GEOLOGIE EN MIJNBOUW VOLUME 49(4), p. 319-328 1970

THE STATISTICAL DISCRIMINATION BETWEEN COVERSAND AREAS IN THE NETHERLANDS

R.D. CROMMELIN[1]) and M. KEULS[2])

ABSTRACT

In 1968 Maarleveld published the results of a research of the coversand area in The Netherlands, based on geomorphological evidence and on the evaluation of macroscopical characteristics of the coversands mainly in the 105–75 micron fraction, resulting in a subdivision into 29 subareas. Each subarea could be described in terms of specified amounts (per mil values) of the three characteristics: white, black and green grains. In this study the analytical results of these principal characteristics are considered in a more objective way, in order to arrive at a numerical comparison for all possible pairs of subareas. For any pair this comparison may be expressed as the sample-size N required to detect a difference as large as the one stated between the respective sample means. The N-numbers – which may be considered as difference- or similarity-coefficients have a practical significance in that they provide a measure for the discrimination between any pair of areas as to the characteristic in question: large N-numbers stand for a high degree of similarity, whereas small N-numbers mean that the areas under consideration are very different and consequently may be distinguished with relative ease. The relation between the N-number and the statistical parameters of the characteristics is discussed. A way is indicated how to arrive at defining areas such that they fit classes with a maximum contrast between the analytical data.

INTRODUCTION

In 1966 Maarleveld published a simple method to characterize sands by means of the proportions of variously coloured grains which may be determined with a binocular stereoscope. In a more recent article (1968) the same author applies this method to arrive at a detailed differentiation of the coversand formation in The Netherlands. For that purpose Maarleveld compared and systematized a great number of analyses referring to per mil-contents of white, black and green grains in the size fraction 105-75 micron. The constituant "white" is subdivided into eight classes[3]) with per mil-limits $\geqslant 120$, 90-119, 75-89, 60-74, 45-59, 30-44, 16-29 and < 16. The constituants "black" and "green" are each subdivided into two classes with per mil-limits $\geqslant 4$, < 4 resp. $\geqslant 2$, < 2.

For a good understanding of what follows it should be emphasized that as to grouping observations into classes, Maarleveld postulated the requirement that only the majority ($\geqslant 55\%$) instead of the total number of observations per characteristic should fit the stated class-intervals.

A combination of three classes, one from each of the three kinds of coloured grains[4]) is defined as an association. It appeared that each of the 29 coversand areas distinguished by Maarleveld (Table 1) could be characterized by one of 15 mutually different associations; not every association is represented by an area, whereas one and the same association is sometimes characteristic for several areas.

From Maarleveld's exposition it is apparent that the delimitation of the various areas has been a

---

[1]) Netherlands Soil Survey Institute, Wageningen.

[2]) Department of Mathematics, Agricultural University, Wageningen.

[3]) In Maarleveld's (1968) article called "groups". The expression "classes" however is more in agreement with the concept of order in ranking.

[4]) The definitions white, black and green are considered here as explicit properties of grains, not interpreted in a geological sense. In the text they are abbreviated as WH, BL and GR.

TABLE I

Coversand areas in The Netherlands distinguished by Maarleveld (1968). n = sample size.

1.	Zeeuws-Vlaanderen	n= 21
2.	Halsteren	n= 10
3.	Woensdrecht	n= 15
4.	Bosschenhoofd	n= 12
5.	Wouw	n= 15
7.	Rucphen	n= 89
8.	Made-Oss	n= 34
9.	Teteringen-De Rips	n=121
10.	Goirle-Venray	n=143
12.	Westerhoven	n= 38
13.	Heythuizen-Gennep	n= 56
14.	Horn-Montfort	n= 17
16.	Maas and Waal	n= 10
17.	Ice-pushed ridges Utrecht-Veluwe	n= 8
18.	Gelderse Vallei	n= 65
19.	Zone bordering areas 17 and 18	n= 46
20.	IJsseldunes	n= 23
21.	Vorden-Dinxperlo	n= 31
22.	Wesepe-Lichtenvoorde	n= 42
23.	Raalte-Laren	n= 33
24.	Enschede	n= 15
25.	Den Ham-Goor	n= 23
26.	Hengelo	n= 35
27.	Dunes in Overijssel	n= 14
28.	Berkel area	n= 11
30.	Dinkel and Vecht area	n= 28
31.	Drachten-Winschoten	n=118
32.	Central Friesland	n= 16
33.	Surhuizum-Assen	n= 18

compromise between field experience and analytical results; on the one hand the endeavour to create areas which are justifiable in a geomorphological sense, on the other hand the restriction imposed by the condition that for each area the association should be reasonably homogeneous with reference to certain predetermined limits. However it is evident from the context that in determining the limits of an area the geomorphological factor has played a greater part than have the analytical results and that the latter have only served to modify the limits of areas that already existed in conception from previous experience.

