REGIONALIZATION REVISITED: AN EXPLICIT OPTIMIZATION APPROACH

Jan W. Owsinski*

October 1980
CP-80-26

*Research Scholar, Systems Research Institute, Polish Academy of Sciences, Warsaw.

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

ABSTRACT

This paper presents an optimization framework for the regionalization problem. Having a set of regions for which an allocation problem is defined in the presence of nonlinear spatial characteristics related to grouping of the regions, one has to solve the allocation-and-grouping problem.

Indications are given as to the nature of nonlinear spatial characteristics, making the grouping aspect of the problem appear, together with an exemplary form of the model.

The complexity of a problem thus conceived makes the direct application of complete enumeration, dynamic programming and branch-and-bound impossible. Thus, an iterative scheme is proposed in which the partial and simplified problems are solved at each step of the iterative procedure.

In order to follow this procedure, a method was required that would yield groupings or clusters optimizing a certain objective function. A new, general, objective function for clustering is formulated, providing both optimal partitions and the number of clusters. It is presented along with the method in the second part of the paper in which other methods of solving the regionalization problem are also reviewed.

CONTENTS

REGIONALIZATION REVISITED: AN EXPLICIT OPTIMIZATION APPROACH

Jan W. Owsinski

INTRODUCTION

Regionalization has long been treated as a mere statistical exercise, mainly with a cognitive purpose. In fact, planning models with spatial dimensions usually take an existing structure as given, or, if a choice can be made, its--rarely explicit--criteria account for the data accuracy and model size. On the output side, these models allocate certain activity magnitudes to regions adopted at the outset. Although it is recognized that the location/allocation procedure is to a large extent recursive, existing multi-level techniques also regard the hierarchical structure as given, focussing on interaction and coordination processes. Methods applied to decompose systems by such techniques use, again, mainly numerical facility criteria.

Having said this, attention should be turned to the theories and analyses which postulate that the existence of non-trivial phenomena related to spatial structure requires a normative analysis of this structure. These views have led to spatial production complex approaches, based upon agglomeration and scale economies. The approaches mentioned have as yet not found any modeling counterparts on the interregional level. The present summary report shows an outline for a framework of such a model and for the algorithmic solution conditions.

PART I. THE REGIONALIZATION-AND-ALLOCATION PROBLEM*

1. Statement of the Problem

Suppose there is a set of basic spatial units to which one has to allocate resources in such a way that a certain output function defined over possible allocation schemes attains its maximum. Suppose also that the function to be maximized has to account for the following items:

a. the units are physically located in geographical and economic spaces so that besides the efficiency parameters there are distances (or proximities) defined for each pair of them;
b. transportation costs are nonlinear with regard to distances among spatial units;
c. there are agglomeration economies, relating output to allocation, efficiency and distance;
d. agglomeration economies, from c., can only be realized fully when appropriate administrative divisions correspond to them; and
e. the creation/modification of administrative divisions mentioned in d. is connected with two types of costs:
    i. cost of running the structure; and
    ii. cost incurred by moving away from optimality when efficiencies and distances change, while administrative divisions remain.

These issues will be commented upon in more detail in I.2.

Thus, an allocation/location problem was linked with regionalization. The problem generally formulated above would, even for a possibly simple form of the objective function, be very complex. It is indeed quite unprecedented in its formulation

*This is a shortened summary of a more detailed report, to be published at a later date, containing all necessary formal considerations. The present report is meant to provide a general idea of the problem formulation and methods of solution, to raise doubts and discussion, and thereby contribute to further work.

since allocation models rarely account for agglomeration economies and it is unusual for regionalization analyses to apply global objective functions, to say nothing of the joint consideration of both.

Before proceeding to a more detailed consideration of the elements of the problem, and therefore the possibilities of its solution, one important remark is necessary: a set of elements, together with their attributes and interrelations, can only be viewed as a system when a goal or purpose has been defined for it (whether within or outside the system).

This quasi-definition stipulates that a regionalization procedure can be meaningful only when it has an explicit purpose or goal.

2. Elements of the Problem

When defining distances among basic units, one should consider the actual geographical situation (e.g. bordering) and the socio-economic links (rather than similarities) such as migratory flows and input/output relations, or simply transportation flows. This is necessitated by the normative and not the descriptive nature of the model. It is therefore not useful to employ a large number of spatial units' characteristics when defining distances. However, the question of distance properties remains open since it is closely related to the objective function formulation and to the solution algorithm.

Each basic spatial unit has an attribute of efficiency parameter relating an output magnitude, say: net product value of the unit, to an input magnitude, say: fixed assets complemented with some current expenditures within the unit.

The transportation cost function is given in such a way that it refers directly to distances appearing in this problem.

The agglomeration economies are basically given through the relations of efficiency parameters to such agglomeration-based magnitudes as, for example, population density, employment, global product, and again, fixed assets. The fixed assets should not entail any undue correlation, since the explicit dependence on it can simply be a constant. Attention should, however, be paid

to product dependence, insofar as distance and therefore also the transportation cost function and the effect thereof, might as well embody product interrelations among basic units.

The hypothesis behind this efficiency function is that it preserves its validity over larger spatial entities and over longer time horizons. Thus, on the inter-unit scale, efficiencies will change depending upon delimitations of multi-unit regions. The problem is to determine the overall composition of regions so as to maximize the global efficiency. This could be done without reference to the actual allocation of resources, were it not for two reasons: first, that any structuralization incurs costs in terms of resources currently available, and second, that actual allocation changes (although to a small degree) the parameters of the system. Of these two questions, the first can be dealt with directly in the model by means of appropriate modifications to the objective function. The modifications would have to be left to the post-optimal analysis phase, whence optima in a broader sense can be found.

3. A Formulation

The considerations presented earlier can now be exemplified for a possibly simple case.

Assume there is just one type of resource to be distributed, whose volume is denoted by x, among n basic spatial units, so that unit i gets x_i, $i \in I = \{1,\ldots,n\}$, of the resource. (It would be simplest to take capital as x.) Distances among units are denoted by d_{ij}.

