

AN AREAL CLASSIFICATION BY MEANS OF CLUSTER ANALYSIS USING ABIOTIC LANDSCAPE DATA

G. T. A. HOF¹, R. REILING¹,
E. ZILVERBERG¹ & Th. W. M. LEVELT¹

ABSTRACT

Hof, G. T. A., R. Reiling, E. Zilverberg & Th. W. M. Levelt 1981 An areal classification by means of cluster analysis using abiotic landscape data. *In*: A. J. van Loon (ed.): Quaternary geology: a farewell to A. J. Wiggers – Geol. Mijnbouw 60: 427-432.

For an area NW of Roermond, province of Limburg, a landscape classification has been developed on the basis of four variables: loam content of the soil, water-table class, soil type and altitude. The presented variables are more or less associated with each other. A hierarchy of the variables, which implicitly means the existence of an overall theory of landscape genesis, has not been used.

The classification was performed by means of cluster analysis and reflects the association between the variables in the structure of the agglomerative legend system. For every desired number of groups the legend system gives the mapping units, which have a maximum homogeneity for the combinations of the variables. With the graphical facilities of the computer programme a map can be drawn at any desired grouping level. As an example, the grouping level 8 is discussed in the text.

INTRODUCTION

The present paper discusses an agglomerative method of physical geographical land classification based on abiotic land characteristics. This method has been tested in a small pilot area as a first step towards the ultimate goal of developing a classification system for the whole of The Netherlands.

The basic data for this classification have been collected from maps on scales 1:50,000 and 1:25,000. The pilot area is situated NW of Roermond in the province of Limburg (Fig. 1). Its size is 18.75 km².

DESCRIPTION OF THE PILOT AREA

Geology

An outline of the regional geology is given in figure 2. In the study area a subsoil of fluvial sediments of the river Meuse (Veghel and Kreftenheye Formations) is overlain by a 10 to 30 m thick series of aeolian coversands, slope deposits and materials of fluvio-periglacial or residual origin (Nuenen Group).

At the surface and directly below it aeolian deposits (Older

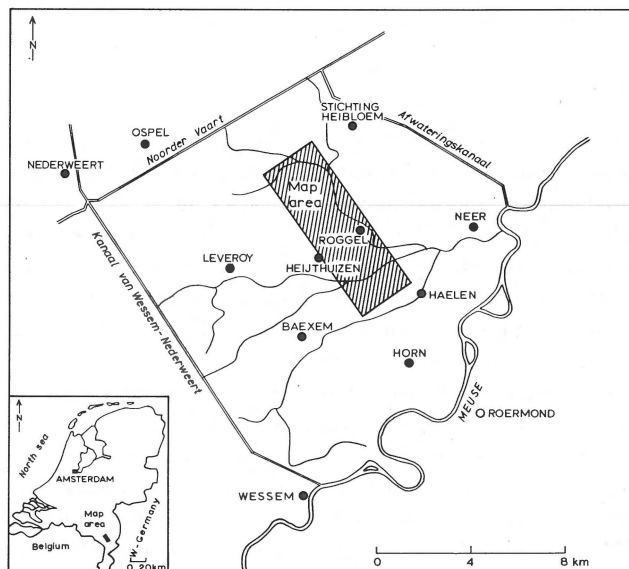


Fig. 1
Location of the study area.

¹ Institute of Earth Sciences, Free University, De Boelelaan 1085, 1081 HV AMSTERDAM, The Netherlands.

