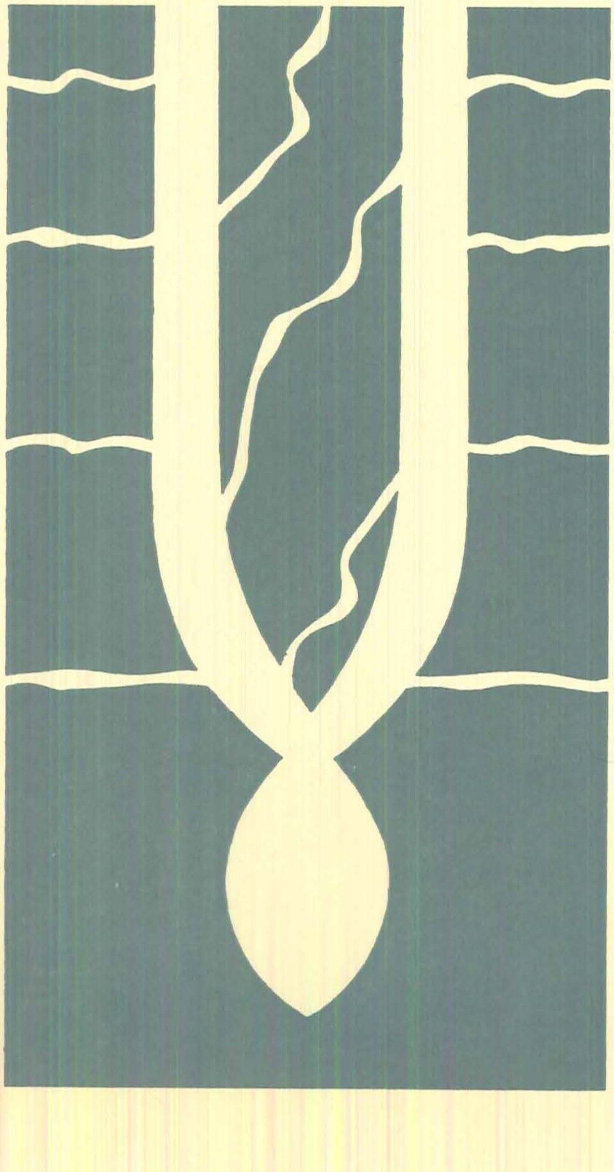
Soil Survey Papers, No. 7

CLUSTER ANALYSIS APPLIED TO MINERALOGICAL DATA FROM THE COVERSAND FORMATION IN THE NETHERLANDS



by R. D. Crommelin
and
J. J. de Gruijter

Netherlands Soil Survey Institute, Wageningen

SOIL SURVEY PAPERS, No. 7

CLUSTER ANALYSIS APPLIED TO MINERALOGICAL DATA FROM THE COVERSAND FORMATION IN THE NETHERLANDS

R. D. Crommelin

and

J. J. de Gruijter

SOIL SURVEY INSTITUTE, WAGENINGEN, THE NETHERLANDS — 1973

CONTENTS

SUMMARY

In an earlier investigation the coversand formation in the Netherlands had been classified by Maarleveld (1968) into 28 subareas on the basis of both detailed geological surveying and the evaluation of macroscopical characteristics primarily in the 105-75 micron size fraction.

This paper describes a versatile classification procedure using solely quantitative data, viz. iterative cluster analysis, which is applied to the same coversand material examined by Maarleveld. The aim was to examine whether both methods provide comparable results and if so, to what extent.

In several trials, using varying numbers of clusters and characteristics in the analysis, it was shown that in general both methods, being very different as to fundamental principles, do give in most cases different, though not necessarily inconsistent, results.

However, it was found that when the cluster procedure was carried out with a relatively small number of clusters (11) and a set of additional variables was used in the analysis, the map images of both methods fit reasonably well.

It is suggested that cluster analysis might well be applied successfully in the future to all sorts of classificatory problems in the earth sciences.

1. INTRODUCTION

During the last thirty years researchers have frequently investigated the geological and pedological aspects of the coversand formation in the Netherlands. However, the drawing of final conclusions about the origin and properties of coversand has been hampered up to now by the fact that researchers have often approached the problem from different points of view. Moreover, the lack of a systematic sampling and a large scale survey over the total coversand region has prevented the examination of how far significant differences exist which would justifiy a subdivision into a number of separate coversand provinces. Because interest from geological and soil science quarters arose about this question, a thorough investigation into the problem started in 1963 in co-operation with Prof. Dr. G. J. Maarleveld and with financial aid from the Netherlands Organization for the Advancement of Pure Research (ZWO).

By combining geomorphological observations and a set of macroscopical data, Maarleveld succeeded in distinguishing 28 different subareas. The results of this investigation have been published in Boor en Spade (Auger and Spade) 16, 1968. Maarleveld's map is reproduced here as Appendix 2. In the subdivision given by Maarleveld, the analytical results were to a certain extent adjusted to recognizable geomorphological elements in the field.