OBJECTIVE

Maarleveld's map reveals – as was expected – that the main trends of the characteristics can in broad outline be correlated with the geomorphology of the coversand formation. More specifically it may be asked in how far the mapped areas may be discriminated on the base of the coloured grains. This question concerns an important point of Maarleveld's research, for the researchworker whose purpose is surveying will not only be interested in the question as to which association his area belongs to, he will particularly be interested in the question in how far the stated association possesses discriminative power with respect to associations of surrounding areas. This aspect has not been discussed by Maarleveld and will now be considered further because it is illustrative for a frequently occurring and general problem in analogous research viz. trying to find a scale for comparing and evaluating stated differences.

Obviously those areas which belong to one and the same association, will be so similar that they are practically indistinguishable. Nor does the other extreme case, when the associations are widely divergent, cause any difficulty; they can readily be recognized from one another. But how is the situation in the intermediate cases – which are by far most frequent – viz. if the associations are more or less different and the areas moreover happen to be adjacent? This is precisely the situation the researchworker has often to deal with. If field distinctives cannot be applied, the alternative is some laboratory-method – in this case Maarleveld's method – in order to decide whether the neighbouring area is sufficiently different qua association to consider it as a separate unity. Then obviously the question arises: what shall be understood by "sufficiently different"? How different should the associations of two areas be, so that the latter may easily be distinguished? In short, can the loose sense of the term different be quantitatively defined? The answer to such sort of questions is here given by expressing the comparison between each pair of areas in a number-scale.

It must be remembered that the concept "association" in the general sense used here above is too complex for our purpose. It may well serve to describe geological conditions but being related to the 3 characteristics jointly, it would be difficult to handle statistically. The numbers in question are instead a function of statistical parameters (mean and variance) of each characteristic separately. Thus in connection with the classification problem to be discussed hereafter, it is the characteristics that will be considered, not the associations.

In figures (1a, b, c) the areas have been arranged according to decreasing mean per mil-value per

Classification on characteristic "WHITE"

I ≥ 120	V 45-59
II 90-119	VI 30-44
III 75-89	VII 16-29
IV 60-74	VIII < 16

See legend page 324.

Class	$\overline{X}$	Area	20	21	22	28	24	23	27	33	26	30	8	1	2	31	25	19	5	12	9	16	32	3	4	7	13	10	18	14
I	154.4	20																												
II	119.3	21	15																											
III	81.9	22	3	6																										
	73.3	28	3	6	52																									
IV	67.1	24	3	4	15	92																								
	64.0	23	2	3	12	46	321																							
	61.5	27	2	3	9	33	101	611																						
	54.0	33	2	3	5	10	15	34	60																					
	52.5	26	2	2	4	7	11	22	35	1093																				
V	51.1	30	2	2	5	9	14	24	39	434	1561																			
	47.9	8	1	2	2	3	3	8	9	35	66	209																		
	47.6	1	2	2	3	5	5	11	14	51	81	244	9416																	
	37.7	2	2	2	2	3	3	5	5	8	9	19	6	14																
	37.6	31	1	1	2	2	2	4	4	7	9	13	12	16	9999															
	36.9	25	1	2	2	3	3	5	5	9	9	17	10	17	2554	3356														
VI	34.7	19	1	2	2	2	2	3	4	5	6	11	5	8	112	171	340													
	34.0	5	2	2	2	2	2	3	3	5	5	10	3	6	44	104	180	1705												
	32.5	12	1	1	1	1	1	2	2	3	4	6	3	4	22	45	59	150	219											
	32.3	9	1	1	1	1	1	2	2	3	4	6	3	4	25	41	53	142	230	9999										
	28.9	16	2	2	2	2	2	3	3	4	4	7	2	5	14	20	30	33	29	52	65									
	28.3	32	1	2	1	2	1	3	2	3	3	6	2	3	7	16	20	21	15	30	41	2149								
	27.1	3	1	2	1	2	1	2	2	3	3	6	2	3	7	13	16	16	11	19	25	244	340							
VII	25.3	4	1	2	1	2	2	2	3	3	3	5	2	3	7	10	14	13	10	13	16	88	81	252						
	24.7	7	1	1	1	1	1	2	1	2	2	3	1	2	4	7	8	8	6	9	11	37	41	100	2435					
	24.1	13	1	1	1	2	2	2	2	3	3	5	2	4	9	9	12	14	14	16	16	72	76	157	1170	2338				
	23.2	10	1	1	1	1	1	1	1	1	2	2	1	2	3	5	5	5	4	6	7	17	18	33	132	207	1110			
	23.1	18	1	1	1	1	1	2	2	2	2	3	2	2	4	6	7	7	6	7	9	24	23	42	174	222	1276	9999		
VIII	15.8	14	1	1	1	1	1	2	2	2	2	3	1	2	3	4	4	4	3	3	3	7	6	8	13	9	24	11	16	
Class			I	II	III	IV				V					VI							VII								VIII

