ANALYSIS OF SOIL SURVEY DATA OF THE HUPSELSE BEEK

J. Simunek

RAPPORT 4

Mei 1990

Vakgroep Hydrologie, Bodemnatuurkunde en Hydraulica
Nieuwe Kanaal 11, 6709 PA Wageningen

Ing. Jiri Simunek
Research Institute for Improvement of Agricultural Soils, Zabrovreská 250, 25580 Praha 5 - Zbraslav, Czechoslovakia.
This report is the result of the work that has been carried out during a four month-fellowship payed by the International Agricultural Centre (IAC) Wageningen.

ACKNOWLEDGEMENTS

This is the report of the work done during my 4 month stay at the Department of Hydraulics and Catchment Hydrology in the Agricultural University at Wageningen, with the financial support of the International Agricultural Centre. It forms a small part of the studies carried out within the frame of the project Spatial Variability of Hydrological and Soil Physical Characteristics led by Ir. J.N.M. Stricker.

I wish to thank Han Stricker, Paul Torfs, Piet Warmerdam, Jacques Kole and other colleagues from the department for their understanding and kind help.

ABSTRACT

The report contains the analysis of the soil survey data of Hupselse Beek. Elementary statistics, geostatistical analysis of the spatial variability and the methods of multivariate analysis are used. The variables describing the geometry of the soil profile, the textural characteristics, the root zone, the organic matter content etc. are investigated. For each variable its statistical moments and semivariograms are calculated. Cluster analysis, principal component analysis and factor analysis are used to investigate interrelation between variables and observations and to reveal the structure in multivariate data.

TABLE OF CONTENTS

1.	Introduction	1
2.	Data collection and preparation	2
3.	Elementary statistics	4
4.	Spatial analysis	14
5.	Multivariate analysis	23
	5.1. Cluster analysis	24
	5.2. Principal component analysis	31
	5.3. Factor analysis	37
6.	Conclusions	42
	References	44

1. Introduction

The Hupselse Beek watershed has long been an experimental area for different hydrological, pedological and geological studies. The survey area covers 650 ha and is situated in the eastern part of the Netherlands near Groenlo. A detailed description of the geological structure can be found in Burrough et al. (1983). A great amount of data covering both soil morfological properties (Burrough et al.,1983; Wösten et al.,1983) and soil physical properties (Wösten et al., 1983; Brom, 1983; Boolting, 1984) was collected at this area.

The measurement of the soil physical data is relatively time consuming and costly. It might be possible to partially circumvent this problem by estimating the soil physical characteristics from easy to obtain parameters, i.e. from soil morfological and textural characteristics (Wösten and van Genuchten, 1988; Haverkamp and Parlange, 1986; Bloemen, 1980). the different techniques can be used for this purpose, i.e. cokriging (Vauclin et al.,1983), kriging with external drift or with a guess field (Ahmed and De Marsily, 1987). The spatial dependency and the variability within and between both groups of data have to be studied before proceeding to the prediction of the soil physical characteristics from easy to obtain soil parameters. The techniques which describe the spatial variation of the soil physical variables were recently studied by Hopmans (1986,1987) and Hopmans and Stricker (1987, 1988). The main objective of this study is to reveal the spatial dependence of the soil morfological, textural and other data and to find interrelations among the soil variables.

2. Data collection and preparation

Two data sets are available for the Hupselse Beek hydrological catchment. The first sampling scheme with the aim to determine optimum survey scales was reported by Burrough et al. (1983). The technique of nested sampling was used to collect the soil survey data from 64 sampling points over an area of 1500*1500 m.

On the basis of this study the optimum sampling density approximately 2 borings per ha was used for the second sampling scheme. A total of 1064 borings were made in a 650 ha study area. Forty profile characteristics for each boring containing informations on the root zone, the groundwater and horizon characteristics were determined (Wösten et al.,1983). This data set contains both quantitative and qualitative variables and different classification codes.

Most soil profiles can be characterized by the main four horizons - the plough A horizon, a B and C sandy or loamy sand horizon and a D horizon consisting of a sandy or silty clay. Wösten et al. (1985) showed that the B and C horizons of all soil types could be combined. This taken into account all soil profiles can be simplified to three textural layers - the A, BC and D horizons. If two or more horizons were combined then the resulting quantitative variable was taken with the value

$$X = \frac{\sum X_i d_i}{\sum_i d_i} , \qquad (1)$$

where d_i is the thickness of the i-th horizon and X_i is the value of the considered variable in this horizon.

After the inspection of all available data sixteen variables (Tab. 1.) were selected for further analysis and the data from the first sampling scheme were included into the second sampling scheme in spite of the fact that not all variables were available. The main reason for including the first sampling scheme was that this scheme comprises also the sampling points

Number	Name of variable
1	Topographic height
2	Depth to a clay layer
3	Annual highest groundwater level
4	Annual lowest groundwater level
5	Rootable depth
6	Root zone observed
7	Thickness - A horizon
8	Organic matter content - A horizon
9	Clay content - A horizon
10	Median sand size fraction- A horizon
11	Thickness - BC horizon
12	Organic matter content - BC horizon
13	Clay content - BC horizon
14	Median sand size fraction - BC horizon
15	Degree of layering - D horizon
16	Resistance to sampling - D horizon

Tab. 1. List of variables

with smaller mutual distances than the second scheme. It can be very useful for spatial analysis.

The first sampling scheme was bored to a depth of 130 cm, the second scheme to a depth of 200 cm or to the upper surface of the boulder clay or the Miocene clay. The depth to a clay layer could not be sometimes properly estimated because the clay layer was not reached within the depth of boring. In such case the entering value was 130 or 200 cm (Burrough et al., 1983).

The resulting data set was created by comprising all sixteen variables together with spatial coordinates. If a particular variable was missing than its value was indicated by -1.

3. Elementary statistics

Fig. 1. shows the histograms of all variables. The histogram is the experimental curve of the frequency of occurrence of the different values of the variable. The variate values are on abscissa, frequency on the ordinate and contiguous bars represent the frequency of the classes. The last bar of the histograms indicates the number of observations with the value of the considered variable either within a given interval or higher. Most of the variables shows unimodal distribution, but some variables as those qualitative variables describing layering and resistance to sampling of the D horizon show bimodal distribution and thickness of the BC horizon even multimodal distribution.

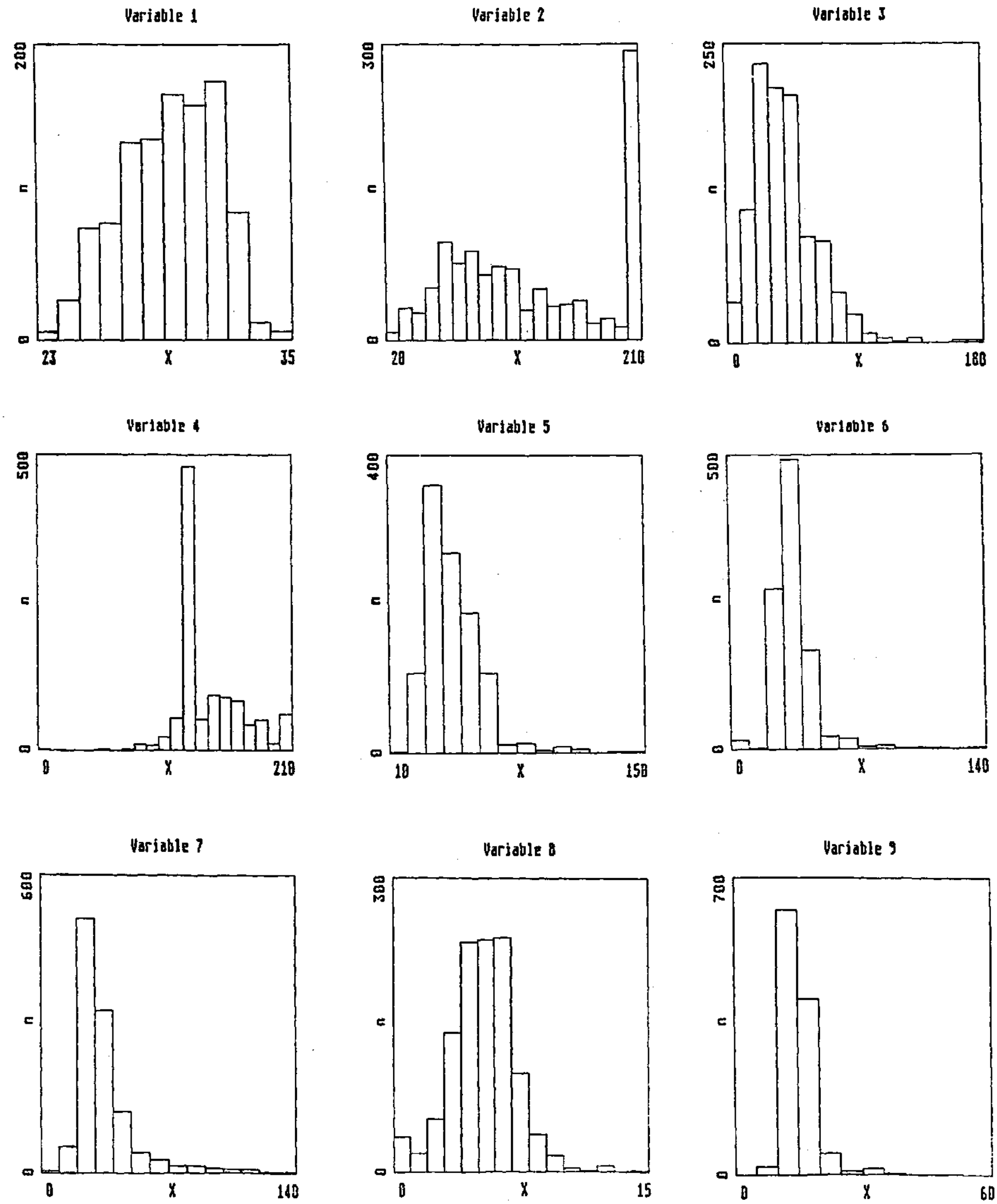

Fig.1. Histograms of the raw data.