It was felt that the large mass of data collected from over a thousand samples should be examined objectively in order to try to solve some important problems not touched upon by Maarleveld.

In order to extract more information from the analytical data, two different procedures were applied to the coversand analyses. In an earlier paper (CROMMELIN and KEULS, 1970) the results of the first method have been described: the discriminative power of Maarleveld's subdivision was tested by a numerical comparison between all possible pairs of the distinguished subareas, using a sort of difference-coefficient based on the values of the three principal variables.

In the present paper the classification aspect of the coversand formation is considered. It seemed interesting to apply a classification procedure independent of preexisting knowledge of the above mentioned subareas and to compare how such an objective classification fits the more subjective classification arrived at by Maarleveld.

2. METHOD

One of the numerous methods of numerical classification is iterative cluster analysis. This aims at partitioning a given set of objects (samples) on the basis of multiple measurements, such that the heterogeneity of the classes is minimized. See for general information about this approach e.g. BALL and HALL (1965), JANCEY (1966), LANCE and WILLIAMS (1967).

The variant, we used in this study, has already been successfully applied to pedological problems at the Netherlands Soil Survey Institute. A detailed account of the procedure will appear in the next future (DE GRUIJTER, in prep.); here it will only be briefly described.

A measure of heterogeneity was adopted: the sum of the squared Euclidean distances between each object and the centroid (i.e. average) of the class to which it has been allocated. Objects and centroids are conceived as points in a multidimensional space of which the mutual perpendicular axes represent the variables. Squared Euclidean distances in such a space can be computed in the same way as in two- and three-dimensional spaces, namely by squaring the differences between corresponding co-ordinates and summation of these over the axes (Pythagoras). Denote this measure with SS_w (sum of squares within classes), and let:
SS_t (total sum of squares) = sum of squared Euclidean distances between all objects and the overall centroid,
SS_b (sum of squares between classes) = sum of squared Euclidean distances between the class centroids and the overall centroid, weighted with the frequencies of the classes.
Then it is easily shown that SS_t, a constant for a given set of data, equals $SS_w + SS_b$. Both SS_w and SS_b may be expressed as a percentage of SS_t.

The search strategy for a grouping with minimum SS_w was as follows.

1. Choose the number of classes.
2. Choose an initial grouping as a starting point.
3. Remove the first object and re-allocate this, to the same or another class, such that SS_w is as low as possible. In case of a switch, re-evaluate the centroids of the old and the new class. Repeat this successively for each object.
4. Iterate step 3, but stop if no more switches occurred during the last iteration.

Ad 1. The choice of the number of classes is up to the researcher; the question is dealt with in the following section.
Ad 2. Suitable starting points have been obtained with the method of Ward (1963).
Ad 3. The best allocation can be easily found, once the distances have been computed between the object concerned and current class centroids.
Ad 4. It should be stated that, due to the switches, the final grouping has a lower SS_w, i.e. is more homogeneous than the initial grouping, the strategy however does not warrant that it is the best of all the possible alternatives.

In order to give the variables equal weight, they were standardized (unit variance) before the analysis.

The calculations have been carried out by the Institute for Mathematics, Information processing and Statistics (TNO) at Wageningen.

3. LABORATORY DATA AND CLUSTER ANALYSIS

MAARLEVELD (1968), using a binocular microscope, examined over a thousand samples. Within the 420-300, 210-150 and 105-75 micron size fractions he determined the contents of 5 types of grains in pro-mil values, viz. white, green, black and red grains and unidentified rock fragments. As only a negligeable amount of green grains occurred in the 420-300 micron size fraction, this variable was deleted. The cluster analysis applied to all samples therefore contains 14 variables.

From the total set of 1056 samples a subset of 465 samples in the 210-150 micron size fraction was examined mineralogically; individual proportions of the heavy minerals, or sometimes natural groups of heavy minerals, expressed as percentages of the total heavy fraction, the percentage of the heavy fraction itself, and the percentages of sodium feldspar and plagioclase were determined. This provided an additional 8 variables.

Thus there were 2 distinct sample sets available for cluster analysis, firstly the total set $n = 1056$ with $m = 14$ variables and secondly the subset $n = 465$ with $m = 22$ variables.

The number of clusters over which the population of objects is divided can in principle be chosen freely. Although it is true that the larger the number of clusters, the larger the SS_b and consequently the larger the percentage of the total difference explained by the clustering, there is an advantage in reducing the number of clusters, as a too detailed classification is often impractical. A suitable compromise number should be small but large enough to explain a reasonable percentage of the total difference due to clustering say 65 %-70 %.

As a first step in the programme it was obviously interesting to execute the grouping on the basis of the subdivision of Maarleveld's classification, by fixing the number of clusters at 28 and examining how well both methods of classifications agreed.