Classification on characteristic
" BLACK "

I ≥ 4
II < 4

See legend page 324.

Class	$\bar{X}$	Area	13	16	14	20	12	10	2	3	9	8	4	21	27	31	30	23	22	26	7	5	28	25	33	32	1	19	24	18
	9.41	13																												
	7.10	16	81																											
	7.00	14	63	8002																										
	5.91	20	29	74	78																									
	5.37	12	18	30	32	321																								
I	5.11	10	14	32	34	193	1742																							
	5.10	2	23	22	20	148	1190	9999																						
	4.47	3	17	16	15	54	120	307	253																					
	4.39	9	10	16	16	50	112	240	222	9999																				
	4.29	8	12	12	12	38	78	176	136	3382	9999																			
	4.08	4	15	9	9	28	51	115	68	657	1133	1922																		
	3.07	21	8	4	4	9	14	25	13	36	53	47	51																	
	2.57	27	8	3	3	7	10	18	7	21	31	25	19	162																
	2.52	31	5	3	3	7	9	15	9	19	25	22	24	178	9999															
	2.36	30	6	3	3	6	8	15	8	17	24	20	19	95	877	2110														
	2.33	23	6	3	3	5	7	13	6	13	21	17	14	73	516	1413	9999													
	2.26	22	5	2	2	4	6	12	4	10	18	13	10	51	238	656	3641	5276												
	2.14	26	5	2	2	5	6	11	5	11	17	13	11	44	147	331	813	831	1779											
II	2.13	7	5	2	2	4	5	10	3	7	15	10	7	31	77	255	518	510	1033	9999										
	2.13	5	7	2	2	5	7	13	5	13	19	15	10	42	90	326	717	643	1194	9999	9999									
	2.09	28	8	3	3	6	8	13	6	15	20	17	14	47	124	296	637	559	861	9999	9361	9999								
	2.04	25	6	2	2	5	6	11	4	10	17	12	8	32	61	200	331	288	403	2238	1542	1567	8790							
	2.00	33	7	2	2	5	6	12	4	11	17	13	9	31	56	177	276	231	293	1154	717	741	2673	7350						
	1.94	32	7	2	2	5	6	11	5	11	17	13	9	31	56	148	222	183	213	625	383	445	1165	1506	4583					
	1.86	1	6	2	2	4	6	10	4	9	15	11	7	24	35	106	136	111	121	283	168	174	391	395	692	2645				
	1.61	19	6	2	2	4	5	9	3	7	12	9	5	15	17	54	57	45	44	75	43	41	80	63	79	139	200			
	1.53	24	6	2	2	4	5	9	3	8	12	9	6	15	14	49	51	38	36	60	31	27	62	42	50	89	107	1583		
	1.30	18	4	1	1	3	4	7	2	5	9	6	4	10	10	28	26	21	20	30	18	18	29	24	27	38	43	124	202	
								I															II							

Classification on characteristic

" GREEN "