CHRONOSTRATIGRAPHY		LITHOSTRATIGRAPHY/LITHOLOGY			
		FLUVIAL DEPOSITS	DEPOSITS OF LOCAL ORIGIN *		
HOLOCENE		BETUWE FORMATION	SINGRAVEN FORMATION	KOOTWIJK FORMATION	
PLEISTOCENE	WEICHELIAN	KREPTENHEYE FORMATION	NUENEN GROUP	Younger Coversands II	
				Allerød Interst.	Usselo Bed or Soil
				Early Dryas St.	Younger Coversands I
				Bölling Interst.	Lower Loamy Bed
					Older Coversands II
	PLENIGLACIAL	Upper Pleniglacial		Beuningen Gravel Bed	
		Middle Pleniglacial		Older Coversands I	
		Lower Pleniglacial		Fluvis-periglacial and eolian deposits	
	EARLY GLACIAL			Peat and soil formation	
	EEMIAN				Fluvis-periglacial and eolian deposits
SAALIAN	LATE	VEGHEL FORMATION			
	MIDDLE				
	EARLY				

- * - eolian deposits
- fluvis-periglacial
- peat
- deposits of small streams

Fig. 2
Stratigraphy and lithology of the investigated area.

Coversand II and Younger Coversand I and II of the Nuenen Group) are the prevailing sediment type. The Older Coversand normally consists of slightly to very loamy fine sand. As a rule it can be distinguished from the Younger Coversand I by the better sorting and less pronounced stratification of this latter type. As compared with the generally level Older Coversand areas the Younger Coversands have a more irregular surface topography. Mineralogically the coversands are very poor since their content in unweatherable minerals is quite high.

Geomorphology

On the whole the study area has a level to gently undulating topography since coversands of the Nuenen Group are the predominant surface deposit. Stronger relief only occurs where wind blown landdunes (Kootwijk Formation) have formed as a result of medieval removal of vegetation. This process has been brought to a halt now by afforestation.

In the south the pilot area is dissected by a W-E running elongated depression which has been interpreted as an abandoned course of the river Meuse, probably of Pleistocene age (STIBOKA, 1972). On the basis of recent observations this interpretation appears to be open to discussion (REILING & ZILVERBERG, 1981). The depression which is now occupied by a small stream (the Heythuyzer beek) has been partly infilled by Holocene local fluvial deposits (Singraven Formation). The Roggelsche beek, another minor water course in the area, has a N-S orientation.

While the study area dips gently from + 31 m N.A.P.² in the NW to + 27 m N.A.P. in the SE incision of the two small

streams, the Heythuyzer beek and the Roggelsche beek reach down to + 22.5 m N.A.P.

Soils

Designation of genetic soil types discussed in this section is according to the Netherlands systems of soil classification (DE BAKKER & SCHELLING, 1966). The relevant taxonomic units of the system are concisely defined in table I.

Table I
The relevant soil orders and an explanation of the prefix xero and hydro.

soil order	description
podzol soil	mineral soil with a prominent podzol B horizon
earth soil	mineral soil with a well-developed humic A1 horizon
vague soil	mineral soil lacking clear soil development
prefix xero	without hydromorphic characteristics
prefix hydro	with hydromorphic characteristics

Hydropodzol soils and hydro-earth soils are formed in the low and level parts of the coversand areas where a high groundwater table prevailed before the introduction of artificial drainage. This implies that the hydromorphic features of these soils are inherited from environmental conditions which have now ceased to exist.

Thick earth soils occur around the Roggel township. These soils have man-made humic A1 horizons with a thickness of 50 cm or more, that is they have a Plaggen epidepon in the sense of the Soil Taxonomy (SOIL CONSERVATION SERVICE U.S. DEPARTMENT OF AGRICULTURE, 1975). The thick, dark surface layer results from the long-continued application of manures with a fairly high admixture of mineral material. Since this manuring practice goes back to medieval times a thick humic A1 horizon was gradually built up. With the advent of chemical fertilizers near the close of the 19th century the Plaggen system was abandoned. In the landscape fields with a Plaggen epidepon—the so-called old arable lands—are readily recognized by their sharp rectilinear boundaries and somewhat higher elevation.

Xerovague soils are dry sandy soils which do not show any sign of pedogenic evolution beyond the presence of a thin humic A1 horizon. They are normally associated with the very young land dunes of the Kootwijk Formation.

Groundwater regime

The Netherlands Soil Survey Institute distinguishes seven water-table classes (symbol: Gt) which reflect the seasonal fluctuation in mean groundwater level in the soil. The classes which are relevant to the study area are given in table II.