The analysis for $k = 28$ was carried out on both the total set data and the subset data. The reproduction of the results on overall maps was not feasible due to printing difficulties but they are summarized and compared in table form with Maarleveld's subdivision (Table 1a and 1b).

Case 1: $k = 28$, $m = 14$, $n = 1056$ (Table 1a, Appendix 2) *

Ideal agreement between survey and clustering would mean that each cluster coincides with one survey area, excepting areas 17 and 19, wich having been taken together, would then be characterized by 2 clusters. Another extreme is conceivable in the sense of a random distribution of the clusters over the areas. Neither of these extremes was to be expected, but it was hoped that clusters and areas would broadly coincide and that only in the border region between areas some overlapping would occur. Table 1a shows a distribution that departs considerably from this desirable situation. The average number of clusters per area is 7 though in some areas we find

* It should be emphasized that in each run of the analysis the computerprogram prints out an entirely arbitrary order of cluster numbers. Hence there exists no relation between corresponding cluster numbers in the four cases next to be treated.

from 10 to 12 clusters. Conversely many clusters are distributed over the total coversand region, e.g. cluster numbers *10, 12, 16, 17, 18* and *20* are represented in 12 to 16 areas.

Though the coincidence between clusters and coversand areas is still far from satisfactory, some general pattern in the cluster arrangement is recognizable if one considers *groups* of clusters: e.g. clusters *1-7* and *24-28* are virtually confined to

Fig. 1. Map of the Netherlands showing the names mentioned in this publication.

Table 1a. Relation between 28 coversand areas (Maarleveld's classification) and 28 clusters (analysis based on 14 variables)

Cluster no. \ Area no.	Noord-Brabant + northern part of Limburg												Central Netherlands		Achterhoek + Overijssel										Northern provinces			Number of samples per cluster	Number of areas per cluster
	2	3	4	5	7	8	9	10	12	13	14	16	18	19+17	20	21	22	23	24	25	26	27	28	30	31	32	33		
1	1																								1			2	2
2	3			1					3	6	6											1						20	6
3									1	8	2																	11	3
4										6																		6	1
5											2																	2	1
6										3		5																8	2
7								1		4																		5	2
8															9	5												14	2
9															10													10	1
10					5	3	7	3		1			3	6				4	2	8	7	2		3	12	1	2	69	16
11	1		1		12		1			4	1														1			21	7
12		1			6	1	8	5					12	8				1		3	3	1		3	8	3		63	14
13				1														1			1		1					4	4
14				1	1	19	5	4	8			4													1			43	8
15						2	7																					9	2
16			1		2		2	2					13	6						1	2			1	1			31	10
17	4	2		8	4				5	2	1										1							27	8
18			3		16		6	10					31	27				3		5	2	1	1	13	76	12	9	215	15
19															3	3	1	1										8	4
20		1		3	2			2	3										1	4	1	1	1		5		1	25	12
21						1										1	12	10	8	1	13	5	4	6	2		4	67	12
22																6	17	8	4		2	2	3	1				43	8
23															1	15	12	2										30	4
24			1		10	2	32	45	2		2										2	1						97	9
25	1	2	5			2	17	31	3	1	1		1	1											3			68	12
26		8		1	31		10	17	2					1											4			74	8
27						3	19	9	5		1	1								1								39	7
28		1					1	11	6	22	1	2												1				45	8
Number of samples per area	10	15	11	15	89	33	115	140	38	57	17	12	60	49	23	30	42	30	15	23	34	14	10	28	114	16	16	1056	
Number of clusters per area	5	6	5	6	10	8	12	12	10	10	9	4	5	6	4	5	4	8	4	7	10	8	5	7	11	3	4		188

Table 1b. Relation between 28 coversand areas (Maarleveld's classification) and 28 clusters (analysis based on 22 variables)

Cluster no. \ Area no	Noord-Brabant + northern part of Limburg												Central Netherlands		Achterhoek + Overijssel										Northern provinces			Number of samples per cluster	Number of areas per cluster
	2	3	4	5	7	8	9	10	12	13	14	16	18	19 + 17	20	21	22	23	24	25	26	27	28	30	31	32	33		
1			2	2	9																							13	3
2		6			18			2	1																			27	4
3					10			6																				16	2
4	1		3		4		19	28	3																			58	6
5		1					1	2	2	7	1																	14	6
6	3								2	3	3																	11	4
7										10																		10	1
8										2	2																	4	2
9										3																		3	1
10											1																	1	1
11				5					8																			13	2
12						1						8												1				10	3
13										2		1																3	2
14						3																						3	1
15						1	15		3																			19	3
16															1	4	2	1										8	4
17																4	13	4	4		3	2	3	2				35	8
18															4													4	1
19															2	9	3											14	3
20						4	2						1	6				1			1							15	6
21													14	17				1		4	2	1		6	5		3	53	9
22						1	5		1				10	6														23	5
23						1	3						6	8				2	1	8	1	2		4	2			38	11
24																						1			2		2	5	3
25							1		1				1	1							1			1	41	8	2	57	9
26															4													4	1
27																1												1	1
28															3													3	1
Number of samples per area	4	7	5	7	41	11	46	38	21	27	7	9	32	38	14	18	18	9	5	12	8	6	3	14	50	8	7	465	
Number of clusters per area	2	2	2	2	4	6	7	4	8	6	4	2	5	5	5	4	3	5	2	2	5	4	1	5	4	1	3		103