I $\geq$ 2

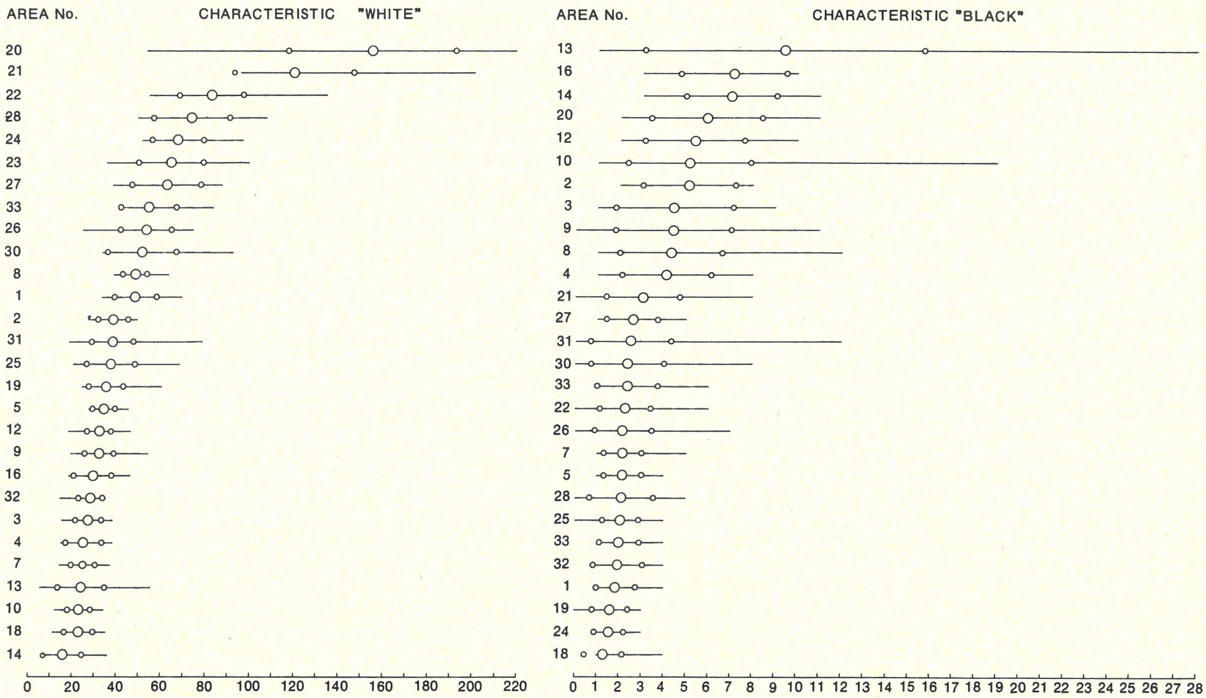
II $<$ 2

See legend page 324.

Class	$\bar{X}$	Area	2	20	13	5	12	14	28	21	16	33	26	24	3	27	7	23	30	31	10	22	32	8	25	4	1	9	19	18
	3.70	2																												
I	3.30	20	553																											
	2.36	13	27	38																										
	2.33	5	54	54	9999																									
	1.84	12	15	15	92	121																								
	1.77	14	20	16	71	104	3383																							
	1.00	28	18	12	19	36	48	74																						
	.90	21	7	5	10	12	18	17	2943																					
	.90	16	12	7	12	17	21	19	4585	9999																				
	.83	33	12	9	14	23	30	41	2239	4995	7830																			
	.71	26	10	9	15	22	31	41	776	988	1358	3926																		
	.67	24	9	6	9	12	14	14	382	196	196	1239	9999																	
	.60	3	8	5	8	10	11	9	230	87	54	551	3258	2034																
	.57	27	8	5	8	10	12	11	225	88	75	482	2183	1291	8306															
II	.53	7	3	3	6	6	10	10	96	94	99	219	768	743	2518	7478														
	.49	23	7	6	9	12	15	18	173	128	163	331	888	849	1909	3686	9999													
	.39	30	5	3	5	6	7	6	69	27	18	115	332	101	101	191	661	2431												
	.33	31	3	2	4	4	6	6	38	34	34	65	148	102	144	192	324	681	2561											
	.32	10	2	2	3	3	4	3	21	16	13	36	96	43	56	76	192	369	842	9999										
	.29	22	4	3	5	5	6	5	40	19	15	63	160	52	52	79	197	455	424	4747	3450									
	.25	32	6	4	5	7	7	5	63	18	10	85	192	49	32	62	166	444	193	1565	804	3676								
	.21	8	4	3	4	4	5	4	34	13	8	49	120	29	22	37	109	237	100	597	297	656	1574							
	.17	25	5	3	5	5	6	4	42	14	9	57	127	33	24	40	99	228	88	407	194	388	665	3296						
	.17	4	7	4	5	7	6	5	59	15	8	72	149	36	20	40	101	260	74	387	172	332	447	1846	9999					
	.05	1	5	3	4	5	5	3	31	9	4	39	83	17	8	16	52	113	24	121	51	68	48	85	161	103				
	.04	9	1	1	2	2	2	1	8	4	2	12	30	6	4	6	28	40	11	70	34	31	23	43	69	54	9999			
	.00	19	3	2	3	3	3	2	15	5	2	21	49	8	3	7	35	61	11	74	32	31	14	29	45	19	110	297		
	.00	18	2	2	3	2	3	2	12	4	1	16	40	6	3	6	30	49	9	66	29	26	11	24	35	14	85	266	9999	
			I				II																							

