MÜNSTERSCHE
GEOGRAPHISCHE
ARBEITEN

Gerhard Bahrenberg / Ulrich Streit (Hrsg.)

German Quantitative Geography

Papers presented at the 2nd European Conference on
'Theoretical and Quantitative Geography'
in Cambridge 1980

SCHÖNINGH

11 1981

German Quantitative Geography

Papers presented at the 2nd European Conference on
'Theoretical and Quantitative Geography'
in Cambridge 1980

Gerhard Bahrenberg / Ulrich Streit (Hrsg.)

FERDINAND SCHÖNINGH · PADERBORN · 1981

Alle Rechte, auch das der auszugsweisen photomechanischen Wiedergabe, vorbehalten.
© 1981 by Ferdinand Schöningh, Paderborn, ISBN-Nr. 3-506-73211-0

Gesamtherstellung: Buchdruckerei Regensberg, Daimlerweg 58, 4400 Münster

PREFACE

This volume contains the contributions of the German participants at the 2nd European conference on "Theoretical and Quantitative Geography" in Cambridge, September 1980.

Although the papers by E. Giese and U. Streit are published in the conference proceedings, and the paper by G. Bahrenberg is going to be published in a special issue of 'Environment and Planning A' we decided to include them in this collection in order to give an as complete as possible impression of recent German work in quantitative and theoretical geography. To avoid double publication the papers are presented here in a slightly different version.

The German working group decided to publish all papers in English since this seems to be the only way to make our ideas and work be known within the international community of geographers. We would like to thank our colleagues very much for taking the burden to produce lengthy papers in English. Only those who ever tried to express themselves by writing in a foreign language are able to appreciate the tremendeous and troublesome work one has to invest. Although every author is fully responsible for his contribution in any respect we would like to apologize to our foreign friends and colleagues for occasional stylistic deficiencies that may make reading sometimes hard.

The papers in this volume are by no means representative for West German quantitative and theoretical geography but may give a good impression of the range of research work in the field. E. Giese presents a systematic outline of the development during the last 15 years in his introductionary paper. The other papers deal with problems of spatial stochastic analysis and modelling, multivariate analysis, policy and planning oriented research, and computer aided cartography thereby indicating the foci of present research interests in West Germany. It should be noted in this context that German quantitative geography received significant stimulation by regional policy and planning.

The publication was possible only because of much help we received from others. We would like to thank the 'Deutsche Forschungsgemeinschaft' for a substantial financial support which allowed our participation at the Cambridge-meeting. Our thanks are due to Bob Bennett and his British colleagues who organized the conference giving us an opportunity to discuss our ideas. Thirdly we would like to thank our secretaries Angelika Rentel, Beate Rolfs (Bremen), and Hildegard Schulz (Münster) who had to struggle with quite unusual manuscripts but succeeded. Finally we thank the editors of 'Münstersche Geographische Arbeiten' who agreed to publish the collection in their series and offered technical and financial support.

Hopefully the volume helps to strengthen international relations in our discipline and to overcome still existing language barriers.

Bremen/Münster
April 1981

Gerhard Bahrenberg
Ulrich Streit

I N H A L T

E R N S T   G I E S E

THE DEVELOPMENT AND PRESENT STATE OF RESEARCH INTO 'QUANTITATIVE GEOGRAPHY' IN THE GERMAN-SPEAKING COUNTRIES*)

From:

BAHRENBERG, GERHARD und STREIT, ULRICH (Eds.):
German Quantitative Geography. Papers presented at the 2nd European Conference on 'Theoretical and Quantitative Geography' in CAMBRIDGE 1980. Münstersche Geographische Arbeiten 11. Schöningh, Paderborn 1981, p. 9 - 26

*) This article is a shortened draft of the paper presented at the symposium in Cambridge. The full essay will be published by Pion, London, 1981 together with the other general papers of the symposium.

Author:

Prof. Dr. Ernst Giese

Geographisches Institut
Universität Giessen

Senckenbergstraße 1

6300 Giessen

ZUSAMMENFASSUNG

Die vorliegende Abhandlung stellt eine gekürzte Fassung eines Beitrages dar, der in voller Länge und deutscher Fassung in der "Geographischen Zeitschrift", 68. Jg., 1980, Heft 4, erschienen ist. Nach einer kurzen Darstellung der Entwicklung der "Quantitativen Geographie" im deutschsprachigen Bereich werden die derzeitigen Schwerpunkte der Forschung der deutschsprachigen Arbeitsgruppe vorgestellt, wobei insbesondere auf Fortschritte in der Theorie- und Modellbildung eingegangen wird. Auf vier Forschungsschwerpunkte der Theorie- und Modellbildung wird näher eingegangen. Es handelt sich

1. um das Gebiet der räumlichen Diffusionsforschung und raum-zeit-varianten-stochastischen Prozeßanalyse
2. um den Komplex der partiellen Standorttheorien,
3. um den Bereich der räumlichen Wachstums- und Entwicklungstheorien und
4. um das Feld der planungsorientierten Modellbildung.

Darüber hinaus wird kurz über Fortschritte in der Methodik und geographischen Datenverarbeitung referiert, die neben der Theorie- und Modellbildung wesentliche Bestandteile der "Quantitativen Geographie" darstellen. Abschließend wird auf Probleme und kritische Momente der jüngeren Entwicklung der "Quantitativen Geographie" hingewiesen.

I Methodological reorientation

When I. BURTON published his paper on the 'Quantitative Revolution' in geography in 1963, there was no trace of such a revolution in any of the German-speaking countries. The publication of a series of methodological papers not till the mid 1960s saw the beginnings of the development of 'quantiative geography' in these countries. Mathematical and statistical techniques began to be applied in geography and were used to try and solve some of the traditional types of questions and problems, in particular those relating to classification and regionalization. Typical of this phase of development are the papers by D. STEINER (1965a,b) on the application of factor analysis in geography, by CHR. WERNER (1966) on the geometry of traffic networks or that by F. AHNERT (1966) in which he pointed to the importance of computers and mathematical models for the future of geomorphology. The first techniques to be applied widely were regression and correlation analysis. There was particular interest in factor analysis and principal component analysis.
Factor analysis introduced many geographers in the German-speaking countries to the 'quantitative phase in their development'. There was an unwritten rule in the geographical world that anyone who was anyone must have completed at least one factor analysis and by the end of the 1960s and especially in the early 1970s the results of these efforts were appearing one after another (see A. KILCHENMANN 1968, 1970, 1971; J. BÄHR 1971a,b, 1974; G. BRAUN 1972, 1975; M. SAUBERER 1972, 1973 a,b, 1974; G. STÄBLEIN 1972, 1973, 1975; R. HERRMANN 1973; etc.). During this period factor analysis became the favourite mathematical technique amongst 'quantitative geographers' in the German-speaking countries.

It was only realized later that they had stumbled on a particularly difficult mathematical-statistical technique, giving rise to heated discussions subsequently (see E. GIESE 1978c, F.J. KEMPER and W. SCHMIEDECKEN 1977); the difficulties played no part in the first innovative phase of 'quantitative geography'. The first concern was to master and tame a new methodological toy and it resulted in a period of experimentation and in the relatively uncritical and blind application of novel mathematical-statistical techniques.

A survey of the application of factor analysis and the closely allied technique of cluster analysis shows clearly how the use of mathematical-statistical techniques slowly spread throughout the German-speaking countries. In this regard the North-American influence should not be overlooked: the same process of adaptation occured as in the United States, but with a time-lag of almost ten years. B.J.L. BERRY published his first factor analysis in 1960; the first factor analyses in the German-speaking countries date from the late 1960s (A. KILCHENMANN 1968), apart from the work of D. STEINER, who was then living in North America.

It would give a false picture of 'quantitative geography' in the German-speaking countries to imply that factor analysis was the only technique examined in the initial phase of development. As well as the techniques of factor and cluster analysis, geographers interested in quantitative methods examined and applied a number of other mathematical-statistical techniques, such asdiscriminant analysis, multiple regression analysis, test and estimation techniques of various different kinds, methods for analysing time series (Markov chains, Fourier analysis), as well as graph theory and matrix techniques. The applications were restricted in the main to the field of intervall/ratio data; techniques for coping with nominal and ordinal data were only taken up later.

II Reassessment of scientific theory

In the first instance there was no revision of the content of geography in the German-speaking countries to accompany the adoption of the mathematical-statistical techniques, even though such a change should have been an intregral part of any such development. A new methodical toy was embraced without any attempt being made to change either the terms of reference, the aims and objectives, or even the fundamental scientific approach. The first decisive step in this direction was taken by D. BARTELS (1968b) with the publication of his paper 'On the reassessment of the scientific underpinnings of human geography', together with the translations he initiated of various studies in English (see D. BARTELS 1970a) and P. HAGGETT's textbook (1965/1973) 'Locational analysis in human geography'. It marked the assumption of the new research direction on a much broader base.

The writings of D. BARTELS are important in two respects for the development of 'quantitative geography'.

1. Previously human geography had stuck doggedly and somewhat one-sidedly to the classical concept of landscape studies and regional geography, but D. BARTELS tried to provide it with a new or at least an alternative theoretical framework. He developed an approach to human geography derived from analytic and pragmatic scientific theory and orientated towards regional science. It is worth noting in this context that even at this early stage D. BARTELS' approach incorporated behavioural variables.

2. D. BARTELS' writings stated clearly that the goal of this approach to human geography ultimately ought to be the development of new theories and models, as well as giving greater precision to existing ones. It meant that the purpose of 'quantitative geography' could not simply be defining, describing and analysing observations more precisely through the application of statistical techniques. Neither could it be the development of quantitative techniques for their own sake, nor the uncritical translation and application of mathematical-statistical techniques to geographical date. Rather the aim had to be to develop more realistic spatial theories and models , much in the same way that regional science had done, or at least to expand on existing theories. For example location theories, like those of J.H. von THÜNEN, W. CHRISTALLER and A. LÖSCH, were originally derived to explain the primary, secondary and tertiary sectors, but more recently regional scientists, such as W. ISARD, E. von BÖVENTER and others, have developed them more in the direction of general economic theory.

This task was tackled enthusiastically by the 'quantitative geographers' and at the same time attempts were made to extend the range of techniques available, and to test their applicability. In the sequel various different theories and models were adopted and attempts made to make them more precise and to develop them further, as well as to test them empirically.

There is a whole series of models dating from the beginning of this phase, all built with the help of regression analysis, factor analysis or discriminant analysis, but with varying specifications with regard to content:

- In physical geography in the field of hydrology, R. HERRMANN (1972,1973,1974, 1976) and U. STREIT (1973) produced models for predicting runoffs as well as models of water pollution in the FRG. In the field of geomorphology F. AHNERT (1971, 1976) and H. ROHDENBURG (1976) produced models of slope degradation.

- In human geography there were three areas. First, models to analyse and describe spatial diffusion processes (the spread of the idea to become guest worker in Turkey with the help of a regression-based gravity model - D. BARTELS (1968c 1970b); the spread of television in Poland using the results of a multiple regression - G. BAHRENBERG/J.LOBODA (1973)). Second, models for analysing central place systems (the analysis of the spatial distribution of central places with the help of quadrat and factor analysis - J. DEITERS (1975a,b), predicting the spatial pattern of consumer demand for goods and services of medium frequency using a beta distribution and multiple regression analysis - J. GOSSEFELDT (1975a,b)). Third, factor analytic models for analysing inner city population movements as well as the social segregation of the urban population (the fields of factorial ecology and social ecology - G. BRAUN (1972,1975,1976)).

Another focus of research in the German-speaking countries 'Quantitative Working Group' was in the field of time series analysis. Primarily using Markov chains various types of time series models were built and used to analyse and predict a number of processes, among them inter-regional migration and regional population growth in the FRG (H.P. GATZWEILER 1975a,b; H.P. GATZWEITLER and R. KOCH (1976). They were also used to predict the runoff patterns of rivers (U. STREIT 1975a). Time series models were also used to generate synthetically more complete data sets, when only short series of observations of the processes in question were available (U. STREIT 1975b, 1976). The application of time series models to spatial variations was a logical extension, but one which only followed later.
Network analysis was a further area of study that emerged fairly early in 'quantitative geography' in the German-speaking countries, but it received little sustained attention.
Reference should nevertheless be made once again to the work of CHR. WERNER (1966) on the geometry of traffic networks, which examined the economic feasibility of traffic networks using graph theory. F. VETTER (1970) extended the work of CHR. WERNER (1966) in his studies of the rail network in Lower Saxony using graph theory. He developed a simulation model for the construction of railway networks, using similar techniques to those adopted by K.J. KANSKY (1963) in his simulation models. The graph theory and matrix studies by I. SCHICKHOFF (1974,1977) using as an example the Dutch rail network, provide a critical appraisal of the simulation techniques used by K.J. KANSKY (1963) and F. VETTER (1970), although to some extent her models are constructed somewhat differently. An interesting study which should stimulate some further work on network theory and its methodical base is done by J. GOSSEFELDT (1978).

The study used graph theory as a means for checking planning concepts, notably development axes and growth poles, both of which have been the subject of much recent discussion and which have already been used in various different aspects of planning in the FRG. The study takes as an example the Land Development Programme for Bavaria and Baden-Württemberg.

III Advances in the formulation of theory and model building

There are four areas of research in the formulation of theory and model building, on which the main interest of the German Working Group is focused:

1. Spatial diffusion and spatio-temporal stochastic processes
2. Partial location theories
3. Theories of spatial growth and development
4. Model building for planning purposes

1. Spatial diffusion research and analysis of spatio-temporal stochastic processes

The logical starting point for any research into the theory of spatial diffusion is the work of T. HÄGERSTRAND in the early 1950s. In the German-speaking countries CH. BORCHERDT (1958,1961) D. BARTELS (1968c, 1970b), G. HARD (1972), G. BAHRENBERG/J. LOBODA (1973) and E. GIESE (1978a, 1979) have worked on problems of spatial diffusion.

There are three significant features which characterize the more recent development of spatial diffusion research in German-speaking countries:

a) the extension of research into spatial diffusion into the more comprehensive field of spatio-temporal stochastic process research;
b) a growing interest in behavioural aspects of the research;
c) the application of the results of research into diffusion to regional planning.

