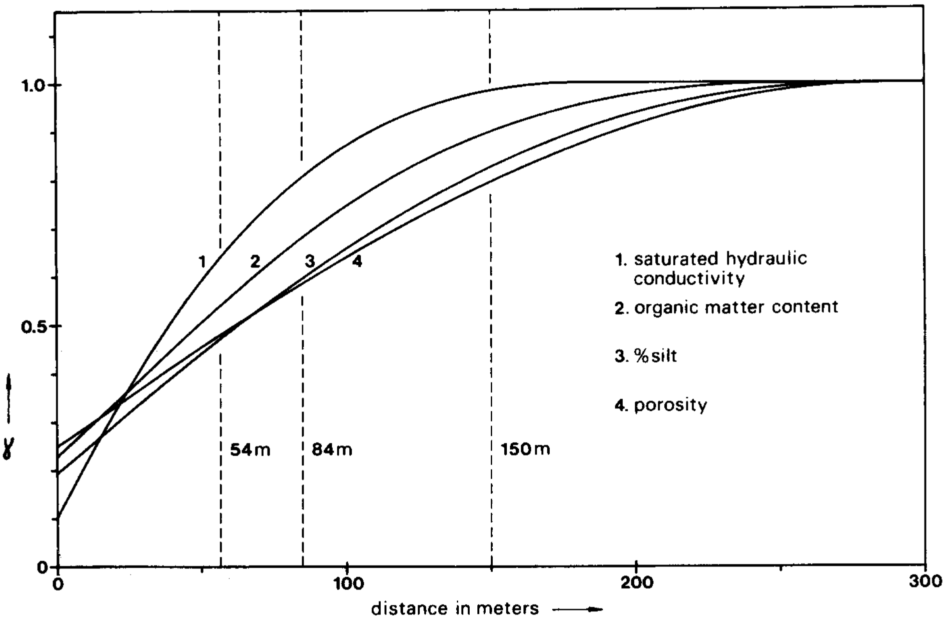

Fig. 6: *Final semivariograms.*

VAN KEULEN 1986).

4 Results

The final semivariograms are presented in fig.6. They are standardized by setting all sill variances to 1 in order to facilitate comparison. The relative errors of the spherical model parameters are about 50% for the nugget variances, 5% for the sill variances and 20% for the ranges. The consequences of these errors for the final results are discussed below. The graphs show some important features:

1. As required the sampling distance (150 m) is within the ranges of all variables. One should keep in mind that the sampling distance is much larger than the maximal and average distances of a block centre to the closest observation, in this case 84 and 54 meters respectively.
2. The nugget variances are roughly 25% of the sill variances. This is a quite reasonable and useful result when compared with other studies (BURGESS & WEBSTER 1980, BURROUGH 1986).
3. The semivariogram of the saturated hydraulic conductivity differs from the others because of its slightly shorter range and relatively small nugget variance. The latter is presumably a result of the support (see section "analyzed variables") of this variable: the value at each location is after all the average of three measurements. This procedure eliminates part of the variance at a microscale.

The kriging maps are shown in figures 7–9. The maps clearly reflect the mutual correlations of the variables, with coefficients ranging from 0.52 for satu-