Brabant and North Limburg, whereas cluster numbers *21-23* occur mainly in the Achterhoek and Overijssel. Cluster numbers *4, 5* and *9* are found exclusively in areas 13 and 14 along the Meuse, in area 20 along the Old IJssel and the Geldrian IJssel. In fact Maarleveld mapped the most conspicuous coversand types of the Netherlands in these regions.

Case 2: k = 28, m = 22, n = 465 (Table 1b, Appendix 2)

The inclusion of the 8 mineralogical characteristics in the analysis helps to derive a more homogeneous distribution of the clusters. This is probably so because if the number of variables is large, the influence of extreme values due to analytical errors is relatively small. Thus the chance that a cluster is formed which has no real significance is lowered.

The average number of clusters per area is now 4, as compared with 7 in the analysis with 14 variables, and there is a more distinct separation between clusters occurring only in Brabant and North Limburg *(1-15)* and those characteristic for the Achterhoek and Overijssel *(16-19)*. Only two clusters *(23* and *25)* are found to be distributed throughout the whole coversand region of the Netherlands.

This analysis, as with the previous analysis, does not delineate specific clusters for the Central Netherlands (viz. Utrechtian Hills + the Veluwe borderland) and the northern provinces. The areas along the Meuse, the Old IJssel and the Geldrian IJssel are however, as in the preceding case, characterized by clusters that occur only there and not in other areas. The overall distribution over the areas, although an improvement over case 1, still remains unsatisfactory.

Evidently the subdivision into 28 clusters is too detailed to derive a homogeneous distribution and so it was decided to execute the program once more on the basis of 11 clusters (a smaller number than 11 would yield an unfavourable ratio of the SS_b to the SS_t and thus a too small percentage of the differences explained by clustering). The results are discussed in case 3 and case 4. Those of case 4 have been reproduced on the accompanying map Appendix 1.

Case 3: k = 11, m = 14, n = 1056 (Table 1c, Appendix 2)

This situation is hardly better than the second one, in that the intended simplification of cluster distribution was not achieved. The average number of clusters per area remains about the same, presumably because the advantage of the smaller number of clusters is undone by there being only 14 variables involved here as opposed to 22 in the preceding analysis. Approximately half the clusters occurs almost exclusively in Brabant and North Limburg *(2, 3, 4, 7, 10* and *11)*. Cluster *9* is specific for the Achterhoek and Overijssel. These are about the same conditions found in case 2.

Case 4: k = 11, m = 22, n = 465 (Table 1d, Appendices 1 and 2)

Agreement with Maarleveld's map is now considerably better, probably as a result of the influence of the additional mineralogical characteristics. The distribution of clusters specific for Brabant, North Limburg, Overijssel, the Achterhoek and the

Table 1c. Relation between 28 coversand areas (Maarleveld's classification) and 11 clusters (analysis based on 14 variables)

Cluster no. \ Area no.	Noord-Brabant + Northern part of Limburg												Central Netherlands		Achterhoek + Overijssel										Northern provinces			Number of samples per cluster	Number of areas per cluster
	2	3	4	5	7	8	9	10	12	13	14	16	18	19 + 17	20	21	22	23	24	25	26	27	28	30	31	32	33		
1	4			1					2	4	1							1			1		1		1			16	9
2										2	4																	6	2
3										7																		7	1
4								1	2	18	2	5																28	5
5															21	11												32	2
6	1	3	1	1	20	6	31	18		1			19	19				5	5	11	13	4		6	21	4	2	191	20
7				1	2	25	29	5	13	1	1	7	4			1				1								90	12
8	1	2	4	1	44	1	15	22	1				37	29			1	5	3	10	5	2	2	15	89	12	12	313	22
9															2	18	41	19	7	1	14	7	6	6			2	123	11
10		10	6	1	17	1	40	93	4	7	3			1											2			185	12
11	4			10	6			1	16	17	6										1	1	1	1	1			65	12
Number of samples per area	10	15	11	15	89	33	115	140	38	57	17	12	60	49	23	30	42	30	15	23	34	14	10	28	114	16	16	1056	
Number of clusters per area	4	3	3	6	5	4	4	6	6	8	6	2	3	3	2	3	2	4	3	4	5	4	4	4	5	2	3		108