² N.A.P. = Dutch Ordnance Level

Table II
Water-table classes, with mean highest and lowest groundwater level.

water-table class (Gt)	mean highest ground water level in cm below local groundlevel	mean lowest ground water level in cm below local ground level
III	< 40	80-120
V	< 40	> 120
VI	40-80	> 120
VII	> 80	> 120

The lowest mean groundwater levels (Gt VII) are found in old arable land (i.e. fields with an overthickened Plaggen epipedon) and in culminations of the gently undulating coversands. Normally the water-table classes in the coversands are V or VI. In valleys and other depressions Gt III is prevalent.

VARIABLES OF THE AREAL CLASSIFICATION

Four variables were selected to subdivide the pilot area: the loam content of the upper 30 cm of the soil profile (loam = fine earth i.e. soil from which particles larger than 2 mm have been removed by sieving; prior to determination of loam content organic matter and pre-lime were removed), the soil type, the water-table class and the altitude respectively.

They represent the lithology, soil condition, hydrology and the topography of the landscape. The loam content variable is determined from the soil map (scale: 1:50,000, sheet 58) and from the provisional version of the geomorphological map (scale 1:50,000, sheet 58). The soil type and water-table class variables are also determined from the soil map. The altitude variable is read from the topographical maps (scale 1:25,000, sheets 58 A, B, D). In practice, the number of classes will often be specified a priori as a separate condition (DE GRUIJTER, 1977). In the present case, each variable is subdivided into four classes (Table III). The number of variables was restricted by the interpretability of the cluster groups.

For sampling purposes the area is subdivided in 300 grid units (250 m by 250 m). The grid size is chosen in accordance with the 1:50,000 map scale. At each grid the data are collected at the central point. Thirteen grids are situated in the built up area. These areas are left out of consideration.

Table III
The subdivision of the variables in classes.

loam content	(% < 50 μ m)	soil type	water-table class (see table II)	altitude in m above N.A.P.
very slightly loamy sand	0	hydropodzol soil	III	> 30.0
slightly loamy sand	10-17.5	thick earth soil	V	27.5-30.0
loamy sand	10-50	hydro-earth soil	VI or V/VI	25.0-27.5
sandy loam	50-85	xerovague soil	VII	< 25.0

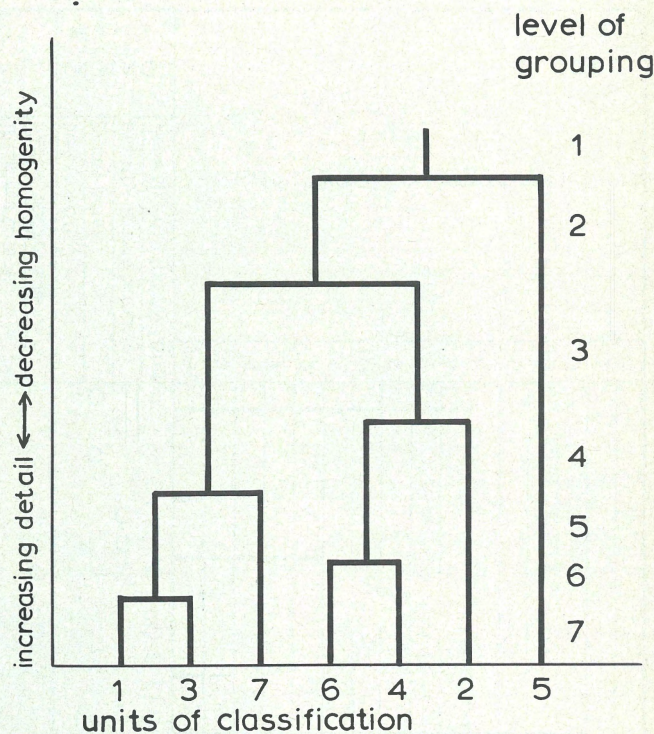


Fig. 3
The hierarchic grouping of 7 units of classification.

THE CLUSTER ANALYSIS

A major problem in numerical classification is how to select the most suitable method from the large number of alternatives (DE GRUIJTER, 1977).