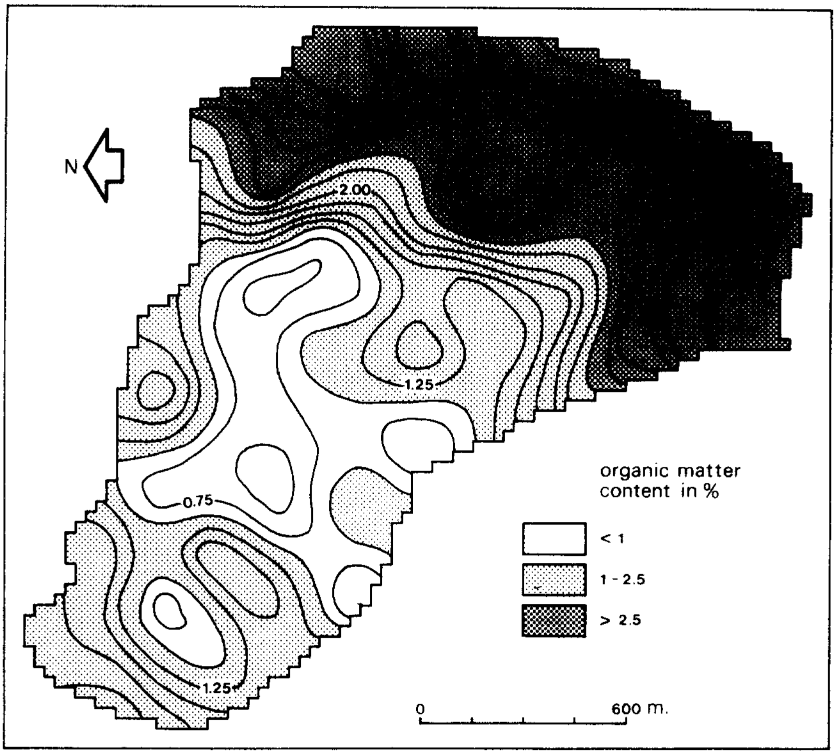

Fig. 7: *Kriging map of organic matter content.*

rated hydraulic conductivity and organic matter content to 0.84 for porosity and organic matter content. More interesting for this study however is the correspondence of the kriging maps to the land unit map. The above mentioned main division between the land units can be easily recognized in the kriging maps. But the land unit map also shows details that cannot possibly be expected in the kriging maps — due to the limited sampling density — and that in reality may very well be non-existent for the considered variables.

Another important feature is the conformation of the map contours to individual samples. The map of the saturated hydraulic conductivity shows this effect distinctly, while the map of the organic matter content has smooth contours. This is a result of the relatively smaller nugget effect for the saturated hydraulic conductivity when compared with the other variables (fig.6). The consequence of this is that the influence of the nearest observation becomes preponderant (JOURNEL & HUIJBREGTS 1978). In addition to this the map of the organic matter content has more missing values. So it is important to realise that the smoothness or sinuosity of the contours in itself does not give an indication of the relative interpolation error. One of the kriging variance maps, in this case referring to the porosity estimates, is presented in fig.10. It illustrates the

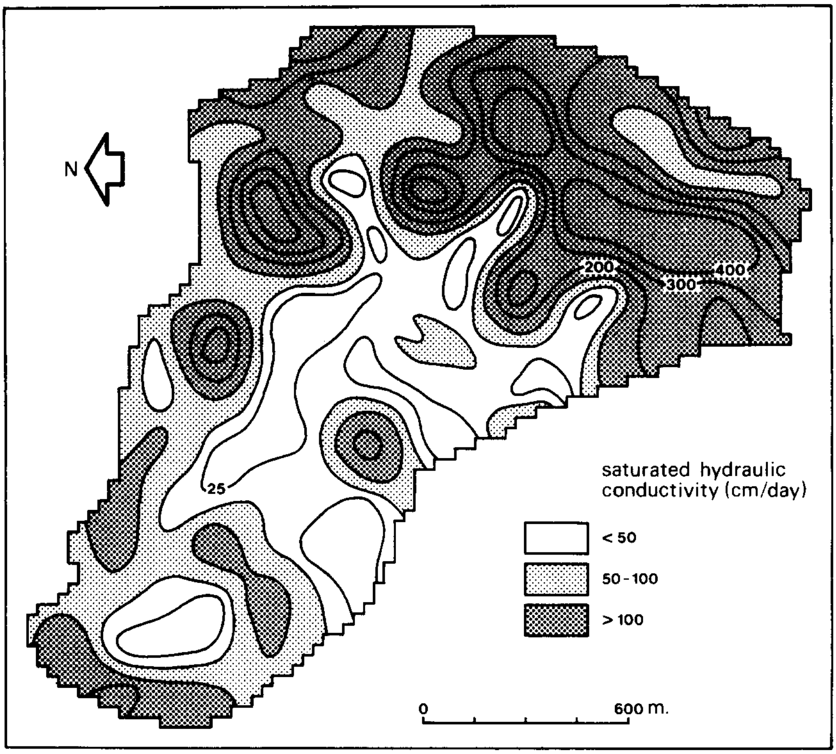

Fig. 8: *Kriging map of saturated hydraulic conductivity.*

range of estimation errors that may be encountered and it also clearly shows the strong influence of edge effects and missing values on the reliability of the map.

The reliability of the kriging interpolations themselves has been tested by means of a jackknifing procedure. To this end a total of 40 sample values was labeled as missing in several kriging runs. It was then calculated which percentage of the actually measured values laid inside the interval defined by the kriging estimate and the associated kriging error. The resulting values are 64% for porosity, 60% for organic matter content and 88% for saturated hydraulic conductivity. In the land unit map this percentage is 68 for each variable, since this is simply the portion of measured values enclosed by an interval of one standard deviation around the mean. These percentage values indicate that the kriging confidence intervals of porosity and organic matter content are only slightly underrated and that the interval of the saturated hydraulic conductivity is overrated. These remarks refer however only to the kriging results at a 150 meter interpolation distance and will not produce more favourable conclusions regarding the application of kriging, so it is concluded that there is no real need for changing the semivariogram model parameters.