Table 1d. Relation between 28 coversand areas (Maarleveld's classification) and 11 clusters (analysis based on 22 variables)

Cluster no. \ Area no.	Noord - Brabant + northern part of Limburg												Centr. Nether-lands		Achterhoek + Overijssel										Northern provinces			Number of samples per cluster	Number of areas per cluster
	2	3	4	5	7	8	9	10	12	13	14	16	18	19 + 17	20	21	22	23	24	25	26	27	28	30	31	32	33		
1	1	7	5	3	41		18	38	4																			117	8
2	3			3					5	20	4																	35	5
3										3																		3	1
4										2	3																	5	2
5										2		9												1				12	3
6				1		11	23		11				1															47	5
7															1	14	13	2	1		1	1	1					34	8
8							3						30	36		1	5	7	4	11	7	5	2	12	4		4	131	14
9							1		1				2	1							1			1	46	8	3	64	9
10															6	3												9	2
11															7													7	1
Number of samples per area	4	7	5	7	41	11	45	38	21	27	7	9	33	37	14	18	18	9	5	11	9	6	3	14	50	8	7	465	
Number of clusters per area	2	1	1	3	1	1	4	1	4	4	2	1	3	2	3	3	2	2	2	1	3	2	2	3	2	1	2		58

northern provinces is more pronounced than in case 3. For the first time a number of areas is discernible which is characterized by the occurrence of only one cluster, though not always a different one (e.g. areas 3, 4, 7 and 10 which are characterized by cluster *1*). On the average there are now 2 clusters per area.

Maarleveld's survey, schematically reproduced in Appendix 1 together with the cluster pattern, shows convincingly that his division in areas fits the pattern of cluster distribution reasonably well; both reflect the larger variation in Brabant and North Limburg as opposed to that in the northern provinces, and both indicate the specific borderzones along the Meuse, Geldrian IJssel and Old IJssel.

For a more detailed comparison between the survey and clustering we need the Euclidean distance matrix which summarizes all possible mutual distances between the 11 clusters (Table 2). It can be seen that clusters *1, 2, 6, 8, 9* and *5* form a submatrix characterized by relatively small mutual distances, i.e. differences. When we also take into account that Maarleveld's areas 3, 4, 7 and 10 belong to cluster *1*, then it follows that the differentiation of the coversand formation in Brabant is in analytical respect not so great as Maarleveld's map suggests.

It is otherwise along the Meuse where Maarleveld observed very different coversand types and where accordingly are found the very divergent clusters *3* and *4*, which are unique for that region. These sands have a variagated appearance which can be accounted for by their high contents of black and green grains and rock fragments. The mineralogical characteristics include the occurrence of chloritoid (mainly in cluster *4*), high percentages of tourmaline and metamorphic minerals and associated low contents of feldspars. Clusters *3* and *4* form a small submatrix; their large mutual distance indicates a large difference, whereas each cluster in turn differs considerably from every cluster belonging to the submatrix characteristic for Brabant.

Table 2. Squared Euclidean distances between cluster-centroids

2	13									
6	16	21								
8	15	28	11							
9	14	28	21	7						
5	41	25	18	31	42					
7	45	50	23	17	36	32				
10	107	93	63	72	98	59	26			
11	185	151	135	145	177	97	77	37		
3	133	89	112	165	170	99	171	184	206	
4	183	132	163	199	204	133	198	186	239	122
Cluster numbers	1	2	6	8	9	5	7	10	11	3

An analogous situation is encountered along the Geldrian IJssel and Old IJssel, where the cluster picture is dominated by the 2 extreme clusters *10* and *11,* which are unique for the region. They are both characterized by high contents of white feldspar grains and green glauconite grains. There is a high content of heavy minerals among which epidote, augite and hornblende predominate but the content of tourmaline and metamorphic minerals is extremely low. Their composition is thus largely the opposite of clusters *3* and *4* along the Meuse and they constitute together with cluster *7* the middlemost submatrix. Clusters *10* and *11* differ considerably from every cluster in both other submatrices, which shows their exceptional character. Cluster *7* is included in this submatrix because it is found in the same region as clusters *10* and *11* to which it bears some resemblance. It is also a transition into the neighbouring cluster *8* with which it has more affinity than with either cluster *10* or cluster *11*. Cluster *7* also forms a transition into the clusters in Brabant, however it differs considerably from clusters *3* and *4* in North Limburg. Clusters *10* and *11* coincide broadly with area 20. The detailed differentiation which Maarleveld mapped in Overijssel eastward of area 20 is not recognizable in the cluster image which is dominated by cluster *8*. Besides occurring in Overijssel we find cluster *8* especially in the Central Netherlands (areas 17, 18 and 19) and in the northern provinces where it gradually gives way to cluster *9*. When we bear in mind that the difference between clusters *8* and *9* is very small (it is in fact the smallest between all possible pairs in the distance matrix of Table 2) we may safely conclude that the coversands north of the line Rhine - Geldrian IJssel - Berkel belong to one large province with little internal variation.