A hierarchic cluster-analysis technique has been selected out of many possible numerical taxonomic cluster procedures. This technique is the Ward's method of cluster analysis which is also called the minimum-variance method. It refers to the fact that within the clusters (c.q. mapping units) there is a minimum of variance for the variables.

Using Ward's method, cluster groups are determined on the basis of a similarity index between the sample points. The analysis is completed using the Clustan programme (WISHART, 1978).

The variables loam content, water-table class and altitude are measured on an ordinal scale, and the variable soil type on a nominal scale. For the cluster analysis all the input data,

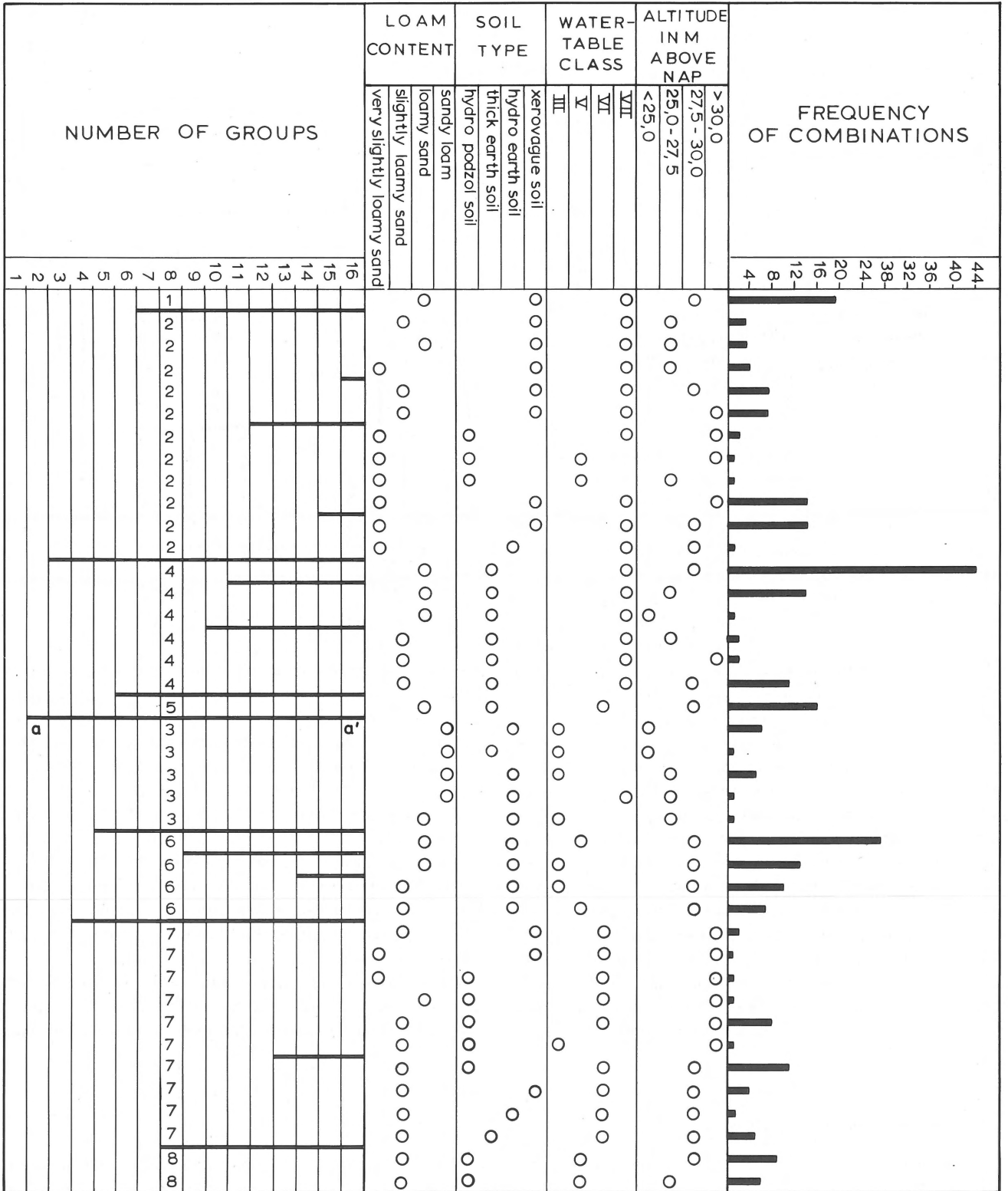


Fig. 4
 Legend system.
 (1, 2, 3, . . . 8) = grouping number in figure 6.