Fig. 1a, b, c
"Difference-coefficient"-matrices showing:
a) sample-size N amply sufficient to distinguish (with a confidence of 90%) a real difference as large as the stated difference between the used samples.
b) classifications as proposed by Maarleveld (1968).

N-numbers shaded: difference is non-significant
N-numbers not shaded: difference is significant
N-encircled : N-numbers $\leqslant 4$
code 9999: $\geqslant 9999$
Sample mean of characteristic ($\bar{X}$) and class-limits in ‰-values.



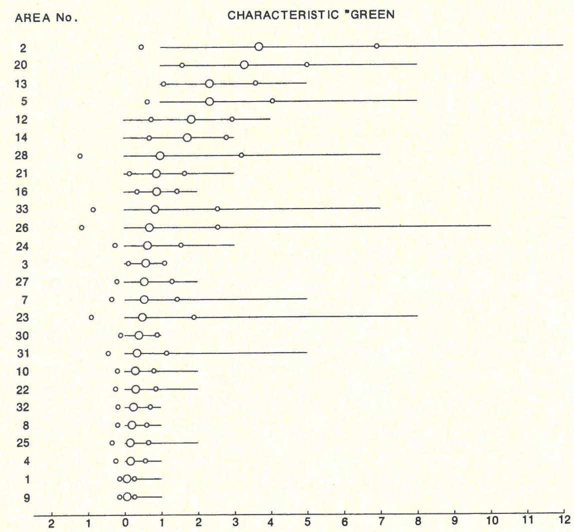


Fig. 2a, b, c
Sample statistics for 28 coversand areas.
O, sample mean ‰ ; o, standard deviation; ———— , range.

characteristic. The entries of the matrices[5]) give the sample size N amply sufficient to distinguish (with a confidence of 90%) a real difference in means as large as the stated difference between the used samples.

The difference between any pair of areas expressed as requisite sample size may be taken as a sort of difference-coefficient and is a useful measure for the practice.

If vital interests are involved and moreover the variable can be measured with great precision it may pay to take large samples in order to detect small but important differences; often however large samples will be prohibitive for reasons of practical and financial nature. This applies particularly to the present research; the worker in the field will only be interested if the magnitude of a difference is such that he can recognize it by comparing not more than a few cases. It seems reasonable therefore to consider about $N = 4$ as a maximum in this respect, so that for each characteristic the matrix was first subdivided into two groups: $N \leqslant 4$ and $N > 4$ with the interpretation: "easily distinguishable" and "hardly or not distinguishable".

The new expression N for the measured difference enables us at the same time to decide whether such a difference is significant, this being the case when the ratio of the *applied* sample size (n_h) to about 1/3 of the *calculated* sample size N is larger than unity. As equivalent to the applied sample size the harmonic mean of the sample sizes for each pair of areas used by Maarleveld was taken. For each characteristic a second matrix (not reproduced here) was now calculated with the above mentioned ratios as entries. N-numbers in the matrices of figures 1a, b, c are either significant or non-significant according as the corresponding entries in these second matrices were > 1 or $\leqslant 1$.[6]) This consideration taken into account, a grouping of the N-numbers into 3 parts rather than into 2 parts (cf. fig. 1a, b, c), is suggested viz.

1° N-numbers $\leqslant 4$; difference always significant.
2° N-numbers from 5 to about 60 à 70; difference for the greater part significant.
3° N-numbers larger than about 60 à 70; difference for the greater part non-significant.