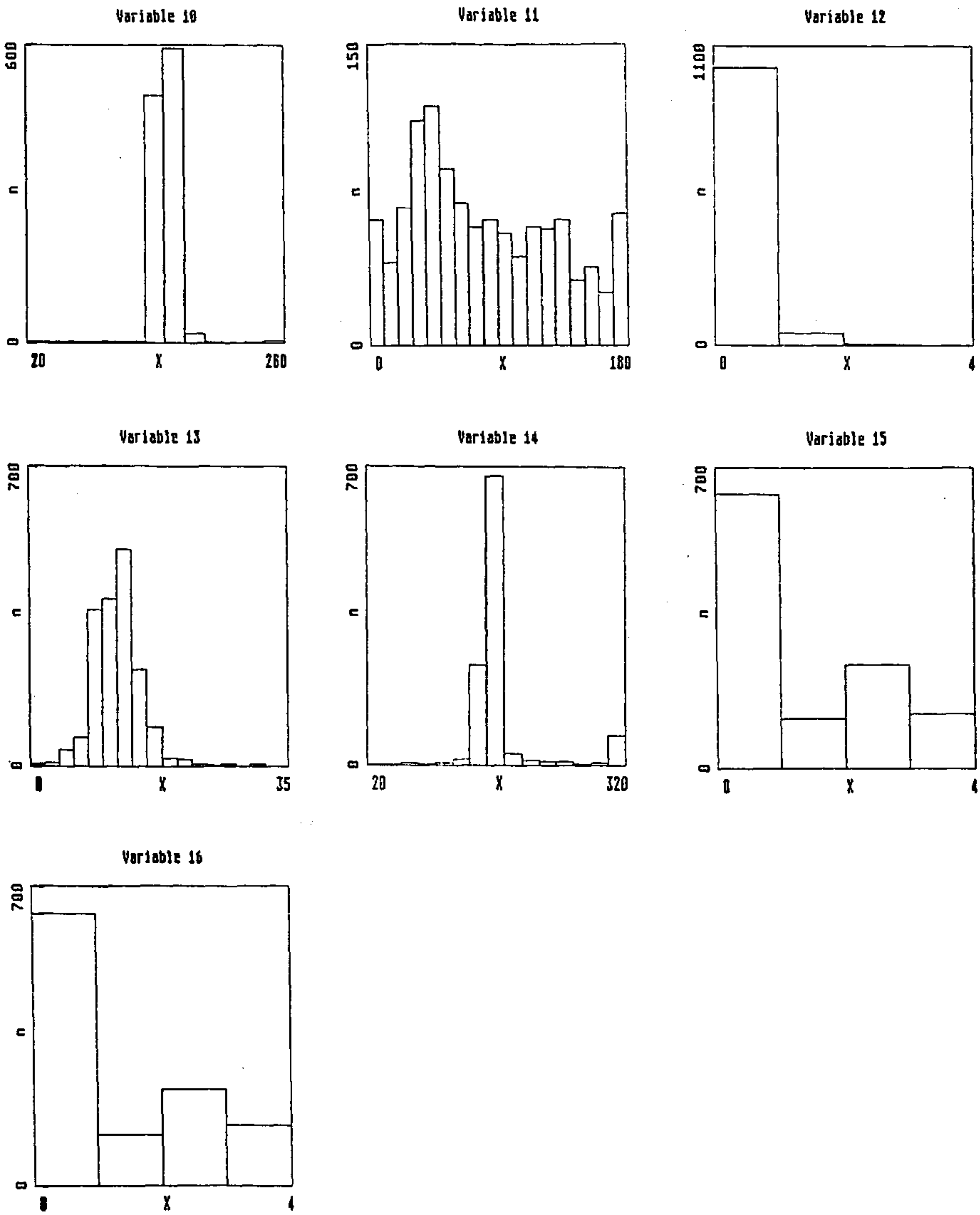

Fig.1. Histograms of the raw data.

There are also variables such as the depth to a clay layer, median sand size fraction of the BC horizon that show unimodal distribution but with many values falling into classes representing the tails of the distribution either on one side or on both sides.

The histogram gives the first informations about the position and spread of a set of values and about the type of distribution of these values. But for further analysis it is useful to summarize these informations by a few numbers that are related to the position, spread and shape of the distribution. Tab. 2. contains the arithmetical mean (AVE), the mean absolute deviation (ADEV) and the standard deviation (SDEV), the variance (VAR), the skewness (SKEW), the kurtosis (CURT) and the minimum (XMIN) and maximum (XMAX) value of measured variables. The great deviations and variance of the median sand size fraction of the BC horizon are caused by the presence of the coarse sand on several sampling sites. Apart from the topographic height all the distributions show the positive skewness, that is an asymmetrical tail extending out towards more positive x. The kurtosis measures the relative peakedness or flatness of a distribution. The depth to a clay layer and the thickness of the BC horizon have a high negative value of kurtosis , that is the distribution has a flat shape without a significant modus. Also the qualitative variables describing the D horizon have relatively flat shape. On the other side the median sand size fraction of the BC horizon has an extremely

I	AVE	ADEV	SDEV	VAR	SKEW	CURT	XMIN	XMAX
1	29.248	1.857	2.221	0.4933E+01	-0.286	-0.729	23.93	34.7
2	122.704	51.038	57.385	0.3293E+04	0.149	-1.398	20.00	200.0
3	37.521	16.147	22.394	0.5015E+03	1.940	7.905	0.00	180.0
4	136.864	22.181	27.148	0.7370E+03	0.609	0.803	9.00	210.0
5	41.244	10.450	14.157	0.2004E+03	1.848	6.534	15.00	140.0
6	31.665	6.521	10.899	0.1188E+03	2.882	20.263	0.00	130.0
7	31.598	9.857	15.405	0.2373E+03	2.683	9.638	0.00	130.0
8	4.831	1.432	1.882	0.3540E+01	0.066	1.381	0.00	14.0
9	14.603	2.056	3.381	0.1143E+02	3.912	34.960	6.00	60.0
10	161.844	9.905	70.756	0.5006E+04	18.829	378.961	25.00	1800.0
11	77.120	42.629	49.775	0.2478E+04	0.379	-0.941	0.00	185.0
12	0.130	0.186	0.295	0.8705E-01	4.354	34.188	0.00	4.0
13	11.691	2.247	3.116	0.9707E+01	0.880	4.714	0.00	30.0
14	221.503	109.572	251.613	0.6331E+05	4.719	22.589	20.20	1980.0
15	0.900	1.018	1.144	0.1308E+01	0.812	-0.766	0.00	4.0
16	0.924	1.038	1.171	0.1371E+01	0.810	-0.781	0.00	4.0

Tab. 2. Statistical characteristics

high positive kurtosis, that is the distribution shows the sharp peak with insignificant tails. None of the variables has the value of kurtosis near zero, what is the value for normal distribution.

All variables show the significant values of skewness and kurtosis, that is their distributions are significantly different from the normal distribution. By using the Kolmogorov-Smirnov test it is possible to evaluate this significance or the significance of the hypothesis that the variables are drawn from any other distribution. In the considered case the null hypothesis is that the data sets have normal, resp. log-normal distribution. The K-S statistic D is calculated as the maximum absolute difference between the data set's cumulative distribution function and the known cumulative distribution

I	Normal distribution		Log-normal distribution	
	D	Significance	D	Significance
1	0.177	0.290E-28	0.059	0.121E-02
2	0.171	0.284E-20	0.130	0.213E-11
3	0.168	0.228E-25	0.148	0.161E-19
4	0.347	0.000E+00	0.298	0.000E+00
5	0.147	0.130E-20	0.225	0.000E+00
6	0.245	0.000E+00	0.314	0.000E+00
7	0.284	0.000E+00	0.203	0.000E+00
8	0.168	0.436E-27	0.189	0.326E-33
9	0.180	0.523E-31	0.126	0.537E-15
10	0.405	0.000E+00	0.336	0.000E+00
11	0.181	0.147E-31	0.170	0.366E-26
12	0.434	0.000E+00	0.181	0.180E-08
13	0.074	0.373E-04	0.122	0.216E-12
14	0.489	0.000E+00	0.413	0.000E+00
15	0.433	0.000E+00	0.324	0.000E+00
16	0.430	0.000E+00	0.301	0.000E+00

Tab. 3. Kolmogorov-Smirnov test

function (normal, resp. log-normal). the significance level Q_{KS} was calculated according to Press et al. (1987):

$$Q_{KS}(\lambda) = 2 \sum_{j=1}^{\infty} (-)^{j-1} \exp(-2 j^2 \lambda^2), \qquad (2)$$

where

$$\lambda = D \sqrt{N} \qquad (3)$$

and N is the number of observations of a particular variable. Tab. 3. summarizes the results of the Kolmogorov-Smirnov test for all sixteen variables. The small value of significance shows that the cumulative distribution function is significantly different from the analysed cumulative distribution function. Only in three cases out of sixteen the null hypothesis that the data set has normal distribution was rejected on higher significance level than the hypothesis that it has log-normal distribution. It is evident for all sixteen variables that the null hypothesis that their distributions are normal or log-normal can be rejected on usually used significance levels (0.01 or 0.05).

During the sampling the whole catchment was divided into twelve sampling units. Six units (from now on IIa) were sampled by the surveyor A and the rest (IIb) by the surveyor B. The first sampling scheme (I), the area of which lays mostly in the area IIb, was made by the surveyor A. It is obvious that the question can be posed what the influence of a surveyor on the resulting data set is. Student's t-test for significantly different means, F-test for significantly different variances and Kolmogorov-Smirnov test for different distributions were used. In case of Student's t-test the two distributions were thought to have either the same variances (t) or significantly different variances (t*). The significance level for Kolmogorov-Smirnov test was calculated according to (2) but with

$$\lambda = D\sqrt{\frac{N_1 \ N_2}{N_1+N_2}} \quad , \qquad (4)$$

where N_1 is the number of data points in the first distribution, N_2 the number in the second distribution. The first two statistical moments for compared data sets are in Tab. 4. and the results from the statistical tests are in Tab. 5. The sampling schemes I and IIa made by the surveyor A are compared with the sampling scheme IIb made by the surveyor B. The depth to a clay layer between the sampling schemes I and IIb was not compared because the profiles in different schemes were bored to different depth. In several cases means, variances and distributions are significantly different. It is difficult, if it is possible, to say what the reason of this difference is; whether the reason is objective - the spatial variability of soil properties or subjective - the surveyor. In all considered cases the rootable depth, the root zone observed and the thickness of the A horizon was strongly underestimated by the surveyor A or overestimated by the surveyor B. It seems that the surveyor B did not take into account the presence of the course sand in the A and BC horizons as did the surveyor A. The differences in the qualitative variables describing the D horizon are also so large that it seems that the reason is not objective but subjective.

I	Sampling scheme					
	I		IIa		IIb	
	Mean	Variance	Mean	Variance	Mean	Variance
1	-----		28.548	0.410E+02	30.739	0.115E+02
2	69.756	0.477E+03	96.701	0.158E+04	95.766	0.174E+04
3	-----		41.121	0.428E+03	33.475	0.554E+03
4	-----		136.179	0.790E+03	137.633	0.677E+03
5	33.313	0.636E+02	31.607	0.129E+03	48.064	0.323E+03
6	-----		26.012	0.572E+02	34.162	0.254E+03
7	25.578	0.459E+02	30.915	0.130E+03	33.134	0.375E+03
8	5.437	0.212E+01	4.589	0.314E+01	5.024	0.403E+01
9	14.125	0.414E+01	14.710	0.509E+01	14.543	0.195E+02
10	153.125	0.266E+02	165.415	0.990E+04	158.955	0.131E+03
11	-----		87.123	0.262E+04	71.727	0.221E+04
12	-----		0.180	0.755E-01	0.089	0.106E+00
13	-----		11.760	0.101E+02	11.609	0.928E+01
14	224.365	0.876E+05	262.039	0.106E+06	173.086	0.569E+04
15	2.078	0.150E+01	0.417	0.735E+00	1.291	0.134E+01
16	1.953	0.198E+01	0.238	0.288E+00	1.563	0.144E+01

Tab. 4. First and second statistical moment

All other differences seems to be explainable by the variability of soil properties in the catchment.

I		I - IIb		IIa - IIb	
		Test	Significance	Test	Significance
2	t	-----		0.319E+00	0.750E+00
	t*			0.319E+00	0.750E+00
	F			0.110E+01	0.343E+00
	KS			0.183E+00	0.437E-05
3	t	-----		0.564E+01	0.219E-07
	t*			0.560E+01	0.281E-07
	F			0.129E+01	0.318E-02
	KS			0.342E+00	0.226E-26
4	t	-----		-0.872E+00	0.384E+00
	t*			-0.875E+00	0.382E+00
	F			0.117E+01	0.774E-01
	KS			0.417E+00	0.000E+00
5	t	0.648E+01	0.201E-09	-0.181E+02	0.000E+00
	t*	0.115E+02	0.784E-22	-0.176E+02	0.000E+00
	F	0.508E+01	0.162E-11	0.251E+01	0.371E-25
	KS	0.718E+00	0.733E-25	0.649E+00	0.000E+00
6	t	-----		-0.108E+02	0.478E-25
	t*			-0.104E+02	0.772E-23
	F			0.443E+01	0.000E+00
	KS			0.430E+00	0.000E+00
7	t	0.309E+01	0.208E-02	-0.231E+01	0.213E-01
	t*	0.624E+01	0.206E-08	-0.224E+01	0.253E-01
	F	0.819E+01	0.582E-17	0.289E+01	0.103E-32
	KS	0.420E+00	0.415E-08	0.385E+00	0.154E-33
8	t	-0.159E+01	0.112E+00	-0.375E+01	0.185E-03
	t*	-0.203E+01	0.446E-01	-0.373E+01	0.206E-03
	F	0.190E+01	0.217E-02	0.128E+01	0.439E-02
	KS	0.241E+00	0.276E-02	0.182E+00	0.486E-07

Tab. 5. Tests for the same means, variances and distributions.

