SPÄTH, Helmut: **Cluster-Formation und -Analyse.** (Cluster formation and analysis). München–Wien: R. Oldenbourg Verlag 1983. 236 p., with many figures, tables and programs, DM 84,–, ISBN 3-486-27441-4.

The formation of classes or 'clusters' of objects which show a maximal internal homogeneity or similarity, is a process which can combine conceptual efforts with mathematical arguments and algorithms. Depending on the purpose of the classification sought as well as on the prior assumptions and information on the underlying objects and on their characterizing features (variables), one or the other method will be preferred. The book of H. Späth treats numerical classification and clustering methods. The use of these methods requires that 'similarity' between objects can be measured numerically. More specifically, considering m objects $k = 1, \ldots, m$, it is supposed that s quantitative (sometimes: ordinal or binary) features have been measured for each object and combined to give m s-dimensional data points $x_1, \ldots, x_s$ in $\mathbb{R}^s$. Similarity between objects is characterized by some distance $d(.,.)$ between corresponding data points. The applied problem of finding an 'optimal classification $\mathscr{C}$' of the set of objects $S = \{1, \ldots, m\}$ is made precise by the mathematical problem of searching for a partition $\mathscr{C} = \{C_1, \ldots, C_n\}$ of S comprizing n classes such that a clustering criterion of the type $W(\mathscr{C}) := \Sigma_i \Sigma_{k \in C_i} d(x_k, z_i)$ is minimized over all possible partitions $\mathscr{C}$; here $d(x_k, z_i)$ denotes some (quadratic) distance between x_k and z_i, a characteristic representative of the class C_i (e.g. the corresponding mean value, or some class-specific regression line). Other criteria are considered, too, e.g. the well-known determinatal criterion or some adaptive distance criterion. A solution is calculated (approximated) by minimum-distance (k-means) algorithms or by an iterative exchange of objects between classes. Neither hierarchical classifications nor probabilistic models or investigations are included.

The book introduces the mathematical concepts and algorithms (Part I, 106 p.), it presents a series of corresponding FORTRAN programs (Part II, 42 p.), and finally gives some illustrative numerical examples for comparing and evaluating the various methods (Part III, 70 p.). It concentrates on the mathematical and algorithmic aspects (i.e. without discussing real life problems or the interpretation of results) and contains some exercises at the end of each chapter. Actually, I know no other book where the topic is presented with the same degree of clarity and internal consistence between the three stages I, II, and III. Given that only matrix algebra is needed as a prerequisite, the book is to be highly recommended not only as an introductory text for students and research workers in statistics or data analysis, but also for practitioners from all fields of applications and concerned with clustering problems.

H.H. Bock

Prof. Dr. H.H. Bock
Inst. f. Statistik u. Wirtschaftsmathematik
RWTH Achen, Wüllnerstr. 3, D-5100 Aachen

Letters to the Editor

Dear Editor!

In I.C. 1983, No. 1 there was a review by Mr Eric Coates of my report FID/CR No 17 "Research on Classification Systems". This review is, however, not based on the FID publication, as a review-copy of this – approved in 1975, published in 1979 – had not been sent to I.C.. The review is, however, based on a mainly identical edition from the Swedish Council for Building Research printed in 1978, as the FID publication was so badly delayed. I have asked Mr Coates myself to write a review as I had confidence in his competence.

Although I found this review not so critical as reviews in I.C. often use to be, I want to take up some questions where my own opinion differs from Coates' or where some misunderstandings occur. As I know that readers generally prefer to observe the critical points in a review, neglecting the positive ones, and as my work in a distant arctic country is not well known on the continent (and apparently seriously disliked in the Indian headquarters), I will try to make my own opinion quite clear.

1) "Wahlin resembles BSO" says Eric de Grolier in his contribution to the FID/CR conference in Augsburg in 1982, where he analyses five Post-World War II universal systems including my US from 1969 (in fact published already in 1963 in Journal of Documentation and in 1966 in American Documentation etc.) and my FS from 1974 (published that year in International Classification No 1). Certainly there is much resemblance between my proposals and BSO, especially if we compare with UDC. The influence, if there is any, could, however, only have gone in one direction. The series Mathematics, Physics, Chemistry, Biology etc. are in broad lines common to our system and also for other systems published in the last decade. In concept terms this corresponds in my US to Number, Space, Time, Motion, Mass, Energy, Matter etc. (not starting from Energy as Coates says).

2) The short description by Coates of my proposal for a universal system with decimal structure (US) is well composed. Even if the following addition has no relation to the review, I ask the editor to give me permission for mentioning the TIM-principle, presented at the Augsburg Conference in 1982, and included in the proceedings part I and II. This includes for my systems a certain alteration in the technical area. T, I and M represent Technology, Industry and Material culture, the latter being the useful products of industry (= all production). This three-part division on the highest level or on branchlevels seems, after several trials, to be a way to bring a better order in the corresponding part of different universal systems, as for every of these three main areas, natural and suitable principles for the subdivision can be attained. The TIM-principle is based on that often neglected idea of activities emanating from something and resulting in something, thus bringing the system in contact with what is going on in our society and also with the statistical systems.