Summarizing we can state that:

1. The analysis on the basis of 28 clusters executed with either 14 or with 22 variables does not fit Maarleveld's survey satisfactorily.
2. The analysis on the basis of 11 clusters with 22 variables is in reasonable accordance with Maarleveld's survey, taking into account that:

a. Both methods show that the most characteristic coversand types occur along the Meuse, the Geldrian IJssel and the Old IJssel.

b. The main body of the Brabant province shows detailed differentiation in coversand types according to both methods. However the analytical method shows that the differences are small.

c. The north western coversand region (Maarleveld's areas 18, 19, 30, 31, 32 and 33) is rather uniform; the differences according to cluster analysis are still less manifest than in the previous case.

d. The differentiation in Overijssel north of the Berkel recorded by Maarleveld cannot be confirmed by this cluster analysis.

4. CONCLUSIONS

An investigation was made to determine the degree of agreement between cluster analysis and the subdivision of the coversand region in the Netherlands proposed by Maarleveld. This means comparing an objective numerical procedure with a more subjective interpretation in which analytical data are combined with field observations. It goes without saying that both methods will give different though not necessarily inconsistent, results. The purpose of the investigation determines to what results the researcher will attach more value.

In a geological sense a classical field survey is preferable because in this case recognizable geological elements and field evidence count heavily in the ultimate interpretation. If, however, one considers a grouping based on maximum differences between analytical properties more important, an objective numerical classification such as the procedure presented here, is superior. Indeed one finds that the range between the average values per variable over the 28 *clusters* is much larger than over the 28 *subareas* of Maarleveld. Even when only 11 clusters are constructed, in which case cluster contrast is obviously less pronounced than with 28 clusters, we still find that the range between the average values of each variable is in most cases larger than over the 28 subareas of Maarleveld. The intensive use of the data undoubtedly makes cluster analysis a valuable method; on the other hand it should be borne in mind, that 'spurious' clusters might occasionally arise from uncontrolled extreme values if the number of variables is small.

Further it must be pointed out that research as described here whereby the distribution of soil types is involved, is characterized by a peculiar restriction in that the objects to be classified have fixed positions in a geographical sense. This implies that it is not always possible to group those objects belonging to one and the same cluster into sensible units without violating the image of a more conventional survey.

The meaning of a numerical classification is the arrival at a balanced distribution of easily distinguishable units. If experimental knowledge plays an important part in the surveying, it is often possible to interpolate and extrapolate between and outside the available field observations. If, however, surveying is based on the interpretation of analytical data, as in cluster analysis, one misses this expedient and can arrive at a justified map image only when a dense pattern of observations is available, with as little hiatus as possible. Obviously whether the structure of the terrain is simple or complicated is important in this respect; one can make do with less observations in the first case than in the second.

It seems safe to say that at least in the Netherlands insufficient experience has been acquired with numerical classification, in particular with cluster analysis applied to soil science and geology, this in contrast with taxonomy where such techniques are applied widely. The research described should be considered as an experiment which is undoubtedly open to further improvement. It is felt that as experience grows, cluster analysis may be applied to various classification problems in the earth sciences.

6. REFERENCES

BALL, G. H. and D. J. HALL — 1965 — Isodata, a novel method of data analysis and pattern classification. Stanford Research Institute, Menlo Park, California.

CROMMELIN, R. D. and M. KEULS — 1970 — The statistical discrimination between coversand areas in the Netherlands. *Geologie en Mijnbouw 49, 4:* 319-328.

GRUIJTER, J. J. DE — in prep. — Application of numerical classification in soil science, with special reference to soil survey. Soil Survey Institute, Wageningen, the Netherlands.