The effect of the spatial organization appears in the objective function through the efficiency functions based upon agglomeration economies and the transportation cost functions.

In order to define the objective function dependent on allocation and spatial structure, let us first introduce hierarchy H, of the spatial units, i.e. the way in which they will be organized. A hierarchy H is a subset of the power set of I, $H \subset 2^I$, of the form $H = \{P^1,\ldots,P^r\}$, where $P^1 = I$, $P^r = \{I\}$, and P^k are partitions of the set I, i.e, $P^k = \{A^{k1},\ldots, A^{kl},\ldots,A^{kL_k}\}$, $\bigcup_{l=1}^{L_k} A^{kl} = I$, $A^{kl} \cap A^{kl'} = \phi$, $l \neq l'$. Furthermore,

for each $k < r$ and for each l such that $A^{kl} \in P^k$ there exists l'' such that $A^{kl} \subseteq A^{k+1,l''}$, and for each P^k there exists an l such that $A^{kl} \subsetneq A^{k+1,l''}$, where l'' is defined as before.*

For any subset A^{kl} of I, the partial objective function $q(A^{kl})$ is defined as:

$$q(A^{kl}) = q^{kl} = \left(\sum_{l' : A^{k-1,l'} \subset A^{kl}} q^{k-1,l'} \right)^{f^k(A^{kl}) \cdot g^k(A^{kl})} ,$$

which for $k = 1$ amounts to

$$q^{1l} = (x_l)^{f^1(l) \cdot g^1(l)} .$$

Thus, independent activities are accounted for in every basic unit and in every grouping of these units; the global output, however, is determined by their overall structure:

$$A = q^r = \left(\sum_{l : A^{r-1,l}} q^{r-1,l} \right)^{f^r(I) \cdot g^r(I)} ,$$

where $q^k = \sum_{l : A^{kl}} q^{kl}$.

Functions f and g correspond to efficiency and transportation costs, respectively. The illustration of their presumed shapes is shown in Figures 1, 2 and 3, together with appropriate clarifications. These figures will not be commented upon, since the assumptions behind them seem to be obvious. Forms of f and g, such as those presented, can be used for a theoretical analysis meant to show the general dependence of the features of optimal hierarchy H and allocation X on parameters of these functions. On the other hand, however, such functions are being identified empirically as, e.g. in the studies of the optimal urban size and can therefore be used in an optimal planning model.

*This definition is more complicated than the general one, stipulating that for $A^l \in H$, either $A^l \cap A^{l'} = \phi$ or $A^l \subsetneq A^{l'}$, $l \neq l'$, but it is more suitable for the problem considered here.

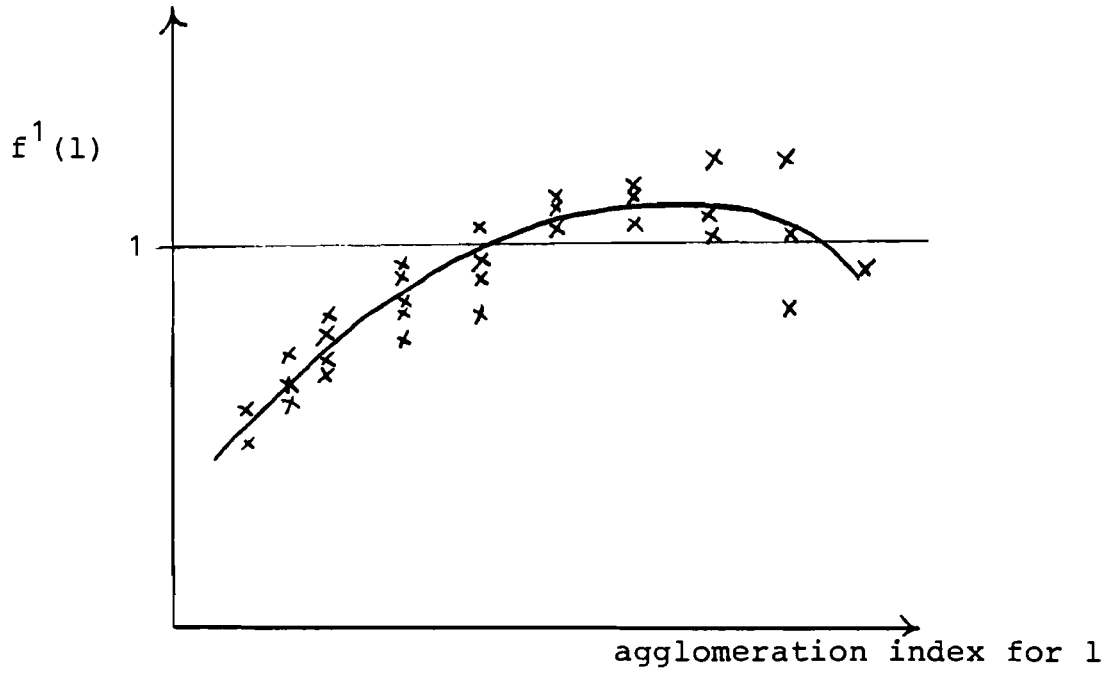

Figure 1. Basic spatial unit's efficiency.