As stated above the average quotients of the kriging estimation variances and the within-unit variances were calculated

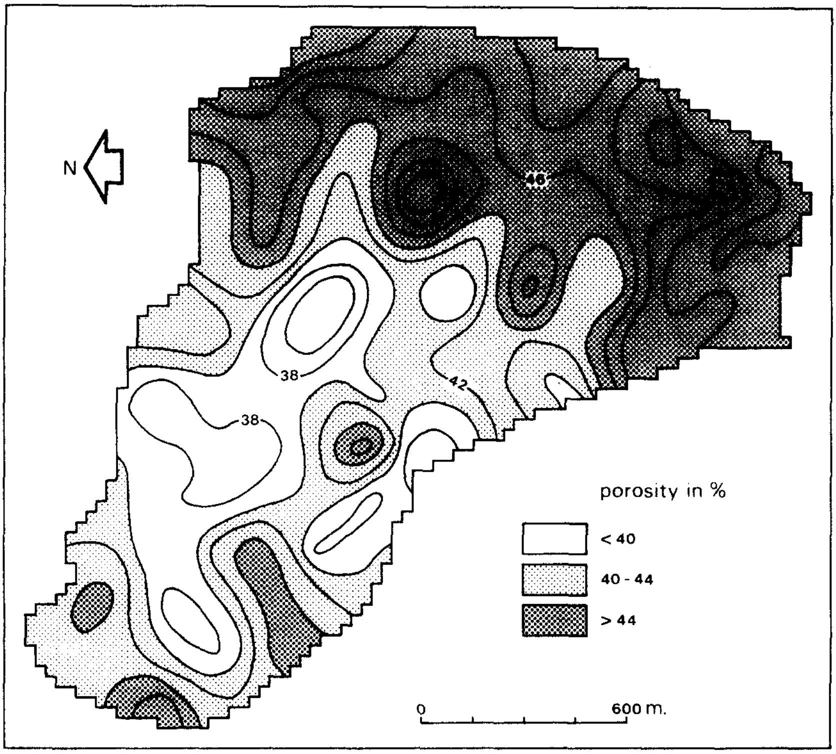

Fig. 9: *Kriging map of porosity.*

for several sampling distances. To this end three maps were constructed from each of the kriging variance maps. The first map indicates the kriging variance over a distance of 30 meter (1 × 1 gridcell). So actually, this map just shows the location of all sampled gridcells with their kriging variances. A second map was constructed by selecting the kriging variances within a window of 3 × 3 gridcells around the sampled cells, thereby simulating a sampling distance of 90 meter. The third map uses the factual sampling distance of 150 meter by doing the same selection with a 5 × 5 gridcell window. By using the environments of sample locations only, edge effects and the effects of missing values are automatically eliminated in all maps. The transformed kriging variance maps were then divided by the land unit-based maps and mean values of the ratio maps were calculated. The map manipulations were performed with the PC-GEOSTAT package (BURROUGH & VAN KEULEN 1986).

The outcome of the above procedure is presented in fig.11. The graphs clearly show that the application of kriging already provides a considerably higher reliability at the actual 150 meter sampling distance. The results are even better at smaller distances. One should however consider that the sampling effort is proportional to the inverse square of the sampling distance and that the kriging error of a variable is the square root of