I		I - IIb Test	I - IIb Significance	IIa - IIb Test	IIa - IIb Significance
9	t	0.746E+00	0.456E+00	0.791E+00	0.429E+00
	t*	0.130E+01	0.197E+00	0.764E+00	0.445E+00
	F	0.471E+01	0.106E-10	0.383E+01	0.000E+00
	KS	0.190E+00	0.337E-01	0.372E+00	0.322E-31
10	t	0.402E+01	0.667E-04	0.144E+01	0.150E+00
	t*	0.708E+01	0.433E-10	0.153E+01	0.128E+00
	F	0.493E+01	0.331E-11	0.755E+02	0.000E+00
	KS	0.796E+00	0.129E-30	0.339E+00	0.782E-26
11	t			0.509E+01	0.424E-06
	t*	-----		0.511E+01	0.373E-06
	F			0.118E+01	0.528E-01
	KS			0.159E+00	0.302E-05
12	t			0.480E+01	0.183E-05
	t*	-----		0.473E+01	0.261E-05
	F			0.141E+01	0.135E-03
	KS			0.877E+00	0.000E+00
13	t			0.759E+00	0.448E+00
	t*	-----		0.762E+00	0.446E+00
	F			0.109E+01	0.361E+00
	KS			0.241E+00	0.631E-12
14	t	-0.306E+01	0.232E-02	0.571E+01	0.152E-07
	t*	-0.137E+01	0.176E+00	0.616E+01	0.135E-08
	F	0.154E+02	0.000E+00	0.186E+02	0.000E+00
	KS	0.429E+00	0.283E-08	0.142E+00	0.913E-04
15	t	-0.509E+01	0.496E-06	-0.141E+02	0.000E+00
	t*	-0.487E+01	0.581E-05	-0.139E+02	0.000E+00
	F	0.112E+01	0.507E+00	0.182E+01	0.279E-11
	KS	0.264E+00	0.739E-03	0.415E+00	0.000E+00
16	t	-0.240E+01	0.168E-01	-0.237E+02	0.000E+00
	t*	-0.212E+01	0.372E-01	-0.227E+02	0.000E+00
	F	0.137E+01	0.724E-01	0.500E+01	0.000E+00
	KS	0.297E+00	0.905E-04	0.554E+00	0.000E+00

Tab. 5. Tests for the same means, variances and distributions.

4. Spatial analysis

The variables are characterized by their position in space. The difference between one variable measured at two different points is dependent on the distance h between these two points, the more closely spaced samples are more correlated to each other than samples farther apart. This dependence is described by semivariograms. The semivariograms, as a quantified summary of all the available structural information, are used to identify the structure of the spatial distribution of the variables considered. They are constructed in order to condense the main structural features of the regionalized phenomenon into an operational form (Journel and Huijbregts, 1978). The semivariograms are usually later used for another analysis, i.e. kriging, isaritmetic mapping, contouring or preparing variance maps etc. The semivariogram $\gamma(h)$ of a variable Z is defined by:

$$2\gamma(h) = Var[\ Z(x+h)-Z(x)\ ], \qquad (2)$$

where h is a separation vector. If the expected value E(Z) is constant in space the equation (2) can be rewritten as:

$$2\gamma(h) = E\ [\{Z(x+h)-Z(x)\}^2] \qquad (3)$$

The full description of the theory can be found in Journel and Huibregts (1978).

At the origine the semivariogram in general increases from the small value known as nugget effect. Beyond some distance called range the semivariogram may become stable at the value named sill. In such complex condition as are in the Hupselse Beek there is no reason to presume that the range and sill are independent from the direction, that is, that the semivariograms will be isotropic. By studying γ (h) in various directions, it can be possible to determine any possible anisotropy.

For all sixteen variables the four semivariograms were calculated for different directions and one semivariogram independent of the direction. The directional semivariograms were constructed in such a way that each data value was associated with every other value located within either a specified distance interval and a specified angle class. The angle class was given by the direction considered and the angle tolerance (22.5°) and the distance interval by the distance together with the distance tolerance (25m). The resulting empirical semivariograms were smoothed by the program (Press et al.,1987) that at first removed any linear trend then used a fast Fourier transform to low-pass filter the data and at the end reinserted the linear trend. The resulting semivariograms are shown in Fig.2. The solid line represents the semivariogram constructed without taking direction into account. The other lines represent the directional semivariograms; the long dashed line the semivariogram for the angle 0° (direction I), the dotted line for the angle 90° (direction II), the dash-dot line for the angle 45° (direction III) and dashed line for the angle 135° (direction IV). The angles defining directions are with respect to the west-east

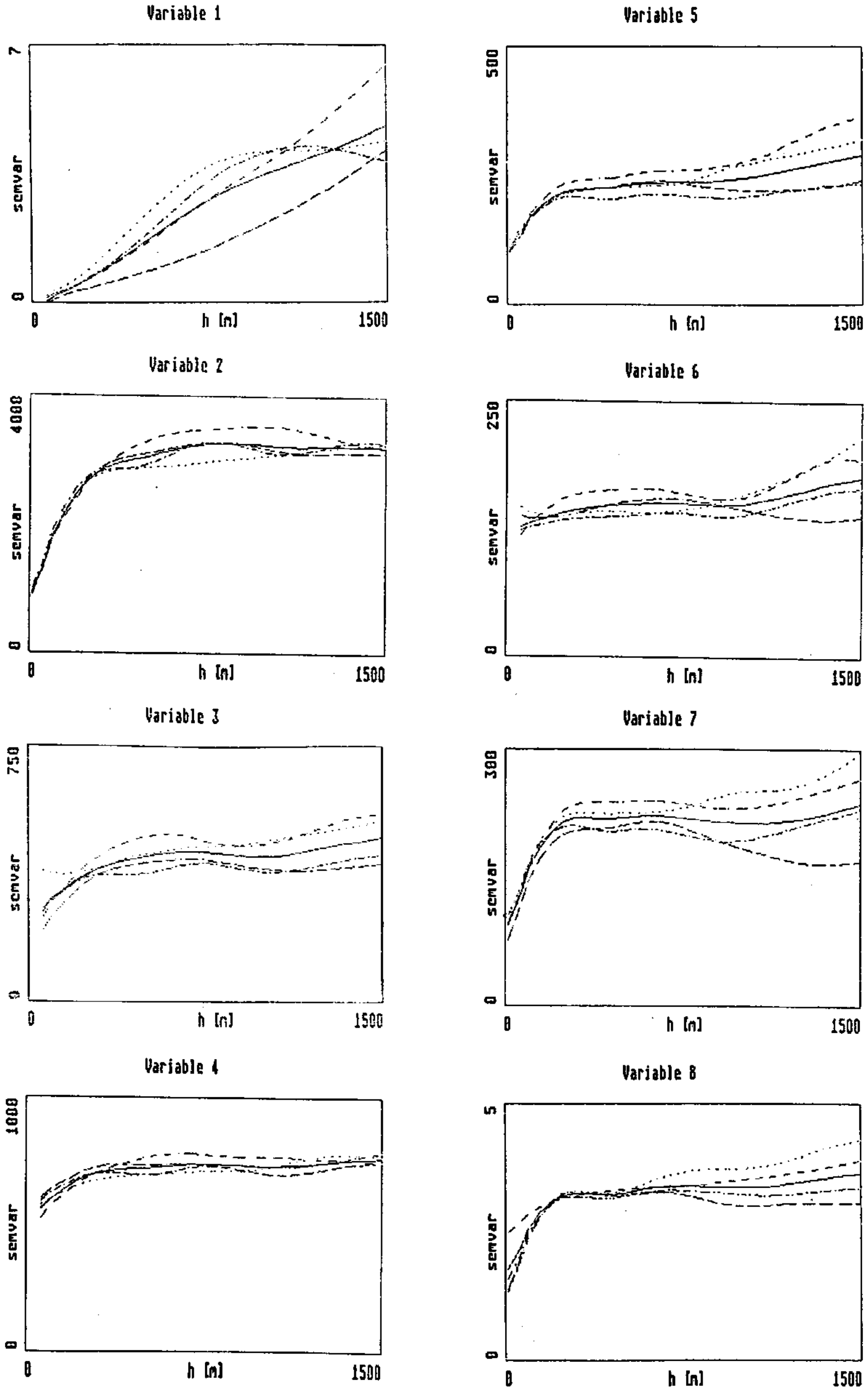

Fig.2. Directional and isotropic semi-variograms.

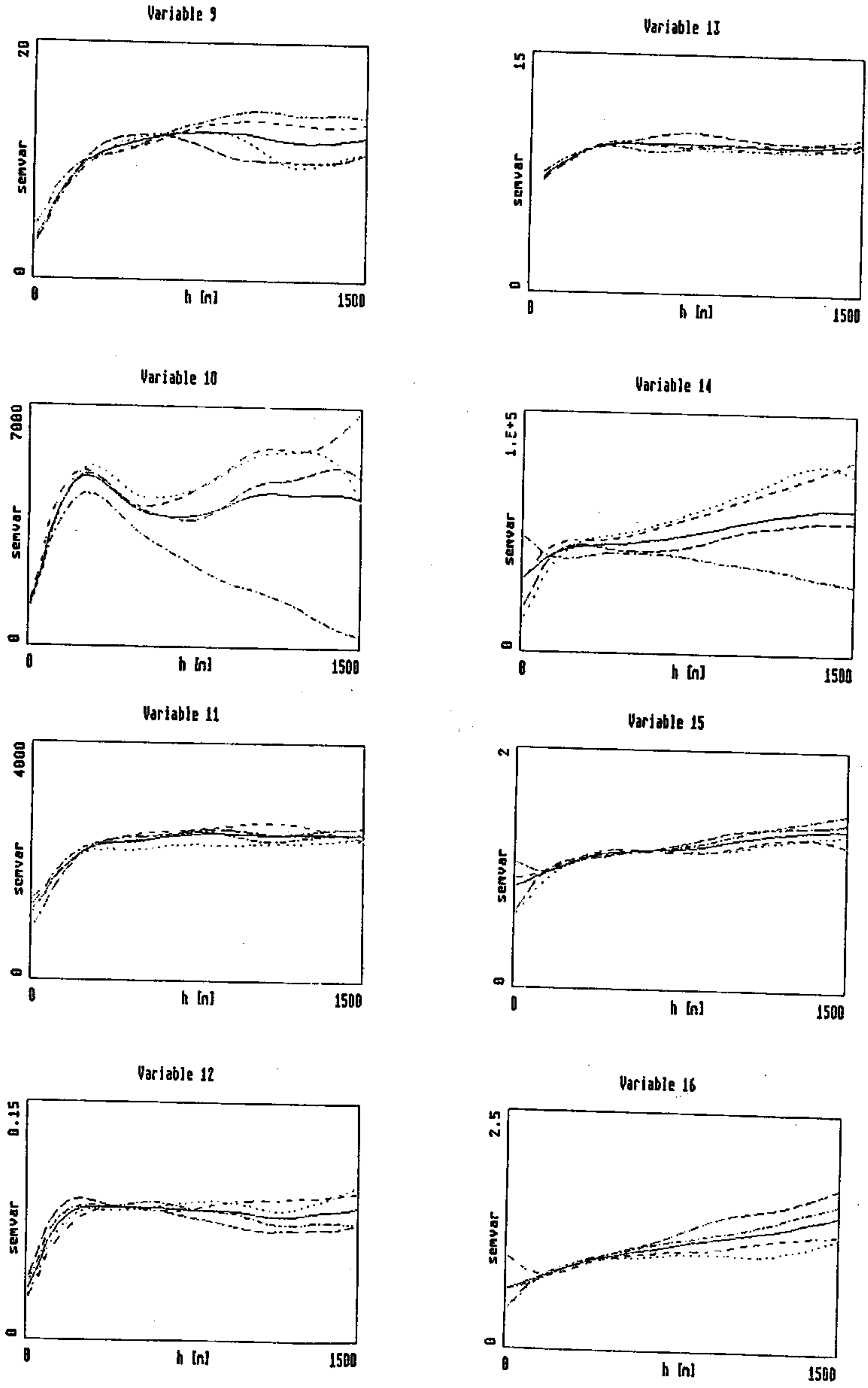

Fig.2. Directional and isotropic semi-variograms.

direction.