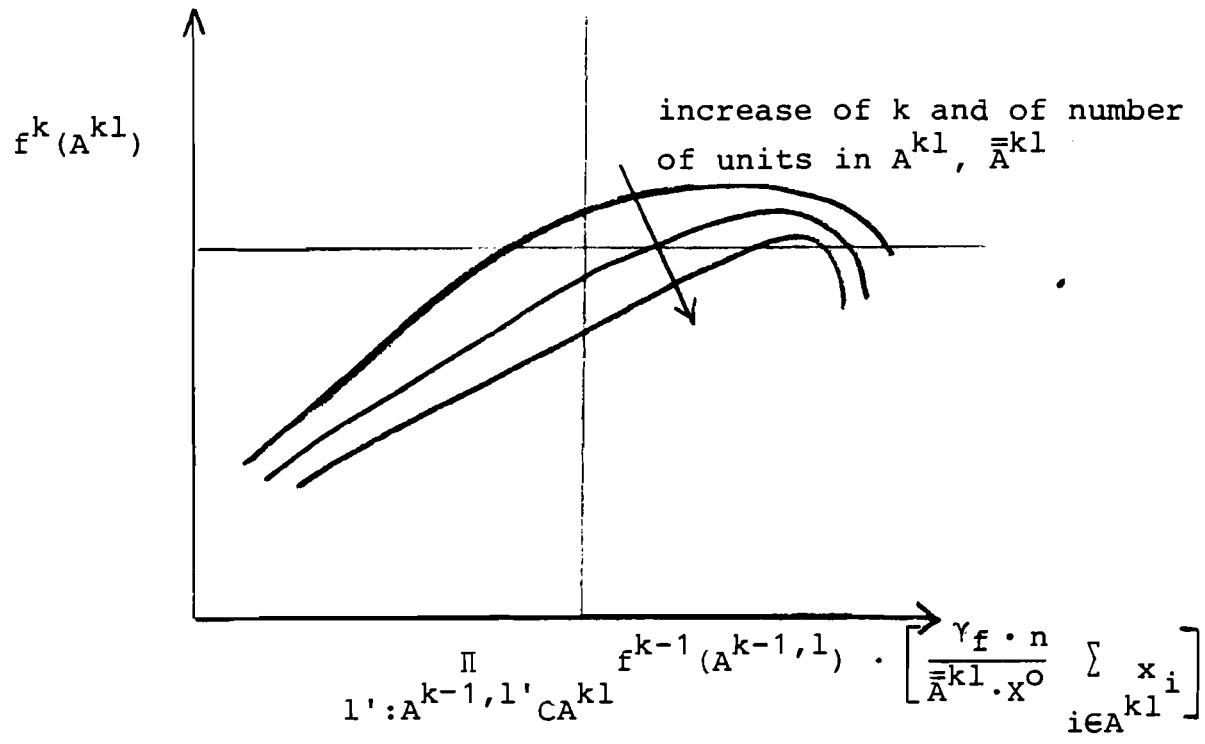

Figure 2. Formation of the aggregate efficiency functions.

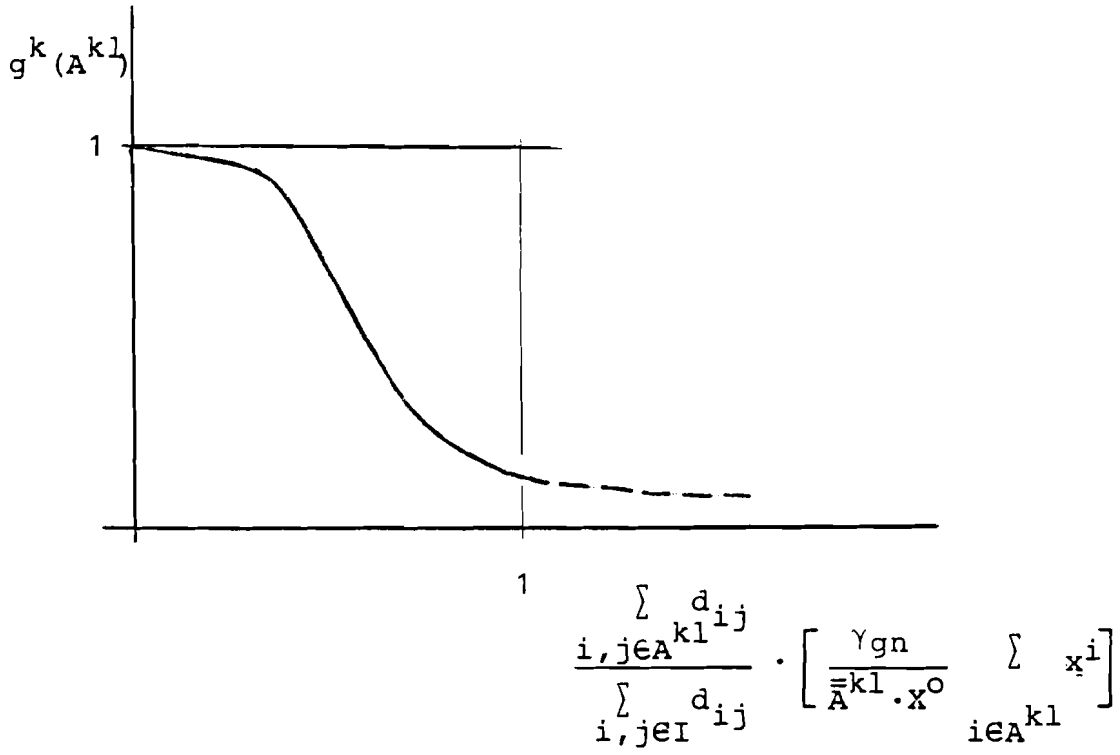

Figure 3. Efficiency decrease due to transportation costs.

The problem is therefore to find H^{opt} and X^{opt} such that $Q(H^{opt},X^{opt}) \geq Q(H,X)$, $\forall$ H,X , $X = \{x_1,\ldots,x_i,\ldots,x_n\}$, subject to

$$\sum_{i\in I} x_i \leq X^o \quad ,$$

$$b_i \leq x_i \leq c_i \quad ,$$

where the second set of constraints stipulates that the spatial units be given at least some minimal (maintenance) level of the resource, and that they cannot be given more than they can consume.

The shapes of f and g, and especially of f depend strongly on the type of output measure to be maximized. Thus, according to various types of such measures, different optimal structures H^{opt} and allocations X^{opt} will appear.

4. Computational Questions

The problem as outlined here cannot be solved by complete enumeration with regard to H, even though for functions f and g independent of x one can obtain solutions for H and X separately. The number of possible hierarchies is absolutely inhibitive for dynamic programming or branch-and-bound techniques, when only the dimensions of the problem exceed those of an academic example. This is caused primarily by the fact that the search has to be performed in the space of hierarchies. However, even when the problem is artificially constrained by setting $r = 3$, i.e. the minimal number of levels for this formulation, for which just one non-trivial partition is sought, P^2, the number of possibilities is still too large for any of these techniques. In fact, even for quite simple objective functions, the optimization methods were used only to find such optimal partitions P^2 which were composed of a given number L_2 of subsets A^{21}.