---

[5]) Area 17 with only 8 samples was left out of consideration, so each matrix has $28 \times 27/2 = 378$ entries.

[6]) For a statistical exposition of the above given line of thought see the Appendix.

Both the first and the second group represent pairs of areas of which in other words it can be said that the sample sizes used by Maarleveld were sufficient. The third group consists of combinations for which the used sample sizes were not sufficient to detect significant differences. Though the intermediate group 2 is like group 1 characterized by combinations with significant differences, the N-numbers are for the greater part too large, which greatly reduces the practical importance of this group.

Thus it is obviously the first group containing the pairs of areas that can be easily discriminated, which interests us directly.

RESULTS

Mutual comparison of fig. 1a, b, c reveals that the matrix for WH has much more small N-values than the matrices for BL and GR. This applies in particular to the numbers $N \leqslant 4$ of the first group. The reason of this difference in distribution becomes apparent from comparing the data of the mean and standard deviation. From fig. 2a, b, c it may be seen that though these parameters for BL and GR are both smaller than for WH, the coefficient of variation (the standard deviation expressed as percentage of the mean) is for BL and GR just larger, sometimes much larger than for WH. Thus the variability of BL and GR *within* areas is on an average larger than that for WH. However the variability *between* areas (cf. the farthest left column of each matrix) is on a relative scale approximately of the same order for each of the 3 characteristics. As in the formula for N the whithin-area variability occurs as the variance (σ^2) in the numerator and the between-area variability stands in the denominator, it is easily seen that N is more often large for BL and GR than for WH.

We have now to consider the relation between the grouping of N-numbers and the classification of the characteristics used by Maarleveld. The class units subdivide the matrices in a number of triangles and rectangles – resp. classes and combinations of classes –. For what follows it is useful to introduce the notion of order and to distinguish 0^{th}, 1^{st}, 2^{nd}, ... order according as the difference between the classes to be compared amounts to 0, 1, 2, ... classes.

Of a suitable classification one may in general require that it has discriminative properties. This implies that the classes should be constructed such that members from one and the same class (0^{th} order) are so similar that they are hard to distinguish, those from the 1st order should be better distinguishable and so on.

With reference to what was said above with regard to the original planning of the research it may be required more particularly that to start with the 1st order the difference should be appreciable and significant; if not the method is of no practical use.

The matrices show that these conditions are poorly satisfied. Thus in the 1st order of fig. 1a (WH) only 15 significant numbers $N \leqslant 4$ occur on a total of 125 entries. In the 2nd order the ratio is better (72/85) but only from the 3rd order all combinations are easily distinguishable. This means that with regard to differentiating areas it does not matter much whether 2 areas belong to the same class or to neighbouring classes, a result that contradicts the postulated relation between the classes. For BL and GR the ratio in the 1st order is larger but still the distinction between 0th and 1st order is far from being evident. The total number of combinations $N \leqslant 4$ in the respective matrices is not only far less than is the case for WH, they are also more irregularly distributed (particularly for GR). This fact is mainly due to the larger within-area variance of these characteristics, the latter condition being moreover the cause that many areas cannot for all practical purposes be distinguished from any other area, regardless the distance between the respective means.

The reason why successive classes generally fail to exhibit sufficient contrast is closely connected with the unfavourable ratio between a) within-area variances and b) differences between area-means.

In the light of these facts the usefulness of the proposed classifications may well be questioned. The chosen intervals should in our opinion be taken only as a documentation serving to describe the areas in terms of the class-limits; not as an argument for that particular choice.

In interpreting M a a r l e v e l d's map (1968) the extra condition, which Maarleveld formulated with regard to fitting observations into classes (cf. Introduction), should be kept in mind. Accordingly we find that actually the number of observations falling outside the stated intervals of the legend-units is quite often as musch as 30%–50%, in a single case even exceeding 50%, a fact that of course seriously vitiates the discriminative power between the areas.

CONCLUSIONS

The question arises whether on the base of the given situation viz. if only the analytical results are considered there are indications that classifications with other class-limits would have been more effective.