In the German-speaking countries U. STREIT (1978,1979) and J. NIPPER (1977,1978) have been particularly interested in stochastic processes and models that vary in both space and time. U. STREIT became interested in them through his work on the runoff characteristics of small rivers. His first analysis was of the temporal variations in runoff data using Markov-chain models, or more accurately various permutations of the ARIMA models described by G.E.P. BOX/G.M. JENKINS (1970) (U. STREIT 1975a,b). With the incorporation of the spatial dimension work began to be done on the spatial persistence of processes, as well as on the statistical modelling of phenomena with spatial autocorrelations. The aim was to extend the ARIMA model, which to that point had been entirely time dependent, so that it could take account of both space and time as the STARIMAR class of models does. It was this problem that initiated the fruitful cooperation, in methodical terms, between U. STREIT and J. NIPPER. The initial interest of the latter was the spatial distribution of points, in particular the question of the connection between spatial form and producing process. The works of M.F. DACEY (1968) and D.W. HARVEY (1968) gave an initial stimulus to his work and led to a greater appreciation of the need for a more strongly process orientated way of looking at things.

The joint work of J. NIPPER and U. STREIT (1977, 1978) is primarily concerned with the formal, theoretical underpinnings of this models. Their thinking was significantly influenced by the studies of A.D. CLIFF/J.K. ORD (1973) and A.D. CLIFF/P. HAGGETT/J.K. ORD/K. BASSETT/R. DAVIS (1975) on spatial autocorrelation coefficients and their possible applications in inferential statistics and in the modelling of stochastic relationships in phenomena which vary in space. Both these studies were true innovations in the field. As well as these studies by geographers, there was also work on the Kriging technique (variogram technique) (G. MATHERON 1971, R.A. OLEA 1975, M. DAVID 1977) which is better known in geology (J. NIPPER 1981).

In addition to the development of diffusion research into the more general analysis of space-time processes progress can also be observed in another direction with the growing interest in the behavioural aspects of the research. Concepts and ideas from the behavioural sciences were adopted and adapted in such a way that there was no attempt to conceptualize the developing diffusion process abstractly in terms of different indicators and indices etc. The aim was rather to concentrate on the potential adopters and then to analyse their reactions to the innovation (H.G. WINDHORST 1979).

Another important development is the fact that the results of spatial diffusion research are increasingly being incorporated into applied research (E. A. BRUGGER 1980). The importance of spatial diffusion research has been recognised particularly in regional planning and regional economics.

Various different concepts based on the theory of polarization have been used, all derived from work on diffusion theory. An initial critical analysis and overview was undertaken by I. SCHILLING-KALETSCH (1976).

2. Partial Location theories

From the beginnings the development or rather refinement and implementation of partial location theories has been a major theme in 'quantitative geography' in the German-speaking countries. The touchstone for this was W. CHRISTALLER's Central Place Theory, which in the 1960s and 1970s stimulated a wealth of empirical and theoretical studies and became one of the most important instruments in regional planning in the FRG. J. DEITERS (1975,1976,1978) and J. GUSSEFELDT (1975,1976,180) have been particularly active in refining and implementing CHRISTALLER's theory. Spurred on by writings of B.J.L. BERRY, L. CURRY and M.F. DACEY in the mid 1960s they have used various different methods and theoretical approaches to test and refine Central Place Theory. Their aim was to produce a more realistic version of the theory on the basis of changes in the pattern of consumer behaviour and use of spatial characteristics. Both have adopted a probabilistic approach to the theory and trace a behaviouristic approach.

The refinement of partial location theories in the tertiary sector was by no means restricted only to W. CHRISTALLER's Central Place Theory. E. GIESE (1978b), for example, used works by W. ALONSO (1960,1964) to examine inner-city land use theory and to study the locational requirements of inner-city businesses and extended the original work by incorporating further important variables.

G. BAHRENBERG (1974, 1976, 1978,1979,1980) has looked in great detail at the location-allocation problems of centrally located public services and, as an extension of this work, at spatial optimization models. Stimulated by the work of C.S. REVELLE, R.W. SWAIN and C. TOREGAS in the early 1970s, interest in spatial optimization models has grown considerably, not least because questions concerning the choice of location and the catchment areas for centralized services, especially centralized public services (hospitals,schools,old peoples' homes, fire stations, emergency centres etc.) are unresolved problems and of great practical importance for planning. G. BAHRENBERG has tried to develop a general statistically discrete optimization model for determining the choice of location and the catchment.area boundaries of centralized public services.

Unlike the tertiary sector, relatively little attention has so far been paid by geography to developing partial location theories in the secondary sector. The studies by W. GAEBE (1976,1978) and W. WITTENBERG (1978) are empirical analyses of the extraordinarily complex sets of factors which determine the locational decision of industrial firms. Both are still in the forefront of the attempts to derive a general theory of location.

The trend towards behavioural research and theory formulation in 'quantitative geography' is particularly apparent in the work of D. HÖLLHUBER (1975,1978), although so far he has mainly been concerned with the choice of residential location by households. To this end he has examined explanations taken from perception and behavioural research and he has then tried to refine them, or rather reformulate them in operational terms. Taking as his starting point the various explanations of how people choose where they live, D! HÖLLHUBER developed the concept of marginal improvement in place of residence, which takes into account the social prestige associated with such an improvement. Certainly this ought to explain a significant part of the decisions of householders about where they live. There would be great value in many more empirical studies of how householders go about choosing where they live, but as yet the fundamental theoretical basis for such explanations is still inadequate.

3. Theories of spatial growth and development

Under the banner of 'engaged geography' attention has turned increasingly in recent years towards trying to understand regional disparities and their underlying causes, as well as trying to contribute to reduce the conflict over the aims and objectives of programmes for equalising living standards on a more rational basis (D. BARTELS 1980, p. 53). As a result there has been renewed interest in theories of spatial growth and development.

'Quantitative geography' has not actually done much to add to the wide variety of theories of regional growth and development already available, being more concerned with putting into practice and testing empirically the very generalised explanations and hypotheses. There has been particular interest in the ideas contained in polarization theory (L. SCHATZL 1973, 1978; I. SCHILLING-KALETSCH 1976; W. TAUBMANN 1979a,b; E.A. BRUGGER 1980).

4. Research orientated towards planning

Interest in developing models, which can be used in regional and area planning and not merely in pure academic research is limited to a relatively small group of geographers working with quantitative methods. In physical geography the work has so far concentrated mainly on models for predicting river runoff rates (R. HERRMANN 1974,1976) and with models of slope degradation (F. AHNERT 1976). In human geography the focus has primarily either been on models for forecasting medium-term population growth and changes in population distribution, especially models for forecasting inter-regional migration (H.P. GATZWEILER 1975b, H.P. GATZWEILER and R. KOCH 1976, R. KOCH 1977, 1978), or urban development models (in particular of the LOWRY type - M.M. FISCHER 1976).

At the present time research is being concentrated on discrete, deterministic regional demographic models. Existing models which were investigated and tried out as possible starting points were at the macro-scale, such as A. ROGERS' multi-regional cohort survival model, or the multi-regional accounting system model devised by P.H. REES and A.G. WILSON (see M.M. FISCHER and M. SAUBERER (1979) for Austria)). However, some new macro-scale models were also developed, like the one by R. KOCH (1977,1978), which was used as a basis for the population forecast in the FRG for 1990. In comparison with the complicated models of A. ROGERS and of P.H. REES and A.G. WILSON, the one devised by R. KOCH was relatively simple, at least for a forecast model dealing with the hard facts of planning in the FRG. There is a great interest in the German-speaking countries in producing models that can be applied as easily and as widely as possible for forecasting regional population growth and changes in population distribution.

In the field of urban planning the situation is rather different in that the main focus of interest has been the building of micro-scale regional demographic models. In this context reference should be made to the small-scale forecast models, or more precisely forecast strategies produced by K.H. DEHLER (1976,1978) and V. KREIBICH (1979, 1980). In the field of small-scale population forecasting there is a growing conviction that the most pressing need is not for more accurate forecast about the future course of events, but rather for techniques which highlight possible alternative strategies. Any pretence that it is possible to forecast the precise course of future events is being abandoned in favour of projections of population growth, that may be used to make clear the effect of land use development planning. As a direct result of this approach now a series of different criteria, used for making forecasts, are much more important than those which are stressed in methodical-statistical research.

As well as population forecasts 'quantitative geography' has examined various other planning concepts (uniform functional areas, growth poles and development axes), which today play an integral part in both the formulation and implementation of regional planning and policy. For example, D. BARTELS (1975) undertook an empirical verification of the concept of 'uniform functional areas' as used in the regional policy of the FRG. J. GOSSEFELDT (1978) tried to find out whether the policy of point-axial settlement development, which forms the basis of the Land Development Programme for Bavaria and Baden-Württemberg, had reduced regional disparities in the level of development, or whether it had made them even more severe. K. GANSER (1972), M. SAUBERER (1976) and others have looked at the reorganisation of local government boundaries and used efficiency analysis to evaluate the viability of local government units. The simplicity and clarity of its mathematics makes efficiency analysis a useful and easily-applied decision-making model in planning.

These summaries of the different studies undertaken by the 'Quantitative Working Group' show that there is a real interest in developing models which can be used in land use and regional planning and not only in pure academic research. Nevertheless the bulk of the work is still dedicated to formal theory building and to pure systems theory. This division within the working group has perceptibly hardened in recent years. It even poses a problem within the I.G.U. Quantitative Methods Working Group because contact between theoreticicans and practitioners threatens to be broken off; the practitioners retreating into the field of applied research, while the theoreticians hide in a destructive, self-imposed isolation. Significantly, after the International Geographical Congress in Moscow in 1976, the I.G.U. Working Group changed its name to Systems Analysis and Mathematical Models. Too few people, at least in the German-speaking world, have tried to adapt the discoveries made in the course of formal model building and theory formulation, or in the course of applying new operational techniques to the needs of either practical planning or education. If the impact of 'quantitative geography' in the German-speaking countries is relatively small, part of the reason is that the Group itself has not bothered to explain clearly enough what it is doing and has not sought to foster 'contacts with ist roots'.

IV Advances in methodology and in the analysis of geographical data

In addition to the developments in theory and model building, there has also been significant progress in the various different spheres of quantitative methodology and in the analysis of geographical data. Although in principle the formulation of theory has nothing to do with methodology, in 'quantitative geography' it is hard to pursue the former without some knowledge of the latter, and without some expertise in the various electronic means of handling data. Without the accurate and critical application of the somewhat complicated mathematical-statistical techniques, it would be hard to produce sufficiently precise research results in 'quantitative geography'. It will therefore always be necessary to try and develop new quantitative techniques, or to 'discover' them in allied disciplines and then to test them for suitability, as has happened recently with the mathematical 'catastrophe theory'. The following are now all proven and accepted working techniques in geography: factor analysis using the varimax rotation, step-wise multiple regression, discriminant analysis, cluster analysis using the WARD algorithm, quadrat analysis, Markov chain and Fourier analysis. Three areas in particular may be picked out, as they have been developed considerably by geographers in the German-speaking countries. The first relates to the extension of multivariate techniques to non-continuous data, especially nominal data (A. KILCHENMANN 1973, F.J. KEMPER 1978a); the second covers the various different permutations on metric and ordinal path analysis (F.J. KEMPER 1978b, H. LEITNER and H. WOHLSCHLÄGL 1980); the third includes the whole range of cluster analytic techniques (M.M. FISCHER 1978, 1980). The problems of regional taxonomy appeared to be solved in geography. M.M. FISCHER identifies the weaknesses of the main techniques and tries to develop alternatives which will be of assistance in helping to choose the right taxonomic technique for solving a given classification or regionalization problem.

The application of mathematical-statistical techniques always requires adherence to certain constraints. In the early stages of the development of 'quantitative geography' mathematical-statistical techniques were applied very uncritically. Quantitative techniques like regression or factor analysis were applied without sufficient attention being paid to the constraints. Without a sufficient chede of the assumptions the results were interpreted unscupulously. They are always perceived as 'right' so long as they correspond with the subjective expected outcomes.
At the same time as the range of available methods was being expanded, a highly critical discussion broke out about the circumstances under which the statistical techniques then in use ought to be applied. More detailed discussions took place about the problems of applying factor and principle component analysis (assumptions about the distribution of the variables, the transformation of the variables, the use of ordinal and nominal data, rotation problems - G. BAHRENBERG and E. GIESE 1975b, F.J. KEMPER 1975 and W. SCHMIEDECKEN 1977, E. GIESE 1978c). There was also discussion of the problems involved in using regression analysis (problems of ecological falsification, problems of 'outliers' - G. BAHRENBERG and E. GIESE 1975a, E. GIESE 1978c).

With reference to geographical data processing, I should like to point to three activities in particular:

- the field of computerized cartography in which A. KILCHENMANN and his colleagues in Karlsruhe have been particularly active, and also D. STEINER in Zürich;
- the field of the quantitative analysis of multi-dimensional remote sensing data (the evaluation of Landsat satellite data - D. STEINER, ETH Zürich); the field of pictorial cover analysis (D. STEINER and his colleagues E. BLUM, B. FLURY, H. GILGEN and G. LENZ).

V Trends in development and their associated problems

The recent development of 'quantitative geography' in the German-speaking countries is well documented in the past symposia and meetings of the Working Group, the proceedings of which have all been published. An initial review of developments was undertaken at the symposium in Giessen in 1974 (see E. GIESE 1975). There followed in quick succession the meetings in Insbruck in 1975 (see D. BARTELS and E. GIESE 1976), in Bremen in 1977 (see G. BAHRENBERG and W. TAUBMANN 1978), in Strassburg in 1977 (see A. KILCHENMANN 1978-79), in Göttingen in 1979 (G. BAHRENBERG and E. GIESE 1980) and in Zürich in 1980 M. OSTHEIDER and D. STEINER 1981).

If one compares the papers presented at the different symposia since 1974, the following clear development trends stand out:

1. If the symposium at Giessen revealed an over-emphasis on methodology and technically orientated studies, the emphasis of research has now shifted towards the formulation of theory and formal model building. However this should not be taken as a sign that the process of adapting and further developing mathematical-statistical techniques is at an end.

2. The interests of the Working Group have branched out into many sub-areas. Themes stemming from a great variety of different questions and based on a wide range of theoretical assumptions have been discussed. There is, however, a lack of any central unifying themes; there is also no coherent research programme. The lack of internal coherence within the Working Group is a source of complaint.

3. Formal-theoretical research seems to be somewhat over-emphasised and this trend has recently intensified with the growing interest in systems theory (see the studies by D. STEINER 1978, 1979a).