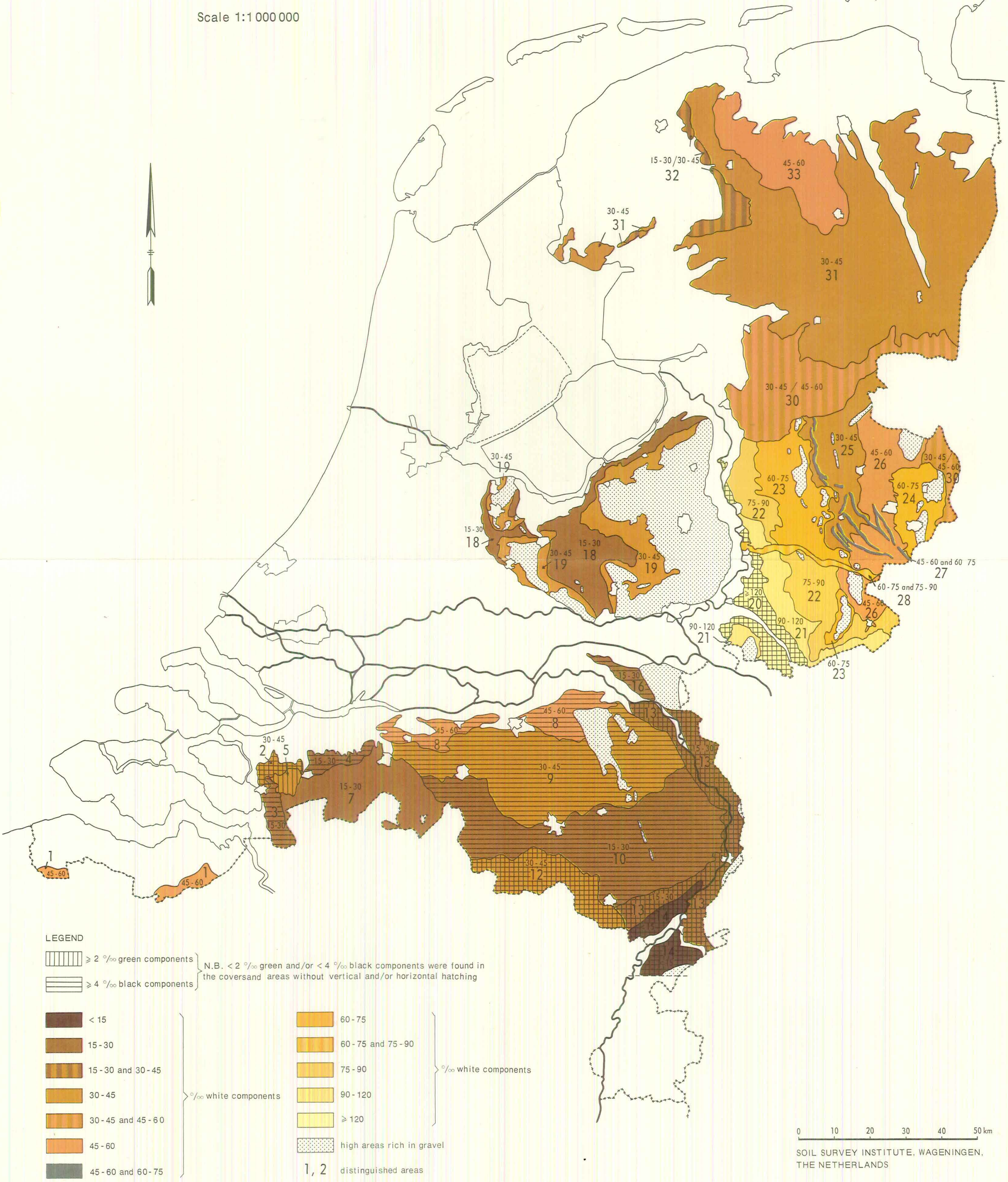
JANCEY, R. C. — 1966 — Multidimensional group analysis. *Austr. J. Bot. 14:* 127-30.

LANCE, G. N. and W. T. WILLIAMS — 1967 — A general theory of classificatory sorting strategies. II. Clustering systems. *Comp. J. 10:* 271-7.

MAARLEVELD, G. C. — 1968 — Voorlopige resultaten van dekzandonderzoek met de plakbandmethode. *Boor en Spade 16:* 38-65.

WARD, J. H. — 1963 — Hierarchical grouping to optimise an objective function. *J. Amer. Stat. Ass. 58:* 236-244.

The Dutch coversand areas classified according to the quantities of characteristic green, black and white components found in the fraction 75 - 105 micron (from Maarleveld,1968)