The practical rules suggested by Journel and Huigbregts (1978) are that an experimental semivariogram should only be considered for a sufficiently large number of data pairs (30, 50) and for small distances h in relation to the dimension of the field on which it has been computed. The former rule is easy to fulfill because only a few points at the origin are supported by less than one thousand data pairs, the latter one means that the distance of reliability is about 1500m. The fact that some points at the origin were estimated with much less data pairs than the rest of the semivariogram can explain some fluctuation and unreliability of this part.

Most variables show more or less isotropic behaviour at the origin with increasing anisotropy with increasing separation vector. Since only slight differences between the directional semivariograms were observed at the origin the different analytical models were fitted to the isotropic semivariograms of those variables that did not show pure nugget effect for the distance interval between 0 to 500 m. The models used are :

- the Gaussian model for the median sand size fraction of the A horizon:

$$\gamma(h) = C_1 \, [\, 1 - \exp(-h^2/a^2) \,] \qquad (7)$$

- the linear model for variables describing the D horizon :

$$\gamma(h) = C_0 + C_1 \, h \qquad (8)$$

- the exponential model for the topographic height :

$$\gamma(h) = C_o + C_1 h^a \qquad (9)$$

- the spherical model for the other variables :

$$\gamma(h) = C_o + C_1 \left| \frac{3h}{2a} - \frac{h^3}{2a^3} \right| \qquad h:[0,a] \qquad (10)$$

$$\gamma(h) = C_o + C_1 \qquad h \geq a \qquad (11)$$

The spherical model reaches a sill value , $C_o + C_1$, for a distance equal to the range a. The Gaussian model reaches a sill asymptotically and can be considered with $a'=a\sqrt{3}$. The linear and the exponential models have no sill. The linear model was fitted by a simple regression and the unknown coefficients for the other three models were optimized for the distances h<500m using the Levenberg-Marquardt non-linear least squares method. The results are shown in Fig. 3. and in Tab. 6.

I	Model	C_o	C_1	a
1	Exponential	0.0404	0.000124	1.5286
2	Spherical	543.86	2351.74	239.00
3	Spherical	0.00	388.43	160.96
4	Spherical	0.00	668.05	106.65
5	Spherical	5.06	205.69	75.88
7	Spherical	16.70	191.67	102.56
8	Spherical	0.00	3.14	59.14
9	Spherical	1.30	9.47	257.42
10	Gaussian	---	4997.12	74.02
11	Spherical	706.08	1469.17	174.40
12	Spherical	0.00	0.08	80.00
13	Spherical	2.10	6.94	175.30
14	Spherical	18746.59	24112.68	136.76
15	Linear	0.8472	0.0007	---
16	Linear	0.6232	0.0008	---

Tab. 6. Semivariograms.

The semivariogram for the topographic height shows the parabolic behaviour without significant nugget effect at the origin and the highest anisotropy of the directional semivariograms of all variables. This significant anisotropy displayed by the semivariograms has an obvious physical explanation. The Hupselse Beek is a valley with the lowest part on the west and the highest part on the south. It means that there is a significant trend in the direction IV. The direction I displays this trend too, but with smaller magnitude of semivariance. The semivariograms for the other two directions show a Gaussian behaviour with a practical range about 1000 m. The semivariograms for the depth to a clay layer and the thickness of the BC horizon show similarly the most regular behaviour without significant anisotropy. The semivariograms for the annual highest groundwater level and for the rootable depth show a transition phenomenon between the origin and a distance of about 1000 m. Beyond this distance there is an increase in the semivariogram values indicating the presence of the trend (quasi-stationarity). For both variables describing the fluctuation of the groundwater level the spherical model was fitted with prescribed zero nugget effect. The semivariograms for the root zone observed show the pure nugget effect. The variables 5, 7, 8 and 9 show very similar spatial behaviour; almost isotropic behaviour at the origin with increasing anisotropy together with increasing distance vector. The growth of the semivariograms for the median sand size fraction of the A horizon

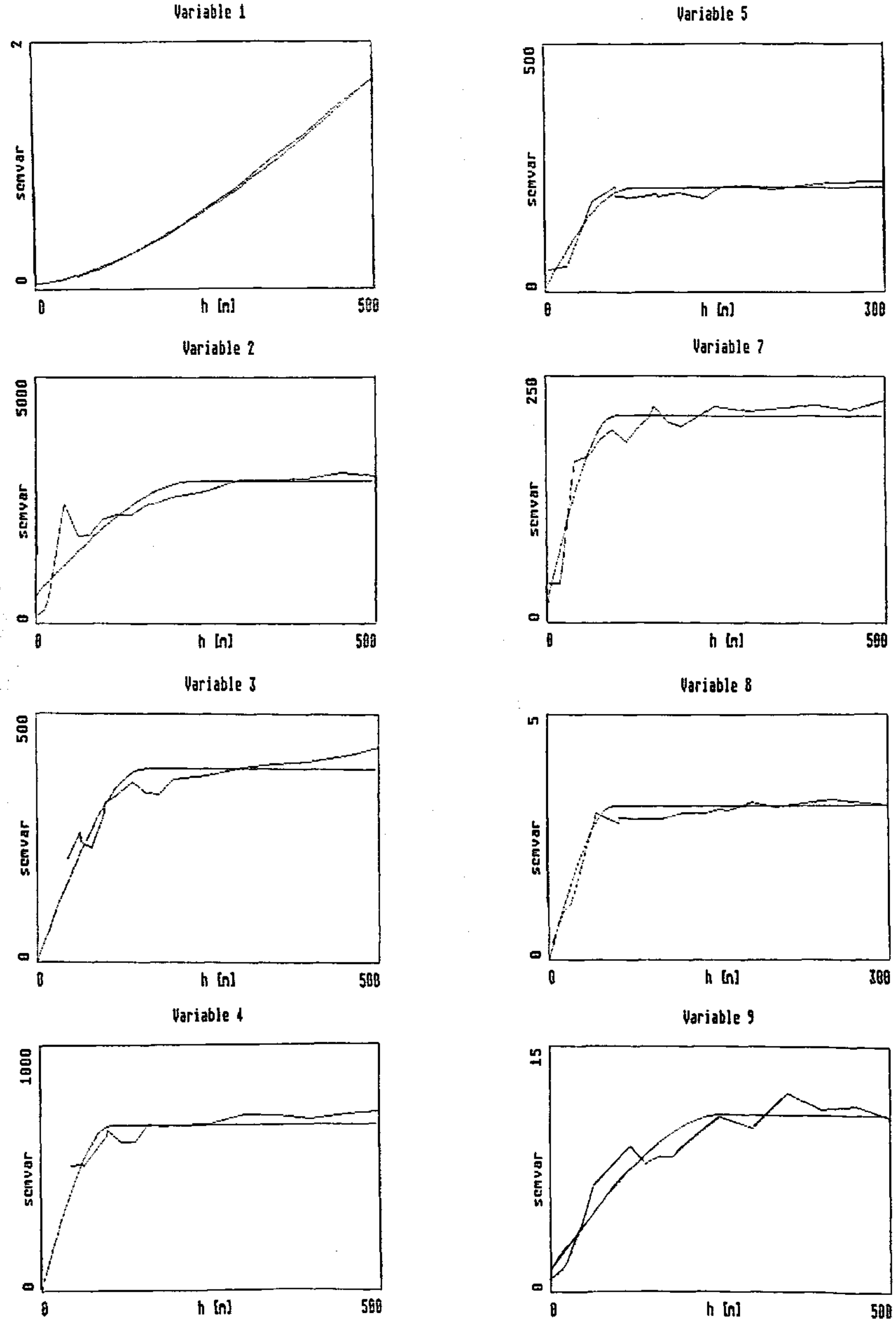

Fig.3. Models fitted to the empirical isotropic semi-variograms.

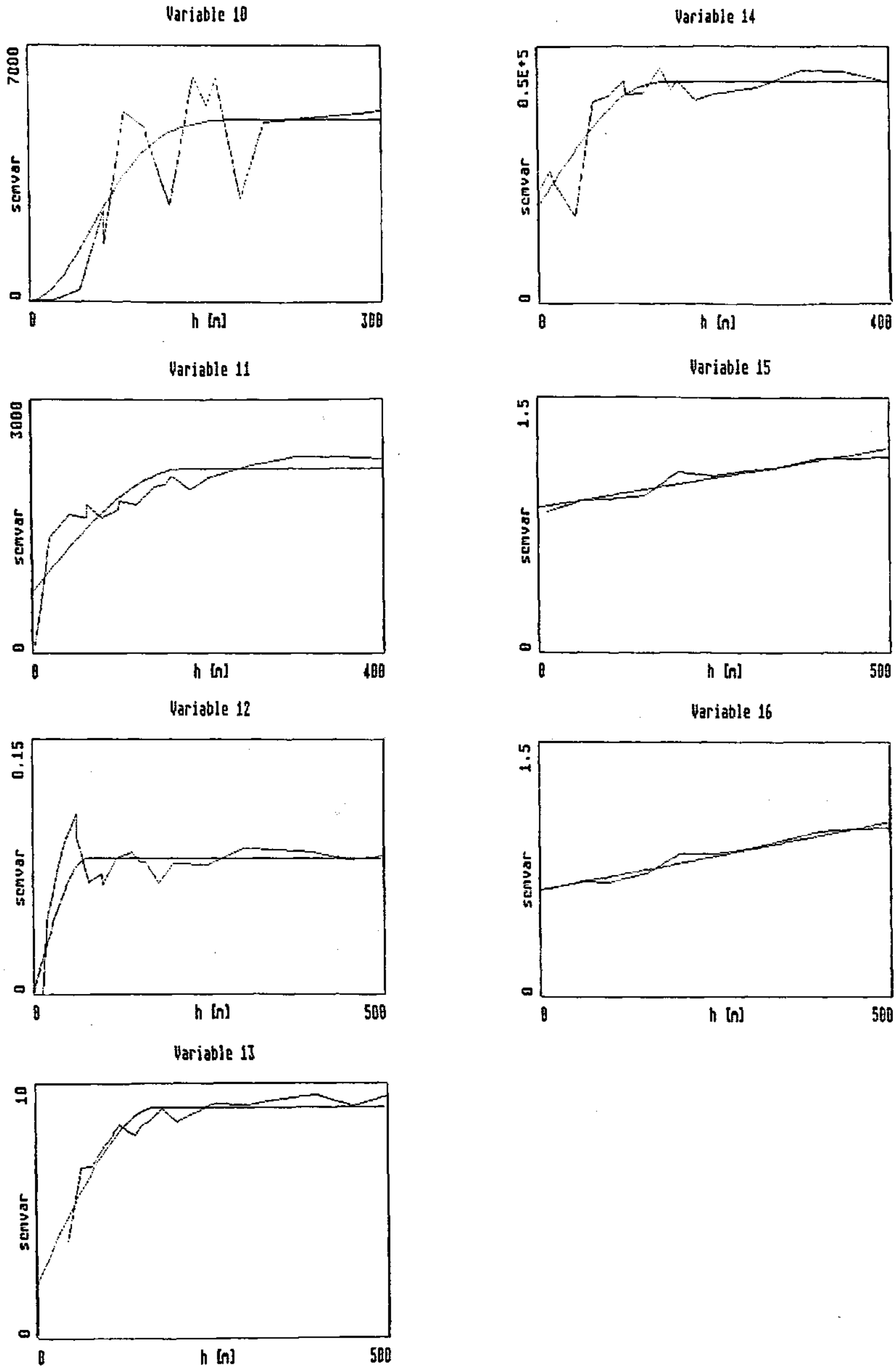

Fig.3. Models fitted to the empirical isotropic semi-variograms.

is not monotonic, the semivariograms show apparent hole effect. But as was shown in the previous chapter this variable was very probably influenced by a subjective factor (by a surveyor) and thus the semivariograms are not reliable especially for the distances for which more pairs are taken from the different sampling schemes. The semivariograms for the variables describing the BC horizon apart from the median sand size fraction show very regular and isotropic behaviour. The median sand size fraction of the BC horizon was influenced in the same way as was in the A horizon. The variables describing the D horizon show a high nugget effect and the linear trend in almost whole range of reliability.