From fig. 1a (WH) it is indeed apparent that e.g. by taking limits at 10‰, 20‰, 40‰ and 80‰, more homogeneous blocks are produced, resulting in a classification with greater discriminative power. Such class-limits have moreover the advantage that they are characterized by equal *ratios* which, when working with proportions (counts) should be preferred to equal *differences.* With regard to BL and GR (fig. 1b, c) it is harder to devise more suitable classifications owing to the rather irregular distribution of $N \leqslant 4$ numbers; yet the matrices suggest that additional limits in the neighbourhood of 1,70‰ for BL resp. 0,35‰ and 0,10‰ for GR would be effective in producing more homogeneous blocks.

The large and unpredictable variability of coversand over otherwise homogeneous geological unities inevitably leads to a sort of controversy in the approach of the classification problem: If the surveying is based on external geological phenomena it may very well be that the resulting map unities show maximum distinction only in that respect but at the same time poor differentiation as to analytical properties. The alternative would be to try to create populations with maximum discrimination in analytical respect, but it remains to be seen whether deliminating such populations would lead to satisfactory results for the geologist.

However a compromise is conceavable between these extreme viewpoints. Prior to start with definite sampling a tentative investigation should be carried out to get an impression about the variability of means and variances to be expected in the research area. These estimates may then serve as a helpful guidance in determining areas such that they may be combined in classes with sufficient discriminative power, thereby taking care not to maintain areas which are too similar mutually or to their neighbours.

For the research in question this would in several instances have implied combining smaller areas into larger unities. In doing so in a suitable way the difference between these larger unities generally increases, whereas the within-variance does not change appreciably. The result is a more significant difference (a smaller N).

It is recognized that some loss of field evidence is probably unavoidable, but is is felt that this loss is compensated for, if we can realize a more effective use of the analytical data.

ACKNOWLEDGEMENTS

This study forms part of a research program sponsored by the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

The first author in particular is indebted to Ir. Ph.Th. Stol of the Institute for Land and Water Management, Wageningen, for his cooperation and helpful advice during the initial stage of this work.

The final programming and calculations were carried out by the Computing Centre of the Agricultural University of Wageningen; the assistence of Dr.Ir. A.M.J. van Montfort is gratefully acknowledged.

The writers wish to thank Prof.Dr. G.C. Maarleveld for permission to make use of unpublished data.

REFERENCES

Maarleveld, G.C. (1966) – A simple method of characterizing sands. Tijdschr. Kon.Ned.Aardr.Gen. 83, 3, p. 272-277.

—— (1968) – Voorlopige resultaten van dekzandonderzoek met de plakbandmethode. Boor en Spade 16, p. 38-65.

Sokal, R.R. and P.H.A. Sneath (1963) – Principles of Numerical Taxonomy. W.H. Freeman and Co., San Francisco, 359 p.

APPENDIX

For two populations, numbered 1 and 2 we measure a variable $\underline{x}$. The means $\underline{x}_1$ and $\underline{x}_2$, obtained from resp. n_1 and n_2 samples are supposed to be normally distributed with expectations μ_1 and μ_2 and variances σ^2/n_1 and σ^2/n_2.

For the difference $\underline{d}$ between the means we then have:

$$\underline{d} = \underline{x}_1 - \underline{x}_2 \cong (\mu_1 - \mu_2) + (1/n_1 + 1/n_2)^{1/2}\ \sigma\underline{\chi}$$

Writing $\mu_1 - \mu_2 = \Delta$ for the expected difference and substituting the harmonic mean

$$n_h = (.5 * (1/n_1 + 1/n_2))^{-1}$$

of the respective samples n_1 and n_2, we have:

$$\underline{d} \cong \Delta + (2/n_h)^{1/2}\ \sigma\underline{\chi} \qquad (1)$$

Thus it is expressed that $\underline{d}$ is a normally distributed stochastic variable with expectation Δ and standard deviation

$\sqrt{2/n_h}\,\sigma$, $\underline{\chi}$ being a standard normal variable (that is N (0,1) – distributed, or normally distributed with expectation 0 and standard deviation 1). Keeping to essentials, we shall in the following assume σ to be known.