That is why an approximate approach must be employed, unless one assumes some very particular properties for the problem.

It is proposed that the procedure would be based upon a grouping (clustering) algorithm, which will optimize an objective function, so that its results can be controlled and relations established with the overall problem. The algorithm will first optimize the structure with regard to g through the values d_{ij} and then modifications will be brought into d_{ij} based upon values of f for groupings previously obtained. In case of inseparability of H and X these modifications could also account for the influence of allocated x_i on both function f and g.

It turns out however, that no explicit optimization approach exists for grouping. Part II comprises a shortened overview of the approaches available and the proposition of the global objective function for grouping together with the optimization algorithm.

PART II. THE GROUPING/CLUSTERING PROCEDURES: A SEARCH FOR AN OBJECTIVE FUNCTION

1. The Nature of the Trouble

Mathematics is a tautological system. Hence, it is often conjectured that when taking intuitively obvious elementary assumptions and rules of reasoning one should reach equally intuitively obvious results. Experience shows that this is not true. The main reason is the incapacity of predicting the far-off consequences of initial assumptions, especially when there are a number of qualitatively similar assumptions, among which a choice should be made. This is especially true when the results have a highly complex and multidimensional nature.

However, for simple cases it is possible to formulate assessments concerning the outlook of the results. When these assessments take on a more precise and general form, i.e. they are analytically expressed in appropriate formulae, they can be utilized throughout the solution of the problem, together with the initial elementary assumptions. Such is, for instance, the correctional sense of some constraints in the economic problems formulated as mathematical programming tasks.

The same applies to the clustering problems, encountered in data analysis, taxonomy, classification, pattern recognition, etc. There is a choice of elementary assumptions concerning either local distance, or a similitude of elements in a population, or more global in-group homogeneity versus inter-group diversity criteria. The clustering methods in constructing their algorithms are based upon these criteria. Over larger and complex populations it is difficult to assess the adequacy of the methods applied, since the results are then by no means intuitively analyzable. When analyzing simple examples, one can easily see the inherent biases of the algorithms. In order to eliminate these biases, or to make them controllable, it is necessary to impose an explicit global criterion. Furthermore, it can be hoped that such a criterion could also help in solving the problem of the optimal number of groups or even optimal hierarchies in addition to the usually solved problems of group composition.

2. The Problem

Having n elements indexed i, $i \in I = \{1,\ldots,n\}$, whose mutual "distances" or "dissimilitudes" $d_{ij} = d_{ji} \geq 0$ define the triangular matrix D, to find partition P^k of the set I, $P^k = \{A^{k1},\ldots,A^{kl},\ldots A^{kL_k}\}$, such that elements i belonging to one group A^{kl} are more "similar" than those belong to various other groups.

3. The Elementary Assumptions

These assumptions do in fact constitute an interpretation of "similarity" or "likeness" mentioned in the problem formulation.

A. The elements i,j are more similar than i,k iff $d_{ij} < d_{ik}$.

B. The elements belonging to groups A^{kl}, $A^{kl'}$ are more similar than those belonging to A^{kl}, $A^{kl''}$ iff $d(l,l') < d(l,l'')$, where [Duran and Odell (1974)]

a. $$d(l,l') = \max_{\substack{i\in A^{kl} \\ j\in A^{kl'}}} d_{ij} \quad ;$$

b. $$d(l,l') = \min_{\substack{i\in A^{kl} \\ i\in A^{kl'}}} d_{ij} \quad ;$$

c. $$d(l,l') = \frac{1}{\bar{\bar{A}}^{kl}.\bar{\bar{A}}^{kl'}} \sum_{\substack{i\in A^{kl} \\ j\in A^{kl'}}} d_{ij} \quad ,$$ [Diday and Simon (1977)] ,

or any other distance-like function of i,j, $i\in A^{kl}$, $j\in A^{kl'}$, e.g.:

d. $$d(l,l') = \frac{\bar{\bar{A}}^{kl}.\bar{\bar{A}}^{kl'}}{\bar{\bar{A}}^{kl}+\bar{\bar{A}}^{kl'}} (\bar{X}^{kl}-\bar{X}^{kl'})^T(\bar{X}^{kl}-\bar{X}^{kl'}) \quad ,$$

where $\bar{X}^{kl} = \frac{1}{\bar{\bar{A}}^{kl}} \sum_{i\in A^{kl}} x_i$, $x_i = \{X_{i1},\ldots,X_{is}\}$ being the vector characterizing point i. Medians or centroids can also be used;

C. The elements i,j are similar iff $d_{ij} \leq \varepsilon$, where

a. ε is arbitraily chosen [Bielecka et al. (1979), Fortier and Solomon (1966), Tremolières (1979)];

b. ε is a function of d_{ij}, $i,j \in I$, e.g., [Bielecka et al. (1979)]

$$\varepsilon = \frac{2}{\bar{\bar{I}}(\bar{\bar{I}}-1)} \sum_{\substack{i,j \in I \\ i<j}} d_{ij} = \bar{d} \quad ;$$

D. The elements i,j are more similar than i,k iff

$$\delta_{ij} < \delta_{ik} \quad , \qquad \text{where for } \forall(i,j,k),\ i,j,k \in I \quad ,$$

$\delta_{ij} \leq \sup (\delta_{ik}, \delta_{jk})$, and a δ_{ij} can be outlined from d_{ij} through a simple algorithm given in Diday and Simon (1977) or in Hartigan (1975).