Nearly all applied variables show spatial dependency within a distance of 75 m at least. Unfortunately this part of semivariograms is supported by the smallest number of the data pairs. Therefore to strengthen this conclusion more information should have been available from shorter distances.

5. Multivariate analysis

After the inspection of the recorded data and the selection of those variables which were available on the majority of points and which showed sufficient variation we obtained the matrix of N=1128 rows and M=16 columns. Every row characterize one bored profile (an observation) and every column one soil property or the geometric distance (a variable). If we wish to cope with this matrix and classify the soil taking into account all these variables we have to use multivariate methods. This methods allow us to manipulate with several variables simultaneously and to consider their changes, to show the relationship between properties and to reveal clustering of observations or variables. The variables are grouped into different sets and the interrelationships between and inside the sets are studied. Consequently a ranking of variables from completely independent to highly correlated can be made (Seyhan, 1981). Cluster analysis, principal component analysis and factor analysis were used to find the relationships and clustering between variables or observations.

5.1. Cluster analysis

The aim of cluster analysis is to investigate the interrelation between observations or between variables and to reveal the structure in multivariate data. The objective is to arrange a suit of observations into a meaningful order so that relationship between one observation and another may be deduced. Observations then can be placed into manageably few more or less homogeneous groups that can be treated uniformly for planning and management purposes.

At first some measure of similarity has to be computed. Many different coefficients of resemblance have been used (Seyhan, 1981), most often the correlation coefficient or a standardized m-space Euclidian distance. The Euclidian distance, resp. the correlation coefficient, seems to be more appropriate as a measure of similarity between observations, resp. variables. For the correlation coefficient the maximum similarity is represented by the value 1, and the maximum dissimilarity by -1; for the distance coefficient maximum similarity has the value zero. The greater the distance coefficient the greater the dissimilarity. The coefficients of resemblance are placed into the matrix of similarity. If the similarity between observations, resp. variables, is computed, then the similarity matrix has dimension N*N, resp. M*M. From that it can be seen that it becomes arduous with the increasing observation size. The next step is to examine the similarity matrix so objects with the highest mutual similarity are placed together. These groups of objects are associated with other groups which they most closely resemble etc. Several clustering techniques are possible. In this analysis the weighted pair-group method (Davis, 1973) was used. Cluster analysis was used to find the interrelations either between variables or observations.

The correlation matrix (Tab. 7.) as well as the dendrogram (Fig. 4.) constructed on the basis of this matrix reveal the interrelationship between the variables and their hierarchical structure. It is possible to find three main clusters. The first main cluster is formed by the topographical height, the organic matter content in the A horizon and by two subclusters; the first describing the fluctuation of the

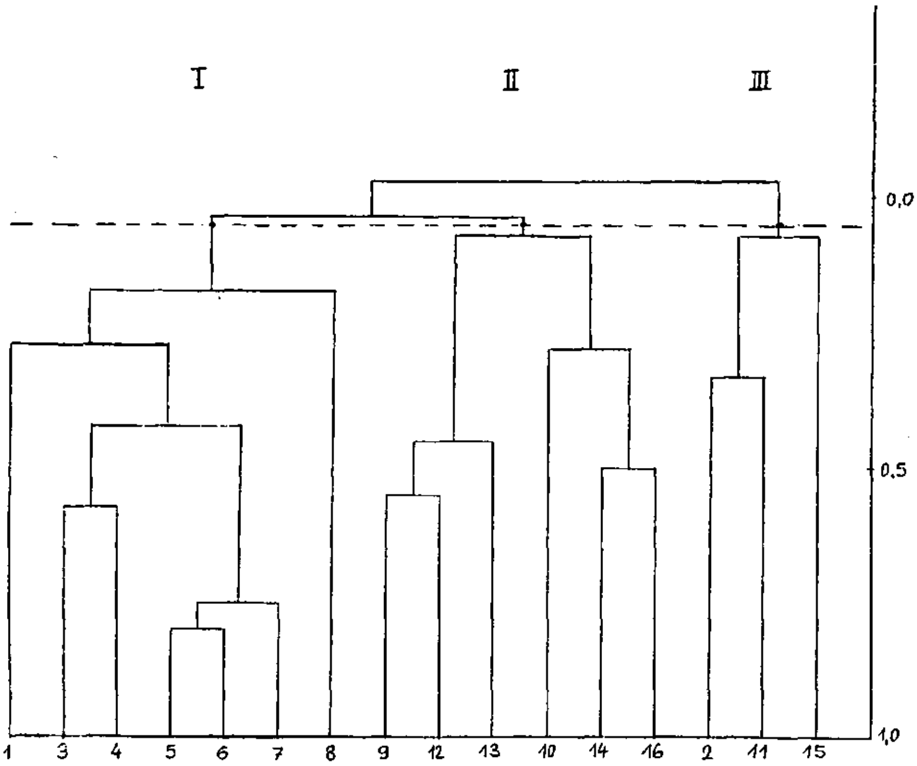

Fig.4. Dendrogram of correlation matrix in Tab.7.

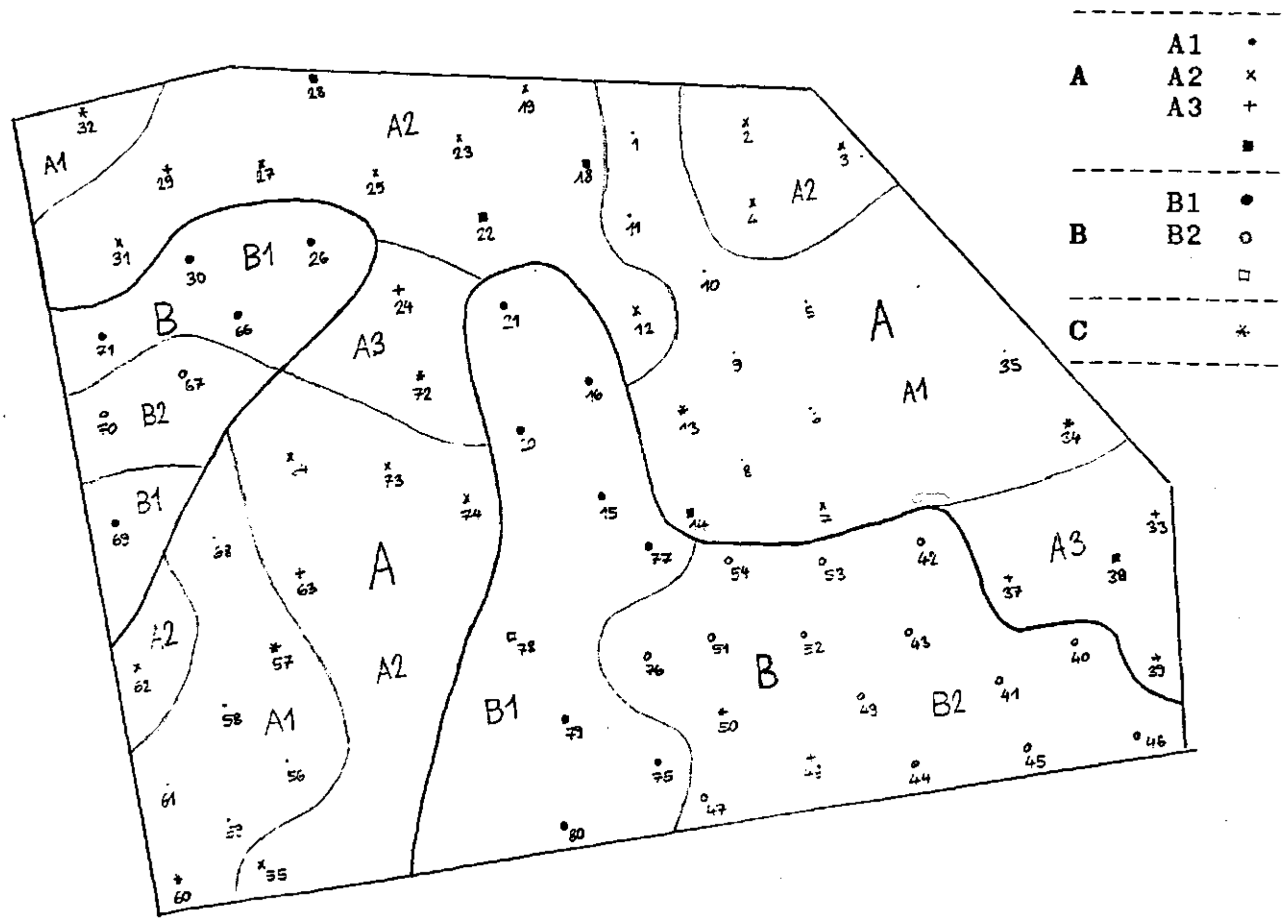

Fig.5. Experimantal area divided into two and five classes.

	1	2	3	4	5	6	7
1	1.0000	0.0840	0.3529	0.1780	0.3013	0.3350	0.2187
2	0.0840	1.0000	0.2239	0.1664	0.1419	0.1457	0.1222
3	0.3529	0.2239	1.0000	0.5713	0.6061	0.6679	0.4565
4	0.1780	0.1664	0.5713	1.0000	0.2896	0.3330	0.2673
5	0.3013	0.1419	0.6061	0.2896	1.0000	0.8076	0.7914
6	0.3350	0.1457	0.6679	0.3330	0.8076	1.0000	0.7229
7	0.2187	0.1222	0.4565	0.2673	0.7914	0.7229	1.0000
8	0.1479	−0.0271	0.1233	−0.0005	0.1947	0.2093	0.4159
9	−0.1438	0.0182	−0.0920	−0.0257	0.0018	−0.0360	−0.0201
10	0.2607	−0.0218	0.1436	0.0785	−0.0708	−0.1770	0.2922
11	−0.1835	0.4314	0.2943	0.3761	−0.1567	−0.3131	−0.4068
12	0.0897	0.0113	−0.1758	−0.1210	0.0590	0.4434	0.1147
13	−0.0715	−0.0281	−0.0210	0.1040	−0.0635	−0.0414	−0.0677
14	−0.2985	−0.0484	0.3760	−0.2039	−0.2206	−0.5096	−0.3394
15	0.0164	0.0994	0.1255	0.1179	−0.1416	−0.1560	−0.1251
16	−0.0356	−0.2176	−0.1050	0.0640	−0.0161	0.1181	0.0926