mapping unit c-q. cluster	loam content		soil type		water-table class		altitude in m above N.A.P.		% of the total area	number of grid-units belonging to the most frequent combination	total number of grid-units belonging to the mapping unit
	slightly loamy sand	very slightly loamy sand	loamy sand	sandy loam	hydropedal soil	thick earth soil	xerovague soil	hydrosarth soil			
1			X				X		7	19	19
2	X							X	20	14	57
3				X					5	6	14
4			X			X		X	26	44	74
5			X				X		6	16	16
6			X			X		X	20	27	57
7		X			X			X	12	11	35
8	X			X			X		5	9	15

Fig. 5

Legend to the grouping level 8.

x = most frequent combination; - = other classes.

therefore, were coded in binary form. With these data a similarity coefficient matrix is developed as based on the four variables determined at each sample point. The squared euclidean distance is selected as the similarity coefficient. The Ward's method of cluster analysis selects the cluster groups from this similarity coefficient matrix. Two members resembling each other are joined to form one group. Next, two units, two groups or one group and one unit are joined together according to the principles of 'least error sum of squares' of euclidean distances. The homogeneity decreases during the grouping process. This grouping process can be visualised in a dendrogram as shown in figure 3. The original dendrogram was used for the fabric of the legend system (Fig. 4).

At each grouping level a map can be drawn with the graphical facilities of Clustan programme, where every unit is labelled by the number of the corresponding group of cluster.

RESULTS

The legend

The cluster analysis results in a legend system presented in figure 4. In this figure, the first column on the left presents the results of the cluster analysis with the order of grouping and the number of groups derived. The central four columns give the composition of the combinations and their relative position in the system. In the last column the frequency of the combinations is given. If, for instance, one wishes to divide the pilot area into two mapping units, the line aa¹ in figure 4 should be elongated throughout the figure. Mapping unit 1 is represented by the combinations of variables and the corresponding frequencies above the imaginary line; mapping unit 2 by those in the lower sector. In the legend the variation of the variables at any desired grouping level can be observed by studying the columns. In this figure the frequency division of the combinations within any group can be observed.

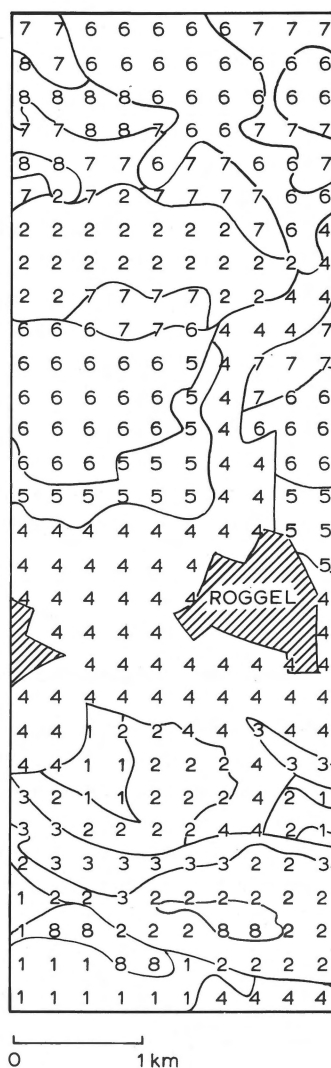


Fig. 6

Map related by the grouping level 8.