E. The elements belonging to the same groups A^{kl} are similar while those belonging to various groups are dissimilar [Diday and Simon (1977)], iff

$$\forall A^{kl} \in P^k \forall i,j \in A^{kl},\ m \notin A^{kl} : d_{ij} \leq d_{im} \wedge d_{ij} \leq d_{jm} \quad ;$$

F. The elements belonging to a group A^{kl} are similar, while those belonging and not belonging to it are dissimilar [Kacprzyk and Stanczak (1975, 1978), iff

$$W(A^{kl}) = \sum_{\substack{i \in A^{kl} \\ j \notin A^{kl}}} w(d_{ij}) < \sum_{\substack{i \in A^{kl} \\ j \in A^{kl}}} w(d_{ij}) = w(A^{kl}) \quad ,$$

where $w(d_{ij})$ is a function of "similarity" or "linkage" decreasing in R_+^1.

4. Construction of a Procedure

The initial assumptions formulated here are for the most part intuitively acceptable, if not obvious--with the exception perhaps of assumption D. which refers to the notion of ultrametric. Having these assumptions, being in fact local similarity/dissimilarity criteria, one can proceed to the construction of groupings $A^{kl} \in P^k$.

The groupings can be constructed directly on the basis of assumptions A.B. The method is very simple; at each step two elements or groupings are merged for which appropriate d is minimal. Or those are separated for which appropriate d is maximal (see Bielecka and Szczotka (1978), Byfuglien and Nordgård (1973) for some such algorithms). In this way, hierarchies H are obtained, although in no way can they be compared for their "adequacy".

In the case of the direct pair-wise application of A.B there is additionally in the definition of a hierarchy:

$$\forall k \; \exists l < l' : P^{k+1} = \{A^{k1},\ldots,A^{k,l-1},A^{kl} \cup A^{kl'},A^{k,l+1},\ldots, A^{k,l'-1},A^{k,l'+1},\ldots,A^{kL_k}\} ,$$

where P^k is a partition of I containing A^{kl} and $A^{kl'}$.

Obviously, it follows directly from the above definition that for each $A^{kl} \in H$ there exists a partition $P^k \ni A^{kl}$, $P^k \subset H$. This condition stipulates a series $\{P^k\}_{k=1}^n$ of partitions, $\{P^k\}_{k=1}^n = H$, $P^n = \{I\}$, $P^1 = I$.

Thus, for given I and D, for each of the local criteria B. a different hierarchy H(B) is obtained. The methods, related to assumptions B. are referred to as "complete linkage" (Ba), "single linkage (Bb), "average linkage" (Bc) and the Ward technique (Bd). Their generalization was proposed in Lance and Williams (1967) and broadened in Wishart (1971) to include the Ward technique. Originally, the Ward technique proceeded by joining these groups, for which the increase of

$$S = \sum_{i \in A^{kl}} (x_i - \bar{x}^{kl})^T (x_i - \bar{x}^{kl}) ,$$

resulting from joining was minimal. By denoting the above value of the "error sum of squares" as S^{kl}, analogous for $A^{kl'}$ as $S^{kl'}$, the one for $A^{kl} \cup A^{kl'}$ as $S^{kll'}$, and the one from Bd as $S^{kl/l'}$ it can be shown that

$$S^{kll'} = S^{kl} + S^{kl'} + S^{kl/l'} \quad .$$

Hence, the original Ward approach is equivalent to sequential hierarchical grouping with Bd.

Still another hierarchy can be obtained for a given set of elements and "distances" when using the ultrametric δ_{ij}, as defined in assumption D. In fact, the ordering of δ_{ij}, when obtained already from d_{ij}, yields directly a hierarchy.

The main advantage of the methods mentioned above is that they are numerically very simple. However, they do not provide any measure for determininiming the "best" k among $k \in N\ [1,\ldots,n]$, for $H = \{P^k\}_1^n$, neither for a choice of an H among a variety of them. In establishing a hierarchy, these methods utilize the wholly local criteria and therefore do not make it possible to compare their P^k's for any k. Thus, in spite of the obvious biases of some of these methods in terms of "propensities" to form e.g. larger or smaller groups, only the intuitively tangible comparisons could be made for simple, unrealistically small examples. Furthermore, many of the sequential hierarchic grouping methods cannot be easily used as classifying devices, i.e. when after a $\{P^k\}_1^n$ had been found, to locate an n+1-st element of $I^1 = I \cup \{n+1\}$, and so on.

Some globality and intuitiveness, although at a loss of numerical efficiency, is introduced by means of methods based on assumptions C. These methods may utilize, in addition to Ca,b, other particular assumptions in order to define operationally appropriate algorithms. The most "conservative" assumption following literally Ca would be that a group $A^{\varepsilon l}$ can be considered as such iff

$$\forall i,j \in A^{\varepsilon l},\ k \notin A^{\varepsilon l} : d_{ij} \leq \varepsilon \wedge d_{ik} > \varepsilon \wedge d_{jk} > \varepsilon \quad , \qquad (1)$$

i.e. all points in an $A^{\varepsilon l}$ are similar and no point similar to them belongs to an other group. Groups formed in this way can be called ε-homogeneous. Application of the criterion of ε-homogeneity has the advantage of clear intuitive meaning for the global solution, and it also yields such unique global solution, a partition $P(\varepsilon)$ composed of $L(\varepsilon)$ groups. There is, on the other hand, the additionally important burden of computations. Moreover, the criterion is rational for points $i \in I$ in a metric space, while d_{ij} may not have anything to do with formal distances. The same can also be said of absolute homogeneity, i.e. the criterion E.

In practice, both Ca with (1) and E are rarely used for reasons mentioned above. The requirements of (1) and E can again be relaxed if an ultrametric δ_{ij} is introduced to (1) and E instead of d_{ij}.

Still, intuitive simplicity and obviousness of assumptions C causes a number of applications based upon them to appear. To make these elementary assumptions operational in the effective algorithms some additional assumptions are introduced. Thus, in the FARRELL and FARRELL-mod [Bielecka et al. (1979)] methods, it is attempted to locate the spheres A^{kl} of a given predefined radius, so that

$$\forall i \in I \quad \exists \; A^{kl} \ni i \quad ,$$

cross sections $A^{kl} \cap A^{kl'}$, $l \neq l'$, are possibly "small" in terms of $A^{kl} \cap A^{kl'}$ and the centers of A^{kl} approximate local gravity centers.