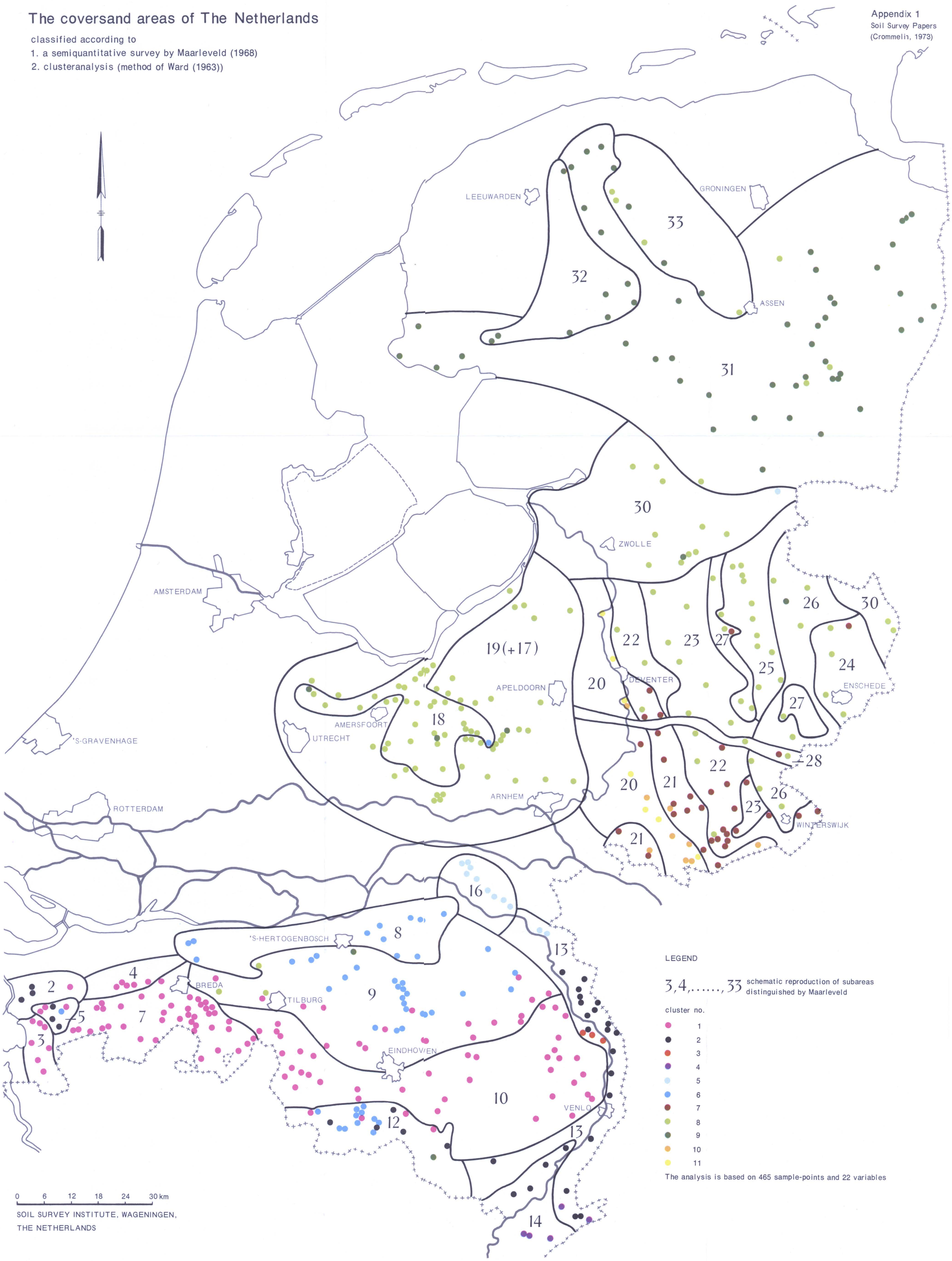


40x51 pdf

The coversand areas of The Netherlands

classified according to
1. a semiquantitative survey by Maarleveld (1968)
2. clusteranalysis (method of Ward (1963))

Appendix 1
Soil Survey Papers
(Crommelin, 1973)



LEGEND

3,4,......,33 schematic reproduction of subareas distinguished by Maarleveld

cluster no.
1
2
3
4
5
6
7
8
9
10
11

The analysis is based on 465 sample-points and 22 variables

0 6 12 18 24 30 km

SOIL SURVEY INSTITUTE, WAGENINGEN, THE NETHERLANDS

SERIES SOIL SURVEY PAPERS

No. 1. *Jongerius, A.* and *G. Heintzberger:* The preparation of mammoth-sized thin sections. 1964 — Nfl. 3.—

No. 2. *Jager, A.* and *W. J. M. van der Voort:* Collection and preservation of soil monoliths from sandy soils and ripened clay soils above and below the water table. 1966 — Nfl. 2.—

No. 3. *Broek, J. M. M. van den* and *L. van der Waals:* The Late Tertiary Peneplain of South Limburg (The Netherlands). Silicifications and fossil soils; a geological and pedological investigation. 1967 — Nfl. 2.—

No. 4. *Brinkman, R.* and *L. J. Pons:* A pedo-geomorphological classification and map of the Holocene sediments in the coastal plain of the three Guianas. 1968 — Nfl. 9.50

No. 5. *Sombroek, W. G.* and *I. S. Zonneveld:* Ancient dune fields and fluviatile deposits in the Rima-Sokoto river basin (N.W. Nigeria). 1971 — Nfl. 14.50

No. 6. *Bal, L.:* Micromorphological analysis of soils. Lower levels in the organization of organic soil materials. 1973 — Nfl. 17.50

No. 7. *Crommelin, R. D.* and *J. J. de Gruijter:* Cluster analysis applied to mineralogical data from the coversand formation in the Netherlands. 1973 — Nfl. 7.—