The map

The units which are presented in the map, consist of groups of the same level of grouping which are inserted in the legend system. Eventually it is possible to make a whole spectrum of maps from one to 'n' maps, in which 'n' is the total number of possible combinations. The number of groups that is chosen depends greatly on the aim of division. A division in 8 mapping units gives a good picture of the landscape pattern of the area. In figure 5, for every map unit the most frequent combination is marked with the symbol X. The other classes also occurring in the considered mapping units are shown by dashes. Combinations occurring only once or twice are omitted. Figure 5 may be considered as a derived version of the legend system on grouping at level 8. The name of the mapping unit is derived from the most frequent combination and is given in table IV. The related map is given in figure 6. The delineations of its mapping units have been slightly

Table IV
Description of the mapping units at level 8

map- ping unit	loam content	soil type	water table class	altitude in m above N.A.P.
1	loamy sand	xerovague soil	VII	27.5-30.0
2	very slightly loamy sand	xerovague soil	VII	27.5-30.0
3	sandy loam	hydro-earth soil	III	< 25.0
4	loamy sand	thick earth soil	VII	27.5-30.0
5	loamy sand	thick earth soil	VI	27.5-30.0
6	loamy sand	hydro-earth soil	V	27.5-30.0
7	slightly loamy sand	hydropodzol soil	VI	27.5-30.0
8	slightly loamy sand	hydropodzol soil	V	27.5-30.0

adapted so as to obtain a better agreement with boundaries of the pre-existing base maps. Thus a certain measure of subjectivity has been introduced.

EVALUATION

Numeric taxonomic procedures may give strongly divergent results between them (SEYHAN & KEET, 1981; SEYHAN, 1981). The groups which are formed by this cluster analysis, cannot be converted into causal, functional or genetic relations. For regionalisation, the cluster analysis has advantages. The traditional, tentative approach of the landscape mapping tends to seek a functional explanation by hierarchic classification of polythetic variables; cluster analysis offers perspectives to this.

Cluster analysis does not weigh its variables in advance and therefore creates no premature hierarchy, but an optimal grouping sequence of the combinations of variables is to be searched. A direct advantage of the use of the legend system is that the comparison of the different phases of grouping always

remains possible for the whole grouping process. The essential information about the variation and frequency of the combinations is maintained. The quantitative way of analysis, achieved with the cluster analysis, guarantees the possibility of reproducing the results and consequently increases the objectivity of the ordering process because the ordering itself is an exact quantitative statistical procedure.

ACKNOWLEDGEMENTS

In preparing this article, Mr. Heine and Mr. Sion drew the figures, Mrs. Snijder typed the manuscript, Drs. Schwan and Dr. Seyhan read the manuscript critically and corrected the English text, and Drs. Ten Cate (Stiboka) made the provisional geomorphological map available to us. We are most grateful to them for their help.

REFERENCES

- Bakker, H. & J. Schelling 1966 *Systeem van bodemclassificatie voor Nederland. De hogere niveaus - Pudoc (Wageningen).*
- De Gruijter, J. J. 1977 *Numerical classification of soils and its application in survey - Agricult. Res. Rep. 855 Pudoc (Wageningen).*
- Reiling, R. & E. Zilverberg 1981 *Veldwerk 1980. Midden Limburg - Unpubl. M.Sc. thesis Inst. Earth Sci. Vrije Univ. Amsterdam.*
- Seyhan, E. 1981 *Introduction to multivariate analysis in hydrology - Inst. Earth Sci. Vrije Univ. Amsterdam (in press).*
- Seyhan, E. & B. Keet 1981 *Multivariate statistical analysis (part 1) - Aqua VU 8 (Amsterdam) (in press).*
- Soil Conservation Service U.S. Department of Agriculture 1975 *Soil Taxonomy. A basic system of soil classification for making and interpreting soil surveys - Agricult. Handb. 436.*
- Stiboka 1972 *Bodemkaart van Nederland. Toelichting bij kaartblad 57 Oost, Valkenswaard en 58 West, Roermond - Stiboka (Wageningen).*
- Wishart, D. 1978 *Clustan, user manual - Inter-Univ. res. council series, Rep. 47.*