Another simple and intuitive method thus derived is the "percolation method", of Tremolières (1979), which defines for each i a set $V(i,\varepsilon)$,

$$V(i,\varepsilon) = \{j \in I \mid d_{ij} \leq \varepsilon\} \quad , \tag{2a}$$

and then a density coefficient $v(i,\varepsilon)$, e.g.

$$v(i,\varepsilon) = \bar{\bar{V}}(i,\varepsilon) \quad . \tag{2b}$$

The procedure chooses sequentially the maximal $v(i,\varepsilon)$ and forms the $A^{\varepsilon l} \in P(\varepsilon)$ from the non-grouped i's corresponding to these $v(i,\varepsilon)$. Again, special assumptions are necessary for classification of boundary points. Variable ε is also utilized in the method proposed by Slater (1976a). This method is used to analyze d_{ij}, which are not necessarily symmetric. It requires however, that these values be doubly standardized, i.e. row and column sums be equalized. Thus defined, the distance matrix serves to develop the hierarchy by sequentially analyzing links whose values are higher than the ε, decreasing from $\max_{i,j \in I} d_{ij}$ to zero. In this way consecutive grouping patterns appear, analogously to "single linkage" or "nearest neighbor" procedure. Soundness of individual groups appearing in the hierarchy is checked via special additional techniques.

The gravity method [Bielecka et al. (1979)] starts with the assumption Cb, and groups the non-grouped i's which fulfill it. Special classification assumptions for i's to be added to existing A^{kl}'s are based upon either variance or arithmetic average statistics.

The local criterion F does not yield, by itself, a unique partition P, but rather a family of groups A^{kl} which can be used to form a hierarchy of partitions P^k. An additional criterion, i.e. that a group A^{kL} contains similar elements, [Kacprzyk and Stanszak (1975, 1978)] iff

$$\forall a^{kl} \subsetneq A^{kl},\ W(a^{kl}) > W(A^{kl}) \quad , \tag{3}$$

is used to make the minimally interconnected subnetworks technique operational. Criterion (3) can hardly be referred to as intuitive. It yields, however, very good computational properties. A hierarchy $H^W = \{P^k\}_1^r$, obtained by using this method, is much "flatter" than the "full" hierarchies obtained from assumptions B, i.e. $r < n$. The method, has in fact, a "bias" towards "greater" groups.

In Slater (1976b), conditions similar to (3) serve to define the so-called nodal regions-groups, i.e. such that they

have weaker links (greater distances) with other regions than the elements being the nodes of the groups. Because of the specific formulation, this problem is approached in a manner similar to that of Slater (1976a).

5. Objective Functions

The methods described above, were referred to as local and this feature was said to be partly offset by the intuitively obvious nature of most of the assumptions on which the appropriate procedure is based. This qualification of locality should be commented upon. Since it is a partition P (or a hierarchy H) that is being sought, the qualifications of locality or globality should refer to a capacity of search in the space of P's (or H's). From this point of view the methods mentioned could not even be called local insofar as they do not offer <u>any</u> possibility of comparison and choice among various P's (H's). They only determine one P (or one H), and those which determine an H usually do not provide any possibility for choosing P^k out of $P^k\{_1^r\} = H$. Certainly, each of these methods can be complemented with measures of the (relative!) groups stability, as it was done by Slater or Tremolieres. Thus, through the methodological back door some possibility of comparison is introduced.

Obviously, the possibility of comparison can only be realized through the simultaneous explicit consideration of all groupings entering a P. Hence, an overall objective function or its proxy should be constructed.

Another comment refers to the nature of the iterative numerical processes leading to the establishment of a solution. As we have seen in the example of assumptions B. and C., the essential change in the character of the solution (H or P) does not necessarily entail a change in the nature of the iterative process (although it may). Thus, in assessing a method, more attention should be paid to the uniqueness and comparativeness of the final results rather than to the course of the procedure.

A number of objective functions have been proposed for solving the grouping problem. Some of these are presented below together with their "natural" extensions based upon previously formulated initial assumptions.

G. Partition P^k is better than partition P^{k*} iff [Duran and Odell (1974), Jensen (1969)]

$$\sum_{m=1}^{L_k} \frac{1}{\bar{\bar{A}}^{kl}} \sum_{\substack{i,j\in A^{kl} \\ i<j}} d_{ij}^2 < \sum_{m=1}^{L_k} \frac{1}{\bar{\bar{A}}^{kl*}} \sum_{\substack{i,j\in A^{kl*} \\ i<j}} d_{ij}^2 \quad ;$$

H. Partition P^k is better than partition P^{k*} iff

$$\sum_{l=1}^{L_k} \frac{2}{\bar{\bar{A}}^{kl}(\bar{\bar{A}}^{kl}-1)} \sum_{\substack{i,j\in A^{kl} \\ i<j}} d_{ij}^2 < \sum_{L=1}^{L_k} \frac{2}{\bar{\bar{A}}^{kl*}(\bar{\bar{A}}^{kl*}-1)} \sum_{\substack{i,j\in A^{kl*} \\ i<j}} d_{ij}^2 \quad ;$$

I. Partition P^k is better than partition P^{k*} iff [Duran and Odell (1974)]

$$\max_l \left(\max_{i,j\in A^{kl}} d_{ij}\right) < \max_l \left(\max_{i,j\in A^{kl*}} d_{ij}\right) \quad ;$$

K. Partition P^k is better than partition P^{k*} iff

$$\sum_{\substack{l,l'=1 \\ l<l'}}^{L_k} d(l,l') > \sum_{\substack{l,l'=1 \\ l<l'}}^{L_k} d(l,l')^* \quad ,$$

where $d(l,l')$, and $d(l,l')^*$ are inter-group distances defined in any of the ways given in assumption B., for partitions P^k and P^{k*} respectively.