	8	9	10	11	12	13	14
1	0.1479	−0.1438	0.2607	−0.1835	0.0897	−0.0715	−0.2985
2	−0.0271	0.0182	−0.0218	0.4314	0.0113	−0.0281	−0.0484
3	0.1233	−0.0920	0.1436	0.2943	−0.1758	−0.0210	0.3760
4	−0.0005	−0.0257	0.0785	0.3761	−0.1210	0.1040	−0.2039
5	0.1947	0.0018	−0.0708	−0.1567	0.0590	−0.0635	−0.2206
6	0.2093	−0.0360	−0.1770	−0.3131	0.4434	−0.0414	−0.5096
7	0.4159	−0.0201	0.2922	−0.4068	0.1147	−0.0677	−0.3394
8	1.0000	0.2647	−0.0609	−0.1491	0.1828	0.0595	0.1100
9	0.2647	1.0000	0.1779	−0.0441	0.5526	0.3791	−0.1262
10	−0.0609	0.1779	1.0000	0.0128	−0.1000	0.2093	0.4052
11	−0.1491	−0.0441	0.0128	1.0000	−0.1845	0.0170	−0.0957
12	0.1828	0.5526	−0.1000	−0.1845	1.0000	0.5313	−0.3371
13	0.0595	0.3791	0.2093	0.0170	0.5313	1.0000	0.0616
14	0.1100	−0.1262	0.4052	−0.0957	−0.3371	0.0616	1.0000
15	−0.1040	−0.0447	−0.2522	−0.0471	0.2078	−0.0104	0.3072
16	0.0240	0.0711	0.1552	−0.1270	0.2921	0.0627	0.5044

	15	16
1	0.0164	−0.0356
2	0.0994	−0.2176
3	0.1255	−0.1050
4	0.1179	0.0640
5	−0.1416	−0.0161
6	−0.1560	0.1181
7	−0.1251	0.0926
8	−0.1040	0.0240
9	−0.0447	0.0711
10	−0.2522	0.1552
11	−0.0471	−0.1270
12	0.2078	0.2921
13	−0.0104	0.0627
14	0.3072	0.5044
15	1.0000	0.1464
16	0.1464	1.0000

Tab. 6. Correlation matrix.

groundwater level (the annual highest and lowest groundwater level) and the second the geometry of the A horizon (the rootable depth, the root zone observed, the depth of the A horizon). This cluster contains the variables which are the most important for the soil productivity. The second main cluster contains the soil properties describing the soil texture of the A and BC horizon. It is formed again by two subclusters. In the first subcluster there are the clay contents of both horizons with the organic matter content of the BC horizon, the second subcluster contains the median sand size fraction of both horizons together with the resistance to sampling of the D horizon. The third main cluster is formed by the subcluster consisting of the depth to a clay layer with the thickness of the BC horizon and the variable that showed the greatest dissimilarity to all other variables, the degree of layering of the D horizon.

To find the interrelation between observations not the whole catchment area was taken into account, but only part of it with the area about 67 ha (Fig. 5.). On this area there are 80 observations but only 77 of them have all measured variables. For three sampling sites the variables describing the root zone observed and the rootable depth are missing. As a measure of similarity the Euclidian distance coefficient was used

$$d_{ij} = \frac{\sum_k (X_{ik} - X_{jk})^2}{m} \qquad (12)$$

where m is the number of variables, X_{ik} denotes the k-th variable measured on the observation i and X_{jk} is the k-th variable measured on the observation j. The resulting dendrogram can be divided into four main clusters. The cluster D containing 3 observations with incomplete number of variables and the cluster C containing 8 observations without the sandy BC horizon are not shown on Fig. 6. The interrelationship between remaining two main clusters A and B and between the clusters at the lower level can be derived from Tab. 8. The main cause of dissimilarity between clusters is the geometry of the soil profile, i.e. the depth to a clay layer, the thickness of the BC horizon and the fluctuation of the groundwater level. The reason why some observations can not be included into the clusters at lower level can be found mainly in the soil properties. In the case of observations 14 and 18 it is the high organic matter content in the BC horizon, for observations 22, 58 and 78 the high median sand size fraction. The high clay content in the A horizon of the observation 28 and the great rootable depth of the observation 38 cause great distance from the other clusters.

Clus.		Depth to a clay layer		Annual highest GWL		Annual lowest GWH		Thickness of BC horizon		Rootable depth	
	A1		44.5		10.0		120.1		25.5		39.1
A	A2	76.1	99.0	20.4	24.7	123.2	121.3	45.3	61.0	45.7	47.3
	A3		104.3		31.6		125.7		57.1		52.9
	B1		144.3		40.0		144.3		105.7		46.4
B	B2	167.3	180.7	42.1	45.3	153.2	162.6	118.1	133.0	49.1	51.7

Tab. 8. Means of some variables in different clusters

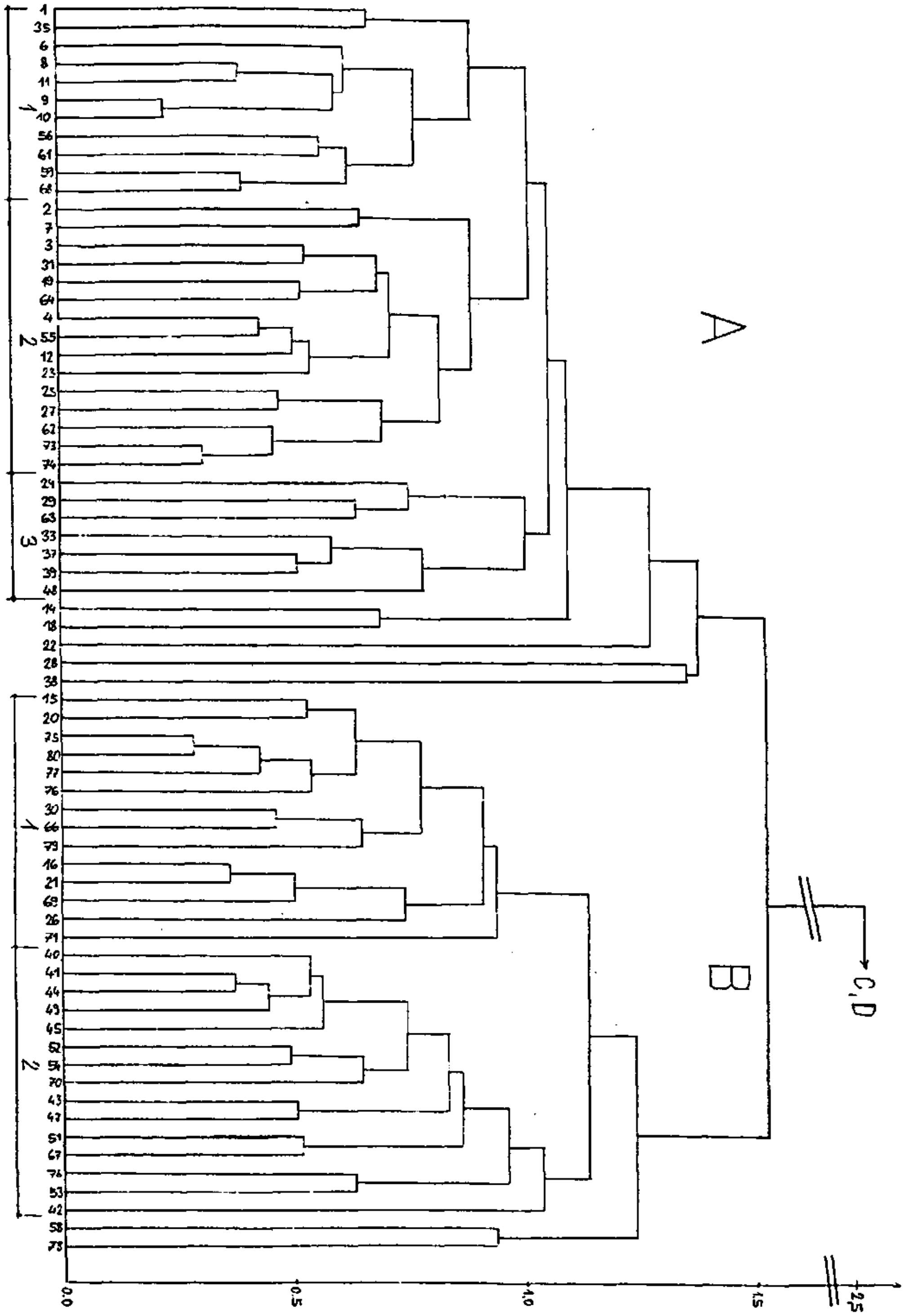

Fig.6. Dendrogram for soil profiles from area on Fig.5.

We can use the results of cluster analysis in delineating different parcels for purposes of land management. It was presumed that the C cluster with the missing BC horizon can be considered as a part of the cluster A1 with very small thickness of the BC horizon and the observations with missing data were neglected. If we take into account only two main clusters A and B we can divide the whole area into subareas and the position of boundaries is very easy to find. Further division can not consider every point if the individual parcels are to be reasonably large and compact and their boundary relatively smooth. Neglecting of some individual points may be reasonable, because there is not great dissimilarity between clusters at lower level. The resulting map is on Fig. 5. In the clustering technique the geometric location of each sample point was not taken into account.

5.2 Principal component analysis

The object of principal component analysis is to interpret the structure within the variance - covariance matrix of a multivariate data collection. The M original variables are linearly transformed to the same number of new variables - principal components, where each new principal component is a linear combination of the original variables. The principal components are arranged into such an order that each new component account for as much of the total variances as possible.

The proportion of the total variance accounted for by the longest principal axis is thus considerably larger then that represented by either of the original axes. It is hoped that the first few principal components will represent a large portion of the total variance. Obviously, we can study only few principal components that account for great amount of total variances and discard the others without loosing much of the variance in the data set and so reduced the dimensionality of the original data.

The first step is to standardize variables to make their variances equal. Otherwise the orientation of the principal axes is controlled largely by those variables with the largest variances. Because the variance - covariance matrix of standardized variables is just the correlation matrix, it was possible to take the similarity matrix from the cluster analysis (Tab. 7.) as a starting point of principal component analysis. To find the principal components is nothing else then to find the eigenvectors and the eigenvalues of a variance - covariance matrix. The eight largest eigenvalues, the percentage of the

Order	Eigenvalue	Percentage of known variance	Cumulative percentage
1	3.7127	23.2041	23.2041
2	2.3470	14.6688	37.8729
3	1.9547	12.2167	50.0895
4	1.7904	11.1902	61.2798
5	1.4250	8.9061	70.1859
6	1.0827	6.7669	76.9528
7	0.9761	6.1007	83.0535
8	0.8159	5.0993	88.1528

Tab. 9. Eigenvalues of the correlation matrix

Variable	Vectors 1	2	3	4
1	0.2538	-0.0456	0.0050	-0.0806
2	0.1085	-0.2212	-0.0408	0.3836
3	0.3519	-0.3248	0.3099	0.1065
4	0.2444	-0.2507	0.1323	0.3174
5	0.4533	-0.0457	-0.0139	-0.0903
6	0.4896	0.0838	-0.1002	-0.0331
7	0.4460	0.0667	0.0521	-0.1957
8	0.1761	0.1972	0.1125	-0.0749
9	0.0181	0.4052	0.1002	0.3400
10	0.0218	-0.0063	0.4382	-0.0384
11	-0.0876	-0.3533	0.0013	0.5212
12	0.1230	0.5301	-0.0096	0.2885
13	-0.0031	0.3194	0.2138	0.3852
14	-0.1946	-0.1216	0.6243	-0.1735
15	-0.0604	-0.0218	0.1940	0.1152
16	-0.0049	0.2022	0.4289	-0.1316

Tab. 10. The principal component matrix : columns - eigenvectors
rows - variables

total variance each eigenvalue accounts for and the cumulative percentage of the total variance are listed in Tab. 9. The percentage of the total variance individual eigenvalue accounts for is possible to calculate since the sum of all the eigenvalues equals the sum of the M original variances, in case of standardized variables it equals directly the number M. We can see that the first eigenvalue is much the largest, and the first three, out of sixteen, account for more then half the variance in the sample and eight eigenvalues account for almost ninety percent of the variance.