4. Now as in the past there is a neglect of applied and practical research. Although some attempts are being made to develop simple useable models and algorithms which can be applied in practical planning, the schism between applied and practically useful research and formal theoretical research cannot be ignored. Within the Working Group it poses a so far unresolved problem. There is almost no contact with teachers in schools.

5. Simple causal models have been overshadowed by complex systems theories with dynamic processes and self-regulating mechanisms. The hypothetical pictures of reality being produced today are ever more complicated. The models are built in the hope that they will represent reality much more accurately than simple causal models, such as were produced when 'quantitative geography' first appeared on the scene. It reveals a fundamental problem in 'quantitative geography' and one that so far has been given little thought. There are two distinct aspects to the problem:

The increase in the empirical explanatory power of models declines as the models become increasingly complex, much as predicted by the law of declining growth of yields (see Fig. 1). At the same time the range (area) of application of the models often become fewer as the models become more complex, because of the limiting constraints governing their use. The closer the models come to reality, the more complex the models and the more complicated the mathematical techniques required to operate them. In many cases the mathematical constraints of these techniques are more restrictive. That leads to the fact, that the areas in which these kind of models and techniques may be applied are more limited, for their assumptions cannot be fulfilled in reality (see Fig.2).

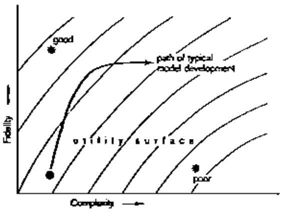


Fig. 1: The interplay of model complexity and fidelity: a typical time path moves from 'simple and dirty' to complex. (After HAGGETT, 1978; BENNETT, 1979)

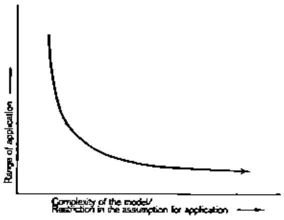


Fig. 2: The interplay of model complexity and restrictions in the assumptions for application respectively and range of application

6. This problem is closely related to another. 'Quantitative geography' cannot quite escape the accusation, that because of the difficulties surrounding the development of models, only those problems have been investigated, which can be easily expressed in formal terms and solved. The decisive problems facing society are still solved in other ways. In other words the sights of 'quantitative geography' may be firmly set on the construction of formal models, but so far this has only led to a very limited perception of reality.

REFERENCES

AHNERT,F. (1966): Zur Rolle der elektronischen Rechenmaschine und des mathematischen Modells in der Geomorphologie. In: Geographische Zeitschrift 54.Jg., S. 118-133.

AHNERT,F.(1971): A General and Comprehensive Theoretical Model of Slope Profile Development.Occasional Papers in Geography,No.I,University of Maryland,College Park

AHNERT,F.(Ed.)(1976): Quantitative Slope Models.Zeitschrift für Geomorphologie, Suppl.-band 25,Bornträger,Berlin/Stuttgart

ALONSO,W.(1960): A Theory of the Urban Market. In:Regional Science Association, Papers and Proceedings 6, S. 149-157

ALONSO,W.(1964): Location an Land Use:toward a General Theory of Land Rent. Havard University Press,Cambridge, Mass.

BÄHR,J.(1971a): Gemeindetypisierung mit Hilfe quantitativer statistischer Verfahren. In: Erdkunde,25, S. 249-264

BÄHR,J.(1971b):Eine Faktorenanalyse zur Bevölkerungsstruktur - dargestellt am Beispiel Südwestafrikas.In: Die Erde, 102,S. 262-285

BÄHR,J.,W.GOLTE(1974): Eine bevölkerungs- und wirtschaftsgeographische Gliederung Chiles.In: Geoforum,17, S. 25-42

BAHRENBERG,G.,J.LOBODA(1973): Einige raumzeitliche Aspekte der Diffusion von Innovationen - Am Beispiel der Ausbreitung des Fernsehens in Polen. In:Geographische Zeitschrift,61.Jg., S. 165-194

BAHRENBERG,G.(1974): Zur Frage optimaler Standorte von Gesamthochschulen in Nordrhein-Westfalen.Eine Lösung mit Hilfe der linearen Programmierung.In: Erdkunde,28,S. 101-114

BAHRENBERG,G.,E.GIESE(1975a): Statistische Methoden und ihre Anwendung in der Geographie. Teubner Studienbücher Geographie, B.G.Teubner,Stuttgart

BAHRENBERG,G.,E.GIESE(1975b): Zum Problem der Normalität und der Transformation bei der Faktoren-bzw. Hauptkomponentenanalyse.In:Giessener Geographische Schriften,Heft 32,Giessen(=Geographisches Institut der Universität), S. 9-29

BAHRENBERG,G.(1976): Ein sozial gerechtes Optimierungsmodell für die Standortwahl von öffentlichen Einrichtungen. In:Tagungsbericht und wissenschaftliche Abhandlungen des 40. Deutschen Geographentages Innsbruck 1975,F. Steiner,Wiesbaden,S. 443-452

BAHRENBERG,G.(1978a): Ein allgemeines statisch-diskretes Optimierungsmodell für Standort-Zuordnungsprobleme. Karlsruher Manuskripte zur Methematischen und Theoretischen Wirtschafts- und Sozialgeographie, Heft 13, Karlsruhe (=Geographisches Institut der Universität)

BAHRENBERG,G.,W.TAUBMANN(1978b): Quantitative Modelle in der Geographie und Raumplanung. Bremer Beiträge zur Geographie und Raumplanung, Heft 1, Bremen (=Schwerpunkt Geographie der Universität) - Symposium Bremen 1977 -

BAHRENBERG,G.,G.MATTHIESSEN,W.STEINGRUBE (1979): STAL. Heuristische Algorithmen zur Lösung statisch-diskreter Standort-Allokationsprobleme mit disjunkten Einzugsbereichen. Universität Bremen - Schwerpunkt Geographie -, Materialien und Manuskripte, Heft 2, Bremen

BAHRENBERG,G.,E.GIESE(1980): Quantitative Modelle und ihre Anwendung in der Anthropogeographie und Raumplanung. Einleitung zur Vortragssitzung. In: Tagungsbericht und wissenschaftliche Abhandlungen des 41. Deutschen Geographentages Göttingen 1979, F.STEINER, Wiesbaden 1980 - Symposium Göttingen 1979 -

BARTELS,D.(1968a): Die Zukunft der Geographie als Problem ihrer Standortbestimmung. In: Geographische Zeitschrift, 56. Jg.,S.124-142

BARTELS,D.(1968b): Zur wissenschaftstheoretischen Grundlegung einer Geographie des Menschen. Erkundliches Wissen. Beihefte zur Geographischen Zeitschrift, 19,F. Steiner, Wiesbaden

BARTELS,D.(1968c): Türkische Gastarbeiter aus der Region Izmir. Zur raumzeitlichen Differenzierung der Bestimmungsgründe ihrer Aufbruchsentschlüsse. In: Erdkunde,22, S.313-324

BARTELS,D.(Hrsg.)(1970a): Wirtschafts- und Sozialgeographie. Kiepenheuer u. Witsch, Köln und Berlin

BARTELS,D.(1970b): Geographische Aspekte sozialwissenschaftlicher Innovationsforschung. In: Tagungsbericht und wissenschaftliche Abhandlungen des 37. Deutschen Geographentages Kiel 1969,F.STEINER, Wiesbaden, S.283-296

BARTELS,D.(1975): Die Abgrenzung von Planungsregionen in der Bundesrepublik Deutschland - eine Operationalisierungsaufgabe. In: Ausgeglichene Funktionsräume - Grundlagen für eine Regionalpolitik des mittleren Weges. Veröffentlichungen der Akademie für Raumforschung und Landesplanung. Forschungs- und Sitzungsberichte, Bd.94, Schroedel,Hannover,S.93-115

BARTELS,D.,E.GIESE(1976): Quantitative Methoden und Beiträge zur Theoriebildung. Einleitung zur Vortragssitzung. In: Tagungsbericht und wissenschaftliche Abhandlungen des 40. Deutschen Geographentages Innsbruck 1975. F. Steiner, Wiesbaden, S.423ff - Symposium Innsbruck 1975

BARTELS,D.(1980): Wirtschafts- und Sozialgeographie. In: Handwörterbuch der Wirtschaftswissenschaft. 23. Lieferung, G.Fischer, Stuttgart,S.44-55

BENNETT,R.J.(1979): Spatial Time Series. Analysis, Forecasting, Control. London

BORCHERDT,Ch.(1958): Über die Vergrünlandung in Bayern. In: Berichte z.Dt. Landeskunde, 21,S.125-129

BORCHERDT,R.J.(1961): Die Innovation als agrargeographische Regelerscheinung. In: Arbeiten aus dem Geographischen Institut der Universität des Saarlandes,6, Saarbrücken (=Geographisches Institut der Universität), S.13-5o

BOX,G.E.P.,G.M.JENKINS(1970): Time Series Analysis. Forecasting and Control. San Francisco

BRAUN,G.(1972): Komplexes Faktorensystem räumlicher und zeitlicher Bewegungen. In: BRAUN,G.(Hrsg.): Räumliche und zeitliche Bewegungen. Würzburger Geographische Arbeiten, Heft 37, Würzburg (=Geographisches Institut der Universität)

BRAUN,G.(1975): Methoden und Modelle zum Schichtungsaufbau und zur räumlichen Mobilität. In: Giessener Geographische Schriften, Heft 32,Giessen (=Geographisches Institut der Universität), S.143-151

BRAUN,G.(1976): Modelle zur Analyse der sozialen Segregation. In: Tagungsbericht und wissenschaftliche Abhandlungen des 40. Deutschen Geographentages Innsbruck 1975.F.Steiner, Wiesbaden 1976,S.455-472

BRUGGER,E.A.(1980): Innovationsorientierte Regionalpolitik. Notizen zu einer neuen Strategie. In: Geographische Zeitschrift, 68. Jg.

BURTON,I.(1963): The Quantitative Revolution and Theoretical Geography, The Canadian Geographers, Vol.7,pp.151-162. (deutsche Obers. in: D.BARTELS(Hrsg.) (1970): Wirtschafts- und Sozialgeographie, Kiepenheuer u. Witsch, Köln, Berlin,S.95-109.

CLIFF,A.D.,J.K.ORD(1973): Spatial Autocorrelation. Monographs in Spatial and Environmental Systems Analysis 5, London

CLIFF,A.D.,P.HAGGETT,J.K.ORD,K.BASSETT, R.DAVIES(1975): Elements of Spatial Structure. A Quantitative Approach. Cambridge University Press. Cambridge

CLIFF,A.D.,J.K.ORD(1975): Space-Time Modelling with an Application to Regional Forecasting. Institute of Brit.Geographers, Transactions, No.66,pp.119-128

DAVID,M.(1977): Geostatistical Ore Reserve Estimation. Developments in Geomathematics 2, Amsterdam, Oxford, New York

DEHLER,K.H.(1976): Zielprognosen der Stadtentwicklung. Schriftenreihe des Bundesinstituts für Bevölkerungsforschung,Bd. 3,H.Boldt,Boppard

DEHLER,M.(1978): Stadt-Entwicklungsplanung Hanau. Planungsbericht 1: Städtische Bevölkerungsentwicklung; Planungsbericht 2: Perspektiven der Bevölkerungsentwicklung. Magistrat der Stadt Hanau, Hanau

DEITERS,J.(1975a): Räumliche Muster und stochastische Prozesse - Lokalisationsanalyse zentraler Orte. In: Giessener Geographische Schriften, Heft 32,S. 122-14o, Gießen(=Geographisches Institut der Universität)

DEITERS,J.(1975b): Stochastische Analyse der räumlichen Verteilung zentraler Orte. Karlsruher Manuskripte zur Mathematischen und Theoretischen Wirtschafts- und Sozialgeographie,Heft 8,Karlsruhe(=Geographisches Institut der Universität)

DEITERS,J.(1976): Stochastische Elemente in der Theorie zentraler Orte. In: Tagungsbericht und wissenschaftliche Abhandlungen des 4o. Dt. Geographentages Innsbruck 1975,F.Steiner,Wiesbaden,S.425-431

DEITERS,J.(1978): Zur empirischen Überprüfbarkeit der Theorie zentraler Orte. Fallstudie Westerwald. Arbeiten zur Rheinischen Landeskunde,Heft 44, F.Dümmler,Bonn

FISCHER,M.M.(1976): Eine theoretische und methodische Analyse mathematischer Stadtentwicklungsmodelle vom Lowry-Tap. Ein methodischer Beitrag zur Regionalforschung.Rhein-Mainische Forschungen,Heft 83,W.Kramer,Frankfurt a.M.