3) My title is not adequate, Coates says. Maybe it de-

pends on the fact that this report was translated from a Swedish manuscript, where information on other systems etc. occurs, useful for Swedish readers, but perhaps by Coates considered unnecessary for an international expert public.

This information comes from my studies of different papers and books for selecting materials that support my theories or perhaps are contradictory to these, selected with the aim of throwing light on the problems. As examples I will point out systems of Vannerus and Tykociner, systems for encyclopedias and statistical systems that I think are not very well known by all classification experts. To bring forward such material should, in my opinion, merit the name of research.

4) I make discoveries fromactionratherthanfrom abstract thinking but "never dwelve in the subterranean body". To the subterranean body of classification I have devoted a lot of interest. An example is the study of the bottom layer of an agricultural system. As you cannot come deeper here, I regard this as a visit, not to Hades, but to a subterranean fruitful garden. The current precoordinated agricultural code F 25 116 for "decomposition of carbohydrates by bacteria in soil" corresponds in the Field System (FS)[1] to a combination of codes from geology (G), biology (B) and chemistry (K) and this subject belongs to each of these sciences. A complex science with two components belongs to both parent sciences, as well as Eric Coates belongs to both his father's and mother's family hierarchy. Sometimes a separation in two part-sciences is generally agreed upon as with physical chemistry and chemical physics, but even if no agreement exists there can often be reasons for discussing the separation of a combination of a and b in a:b and b:a, these two being part-sciences in a and b respectively, as shown with the example on soil and microorganism mentioned in the review.

5) In BSO genetics is included under 310 Biological sciences, 320 Microbiology, 320 Botany, 340 Zoology, 360 Agriculture and 410 Biomedical science. As the very important science genetics is the summary of these five subject areas, why not present genetics under Biological Sciences as such a summary with (or without) corresponding reversed codes in the positions concerned, etc.

Comparison	*FS*	*BSO*
Genetics	B15	313,70
Microorganisms	B15:B2	320,37
Plants	B15:B3	330,37
Animals	B15:B4	340,37
Man	B15:B5	413,7

By this method we give preference to the pure science, by Coates' arrangement to the application or perhaps to the idea that any complex subject should be introduced only when both components have appeared in the system.

In my opinion the combined codes are useful also as a means for consistency in the sometimes obscure flora of Greek or Latin worded disciplines. The whole system gets a twodimensional character, and they bring about a diminishing of the number of simple codes.

Regarding Coates' comments on the distinction between Social Psychology and Psychology of Societies I want to ask if the latter science has any meaningful bearing as psychology is connected with individuals.

6) The codes K for chemistry and F for physics are, Coates says, meaningful only in Swedish, but please observe that with regard to the pronounciation they fit well also for the English language. Even if English is rapidly going to be an international language it should not be necessary that other people should suffer too hard from its lack of consistency between pronounciation and spelling.

7) Regarding the AR-principle, there is a misunderstanding for which I am responsible myself. What I want to say is, that for a specialized system, adapted for a certain branch or area (A-systems), one can avoid going deeply and include a lot of special subheadings if one has a reference system (R-system) of universal type, containing all the basic concepts. We get a special advantage if the same R-system is used for many A-systems, when the R-system is a common link between the A-systems. Also different universal systems can be connected in this way, with the R-system as a "standard reference code". The R-system can also be used independently as a universal system. The AR-idea is better explained – for German-reading people – in my article in DK-Mitteilungen 1979, No 1/2, where also application for product and occupation classification is accounted for.

8) A rather hard criticism is expressed by saying that I "offer neither practical solutions nor an embracing theory". Coates alludes here, I suppose, to a lack of strictly formulated and detailed rules instead of giving different alternatives. I give alternatives as every method or system has its "pro et contra" and I prefer to leave the questions open. In the report my systems are presented in broad lines, detailed structures are, however, elaborated both for document and product systems. As an embracing theory I myself have considered the widening of the concept classification outside the documentation area to systems for e.g. products, patents, standardisation, statistics, education, term lists, editorial work, for encyclopedias etc.

In classification theory agreement is rare and there seem to exist perhaps as many opinions as there are classificationists. Seen against this background it is not surprising that in the review and in my reply different opinions come out – even if a certain agreement exists. Could FID/CR with the help of I.C. keep hold of the differences and act for clearing up different problem areas by analysing the reasons for divergencies (different aims, background etc.) this should, I think, be a very useful activity.

Ejnar Wåhlin
Birger Jarlsgatan 102
11420 Stockholm, Sweden

1 As obviously very few readers have the report at hand I refer to I.C. 1974, No 1 or Zagadnienia Naukoznawsta 1973, No. 4 concerning FS.