L. Partition P^k is better than partition P^{k*} iff [Mulvey and Crowder (1979)]

$$\sum_{i\in I} \sum_{j\in J} d_{ij}x_{ij} < \sum_{i\in I} \sum_{j\in J} d_{ij}x_{ij}^* \quad ,$$

where JCI is a set of eligible centers (usually J=I), and

$$\sum_{j \in J} x_{ij} = 1 \qquad \forall i$$

$$\sum_{j \in J} x_{jj} = L_k$$

$$x_{ij} \leq x_{jj} \qquad \forall i,j$$

$$x_{ij} \in \{0,1\} \qquad \forall i,j \quad ,$$

which indicates a 0-1 programming problem. It is being solved using a sub-gradient method applied to the relaxed Lagrangian form of the initial problem.

M. Partition P^k is better than partition P^{k*} iff [Bielecka and Szczotka (1978)],

$$\frac{S_o^2}{S_I^2} > \frac{S_o^{*2}}{S_I^{*2}}$$

where $S_o^2 = \sum_{\substack{l,l'=1 \\ l < l'}}^{L_k} d^2(l,l')$, and $S_I^2 = \sum_{l=1}^{L_k} \frac{1}{\bar{A}^{kl}} \sum_{\substack{i,j \in A^{kl} \\ i < j}} d_{ij}^2$;

N. Partition P^k is better than partition P^{k*} iff [Diday and Simon (1977)]

$$W(R^k,P^k) < W(R^{k*},P^{k*})$$

where R^k is a set $\{r^{k1},\ldots,r^{kL_k}\} \subset I$ of the "representatives" of groups $A^{k1},\ldots,A^{kL_k}$, such that $r^{kL} \in A^{kl}$ (in particular $\bar{r}^{kl} = 1$) and the function W can, e.g., be defined by

$$W(R,P) = \sum_{l=1}^{L_k} \sum_{i \in r^{kl}} \sum_{j \in A^{kl}} d_{ij} \quad .$$

All the objective functions presented provide a comparison of "goodness" of partitions P^k, i.e. partitions of I, into a given number of groups L_k. This is caused by the fact that the functions G. through N. display similar characteristics as the procedures built upon assumptions B. through F., i.e. they refer to only one side of the initial grouping problem, either internal homogeneity of groups or external dissimilarity. The same applies to M. since the quotient imposed amplifies the one-sided effect rather than balances the two. Thus, the values of these objective functions for optimal P^k are monotonic with regard to L_k.

A point which becomes increasingly important with the introduction of the objective functions is the numerical efficiency of algorithms. The merit of most of the procedures based upon the local assumptions A. through F. lies in their simplicity. The same can hardly be said of optimization procedures with regard to G. through N. Some refer to the dynamic programming philosophy [Jensen (1969)], others utilize special iterative algorithms based upon properties connecting local (with respect to A^{kL}) and global (with respect to P) optima [Diday and Simon (1977)]. The algorithm proposed by Diday has essential numerical advantages, especially from the point of view of memory requirements. However, it requires a good initial guess and does not safeguard against cycling phenomena. Dynamic programming, although robust in reaching solutions, is more cumbersome in calculations.

Thus, in order to solve the grouping problem in its absolute form, i.e. together with L_k, different objective functions have to be developed. This, however, can make the computational problems even more difficult; therefore, very little effort has been made to construct such global objective functions and the corresponding algorithms. An example of such a function is verbally cited in Fortier and Solomon (1966) after Holzinger and Tyron. It can be formulated as:

O. Partition P is better than partition P^* iff

$$\frac{\bar{D}_O}{\bar{D}_I} > \frac{\bar{D}^*_O}{\bar{D}_I} \quad ,$$

where

$$\bar{D}_o = \frac{2}{\sum_{l=1}^{L} \bar{\bar{A}}^l (n-\bar{\bar{A}}^l)} \sum_{l=1}^{L} \sum_{i \in A^l} \sum_{\substack{j \in A^{l'} \\ l'>l}} d_{ij} \quad ,$$

is the inter-group average distance among pairs ij, and

$$\bar{D}_I = \frac{2}{\sum_{l=1}^{L} \bar{\bar{A}}^l (\bar{\bar{A}}^l-1)} \sum_{l=1}^{L} \sum_{\substack{i,j \in A^l \\ i<j}} d_{ij} \quad ,$$

is the intra-group average distance among pairs i,j.

This global function has been abandoned because of the computational difficulties resulting from its form which hardly lends itself to algorithmic simplifications. Hence, Fortier and Solomon have tried another approach, which consists in constructing the objective function such that in some way it utilizes assumption C. Thus,

P. Partition P is better than partition P* iff

$$C(P) > C(P^*) \quad ,$$

where $C(P) = \sum_{\substack{i,j \in I \\ i<j}} g_{ij}$, and $g_{ij} = (d_{ij}-\varepsilon) \cdot \gamma_{ij}$, with $\gamma_{ij} = +1$ when i and j are in the same A^l and $\gamma_{ij} = -1$ otherwise.

It can easily be shown that function C has the basic property that ensures the existence of a non-trivial maximum over its maximal values for various L, i.e.

$$L^{opt} \neq \{1\} \quad \text{and} \quad L^{opt} \neq \{n\} \quad ,$$

where $L^{opt} = \{L_{opt} | C^{opt} = \max_L (C^{max}(L)) = \max_L (\max_{P^L} C(P^L))\}$.

Furthermore, this function is very simple and can be constructed for various ε values directly from the appropriately modified matrices $D(\varepsilon)$. Fortier and Solomon investigated the application of this function for a relatively small example (n=19) and focussed on $\varepsilon = 0.5$. An extension provided in their study did not consider

the elaboration of a more efficient algorithm for optimizing the function proposed bur rather its application for a very specific purpose, related to factor analysis.

Thus, the present author proposes another, general, form of the global objective function.