FISCHER,M.M.(1978): Zur Lösung funktionaler regionaltaxonomischer Probleme auf der Basis von Interaktionsmatrizen: Ein neuer graphentheoretischer Ansatz. Karlsruher Manuskripte zur Mathematischen und Theoretischen Wirtschafts- und Sozialgeographie,Heft 25,Karlsruhe(=Geographisches Institut der Universität)

FISCHER,M.M.,M.SAUBERER(1979a): Neuere Entwicklungen in der regionaldemographischen Modellbildung. Modellstruktur und Anwendungsmöglichkeiten am Beispiel Österreichs. Österreichisches Institut für Raumplanung,Materialien,Nr.11,Wien

FISCHER,M.M.(1980): Regional Taxonomie: A comparison of some hierarchic and nonhierarchic strategies.In: Regional Science and Urban Economics,Heft 2

GAEBE,W.(1976): Analyse mehrkerniger Verdichtungsräume. Das Beispiel des Rhein-Ruhr-Raumes, Karlsruher Geographische Schriften,Heft 7,Karlsruhe(=Geographisches Institut der Universität)

GAEBE,W.(1978): Erklärungsversuche industrieller Standortentscheidungen.In: Seminarberichte der Gesellschaft für Regionalforschung,13,S.161-184

GANSER,K.(1972): Nutzwertanalyse zur Bestimmung des Standortes der Kreisverwaltung.Fallstudie für den neuen Kreis Lippstadt-Soest.Bundesforschungsanstalt für Landeskunde und Raumordnung, Bonn-Bad Godesberg

GATZWEILER?H.P.(1975a): Die Anwendung von regulären,homogenen Markov-Ketten-Modellen erster Ordnung zur Deskription und Analyse von Wanderungen.In: Giessener Geographische Schriften,Heft 32,Gießen (=Geographisches Institut der Universität),S.156-164

GATZWEILER,H.P.(1975b): Zur Selektivität interregionaler Wanderungen. Ein theoretisch-empirischer Beitrag zur Analyse und Prognose altersspezifischer interregionaler Wanderungen. Forschungen zur Raumentwicklung,Bd.1,Bonn-Bad Godesberg

GATZWEILER,H.P.,R.KOCH(1976): Makroanalytisches Simulationsmodell der regionalen Bevölkerungsentwicklung und -verteilung in der Bundesrepublik.In:Tagungsbericht und wissenschaftliche Abhandlungen des 4o. Deutschen Geographentages Innsbruck 1975, F.Steiner,Wiesbaden,S.489-5ol

GIESE,E.(Hrsg.)(1975): Symposium "Quantitative Geographie" Gießen 1974. Möglichkeiten und Grenzen der Anwendung methematisch-statistischer Methoden in der Geographie. Giessener Geographische Schriften,Heft 32, Gießen (=Geographisches Institut der Universität) - Symposium Gießen 1974 -

GIESE,E.(1978a): Räumliche Diffusion ausländischer Arbeitnehmer in der Bundesrepublik Deutschland.In: Die Erde,1o9,Jg.,S.92-11o

GIESE,E.(1978b): Weiterentwicklung und Operationalisierung der Standort- und Landnutzungstheorie von ALONSO für städtische Unternehmen.In:Bremer Beiträge zur Geographie und Raumplanung, Bremen(=Universität Bremen,Schwerpunkt Geographie), Heft 1,S.63-79

GIESE,E.(1978c): Kritische Anmerkungen zur Anwendung faktorenanalytischer Verfahren in der Geographie.In:Geographische Zeitschrift,66.Jg.,S.161-182

GIESE,E.,J.NIPPER(1979): Zeitliche und räumliche Persistenzeffekte bei räumlichen Ausbreitungsprozessen - analysiert am Beispiel der Ausbreitung ausländischer Arbeitnehmer in der Bundesrepublik Deutschland. Karlsruher Manuskripte zur Mathematischen und Theoretischen Wirtschafts- und Sozialgeographie.Heft 34,Karlsruhe (=Geographisches Institut der Universität)

GÜSSEFELDT.J.(1974a): Zu einer operationalisierten Theorie des räumlichen Versorgungsverhaltens von Konsumenten. Giessener Geographische Schriften,Heft 34,Gießen (=Geographisches Institut der Universität)

GÜSSEFELDT,J.(1975b): Über ein probabilistisches Simulationsmodell versorgungswirtschaftlicher Interaktionsmuster.In: Giessener Geographische Schriften,Heft 32, S.141-156,Gießen (=Geographisches Institut der Universität)

GOSSEFELDT,J.(1976): Der Einfluß raumdifferenzierender Strukturen auf die Ausprägung menschlicher Interaktionssysteme.In: Tagungsbericht und wissenschaftliche Abhandlungen des 40. Dt. Geographentages Innsbruck 1975,F.Steiner, Wiesbaden S.432-442

GOSSEFELDT,J.(1978): Die Graphentheorie als Instrument zur Beurteilung raumordnungspolitischer Konzepte, dargestellt am Beispiel der Entwicklungsachsen von Baden-Württemberg und Bayern. In: Geographische Zeitschrift, 66. Jg.,S.81-105

GOSSEFELDT,J.(1980): Konsumentenverhalten und die Verteilung zentraler Orte. In: Geographische Zeitschrift, 68. Jg.,S.33-53

HAGGETT,P.(1965): Locational Analysis in Human Geography.E.Arnold Ltd., London(Deutsche Obersetzung durch D.BARTELS,B. und V.KREIBICH (1973): Einführung in die kultur- und sozialgeographische Regionalanalyse. W.de Gruyter,Berlin,New York)

HAGGETT,P.(1978): Spatial Forecasting: a View from the Touchline. In: MARTIN, R.L.,THRIFT,N.J., BENNETT,R.J.: Towards the Dynamic Analysis of Spatial Systems. London

HARD,G.(1972): Ein geographisches Simulationsmodell für die rheinische Sprachgeschichte. In: ENNEN,E., G.WIEGELMANN (Hrsg.): Festschrift Matthias Zender,L.Röhrscheid-Verlag Bonn,S.25-58

HERRMANN,R.(1972): Ein multivariates Modell der Schwebstoffbelastung eines hessischen Mittelgebirgsflusses. In: Biogeographica, 1, S. 87-95

HERRMANN,R.(1973): Eine multivariate statistische Klimagliederung Nordhessens und angrenzender Gebiete. In: M.BORN(Hrsg.): Beiträge zur Landeskunde von Nordhessen. Marburger Geogr. Schriften,Heft 60,Marburg (=Geographisches Institut der Universität Marburg),S.37-55

HERRMANN,R.(1974): Ein Anwendungsversuch der mehrdimensionalen Diskriminanzanalyse auf die Abflußvorhersage. In: Catena 1, S. 367-385

HERRMANN,R.,E.SCHRIMPF(1976): Zur Vorhersage des Abflußverhaltens in tropischen Hochgebirgen West- und Zentralkolumbiens. In: Tagungsbericht u. wissenschaftliche Abhandlungen des 40. Deutschen Geographentages Innsbruck 1975,F.Steiner,Wiesbaden,S. 750-770

HÖLLHUBER,D.(1975): Die Mental Maps von Karlsruhe - Wohnstandortpräferenzen und Standortcharakteristika. Karlsruher Manuskripte zur Mathematischen und Theoretischen Wirtschafts- und Sozialgeographie, Heft 11, Karlsruhe (=Geographisches Institut der Universität)

HÖLLHUBER,D.(1978): Sozialgruppentypische Wohnstandortpräferenzen und innerstädtische Wohnstandortwahl.In: Bremer Beiträge zur Geographie und Raumplanung, Heft 1,Bremen (=Schwerpunkt Geographie der Universität),S.95-105

KANSKY,K.J.(1963): Structure of Transportation Networks.Relationships between Network Geometry and Regional Characteristics.Research Paper No.84,Chicago

KEMPER,F.J.(1975): Die Anwendung faktorenanalytischer Rotationsverfahren in der Geographie des Menschen.In: Giessener Geographische Schriften, Heft 32, Giessen (=Geographisches Institut der Universität),S.34-47

KEMPER,F.J.,W.SCHMIEDECKEN(1977): Faktorenanalyse zum Klima Mitteleuropas. Ein Beitrag zum Problem der Kontinentalität sowie zur Aussagefähigkeit von Faktorenwerten.In: Erdkunde, 31, S. 255-272

KEMPER,F.J.(1978a): Multivariate Analyse nominalskalierte Daten.In:Mitteilungen des Arbeitskreises für neue Methoden in der Regionalforschung,8.Jg., S. 3-36

KEMPER,F.J.(1978b): Ober einige multivariate Verfahren zur statistischen Varianzaufklärung und ihre Anwendung in der Geographie. Karlsruher Manuskripte zur Mathematischen und Theoretischen Wirtschafts- und Sozialgeographie, Heft 28,Karlsruhe (=Geographisches Institut der Universität)

KILCHENMANN,A.(1968): Untersuchungen mit quantitativen Methoden über die fremdenverkehrs- und wirtschaftsgeographische Struktur der Gemeinden im Kanton Graubünden (Schweiz), Juris Druck und Verlag, Zürich

KILCHENMANN,A.(1970): Statistisch-analytische Arbeitsmethoden in der regionalgeographischen Forschung. Untersuchungen zur Wirtschaftsentwicklung von Kenya und Versuch einer Regionalisierung des Landes auf Grund von thematischen Karten. Ann Arbor, Michigan

KILCHENMANN,A.,W.MOERGELI (1970): Typisierung der Gemeinden im Kanton Zürich mit multivariaten statistischen Methoden aufgrund ihrer wirtschaftsgeographischen Struktur. In: Vierteljahresschrift d. Naturforschenden Gesellschaft Zürich,115,S.369-394

KILCHENMANN,A.(1971): Statistisch-analytische Landschaftsforschung. In: Geoforum,7, S. 39-53

KILCHENMANN,A.(1973): Die Merkmalsanalyse für Nominaldaten. Eine Methode zur Analye von qualitativen geographischen Daten basierend auf einem informationstheoretischen Modell. In: Geoforum,15,S.33-45

KILCHENMANN,A.(1975): Zum gegenwärtigen Stand der "Quantitativen und Theoretischen Geographie".In: Giessener

Geographische Schriften, Heft 32, Giessen(=Geographisches Institut der Universität),S.194-208

KILCHENMANN,A.(1978-1979): Karlsruher Manuskripte zur Mathematischen und Theoretischen Wirtschafts- und Sozialgeographie, Heft 24 (1978)-Heft 34 (1979) - Symposium Straßburg 1977

KOCH,R.(1977): Wanderungen und Bevölkerungsentwicklung in der Raumordnungsprognose 1990. In: Informationen zur Raumentwicklung Heft 1/2, S.13-25

KOCH,R.(1978): Ein Beitrag zur Weiterentwicklung regionaler Bevölkerungsprognoseansätze. In: Bremer Beiträge zur Geographie und Raumplanung, Heft 1, Bremen (=Schwerpunkt Geographie der Universität),S. 107-115

KREIBICH,V.,D.REICH(1979): Analyse und Projektion der kleinräumigen Bevölkerungsentwicklung auf der Grundlage des Gebäudebestandes und der Bautätigkeit. In: E. ELSNER (Hrsg.) Demographische Planungsinformationen, Berlin, S. 168-182

KREIBICH,V.(1980): Wohnungsbelegmuster als Grundlage kleinräumiger Bevölkerungs prognosen. In: OSTHEIDER,M.,D. STEINER: Theorie und Quantitative Methodik in der Geographie, Zürich (= Geographisches Institut der ETH Zürich)

LEITNER,H.,H.WOHLSCHLÄGL (1980): Metrische und ordinale Pfadanalyse: Ein Verfahren zur Testung komplexer Kausalmodelle in der Geographie. In: Geographische Zeitschrift, 68, Jg., S. 81-106

LICHTENBERGER,E.(1978a): Klassische und theoretisch-quantitative Geographie im deutschen Sprachraum. In: Berichte zur Raumforschung und Raumplanung, H.1, 22.Jg.,S.9-20

LICHTENBERGER,E.(1978b): Quantitative Geography in the German-Speaking Countries. In: Tijdschrift voor Economische en Sociale Geografie, 69, S. 362-373

MATHERON,G.(1971): The Theory of Regionalized Variables and its Applications. Les Cahiers du Centre de Morphologie. Mathematique de Fontainebleau, No.5, Fontainebleau

NIPPER,J.,U.STREIT(1977): Zum Problem räumlicher Erhaltensneigung in räumlichen Strukturen und raumvarianten Prozessen. In: Geographische Zeitschrift, 65.Jg.,S. 241-263

NIPPER,J.,U.STREIT(1978): Modellkonzepte zur Analyse, Simulation und Prognose raum-zeit-varianter stochastischer Prozesse. In: Bremer Beiträge zur Geographie und Raumplanung, 1,Bremen(=Schwerpunkt Geographie der Universität),S. 1-17

NIPPER,J.(1981): Autoregressiv- und Kriging-Modelle. Zwei Ansätze zur Erfassung raumvarianter Strukturen. In: OSTHEIDER,M.,D.STEINER: Theorie und Quantitative Methodik in der Geographie, Zürich (=Geographisches Institut der ETH Zürich

OLEA,R.A.(1975): Optimum Mapping Techniques Using Regionalized Variable Theory. Kansas Geological Survey, Series on Spatial Analysis, No.2, Lawrences, Kansas

OSTHEIDER,M.,D.STEINER(1981): Theorie und Quantitative Methodik in der Geographie Zürich(=Geographisches Institut der ETH Zürich)

ROHDENBURG,H.,U.SABELBERG u.H.WAGNER(1976): Sind konkave und konvexe Hänge prozesspezifische Formen? - Ergebnisse von Hangentwicklungssimulationen mittels EDV.In:Catena,3,S.113-136

SAUBERER,M.,C.CSERJAN(1972): Sozialräumliche Gliederung Wiens 1961.Ergebnis einer Faktorenanalyse.In: Der Aufbau, S.284-306

SAUBERER,M.(1973a): Anwendungsversuche der Faktorenanalyse in der Stadtforschung - Sozialräumliche Gliederung Wien. Seminarberichte der Gesellschaft für Regionalforschung,Bd.7,Heidelberg

SAUBERER,M.(1973b): Quantitative Methoden in der Geographie und Raumforschung. Überblick und Arbeitsbeispiele. Diss. Geogr. Institut der Universität Wien

SAUBERER,M.(1974): Zur Planungsrelevanz der Faktorenanalyse.In: AMR INTERN, Mitteilungen des Arbeitskreises für neue Methoden in der Regionalforschung Wien,Nr.15,S.15-19

SAUBERER,M.(1976): Bewertung von Gemeindezusammenlegungen mit Hilfe von Nutzwertanalysen. In: Tagungsbericht und wissenschaftliche Abhandlungen des 40. Deutschen Geographentages Innsbruck 1975,F.Steiner, Wiesbaden, S. 455-472

SCHÄTZL,L.(1973): Räumliche Industrialisierungsprozesse in Nigeria. Giessener Geographische Schriften,Heft 31,Giessen(=Geographisches Institut der Universität)

SCHÄTZL,L.(1978):Wirtschaftsgeographie 1, UTB Schöningh, Paderborn

SCHICKHOFF,I.(1974): Graphentheoretische Untersuchungen am Beispiel des Schienennetzes der Niederlande. Ein Beitrag zur Verkehrsgeographie. Dissertation Münster 1974 sowie Duisburger Geographische Arbeiten,Heft 1, Duisburg (= Seminar für Geographie der Gesamthochschule Duisburg) 1978

SCHICKHOFF,I.(1977): Matrizentheoretische Verfahren zur Bestimmung der Zugänglichkeit von Knotenpunkten eines Verkehrsnetzes, aufgezeigt am Beispiel "Eisenbahnnetz Randstad Holland". In: Tijdschrift voor Economische en Sociale Geografie,Bd.68, S.152-167

SCHILLING-KALETSCH,I.(1976): Wachstumspole und Wachstumszentren. Untersuchung zu einer Theorie sektoral und regional polarisierter Entwicklung. Arbeitsberichte und Ergebnisse zur wirtschafts- und sozialgeographischen Regionalforschung, Heft 1, Hamburg (=Geographisches Institut der Universität)