The part of the principal component matrix containing only the first four eigenvectors is in Tab. 10. The eigenvectors are in columns, the relative contributions of each variable to the principal components, called loadings, are in rows. This matrix contains the important information for the interpretation of the component axes. If the absolute value of the relative contribution is near 1, it means that the axis representing the original variable is closely aligned to the given component axis. On the other side if the loading is near zero the two axes are nearly at right angles and the contributions of this variable to the principal component is small. From Tab. 10. or better from a projection of the vectors on to a plane (Fig. 7.) one can try to give some meaning to the component axes. Graphs of vectors showing the contributions the variables listed in Tab. 1. make to the first, resp. second, two components are in Fig. 7a, resp Fig. 7b. Variables describing the geometry of the A horizon and the fluctuation of the groundwater level contributes strongly to the component 1. These variables are exactly those the first main cluster in cluster analysis consists of. The organic matter content of the BC horizon together with the clay content of both A and BC horizons contributes strongly to the component 2, but so does thickness of the BC horizon in the opposite sense. Again one can find similarity with the cluster analysis. The principal component 3 consists mainly of the median sand size fraction of both A and BC horizons together with the resistance to sampling of the D horizon and the component 4 consists of the thickness of the BC horizon and the depth to a clay layer. Also in the case of these two components it is possible to discover close agreement with the results of cluster analysis. It is not surprising since the starting point of both analysis, either principal component

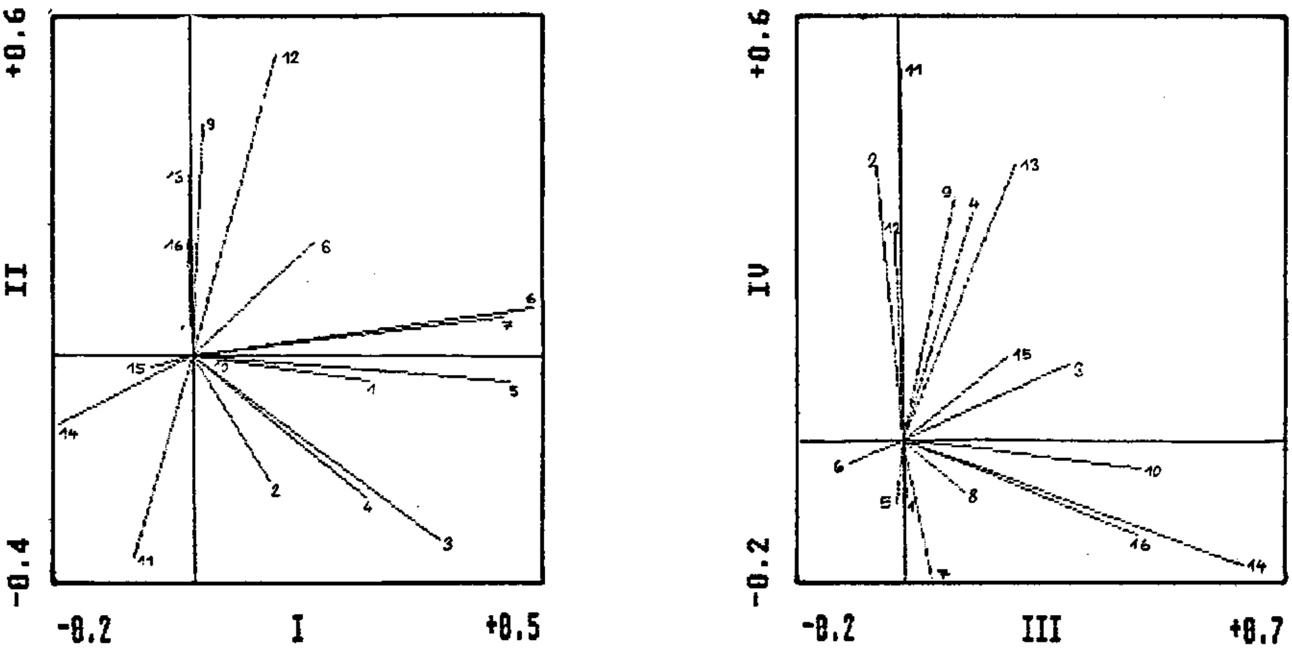

Fig.7. Plot of loadings on the first four principal components.

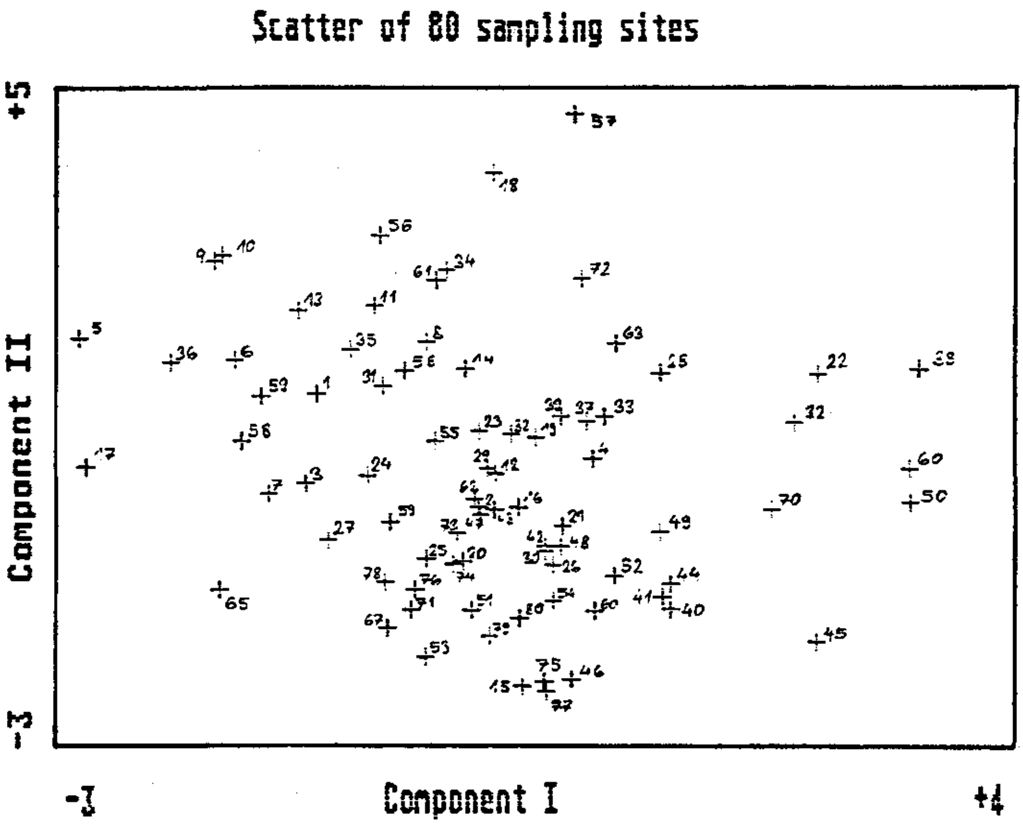

Fig.8. Scatter of 80 sampling sites plotted on the plane of the first two principal components.

analysis or cluster analysis, was the same - the similarity (correlation) matrix.

Multiplying the principal component matrix by the matrix containing the original data we get the principal component scores. For presenting the results again only the small part of the catchment the same like in the cluster analysis was selected. The projection of the principal component scores of 77 sampling sites on first two principal components is in Fig. 8. Since these two components account for almost 40 percent of the total variance, this projection gives the most informative single display of the relations in the whole space. To understand this scatter the meaning attached to the component axes must be taken into account. An examination of Fig. 8. shows that the observations with thick A horizon and with the deep root zone are placed far to the right, whereas the observations with the shallow A horizon and the shallow root zone are placed to the left. Along the second principal component the observations are sorted according to the clay content in both A and BC horizon, with the highest contents at the top and with the smallest at the bottom.

It seems to be slightly in the contradiction with the results of cluster analysis where the main reason of dissimilarity was found mainly in the depth to a clay layer, the thickness of the BC horizon and the fluctuating of the groundwater level that the first two properties contribute strongly only to the forth principal component. But by further inspection of the principal axis matrix and eigenvalues we can find that there is not great difference in the magnitude of the second, third and forth eigenvalue and that these properties contribute strongly also to the second component. So their influence is important in spite of the fact that they have not the decisive influence on neither of the first three components. Another rotation of the component axes would help if more exact interpretation is required.

5.3 Factor analysis

By applying the principal component analysis to the correlation matrix we got the M principal components which account for all of the original variance. To explain the structure of the original data we do not need all M components since the first few account for great amount of the total variance. But the position in the space of these first few components is strongly influenced by the presence of all the other axes. If we choose only the first few components and neglect all the others, it is possible to rotate them and to find a new position for them that is much easier to interpret.

Since we used as a starting point for principal component analysis the correlation matrix (Tab. 7.) and the eigenvectors (Tab. 9.) were computed in normalized form (they define a vector of unit length) we can easily convert the principal component vectors into factors by multiplying every element in the normalized eigenvector by the square root of the corresponding eigenvalue. The factor is then weighted proportionally to the square root of the amount of the total variance which it represents and consequently each factor loading is weighted proportionally to the square root of variance contributed by that variable to the factor. The part of the factor matrix containing only the first four factors is in Tab. 11.

For rotation of the first four factors the technique called Kaiser's varimax was used (Davis, 1973). This method rotate the factors so that each original variable is closely aligned to one of the new factor axes and at right angles to all others,if it is possible. There is then for each factor a few significantly high

Variable	Vectors			
	1	2	3	4
1	0.4891	-0.0698	0.0070	-0.1079
2	0.2090	-0.3388	-0.0571	0.5133
3	0.6780	-0.4976	0.4332	0.1425
4	0.4710	-0.3841	0.1850	0.4246
5	0.8735	-0.0700	-0.0194	-0.1208
6	0.9434	0.1284	-0.1401	-0.0443
7	0.8593	0.1021	0.0729	-0.2619
8	0.3393	0.3022	0.1573	-0.1002
9	0.0349	0.6207	0.1401	0.4550
10	0.0420	-0.0096	0.6126	-0.0514
11	-0.1687	-0.5413	0.0018	0.6975
12	0.2370	0.8121	-0.0134	0.3860
13	-0.0059	0.4893	0.2990	0.5155
14	-0.3749	-0.1863	0.8728	-0.2321
15	-0.1164	-0.0334	0.2713	0.1542
16	-0.0095	0.3098	0.5996	-0.1761

Tab. 11. The factor matrix : columns – factors
rows – variables

Variable	Vectors			
	1	2	3	4
1	0.4975	-0.0688	-0.0262	0.0534
2	0.0717	0.0390	-0.1216	0.6354
3	0.6423	-0.1941	0.3677	0.5743
4	0.3626	0.0080	0.1036	0.6645
5	0.8716	-0.0416	-0.0899	0.1159
6	0.9194	0.1531	-0.2361	0.0603
7	0.9020	0.0249	0.0103	-0.0912
8	0.3732	0.2373	0.1161	-0.1791
9	-0.0392	0.7807	0.0453	0.0006
10	0.0977	0.0439	0.6064	0.0247
11	-0.3387	-0.0466	-0.0363	0.8305
12	0.1672	0.8947	-0.1288	-0.1404
13	-0.0848	0.7265	0.2059	0.1310
14	-0.2523	-0.2105	0.9371	-0.0692
15	-0.1288	0.0850	0.2624	0.1392
16	0.0850	0.2207	0.5966	-0.2734

Tab. 12. The rotated factor matrix : columns - factors
rows - variables

contributions from the original variables and many insignificant contributions. The factor axes are now simpler to interpret in terms of the original variables.