Q. Partition P is better than partition P* iff

$$Q(P) > Q(P^*)$$

where $$Q(P) = (1-\rho) \sum_{A^1} \sum_{\substack{i,j \in A^1 \\ i<j}} w_{ij} + \rho \sum_{A^1} \sum_{\substack{i \in A^1 \\ j \notin A^1}} d^*_{ij}$$

w_{ij} is a function $w(d_{ij})$, $R^1_+ \cup \{0\} \to R^1_+ \cup \{0\}$, such that

1^o. $d_{ij} < d_{ik} \Leftrightarrow w(d_{ij}) > w(d_{ik})$

2^o. $\bar{d}^*_{ij} = \bar{w}_{ij} = 1$, with bars denoting averages

d^*_{ij} obtained from the initial d_{ij} through

$$d^*_{ij} = \frac{d_{ij}}{\bar{d}_{ij}} = \frac{n(n-1)}{2} \cdot \frac{d_{ij}}{\sum_{\substack{i,j \in I \\ i<j}} d_{ij}}$$

and $\rho \in (0,1)$.

The intuitive sense of this function is obvious - it requires maximization of the intra-group "similarities" together with maximization of inter-group "dissimilarities".

We can, for instance, set

$$w(d_{ij}) = \frac{d^{*max}_{ij} + d^{*min}_{ij} - d^*_{ij}}{d^{*max}_{ij} + d^{*min}_{ij} - 1} ,$$

where d^{*max}_{ij} and d^{*min}_{ij} denote maximum and minimum distances in D^*, respectively.

It can easily be shown that for the non-trivial case, i.e. when only $d^{max}_{ij} > \bar{d}_{ij} > d^{min}_{ij}$, the objective function $P(Q)$ reaches a maximum for some $L^{opt} \in N(1,n)$, provided a certain simple condition on ρ holds. When $\rho = \frac{1}{2}$ and $w(d_{ij})$ is defined as above, $Q(I) = Q(\{I\}) = \frac{1}{4} n\,(n-1)$.

6. The Algorithm

First, the triangular matrices D* and W are formed, according to the averaging formula given earlier. Then another triangular matrix is formed, WD*, composed of the elements

$$wd_{ij} = \frac{w_{ij}}{w_{ij} + d^*_{ij}} \quad .$$

These elements are then ordered in the decreasing sequence, so that

$$wd^1 = \max_{i,j \in I} wd_{ij} \quad \text{and} \quad wd^{\frac{1}{2}n(n-1)} = \min_{i,j \in I} wd_{ij}$$

Obviously, the sequencing thus obtained is identical with that for w_{ij}.

The algorithm consists in consecutive determination of optimal P's for decreasing values of ρ. For $\rho = 1$, $P^{opt} = I$. At each consecutive $\rho = wd^k$, the condition is checked

$$(1-\rho) \sum_{\substack{i \in A^l \\ j \in A^m}} w_{ij} > \rho \sum_{\substack{i \in A^l \\ j \in A^m}} d^*_{ij} \quad ,$$

where $wd^k = wd_{ab}$, $a \in A^l$, $b \in A^m$. When this condition is fulfilled, the groupings A^l, A^m are merged, when not, the check will be repeated at every subsequent wd^k, until it is fulfilled. Note: in this way, the optimal partitions $P^{opt}(\rho)$, which form a hierarchy, but not an optimal hierarchy H^{opt} are obtained. There is, indeed, no search in the space of H's. The simple algorithm outlined here can, in fact, be easily adopted to the search of H^{opt} provided the function Q(H) is defined, analogous to Q(P), in which the distances on various hierarchy levels are accounted for, or an ultrametric measure δ^*_{ij} is used together with its counterpart ω_{ij}, defined as for d^*_{ij} and w_{ij}.

REFERENCES

Bielecka, K., M. Paprzycki, and Z. Piasecki. 1979. Proposal of new taxonomic methods for agricultural typology. Geographia Polonica 40. Warsaw.

Bielecka, K., and F.A. Szczotka. 1978. Study on evaluation of quantiative methods in agricultural typology (in Polish). Biuletyn Infromacyjny, z. 23. Instytut Geografii i Przestrzennego Zagospodarowania PAN. Warsaw.

Byfuglien, J., and A. Nordgard. 1973. Region-Building - A comparison of methods. Norsk Geografisk Tijdsskrift 27.2.1973.

Diday, E., and J.C. Simon. 1977. Clustering analysis. Digital Pattern Recognition, Fu, K.S., ed., Berlin, Heidelberg, New York: Springer Verlag.

Duran, B.S., and P.L. Odell. 1974. Cluster Analysis. A Survey. Berlin, Heidelberg, New York: Springer Verlag.

Fortier, J.J., and H. Solomon. 1966. Clustering procedures. Multivariate Analysis I. Krishnaiah, P.R., ed., New York, London: Academic Press.

Hartigan, J.A. 1978. Clustering Algorithms. New York: John Wiley & Sons.

Jensen, R.E. 1969. A dynamic programming algorithm for cluster analysis. Operations Research, 12:1034-1057.

Kacprzyk, J., and W. Stanczak. 1975. Application of the method of minimally interconnected subnetworks to division of a group of enterprises into subgroups (in Polish). Archiwum Automatyki i Telemechaniki. Vol. XX, No. 4.

Kacprzyk, J., and W. Stanczak. 1978. On a further extension of the method of minimally interconnected subnetworks. Control and Cybernetics 7, No. 2.

Lance, G.N., and W.T. Williams. 1967. A general theory of classificatory sorting strategies. Computer Journal 10, No. 3 and 4.

Mulvey, J.M., and H.P. Crowder. 1979. Cluster analysis: an application of Lagrangian relaxation. Management Science 25, No. 4:329-340.

Slater, P.B. 1976a. A hierarchical regionalization of Japanese prefectures using 1972 interprefectural migration flows. Regional Studies 10:123-132.

Slater, P.B. 1976b. A multiterminal network-flow analysis of an unadjusted Spanish interprovincial migration table. Environment and Planning A, 8:875-878.

Tremolières, R. 1979. The percolation method for an efficient grouping of data. Pattern Recognition, 11.

Wishart, D. 1971. Some Problems in the Theory and Application of the Methods of Numerical Taxonomy. PhD Thesis. University of St. Andrew's, Scotland.