STÄBLEIN,G.(1972): Modellbildung als Verfahren zur komplexen Raumerfassung. In: Würzburger Geographische Arbeiten, Heft 37,(Gerling Festschrift), Würzburg(=Geographisches Institut der Universität),S.67-93

STÄBLEIN,G.,G.STÄBLEIN-FIEDLER(1973): Faktorenanalytische Untersuchungen zur fremdenverkehrsgeographischen Struktur der Provinzen Spaniens. In: Marburger Geographische Schriften, Bd. 59,Marburg(=Geographisches Institut der Universität), S.145-161

STÄBLEIN,G.(1975): Der Einsatz der Faktorenanalyse und Printermap für die Regionalanalyse und die Regionalisierung. In: Giessener Geographische Schriften,Heft 34, Giessen(=Geographisches Institut der Universität), S.143-151

STEINER,D.(1965a): Die Faktorenanalyse; ein modernes statistisches Hilfsmittel des Geographen für die objektive Raumgliederung und Typenbildung. In: Geographica Helvetica, 20,S.20-34

STEINER,D.(1965b): A Multivariate Statistical Approach to Climatic Regionalization and Classification.In: Tijdschrift van het Koninklijk Nederlandsch Aardrijkskundig Genootschap,82,S.329-347

STEINER,D.(1978): Modelle zur Darstellung geographischer Systeme. In: Veröffentlichungen der Geographischen Kommission der Schweizerischen Naturforschenden Gesellschaft,Nr.5, S.79-116,Lausanne

STEINER,D.(1979a): Systemtheorie/Systemanalyse und Geographie. Zum Buch von G.P.CHAPMAN "Human and Environmental Systems- A Geographers Appraisal". In:Geographische Zeitschrift,67.Jg., S.185-210

STEINER,D.(1979b): A minicomputer-based geographical data processing system. Proceedings of the NATO Advanced Study Institute on Map Data Processing, Marata, Italien

STREIT,U.(1973): Ein mathematisches Modell zur Simulation von Abflußganglinien(Am Beispiel von Flüssen des Rechtsrheinischen Schiefergebirges). Giessener Geographische Schriften, Heft 27, Giessen(=Geographisches Institut der Universität)

STREIT,U.(1975a): Zeitreihensimulation mit Markov-Modellen, dargestellt an Beispielen aus der Hydrologie. In: Giessener Geographische Schriften Heft 32,Giessen(=Geographisches Institut der Universität),S.165-180

STREIT,U.(1975b): Erzeugung synthetischer Abflussdaten mit Hilfe eines zeit- und raumvarianten Modells im Einzugsgebiet der Lahn. In: Erdkunde,29,S. 92-105

STREIT,U.(1976): Über eine Anwendungsmöglichkeit von Simulationsmodellen bei der wasserwirtschaftlichen Planung. In: Tagungsbericht und wissenschaftliche Abhandlungen des 40. Deutschen Geographentages Innsbruck 1975, F. Steiner,Wiesbaden,S.503-514

STREIT,U.(1978): Räumliche und zeitliche Erhaltensneigung als Ansatzpunkt stochastischer Modellbildung in der Physischen Geographie. Karlsruher Manuskripte zur Mathematischen und Theoretischen Wirtschafts- und Sozialgeographie, Heft 30,Karlsruhe (=Geographisches Institut der Universität Karlsruhe)

STREIT,U.(1979): Raumvariante Erweiterung von Zeitreihenmodellen: Ein Konzept zur Synthetisierung monatlicher Abflußdaten von Fliessgewässern unter Berücksichtigung von Erfordernissen der wasserwirtschaftlichen Planung. Giessener Geographische Schriften, Heft 46,Giessen(=Geographisches Institut der Universität Giessen)

TAUBMANN,W.(1979a): Erscheinungsformen und Ursachen sozioökonomischer Disparitäten am Beispiel von Jütland/Dänemark. In: Münstersche Geographische Arbeiten,Heft 4,Münster (=Geographisches Institut der Universität), S. 149-183

TAUBMANN,W.(1979b): Zur Abgrenzung von sogenannten Problemgebieten - Bemerkungen zu Indikatoren und Verfahrensweisen. Karlsruher Manuskripte zur Mathematischen und Theoretischen Wirtschafts- und Sozialgeographie,Heft 32,Karlsruhe(=Geographisches Institut der Universität)

VETTER,F.(1970): Netztheoretische Studien zum Niedersächsischen Eisenbahnnetz Ein Beitrag zur angewandten Verkehrsgeographie. Abhandlungen des 1. Geographischen Instituts der Freien Universität Berlin,Bd.15,Verlag Reimer, Berlin

WERNER,CHR.(1966): Zur Geometrie von Verkehrsnetzen. Die Beziehung zwischen räumlicher Netzgestaltung und Wirtschaftlichkeit. Abhandlungen des 1. Geographischen Instituts der Freien Universität Berlin,Bd.10,Verlag Reimer Berlin

WINDHORST,H.-W.(1979): Die sozialgeographische Analyse raum-zeitlicher Diffusionsprozesse auf der Basis der Adoptorkategorien von Innovationen. Die Ausbreitung der Käfighaltung von Hühnern in Südoldenburg. In: Zeitschrift für Agrargeschichte und Agrarsoziologie,27.Jg.,1979,S.244-266

WITTENBERG,W.(1978): Neuerrichtete Industriebetriebe in der Bundesrepublik Deutschland 1955-1971. Giessener Geographische Schriften,Heft 44, Giessen(=Geographisches Institut der Universität)

J O S E F   N I P P E R

ON THE MEANING AND INTERPRETATION OF SPATIAL AUTOCORRELATION IN THE ANALYSIS OF SOCIO-ECONOMIC SPATIAL STRUCTURES AND SPATIO-TEMPORAL PROCESSES

From:

BAHRENBERG, GERHARD und STREIT, ULRICH (Eds.):
German Quantitative Geography. Papers presented at the 2nd European Conference on 'Theoretical and Quantitative Geography' in CAMBRIDGE 1980. Münstersche Geographische Arbeiten 11. Schöningh, Paderborn 1981, p. 27 - 34

Author:

Dr. Josef Nipper

Geographisches Institut
Universität Giessen

Senckenbergstraße 1

6300 Giessen

ZUSAMMENFASSUNG

Eine optimale Analyse und Modellierung räumlicher Strukturen und raum-zeit-varianter Prozesse kann nur in Form integrierter Raum-Zeit-Modelle erfolgen. Ein wesentlicher Bestandteil solcher Modelle ist deren autoregressive (autokorrelative) Komponente. Neben deren Identifikation und Schätzung - hier sind eine Reihe von Methoden bekannt - ist deren inhaltliche Interpretation für die Erklärung des Prozesses von ausschlaggebender Bedeutung.

Im Hinblick auf eine adäquate Interpretation sind zunächst zwei Dinge zu beachten:

1. Ähnlich wie der Begriff "Distanz" haben die statistischen Begriffe "Autoregression" bzw. "Autokorrelation" a priori keine konkrete inhaltliche Bedeutung. Sie fungieren als Stellvertreter für (in der Regel) komplexe Sachverhalte.
2. Autoregressions- bzw. Autokorrelationsstrukturen sind das Ergebnis solcher komplexen Faktoren. Bei der Distanz hingegen verläuft die Wirkung der Faktoren nur parallel.

Eine nähere Betrachtung der räumlichen Autokorrelation führt dabei zu folgenden Ergebnissen:

- Die durch den statischen Parameter "räumliche Autokorrelation" gemessene Ähnlichkeit räumlich benachbarter Strukturwerte bzw. Prozeßrealisationen kann primär zwei Ursachen haben:
  1. Allgemeine großräumige Einflußgrössen, die innerhalb des Gesamtgebietes höchstens gering variieren, führen zu ähnlichen Werten bei Nachbarn.
  2. Der räumliche Transfer von Masse, Energie und/oder Information führt zu einer Angleichung der räumlich benachbarten Werte.
- Eine befriedigende kausale Interpretation räumlicher Autokorrelation sozioökonomischer Prozesse ist häufig recht schwierig, da die den Prozeß tragenden und steuernden Entscheidungseinheiten unterhalb der der Analyse zugrundeliegenden räumlichen Aggregationsebene liegen. Diese Entscheidungseinheiten weisen bei ihrem Handeln hohe Freiheitsgrade auf. Die Wertebildung auf der betrachteten räumlichen Aggregationsebene erfolgt durch Summierung bzw. Mittelbildung der individuellen Entscheidungen.

Ein besonderes Problem stellt die Interpretation räumlicher Autokorrelation dar, wenn raum-zeit-variante Prozesse nicht durch ein integriertes Raum-Zeit-Modell dargestellt werden bzw. werden können, sondern rein räumliche Modelle zu verschiedenen Zeitpunkten erstellt werden. Es zeigt sich, daß die Ursachen für das Auftreten räumlicher Autokorrelation dann (als Folge der Suboptimalität des Vorgehens) vielfältiger Art sein kann, wobei die einzelnen Ursachenquellen nicht eindeutig voneinander unterschieden werden können.

Am Beispiel der Ausbreitung der Gastarbeiterbeschäftigung in der Bundesrepublik Deutschland lassen sich so mindestens vier verschiedene Ursachen angeben, die an dem Entstehen räumlicher Autokorrelation in rein räumlichen zeitdiskreten Modellen beteiligt sind.

1. INTRODUCTION

In the analysis and modelling of spatio-temporal processes, the concepts space, time and process should not be separated. On the contrary, they are internally related (BLAUT 1961, HÄGERSTRAND 1973, SAYER 1977, ULLMANN 1974, BENNETT/CHORLEY 1978). If we consider that spatial patterns can really be understood as realizations of spatio-temporal processes, then this basic interrelation should also be attended to the explanation of spatial structure, and - if possible - be discussed explicitly. An optimal analysis and modelling of socio-economic structures and processes can then surely only be completed in the form of integrated spatio-temporal models. In a general form they can be represented as:

$$Y_t = S_A Y_t + S_R X_t$$

(S_A, S_R = autoregressive or regressive transformation respectively)

Individual specific models as well as the procedure for identifying the models and estimating are comprehensively set out in BENNETT's book (1979) "Spatial time series".

An essential part of these models is the spatio-temporal autoregressive structure S_A or the spatio-temporal autocorrelation function respectively. Apart from the formal identification and estimation of this phenomenon - a number of methods are known pertaining to this (see CLIFF/ ORD 1973, ORD 1975, MARTIN/OEPPEN 1975, BENNETT 1979) - its interpretation with regard to the contents is of decisive importance for the explanation of the process. Initially, two points seem to be of significance concerning an adequate interpretation.

1. Autoregression and autocorrelation are concepts which have no concrete a priori meaning. To a certain extent they are closely connected to the concept "distance". Distance can't be interpreted causally as an explanatory variable in most cases. Rather, it functions as an indicator for circumstances which are (usually) highly complex but coincide to a large extent with distance. Besides its good descriptive qualities distance only obtains a meaning for the process through this coincidence.

2. Autoregression structures or autocorrelation structures are the result (effect) of the underlying factors mentioned in point 1, whereby the result is reflected in the similarity of the spatially and/or temporally neighbouring values. In comparison to the indicator "distance", a considerable difference can be observed here. Distance is not the measureable result of the influencing factors. On the contrary the effect of the factors runs only parallel to distance.

2. ON THE INTERPRETATION OF SPATIAL AUTOCORRELATION IN SPATIO-TEMPORAL PROCESSES

Temporal autoregression structures can generally be established as the effect of reaction time, consistency, experience, tradition, development, feedback, memory (VRIES-REILINGH 1968, BARTELS 1970). It seems immediately evident that a river, which now has a water-level of 10 meters, one hour later will have more probably one of 9,5 meters than one of 5 meters. The relatively constant flowing speed and its steady change can be stated as reasons. The annual birth-rate in a certain town will waver only slightly from year to year. Traditional social patterns of behaviour, religious norms, which alter only slowly, can be put forward as explanations. They don't make a more random sequence of higher or lower birth-rate feasible. This kind of temporal stability, which is defined as dependency of present and future process realizations on the past is recognized without exception as a causal reason.

Spatial autocorrelation is shown by the fact that spatially neighbouring values or process realizations cannot be proved, statistically speaking, as independent, but rather the probability is very high, that spatial neighbours have similar values. The causes of such similarities can be divided into two categories:

1. General large scale variables showing at the most only slight variation in the whole area, are involved in the control of the process and lead to similar process realizations in neighbouring area units (e.g. flat land causes an even distribution of rainfall; traditional social norms result in similar birth-rates in neighbouring parishes). These quantities are actually exogenous factors which remain more or less constant for individual spatial units.

2. Spatial transfer of mass, energy and/ or information leads to an adjustment of the spatially neighbouring values. On the regarded spatial aggregation le-

vel, such exchange processes give rise to a kind of self-propelling of the process. The influence of these endogenous components are measured statistically as spatial autocorrelation.

The causal reason and explanation of the occurrence of spatial autocorrelation often seems to be less concrete in practice than for temporal autocorrelation. It arises initially as a formal statistical parameter. A significant reason for this can be found in the fact that time contains a natural order relation (before/after) which compels a clear cause-effect direction. In the case of space, on the other hand, such a relation does not exist. On the contrary, it can only be defined through hypotheses concerning cause-effect relations. Only then can spatial autocorrelation be explained.

For physio-geographical processes two facts can often be determined which support a satisfactory causal interpretation of spatial autocorrelation:

1. In physio-geographical processes, large-scale variables determine to a great extent the course of the process and hereby the spatial variation of the values. Local individualities and influences can only give rise to modification.

2. The physical-chemical exchange-processes underneath or on the basic aggregation level, which cause similar values in neighbouring spatial units (e.g. measuring stations), can very often be identified in a concrete way and are known quite exactly.

An example may back up these ideas. The clear autocorrelation structure in the map of rainfall distribution as drawn up by NIPPER/STREIT (1977) can be well explained, according to SCHIRMER (1973), by the west wind drift as a stimulus for exchange processes and by a large scale formation of hexogonal cell structures to be found in the atmosphere.