We call the observed difference d significant, if it satisfies $|d| \geqslant D_\alpha$, where D_α is to be chosen such that in the absence of a real difference ($\Delta = 0$), the probability of the statement: "$\underline{d}$ is significant" is at most 2α (does not exceed 2α), which can be written:

$P[\ |\underline{d}| \geqslant D_\alpha\ |\Delta = 0\ ] = 2\alpha$ or after substitution of (1): $P[\ (2/n_h)^{1/2}\sigma|\underline{\chi}| > D_\alpha\ ] = P[\ |\underline{\chi}| > D_\alpha/(2\sigma^2/n_h)^{1/2}\ ] = 2\alpha$

We also have

$P[\ |\underline{\chi}| > \chi_\alpha\ ] = 2\alpha$, χ_α being found from a table of the standard normal distribution from which follows:

$$D_\alpha = (2/n_h)^{1/2}\sigma\chi_\alpha \qquad (2)$$

Now we call N (β) a (β–) sufficient sample size with respect to a real difference Δ, if the probability is at least β to obtain (in a sample of size N) a significant observation d.

Taking $\Delta > 0$ we have:

$$P[\ |\underline{d}| > D_\alpha\ |\ \Delta\ ] = P[\ |\Delta + (2/N)^{1/2}\ \sigma\underline{\chi}\ | > D_\alpha\ ] =$$
$$= P[\ |\underline{\chi} + \delta\ | > \chi_\alpha\ ] > P[\ \underline{\chi} + \delta > \chi_\alpha\ ] =$$
$$= P[\underline{\chi} > \chi_\alpha - \delta\ ] \text{ where } \delta = \Delta/(2\sigma^2/N)^{1/2}$$

Putting $P[\underline{\chi} > \chi_\alpha - \delta\ ] = \beta$, we have $\chi_\alpha - \delta = \chi_\beta$.

Moreover: $P[\ |\underline{d}| > D_\alpha\ |\Delta\ ] > P[\ \underline{\chi} > \chi_\alpha - \delta\ ] = \beta$, so that in this case N is a sufficient sample size with respect to δ.

We also have $\chi_\alpha - \chi_\beta = \delta = \Delta/(2\sigma^2/N)^{1/2}$, from which follows:

$$N = \frac{2\sigma^2}{\Delta^2}\ (\chi_\alpha + \chi_{1-\beta})^2 \qquad (3)$$

Choosing $\alpha = 0{,}05$ as a norm for "significant", $\beta = 0{,}50$ for "sufficient" or $\beta = 0{,}90$ for "amply sufficient" we have:

$$\chi_{.05} = 1.645,\ \chi_{.50} = 0,\ \chi_{.10} = 1.280$$

N_{50} is "sufficiently large", N_{90} is "amply sufficiently large", where

$$N_{50} \text{ is } 2\sigma^2/\Delta^2 * 1{,}645^2 = 5{,}412\sigma^2/\Delta^2$$
$$N_{90} \text{ is } 2\sigma^2/\Delta^2 * (1{,}645 + 1{,}280)^2 = 17{,}112\sigma^2/\Delta^2 \qquad (4)$$

$$N_{50}/N_{90} = 0{,}3163.$$

With formula (3) we can also express the difference Δ by N_{90}, viz. that particular sample size which is amply sufficient to "distinguish" the difference Δ. This sample size may be regarded as a measure of the costs involved in procuring the samples.

In the same way we can with formula (3) express the observed difference $\underline{d}$ in a new stochastic variable $\hat{N}$:

$$\hat{N} = \frac{2\sigma^2}{d^2}\ (\chi_\alpha + \chi_{1-\beta})^2 \qquad (5)$$

$\hat{N}$-values have been tabulated in figures 1a-c with $\alpha = 0{,}05$ and $\beta = 0{,}90$. The number $\hat{N}_{90}$ gives the sample size amply sufficient to distinguish a real difference Δ as large as the observed difference.

If on the basis of the new expression $\hat{N}$ for the observed (= measured) difference one wishes to decide if there is question of a significant difference, it may be observed (by substituting (5) and (2)) that:

$|d| > D_\alpha$, if and only if:

$$\frac{1}{\hat{N}_{50}} = \frac{d^2}{2\sigma^2} \cdot \frac{1}{\chi^2_\alpha} > \frac{D^2_\alpha}{2\sigma^2} \cdot \frac{1}{\chi^2_\alpha} = \frac{2\sigma^2\chi^2_\alpha/n_h}{2\sigma^2\chi^2_\alpha} = \frac{1}{n_h}$$

that is if $n_h > \hat{N}_{50} = .3163\ \hat{N}_{90}$ or in other words if the applied (harmonic mean) sample size n_h is larger than $\hat{N}_{90}$, calculated from the observed difference.