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-