In the explanation of spatial autocorrelation in socio-economic structures and processes, we can take the following as a starting point:

The explanation of socio-economic processes proceeds largely, in the sense of systems theory through the identification of the decision units which convey and direct the process as well as their reaction. We can conclude that such decision units are necessarily established underneath the basic spatial aggregation level (e.g. city area, parish, county) (see FEIBLEMAN 1954, CHAPMAN 1977 and STEINER 1979). People, companies and institutions are examples of such units. The constitution of values on the examined spatial aggregation level can be seen as a sort of summing up or averaging the results of individual decisions. Often there is only a relatively small amount of conclusive knowledge available which concerns the mechanism of individual decision processes, as well as that which concerns the interrelation between the individual ones. Often only hypotheses are to hand. We can also proceed on the assumption that the degree of freedom in the individual decision units is generally higher than it is in the case of physio-geographical processes. So therefore general psychologic-sociologic factors as, for example, fashion, competition, religious norms are less significant due to the higher degree of freedom than the corresponding factors in physio-geographical processes. These circumstances however by no means indicate that spatial autocorrelation in socio-economic processes is only a formal statistical component and has absolutely no meaning as regards the content. Rather, the strong stochastic character of spatio-temporal socio-economic processes can be observed.

All in all, these statements show very clearly that a satisfactory interpretation of spatial autocorrelation in socio-economic processes can only be worked out with difficulty.

3. PROBLEMS FOR THE INTERPRETATION OF SPATIAL AUTOCORRELATION IN PURELY SPATIAL MODELS

In practice the situation is often seen whereby the spatio-temporal process is not represented by an integrated space-time model, but only purely spatial models can be provided for different points in time. The general theoretic-philosophic problems of purely spatial models are not insignificant as BENNETT (1979) and BENNETT/CHORLEY (1978), have demonstrated in detail. In the analysis of spatio-temporal socio-economic processes, the situation in the data basis (length of time too short, time lag too great, missing temporal equidistance, changing of the spatial basis through reforms in the area) often makes the proceedings not quite up to optimum.

In such case the following possibilities can be principally indentified with reference to the interpretation of the spatial autoregressive structure:

A) Spatial autocorrelation is a purely formal model component:
   - which is statistically necessary to obtain an optimal model building.
   - Which is used as explanatory variables are not available and also because no ideas or hypotheses can be produced which could identify the variables.

   The aim of the model is therefore solely an optimal formal description of the process, really for the aim of simulation or forecasting not concerned with the identification of variables and their effects (explanation). The model is purely a black box model.

B) Spatial autocorrelation is established with regard to its content. Apart from the two categories already mentioned, two further possibilities are already conceivable.

   The four possibilities can be briefly characterized as follows:

   a) An endogenous spatial dynamic exists on the grounds of spatial interactions, i.e. a transfer from mass, energy and/or information (e.g. diffusion or migration processes). This spatial exchange has the result that the process realizations of neighbouring, regions are similar, i.e. the spatial structure is spatially autocorrelated. This spatial dynamic operates horizontally. The available spatial autocorrelation is the result of a primarily spatial component.

   b) Large scale variables which vary only slightly on the considered aggregation level within the examined area, are involved in the process and have the result that the realizations of the process of neighbouring regions are similar. The spatial autocorrelation which has thus arisen can also be regarded as the result of a predominantly spatial component.

   c) Spatial autocorrelation measures the temporally persistent result of a spatially autocorrelative pattern established some time before. There are not so much spatial components which caused the spatial autocorrelation as the temporal tendency towards maintenance which conserves a past spatial structure showing spatial autocorrelation. Spatial autocorrelation is in this case primarily the result of a process with temporal autocorrelation, i.e. the result of a primarily temporal component.

   d) Spatial autocorrelation is regarded as an indicator for variables which are known but not quantified/quantifiable. Therefore it explains something of the residual component. Different from Bb) the values of these variables in the examined area are not necessarily relatively constant. Spatial autocorrelation is used here, as in possibility A) as a formal model component. However, it is not a pure black box model, as the explanatory variables are known.

For the modelling of concrete socio-economic structures or processes, a combination of interpretation possibilities mentioned above is likely to be the case. For the clarity of the interpretation, though, we arrive at the following result: Even in the case when the combination of sources for spatial autocorrelation can be exactly identified, a quantitative division of the autocorrelative model component into its single origins is not possible. This implies however that a more exact determination of the extent of the influence is not feasible for the individual causes. In particular it is not possible to filter out the contribution of spatial autocorrelation which is brought about by temporal autocorrelation (Bc). This very point would be desirable as only the imperfect analysis of spatio-temporal processes through time-discret purely spatial models give rise to this phenomenon.

4. SOURCES OF SPATIAL AUTOCORRELATION IN THE SPREADING OF THE EMPLOYMENT OF FOREIGN WORKERS IN THE FEDERAL REPUBLIC OF GERMANY

The variety of factors simultaneously contributing to the formation of spatial autocorrelation in the case of analysing spatio-temporal processes by purely spatial models (⟶ cross sectional analysis) can be demonstrated by the example of the spreading of the employment of foreign workers in the Federal Republic of Germany on the spatial aggregation level of the 141 labour exchange districts (LED).

1. Precondition for the spread of employment of foreign workers is the number of jobs available at the beginning of the 1960s which could not be filled by Germans. This demand is first evident in regions of Baden-Württemberg. So, a predefined pattern is available on which the expansion starts. This pattern is spatially autocorrelated as the autocorrelation coefficient R(1) = 0,82 of the rate of unemployment for 1960 shows - at this time an accurate indicator for the demand for labour.
2. The need for labour is strongly dependent on the economic structure in the individual LED, as well as on its development. On the one hand the economic structure changes relatively slowly within a small number of years. On the other hand the economic development in the regions is closely connected to the general economic development in the Federal Republic of Germany. Both have as a consequence that the spatial demand structure for labour as shown in 1960 does not alter abruptly.
3. On the basis of the pattern shown in point one, the employment of foreign workers increases and spreads out. In 1960 a spatial autocorrelation coefficient of R(1) = 0,74 for the foreign worker ratio can be calculated. The further increase of employment of foreign workers in each single LED is influenced, as shown in point 2 by a spatial demand structure which is relatively constant over a period of time. Altogether an autoregressive spatial structure can be expected, also in the following years, for the spatial distribution of foreign workers. According to this, the cause is the temporal consistency in the spatial demand structure (⟶ Bc).
4. The steady general economic growth in almost all branches of the West German economy leads to the increased demand for labour in more and more regions, as is demonstrated in points 1 and 2, especially in such areas which border on LEDs already employing a large number of foreign workers. Information passed from one firm to another concerning taking on foreign workers leads to their increased employment in "new regions". This means: the exchange of information (diffusion of the idea of employing foreigners) leads to the formation of an autocorrelative structure of foreign employment or this structure remains or develops (⟶ Ba).
5. As mentioned above the economic development of each LED is closely connected with the general economic development of the Federal Republic. That means, a general large scale variable can be identified, which governs to some extent the spreading process of foreign worker employment (⟶ Bb).
6. Apart from the data for the development of unemployment and for the development of available jobs (since 1975) there is no data on the spatial aggregation level of the LEDs. So the economic situation and development of the LEDs, which explains the spreading process to a great extent, could not fully be quantified. Therefore, probably some amount of the measured autocorrelation

structure will have been generated by these missing variables (⟶ Bd).

REFERENCES

BARTELS,D. (ed.): Wirtschafts- und Sozialgeographie. Köln/Berlin 1970

BENNETT,R.J. and R.J.CHORLEY: Environmental systems. Philosophy, analysis and control. London 1978

BENNETT,R.J.: Spatial time series. Analysis, forecasting, control. London 1979

BLAUT,J.M.: Space and process. The Professional Geographer, vol. 13, No. 4, 1961, pp.1-7

CHAPMAN,G.P.: Human and environmental systems. A Geographer's appraisal. London/New York/San Francisco 1977

CLIFF,A.D. and J.K.ORD: Spatial autocorrelation. London 1973

FEIBLEMAN,J.K.: The theory of integrative levels. British Journal for the Philosophy of Science, vol. 5, 1954, pp. 59-66

HÄGERSTRAND,T.: The domain of human geography. In: CHORLEY,R.J. (ed.): Directions in geography. London 1973, pp.67-87

MARTIN,R.L. and J.E.OEPPEN: The identification of regional forecasting models using space-time correlation functions. Inst. of Brit. Geographers, Transactions, No. 66, 1975, pp. 95-118

NIPPER,J. and U.STREIT: Zum Problem der räumlichen Erhaltungsneigung in räumlichen Strukturen und raumvarianten Prozessen. Geographische Zeitschrift, Jg. 65, 1977, pp. 241-263

ORD,J.K.: Estimation methods for models of spatial interaction. Journal of the American Statistical Association, vol. 70, 1975, pp. 120-126

SAYER,R.A.: Gravity models and spatial autocorrelation, or atrophy in urban and regional modelling. Area, vol. 9, 1977, pp. 183-189

SCHIRMER,M.: Die räumliche Verteilung der Bänderstruktur des Niederschlags in Süd- und Südwestdeutschland. Forschungen zur Deutschen Landeskunde, Bd. 205, Bonn-Bad Godesberg 1973

STEINER,D.: Systemtheory/Systemanalyse und Geographie. Geographische Zeitschrift, Jg. 67, 1979, pp. 185-210

ULLMAN,E.D.: Space and/or time: opportunity for substitution and prediction. Inst. of Brit. Geographers, Transactions, No. 63, 1974, pp. 125-139

VRIES-REILINGH,H.D.de: Gedanken über die Konsistenz in der Sozialgeographie. In: RUPPERT,K. (ed.): Zum Standort der Sozialgeographie. Münchener Studien zur Sozial- und Wirtschaftsgeographie, Bd. 4, Kallmünz/Regensburg 1968, pp. 109-117

U L R I C H   S T R E I T

ANALYSING SPATIAL DATA BY STOCHASTIC METHODS: SOME EXAMPLES FROM PHYSICAL GEOGRAPHY

From:

BAHRENBERG, GERHARD und STREIT, ULRICH (Eds.):
German Quantitative Geography. Papers presented at the 2nd European Conference on 'Theoretical and Quantitative Geography' in CAMBRIDGE 1980. Münstersche Geographische Arbeiten 11. Schöningh, Paderborn 1981, p. 35 - 44

Author:

Prof. Dr. Ulrich Streit
Institut für Geographie
Universität Münster
Robert-Koch-Straße 26-28
D-4400 Münster

ZUSAMMENFASSUNG

Ebenso wie Zeitreihen zeigen auch räumlich lokalisierte Daten häufig Persistenzeffekte, also regelhafte Ähnlichkeiten in den Variablenwerten benachbarter Raumeinheiten. Sie können als endogene Steuerungsmechanismen im zugrunde liegenden räumlichen stochastischen Prozeß (Zufallsfeld) angesehen und auf Speicherungs- und Transfervorgänge von Massen und Energien zurückgeführt werden. Zur Analyse derartiger distanz- und richtungsabhängiger Persistenzeffekte eignen sich räumliche Autokorrelogramme und Variogramme, die für verschiedene Beispiele (Niederschläge, Grundwasserstände,Schwermetallgehalte, geobotanische Daten) geschätzt und interpretiert werden. Die auf der Variogrammanalyse aufbauende Kriging-Methode zur Interpolation räumlicher Daten wird an einem Beispiel verdeutlicht.

1. INTRODUCTION

Methods of analysing and modelling stochastic processes are gaining increasing interest in certain areas of physical geography, in particular in hydrology and climatology, partly to supplement and also to replace deterministic techniques. An excellentsurvey covering a remarkable number of relevant papers published in English is given by UNWIN (1977).

While classical statistics rely on stochastically independent variables, frequently those processes showing intrinsic interdependences are of particular interest in physical geography. Such more or less strong persistences may be regarded as a consequence of temporal and/ or spatial transfer and storage processes of mass and energy. Well-known examples are the slow temporal variations of sea levels with their buffering capacity against stochastically independent precipitations of the relatively smooth surfaces in the isobaric relief of spatial atmospheric pressure distributions.

This article presents some examples from physical geography applying spatial stochastic techniques to the analysis of spatial data for persistences and anisotropic effects and -in the last part- to spatial forecasts, i.e. the spatial interpolation and extrapolation of missing data.

2. SPATIAL AUTOCORRELOGRAMMS

The concept of spatial autocorrelation based largely on the work of CLIFF and ORD (1973) relies on the assumption of a weak stationary process where the mean values and variances are equal in all units of space and where the covariance, although being dependent on the distance and direction of two units of space, is not dependent on their absolute position.

The fundamental problem in constructing spatial autocorrelation coefficients for units of space of deliberate shape and arrangement is the necessity of defining neighbourhood relations over various spatial lags. Formally, it can be solved by defining weights

$$w_{ij}^{(k)} \begin{cases} >0, \text{if the units of space i and j are neighbours with respect to the spatial lag k;} \\ =0 \text{ otherwise} \quad (i,j=1,\dots,n;\ k=1,2,\dots) \end{cases} \qquad (1)$$

for each pair of spatial sites (i,j). The freedom of choice in defining these spatial neighbourhood relations is, at least theoretically, a significant advantage of this concept of spatial autocorrelation: Theories or hypotheses prevailing in a specific discipline and concerning endogenous control mechanisms of spatial processes can and should be taken into consideration in the weight definition (NIPPER and STREIT, 1977).

A spatial autocorrelation coefficient of order k for metric data can be estimated as a normalized covariance:

$$I(k)= \sum_{i=1}^{n} z_i z_i(k) \;/\; \sum_{i=1}^{n} z_i^2 \qquad (2)$$

with: $z_i=y_i-\bar{y}$, $\bar{y}$= mean of the measured values $\{y_i \mid i=1,\dots,n\}$;

$z_i(k)= \sum_{j=1}^{n} w_{ij}^{(k)} \cdot z_j$, being $\sum_{j=1}^{n} w_{ij}^{(k)}= 1$ for all i and k.

Its significance has to be checked approximately using a normal distribution according to CLIFF and ORD (1973). Here, a slightly modified coefficient

$$R(k) = I(k)\cdot((Var(z_i)/Var(z_i(k)))^{1/2} \quad (3)$$

is used being always within the interval $[-1,+1]$ (HAGGETT, CLIFF and FREY, 1977, 375).

In the first example, the 'Münsterland' (NW-Germany) precipitation regime is to be analysed for spatial effects of persistence and their variabilities with time. For this purpose, the monthly precipitation amounts (mm) as measured at 65 test points have been analysed for each year of a series covering 17 years. A definition of binary neighbourhood weights $w_{ij}^{(k)}$ is based on the Thiessen polygon method frequently used in hydrology and climatology and assigning to each test point a surrounding as a reference area (Fig.1).