The results after the rotation are shown in Tab. 12., Fig. 9. and Fig. 10. The relationship of the individuals to one another in the four dimensional space are exactly retained but their position to factor axes is changed. For example the annual highest and lowest groundwater levels that contribute to all first four principal components now contribute significantly only to the first and forth factor. The first factor is formed by the same variables as is the first principal component, it means by the thickness of the A horizon, the rootable depth and the root zone observe together with the annual highest and lowest

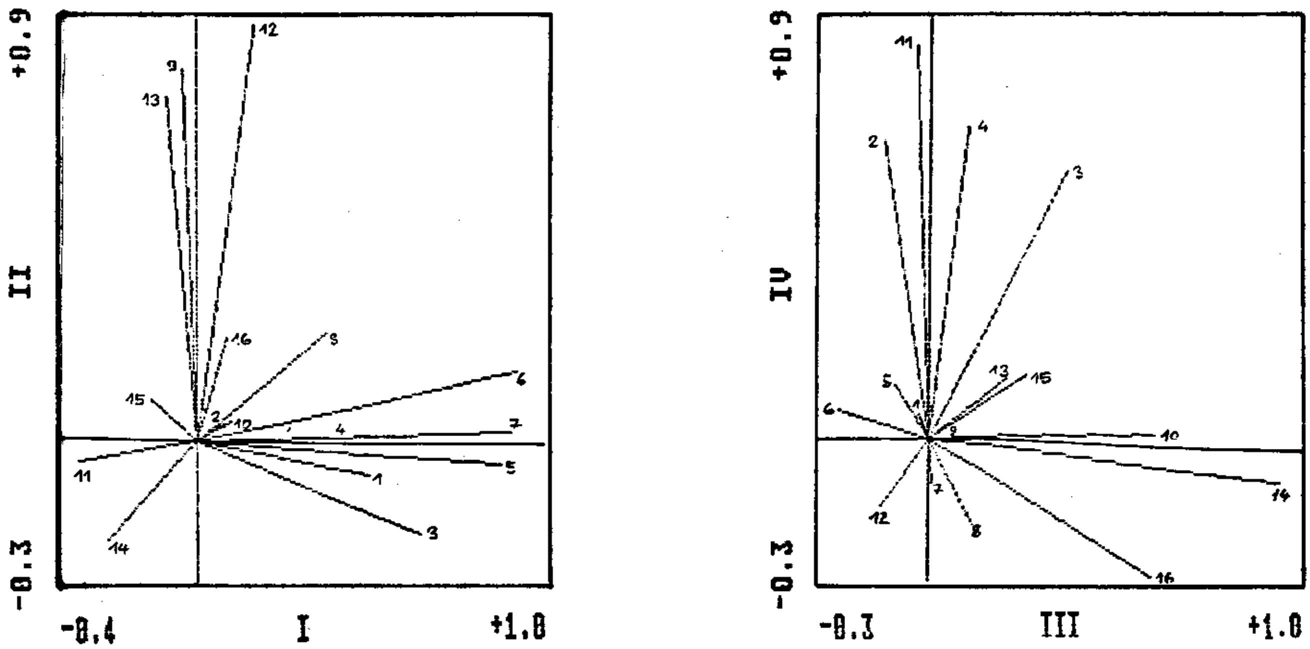

Fig.9. Plot of loadings on the first four rotated factors.

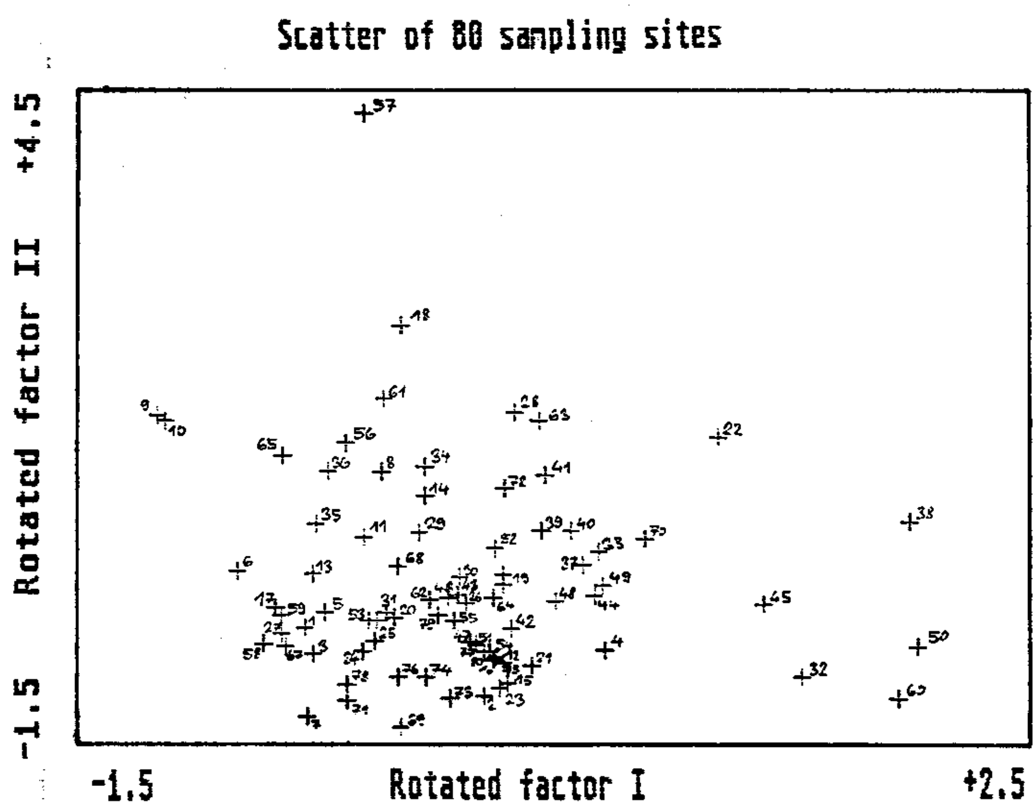

Fig.10. Scatter of 80 sampling sites plotted on the plane of the first two rotated factors.

groundwater levels and the organic matter content of the A horizon. In the interpretation of the second principal component there was great uncertainty since apart from the main contribution from the organic matter content of the BC horizon there were another seven variables with the significant contribution to this component. This uncertainty was removed after rotation. There are only three significant contributions to the second factor - the clay contents of both A and BC horizon together with the organic matter content of the BC horizon. It means that this factor reflects the texture of the soil profile. The meaning of the third factor is the same as the meaning of the third principal component, i.e. it describes the composition of the sand fraction. The main benefit of the rotation was achieved in case of the fourth factor which is now clearly defined by the annual highest and lowest groundwater levels and by the depth to a clay layer together with the thickness of the BC horizon. It means that this factor represents the geometric variables that are influenced by the depth to a clay layer. The projection of the factor scores of 77 sampling sites on first two factor is in Fig. 10. The interpretation of the horizontal axis is the same as in principal component analysis, i.e. the soil profiles with the deep plough horizon and the deep root zone are placed to the right, whereas the profiles with the shallow A horizon and the shallow root zone are placed to the left. Along the second factor the observations are sorted according to the texture of both A and BC horizons, with the heaviest texture profiles at the top and the lightest at the bottom. It is worth noticing how the soil profiles are placed into individual quadrants. More then one third of them falls to the right of centre with the negative value of the second factor, the rest is placed almost equally in the remaining quadrants. In the interpretation of factors stated former there are more light-textured profiles with the deeper plough horizon and the deeper root zone then with the shallow A horizon. If the soil is light it is likely to have the deep plough horizon with the deep root zone, whereas heavier textured soil does not show any such dependence.

6. Conclusions

The main object of this study was to investigate the spatial variability of the soil survey properties and to find interrelations among soil variables. The variability of the sixteen variables describing the geometry of the soil profile, the textural characteristics, the organic matter contents etc. were studied in chapter 3 by classical statistics (statistical moments, the law of distribution), in chapter 4 by geostatistics (the semivariograms) and in chapter 5 by multivariate analysis (cluster analysis, principal component analysis and factor analysis).

In chapter 3 it was concluded that none of all sixteen variables follows either normal or log-normal law of distribution and that some variables as the rootable depth, the root zone observed, the thickness of the A horizon and the median sand size fraction of the A and BC horizon were biased by the subjective factor - a surveyor.

The spatial analysis reveal that apart from the qualitative variables describing the D horizon and the root zone observed all applied variables show the spatial dependency within a distance of 60 m at least. To strengthen this conclusion more information should be available for the shorter distances.

The methods of multivariate analysis show that applied soil survey variables can be divided into four groups. The first group comprises the variables closely related to soil productivity of the soil profile. The second group reflects the texture of the soil profile and the third one the composition of the sand fraction. The second and third group contains the data that can be expected to be most relevant for soil physical properties. In the fourth group there are variables that are influenced by the underlying geological structure. The next step in the research should be to perform the methods of multivariate analysis on the data containing either the soil survey and the soil physical data in order to find out the easily measured soil variables, which may be used to estimate the soil physical characteristics.

References

Ahmed,S., G.De Marsily. 1987. Comparison of geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity. Water res.res.,vol.23., no.9., p. 1717-1737.

Bloemen,G.W.1980. Calculation of hydraulic conductivities of soils from texture and organic matter content. Z.Pflanzennahr. und Bodenkd., 143(5). p.581-615.

Booltink,P.I.,1985. Bodemfysische eigenschappen op zeer korte afstand Hupsel. Concept raport. Vakgroep Hydraulica en afvoerhydrologie, Landbouwuniversiteit Wageningen.

Brom,A.,1983. Bodenfysisch-, hydrologisch veldonderzoek op een 7-tal lokaties in Hupsel in 1982. Rapport no. 2, Studiegroep Hupselse Beek. Vakgroep Hydraulica en afvoerhydrologie, Landbouwuniversiteit Wageningen.

Burrough,P.A.,G.Oerlemans,G.Stoffelsen,J.V.Witter.1983. Reconnaissance soil studies to determine optimum survey scales and mapping legend for soil moisture research in the Hupselse Beek hydrological catchment; Oost Gelderland. Report no. 1 of study group "Soil Variability Hupselse Beek". Vakgroep Hydraulica en afvoerhydrologie, Landbouwuniversiteit Wageningen.

Davis,J. 1973. Statistical Analysis in Geology. Wiley.

Haverkamp,R.,J.Y.Parlange.1986. Predicting the water-retention curve from particle-size distribution: 1. Sandy soils without organic matter. Soil Sci., 142(6). p. 325-339.

Hopmans,J.W., 1986. Application and evaluation of techniques which describe the spatial variability of soil physical and hydrological variables. Publication 75. Dept. of Hydraulics and Catchment hydrology. Agricultural University Wageningen.

Hopmans,J.W., 1987. A comparison of various methods to scale soil hydraulic properties. Journal of hydrology,93.,p.241-256.

Hopmans,J.W., J.N.M.Stricker.,1987. Soil hydraulics properties In the study area Hupselse Beek as obtain from three different scales of observation: An overview. Report Dept. of Hydraulics and Catchment hydrology. Agricultural University Wageningen.

Hopmans,J.W.,J.N.M.Stricker.1988. Stochastic analysis of soil-water regime in a watershed. Manuscript submitted to J. of Hydrology.

Journel,A.G.,Ch.J.Huibregts.1978. Mining geostatistics, Academic Press; London.

Press,W.H., B.P.Flannery, S.A.Teukolsky, W.T.Vetterling.1987. Numerical Recepies - The Art of Scientific Computing. Cambridge University Press.

Seyhan,E.1981. Introduction to multivariate statistical analysis in hydrology. Free University. Institute of earth science de Boelelaan, Amsterdam.

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

Webster,R.1977. Quantitative and numerical methods in soil classification and survey. Clarendon Press; Oxford.

Wösten,J.H.M.,G.H.Stoffelsen,J.W.M.Teunissen,A.F.Holst,J.Bouma. 1983. Proefgebied Hupselse Beek; regionaal bodemkundig en bodemfysisch onderzoek. Rapport. no. 1706. Stichting voor Bodemkartering, Wageningen.

Wösten,J.H.M.,J.Bouma,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.Th.van Genuchten. 1988. Using texture and other soil properties to predict the unsaturated soil hydraulics functions.