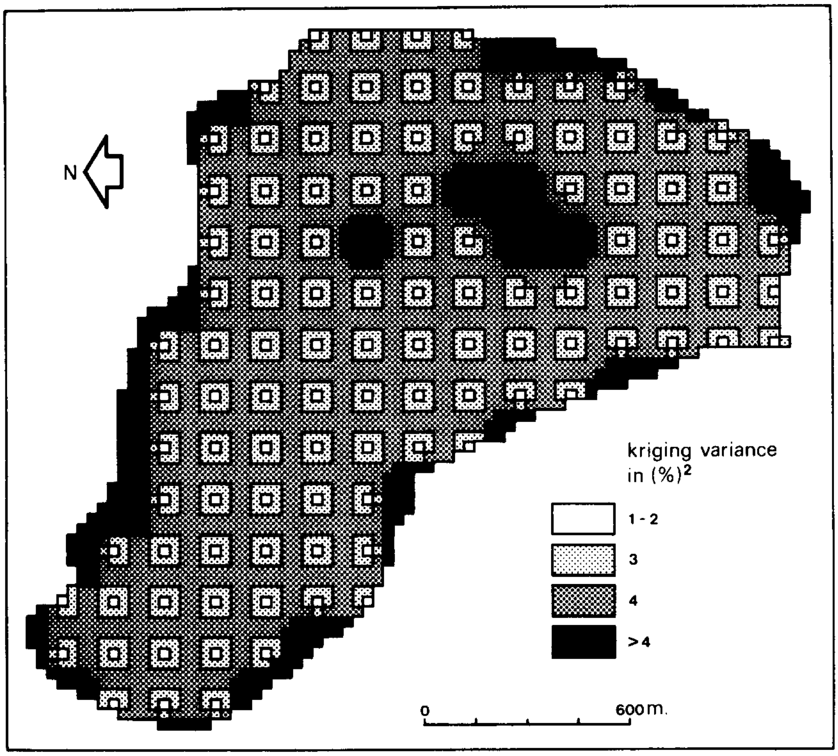

Fig. 10: *Kriging variance map belonging to the porosity map.*

the variance. This means for example that an increase in sampling effort by a factor 9 (= 50 meter sampling distance) for the kriging of the % clay will decrease the error with 60% when compared to the result at a 150 meter sampling distance. From this it was concluded that a much smaller sampling distance would not have made any sense in this particular study.

One may be surprised to find that the results for the saturated hydraulic conductivity are the poorest in view of the semivariograms shown in fig.6. But that graph does not indicate the absolute values of the semivariances, which are relatively high for the saturated hydraulic conductivity when compared to the other variables. The reliability of the graphs in fig.11 depends largely on the reliability of the semivariograms. Therefore a test of the sensitivity of the kriging variance to changes in the semivariogram model parameters was done for the saturated hydraulic conductivity. It was found that the probability of an average 25% increase of variances at the 150 meter sampling distance is less than 5%.

Finally the error in the USLE K-factor has been calculated by means of conventional error propagation theory, based on partial differentials of the K-factor formula (WISCHMEIER & SMITH 1978). It was found that the variance quotient at a sampling distance of 150 meter is 66%. This means that the average abso-

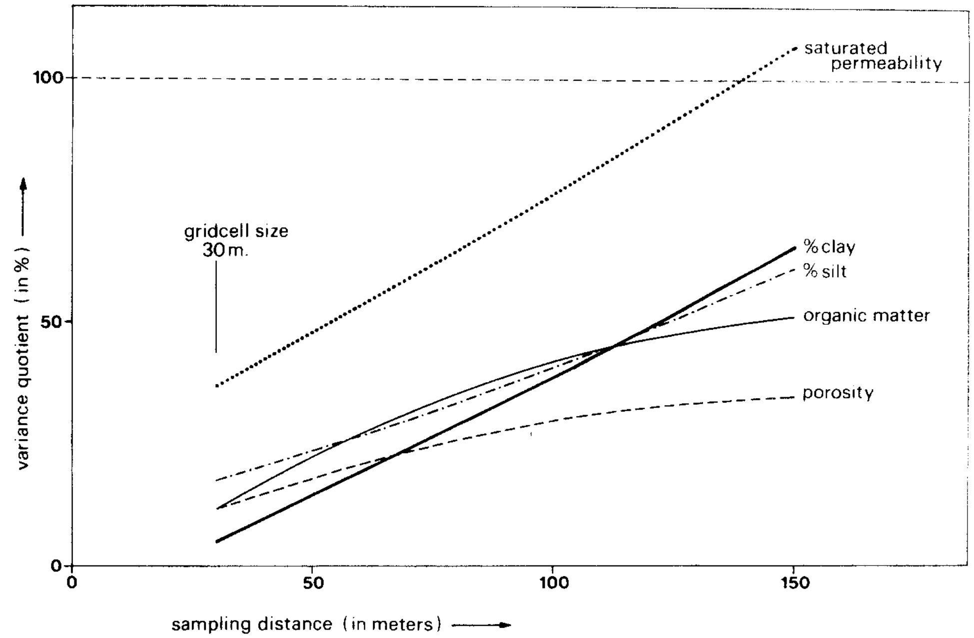

Fig. 11: *Quotients of kriging variance and within-unit variance for a range of sampling distances.*

lute error in the K-factor map is reduced by about 20% when kriging is applied to exactly the same number of observations that are used for the land unit-based extrapolation.

5 Discussion and conclusions

Kriging was found to be a good alternative for the conventional mapping of erodibility by means of land units. The disadvantages of kriging, namely the necessity of a reconnaissance sampling and the need to satisfy the theoretical requirements, are outweighed in this investigation by the following considerations:

- estimation variances are reduced considerably.
- the production of a detailed land unit map will take an effort comparable to that of a reconnaissance sampling.
- the mapping scale (1:5000) allowed kriging to be used within the framework of large geological and geomorphological structures, which are delineated easily by means of existing maps, so that the satisfaction of the theoretical requirements was practically guaranteed.

An important additional consideration in the evaluation of the results is that the achieved variance reduction may be negligible when compared with other sources of errors. Examples are systematic errors that are inherent in the applied erosion model or large errors of factors or variables other than the erodibility. A well known systematic error is the availability of soil particles for transport, which is usually neglected in ero-