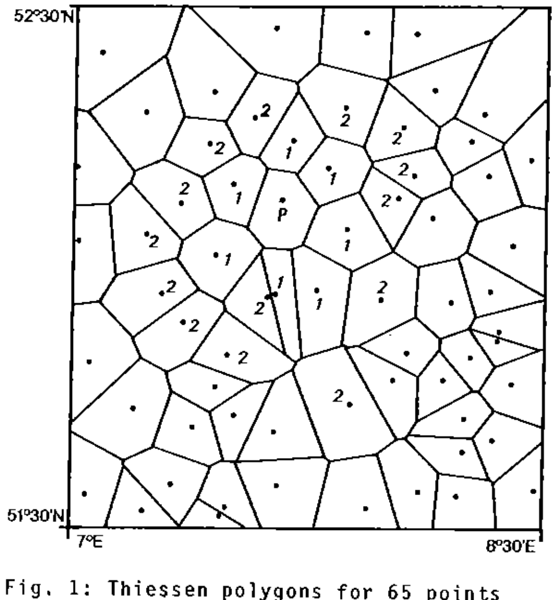


Fig. 1: Thiessen polygons for 65 points with precipitation measurements in the 'Münsterland': Neighbours of order 1 and 2 for P.

Two points i and j shall be first order neighbours, if they share a common polygon segment; they are to be neighbours of the spatial lag k (>1), if i has a neighbour of (k-1)-th order being at the same time first order neighbour of j. This recursive definition according to the criterion of contingency is based on the experience in climatology that, at least in the plains and with means obtained over a longer period of time, the amount of precipitation measured at one point can be regarded as being representative for a larger surrounding area.

Table 1: Spatial autocorrelation coefficients R(k) of precipitation data, basing on 65 meteorological stations in the 'Münsterland'
(+ =significant at 0.05 -level)

Month,year	R(1)	R(2)	R(3)	R(4)	R(5)
Jan.1970	.86$^+$	.85$^+$	.87$^+$	.23	-.06
Jan.1971	-.01	-.10	-.05	-.05	-.05
Jan.1959	.64$^+$	.45$^+$	.24$^+$	-.09	-.45$^+$
Jul.1959	.40$^+$	.30$^+$	-.24$^+$	-.31$^+$	-.53$^+$

On the basis of this special structure, spatial autocorrelation coefficients up to a lag of k=5 have been calculated. Table 1 shows that extreme changes in the spatial persistence may occur from one year to another: While in January 1970 a strong spatial clustering of similar values of precipitation may be supposed with a neighbourhood radius of 3 spatial lags (~22 km), precipitations in the following year show no intrinsic interdependence pattern whatsoever ('white noise'). General rules for the temporal variations of the spatial correlograms, e.g. depending on wet and dry years, could not be detected. A comparison of different months of the same year (1959 being a characteristic example) gives an indication of a tendency towards higher persistences during the winter months; this obviously reflects a smaller percentage of convective precipitation.

The second example discussed here demonstrates a different kind of defining neighbourhood weights between points in space. SYMADER (1979) has analysed at 208 sampling points located in the Northern Eifel (between Aachen and Bonn) the heavy metal content in the soil surface layer. For manganese (Mn), he found concentrations between 10 and more than 900 ppm and a mean value of approximately 200 ppm. To examine whether in the covariation of these data there exist also directional differences besides differences in distances, sector specific correlograms were calculated. The value of 1 is assigned to the neighbourhood weight $w_{ij}^{(k)}$ of two points i and j, if j with respect to i is located in the sector observed, as well as being in the distance ring k. In WE-direction, the exponential decay shows a good approximation to the theoretical autocorrelation function

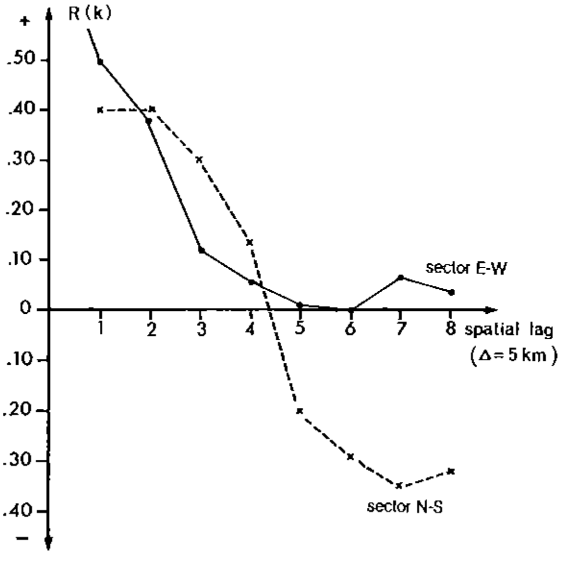


Fig. 2: Spatial autocorrelation coefficients R(k) of Mn-concentration in the Northern Eifel

of the 1st order AR (autoregressive) model (Fig.2). In NS-direction, however, there is a rapid change from positive autocorrelation in the first three spatial lags to negative values;this points to a sort of 'plateau effect' in the spatial distribution of the Mn-contents. An anisotropy of this kind renders the construction of a suitable stochastic model more difficult.

The third example is taken from a mapping of the blackberry flora (*rubus*) performed by WITTIG and WEBER (1978) in the 'Münsterländer Bucht' on a regular grid basis (grid size 2.8 km). Numerical analysis was based on a portion comprising 225 of these grid areas. Neighbourhood weights are determined according to the criterion of common edges and, for higher spatial lags, according to the principle of shortest paths. This example differs from the two proceeding ones; here, the spatial autocorrelation of a binary variable is subjected to analysis,i.e. existence or non-existence of the respective species.

To estimate spatial autocorrelation in binary variables, CLIFF and ORD (1973) recommend the 'black-white' statistic. Figure 3 shows these normalized statistics for two of the most abundant *rubus* species in the 'Münsterland': *Rubus adspersus* and *rubus gratus*, both occur preferentially in areas occupied by the *Quercion roboris-petraeae* (oak forests on acidic soils).

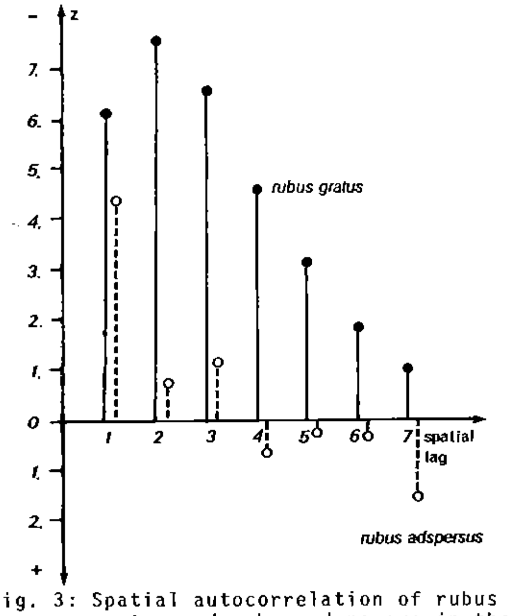


Fig. 3: Spatial autocorrelation of rubus gratus and rubus adspersus in the 'Münsterländer Bucht' (z-scores of the black-white statistic)

In the first spatial step, both show a high positive autocorrelation; however, while *rubus gratus* preserves this spatial persistence even over large spatial lags, it rapidly disappears in the case of *rubus adspersus*. This strongly indicates that physiological or ecological factors acting in a differentiating way might be involved. More careful geobotanical analyses have shown that *rubus gratus* may in fact even grow on very poor quartz sandy soils or on pseudogleys while *rubus adspersus* rarely spreads to different habitats.

3. SPATIAL VARIOGRAMS

An alternative approach towards the concept of autocorrelation is the stochastic theory of 'regionalized variables' and its application in 'Kriging' developed by the French geostatistician MATHERON (1963). Let us consider a Euclidean space where n sample points are represented by vectors $\vec{x}_1, \ldots, \vec{x}_n$. Instead of the stringent requirement of stationarity, we make the somewhat weaker assumption that the spatial process $Y(\vec{x})$ is 'intrinsic'; for each directional distance $\vec{h}$ the following applies (DAVID,1977):

$$E(Y(\vec{x}) - Y(\vec{x}+\vec{h})) = 0, \qquad (4)$$

i.e. the process is on the same level at all points;

$$Var(Y(\vec{x}) - Y(\vec{x}+\vec{h})) = 2\,\gamma(\vec{h}) \qquad (5)$$

i.e. the variance of the first differences depends on the increment $\vec{h}$, however not on the absolute position.

The function $\gamma(\vec{h})$ is called a (semi-) variogram; for stationary processes, it equals the difference of variance and autocovariance function.
To obtain an estimate for this two-dimensional function from the localised sampling values, the infinite planar directions are divided into regular sectors s, and distances are grouped into intervals k of equal length. The empirical variogram value $\gamma(s,k)$ can be calculated according to

$$\gamma(s,k) = \frac{1}{2N} \sum_{i=1}^{n} \sum_{j=1}^{p_i} \left[y(\vec{x}_i) - y(\vec{x}_i+\vec{h}_j)\right]^2 \qquad (6)$$

with: p_i = number of sample points $(\vec{x}_i+\vec{h}_j)$, located with respect to $\vec{x}_i$ in sector s and in the distance interval k;

$N = \sum_{i=1}^{n} p_i$; n=total number of sample points.

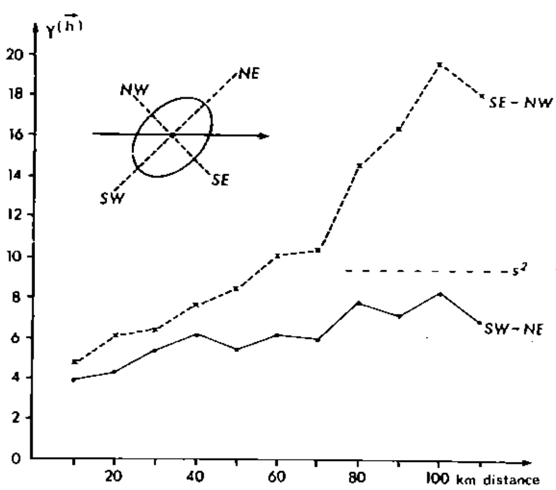


Fig. 4: Directional variograms of daily precipitation on July 16, 1979, measured at 150 stations in the 'Münsterland' and 'Rheinisches-Schiefergebirge'.

Figure 4 shows sector specific empiric variograms of daily precipitations (mm), measured on 16th July 1979 at 150 meteorological stations located in the 'Münsterland' and the adjacent 'Rheinisches Schiefergebirge'. Small variogram values point ot a positive persistence, while above the variance negative autocorrelation occurs.
In the SE-NW direction the similarity of the precipitations decreases more rapidly than in the SW-NE sector. Hence, one may conceive the neighbourhood relations for each point to be shaped elliptically. The ellipses are tilted by approximately 45° against the horizontal line and compressed along the SE-NW axis by a factor of approximately 2.0.This points to a street of showers moving in the SW-NE direction.

A characteristic feature of quite a number of variograms is their noticeable discontinuity at zero: For small distances $|\vec{h}|$ they show a tendency towards values >0. This so-called 'nugget effect' can be regarded as the variance of a purely random component superimposed on the otherwise continuously varying process.

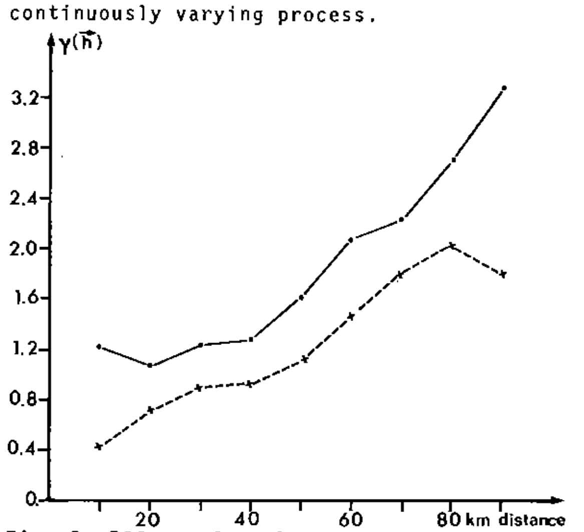


Fig. 5: Effect of an incorrect value on the variogram (Daily precipitation on July 16,1979; 'Münsterland').

•——• variogram for 47 stations, including an incorrect value of 0.0 (spatial mean=5.1mm)

x----x variogram for 46 stations, excluding the incorrect value.

As demonstrated in Fig. 5, the nugget effect might also be due to single erroneous data.

Estimating the variogram function as well as Kriging itself becomes increasingly problematic, if the intrinsic hypothesis is violated, for example if there is a trend, Figure 6 shows an example of this kind.

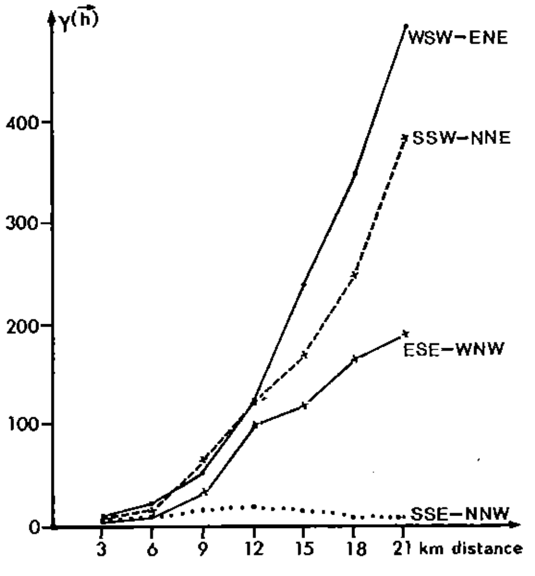


Fig. 6: Directional empiric variograms of groundwater levels (Dec.1970), measured at 80 wells in the 'Hessisches Ried'.

The groundwater levels as measured at 80 observation wells in the 'Hessisches Ried' (between Frankfurt and Darmstadt) show coinciding with the relief a steep gradient in WSW direction towards the River Rhine. Accordingly, the empiric variogram of this sector shows a parabolic increase. However, the variogram for the direction perpendicular to that (i.e. SSE-NNW) is for all distances close to zero because of the close similarity of neighbouring values. In this case, spatial increments of a sufficient high order may be used; the respective theory will not be discussed in this context (e.g. DELFINER, 1976; GAMBOLATI and VOLPI,1979)

4. KRIGING

The examples employing spatial correlograms and variograms given so far strongly emphasize the explorative character of stochastic methods. In the last part of this paper an example demonstrating the applicability of spatial stochastic models is presented.

The possibility of interpolating and extrapolating spatial data is of practical importance to a number of problems in the geosciences and regional planning. To give but one example: Stimulated by the Canadian Geographical Information System CGIS and the American GRID and IMGRID systems, computer-aided 'Landschaftsinformationssysteme' (landscape information systems) are under examination also in German speaking countries, and they are increasingly used in regional planning, above all in ecology-oriented landscape planning. Filling up spatial information gaps and smoothing information obtained from irregularly distributed sampling points presents a special problem. Conventional procedures such as ordinary weighted spatial means and polynomial trend surfaces may be applied, but they lead to more or less serious disadvantages (e.g. no exact interpolation; no reliable information concerning the errors of estimation).

Kriging, based on variogram analysis, offers an interesting alternative. This paper is restricted to giving just a brief outline of 'point Kriging', whereas the publications by OLEA (1975) and DAVID (1977) deal comprehensively with that topic. Formally, Kriging is a refined method of obtaining spatial means over next neighbour points:

$$y(\vec{x}_0) \sim y(\vec{x}_0) = \sum_{i=1}^{m(\vec{x}_0)} \lambda_i(\vec{x}_0) \cdot y(\vec{x}_i) \qquad (7)$$

$$\vec{x}_i \in U(\vec{x}_0) \quad \text{surrounding of } \vec{x}_0.$$

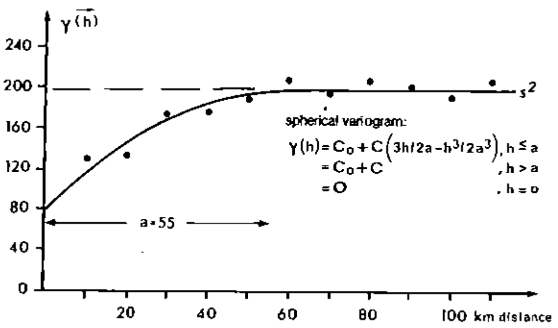


Fig. 7: Isotropic variogram of daily precipitation on June 5,1979 (150 stations in the 'Münsterland' and 'Rheinisches Schiefergebirge')

.... empiric —— theoretic

(C_0=80, C=119, a=55)

The λ-weights for the m neighbour points $\vec{x}_i$ are to be determined such that the estimate of the unknown value $y(\vec{x}_0)$ is the best linear unbiased one. This postulate leads to a system of linear equations which can be used to calculate the λ-values provided the theoretical variogram is known. The procedure has to be repeated, generally, for each point to be estimated.

To obtain the theoretical variogram, a suitable analytical function is fitted to the empirical values; in practice, one takes into consideration only a few standard types, such as the spherical variogram in figure 7 or linear variograms. For anisotropic processes, more complicated functions are needed. In performing the fitting, one can restrict oneself to the first 4 to 10 spatial lags, since with increasing distances persistence may rapidly decrease and since for estimating the unknown value $y(\vec{x}_0)$, usually only the nearest neighbours are used.
The question as to how many neighbouring points are to be included in the estimation can, in most cases, only be answered in a pragmatic way.

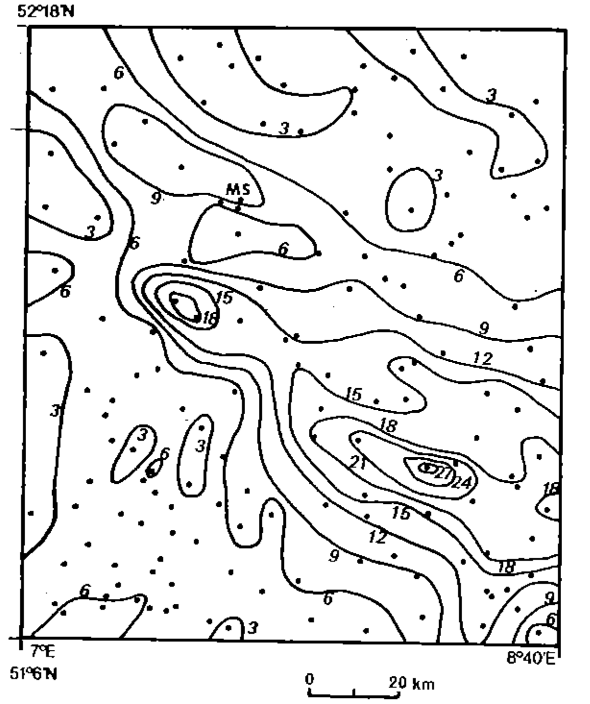


Fig. 8.: Measured values of daily precipitation on July 7,1979 (mm).
• 150 stations in the 'Münsterland' and 'Rheinisches Schiefergebirge'; MS = Münster.

With spherical variograms converging against the variance, a range can be determined (in Fig.7 for instance a range of approx. 55 km) allowing that all neighbouring points within that range are taken into consideration. With linear or otherwise infinitely increasing variograms, however, it is to be recommended to present a maximum number. This number can, because of the 'screen effect', be preset the lower, the smaller the nugget effect is. This screen effect causes a strong reduction of the λ-weights for all those points being shadowed by closely located neighbours.

As an example demonstrating the application of Kriging, an attempt to obtain a spatial interpolation of daily amounts of precipitation for 7th July 1979 is presented (Fig.8). The area investigated by means of 150 sample points is once again the 'Münsterland' and the adjacent 'Rheinisches Schiefergebirge'. On this particular day a cold front spread to Central Europe, leading to rain showers amounting on average to approx. 8 mm. A field of maximum precipitation reaches from WNW to ESE seperating the two spatial minima being ~130 kms apart. Correspondingly, the empiric variogram (Fig. 9) shows a significant anisotropy with a more rapid decrease of persistence in SSW-NNE direction and a reapproach to smaller variogram values at a distance of ~130 kms.

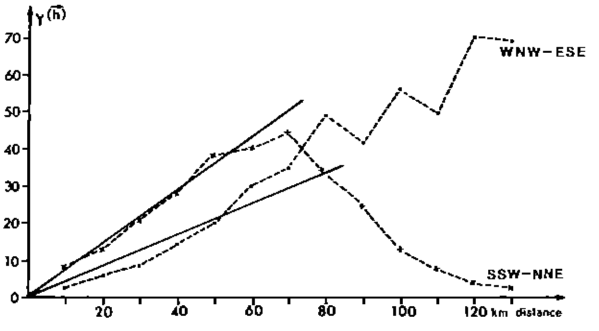


Fig. 9: Linear anisotropic variogram of daily precipitation on July 7, 1979.

----- empiric —— theoretic

In this case, a cone function has been chosen as a theoretical variogram, in horizontal sections represented by ellipses (see Fig. 4) and in vertical sections by linear variograms of directional specifity (DAVID,1977,137).

Interpolation on a square grid with a grid length of 10 km(Fig. 10) agrees fairly well with the isoline-map of the measured data. Singular extreme values have been slightly smoothed as being estimated as weighted means of neighbouring values. Attention should also be given to the map showing the errors of estimation (Fig.11). Mimimum values of 1.0 mm and less always occur where the information density is particularly high; the smaller number of sample points in the western part of the 'Münsterland', however, gives rise to standard errors larger than 2.5 mm. Therefore an installation of additional meteorological stations may start at these maxima of the standard error surface.

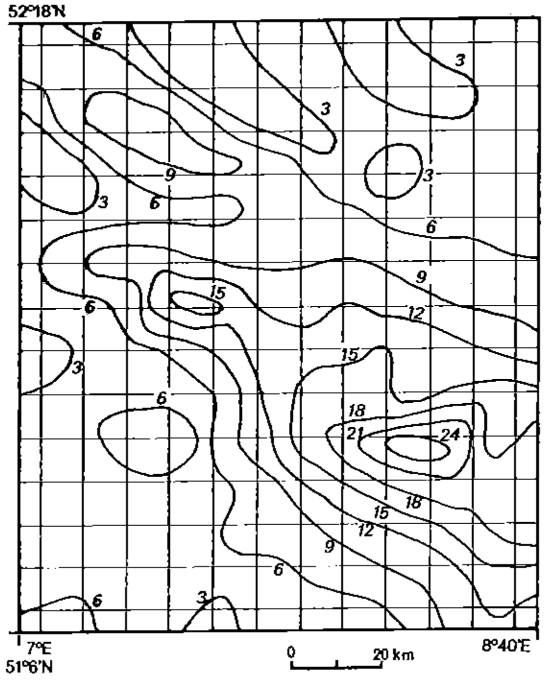


Fig. 10: Kriging: Estimated values of daily precipitation on July 7, 1979 (mm) for a quadratic grid of 10 km. (see Fig.8)

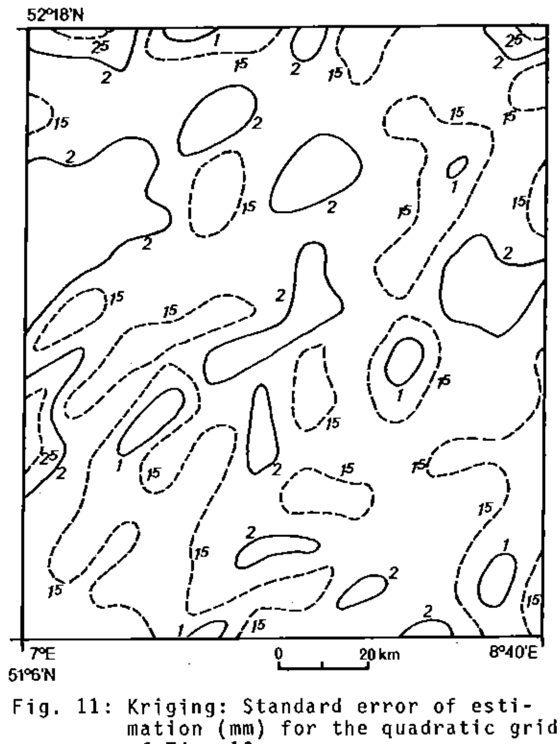


Fig. 11: Kriging: Standard error of estimation (mm) for the quadratic grid of Fig. 10.

REFERENCES

CLIFF,A.D. and ORD,J.K.: Spatial autocorrelation. Pion,London. 1973;178pp.

CLIFF,A.D. and ORD,J.K.: Space-time modelling with an application to regional forecasting.Transact.Inst.Brit.Geograph. 1975/64,p.119-128.

DAVID,M.: Geostatistical or reserve estimation.Elsevier,Amsterdam.1977.364 pp.

DELFINER,P.: Linear estimation of non stationary spatial phenomena. In: GUARASCIO,M. et al.(Eds): Advanced Geostatistics in the Mining Industry; Reidel,Dordrecht.1976,49-68.

GAMBOLATI,G. and VOLPI,G.: Groundwater contour mapping in Venice by stochastic Interpolators.Water Resources Research, 1979,281-290.

HAGGETT,P.,CLIFF,A.D.,and FREY,A.: Locational analysis in human geography. Vol. 2=Locational methods.Arnold,London. 1977,605pp.

MATHERON,G.: Principles of geostatistics. Economic Geology, 1963/58,1246-1266.

NIPPER,J. and STREIT,U.: Zum Problem der räumlichen Erhaltensneigung in räumlichen Strukturen und raumvarianten Prozessen.Erdkunde 1977/65,4,241-263.

OLEA,R.A.: Optimum mapping techniques using regionalized variable theory. Kansas Geolog.Survey,1975/2,137 pp.

SYMADER,W.: Trendflächenanalyse zur Beurteilung raumvarianter Prozesse am Beispiel des Schwermetallgehaltes in Oberböden. Verhandl.d.Gesellsch. f.Ökologie,Münster,1979,209-214.

UNWIN,D.J.: Statistical methods in physical geography.Progress in Physical Geography,1977/1,2;185-221.

WITTIG,R. and WEBER,H.E.: Die Verbreitung der Brombeeren (Gattung Rubus L., Rosaceae) in der Westfälischen Bucht. Decheniana,Bonn.1978/131;87-128.

F R A N Z - J O S E F   K E M P E R

ON SOME MODELS OF CATEGORICAL DATA ANALYSIS AND THEIR APPLICATION TO GEOGRAPHICAL PROBLEMS

From:

BAHRENBERG, GERHARD und STREIT, ULRICH (Eds.):
German Quantitative Geography. Papers presented at the 2nd European Conference on 'Theoretical and Quantitative Geography' in CAMBRIDGE 1980. Münstersche Geographische Arbeiten 11. Schöningh, Paderborn 1981, p. 45 - 58

Author:

Dr. Franz-Josef Kemper

Geographisches Institut
Universität Bonn

Franziskanerstraße 2

5300 Bonn 1

ZUSAMMENFASSUNG

Eine der wichtigsten Entwicklungen der multivariaten Statistik in den siebziger Jahren war die Ausarbeitung von Modellen zur Analyse kategorialer, insbesondere nominaler Daten. Da nominale Variable, für die weder Ordnungsrelationen noch metrische Eigenschaften gelten, in der Geographie sehr häufig vorkommen, sind diese Verfahren von großer Bedeutung für die quantitative Analyse geographischer Fragestellungen. In diesem Beitrag wird keine systematische Darstellung der verschiedenen Verfahren angestrebt, sondern es geht um eine vergleichende Bewertung wichtiger Ansätze hinsichtlich der statistischen Voraussetzungen und der Interpretationsmöglichkeiten der Ergebnisse. Ein solcher Vergleich ist besonders im Hinblick auf die adäquate Verknüpfung von inhaltlichen Problemstellungen mit geeigneten Methoden erforderlich.

Zunächst werden symmetrische Ansätze behandelt, bei denen nicht zwischen explikativen und abhängigen Variablen unterschieden wird. Im Vordergrund steht das loglineare Modell. Besondere Aufmerksamkeit gilt den Fragen der Nullbesetzungen in Kontingenztafeln, der Beschränkung auf hierarchische Modelle und der Stichprobengröße. Von den asymmetrischen Ansätzen werden die lineare logit Analyse und das allgemeine lineare Modell von GRIZZLE, STARMER und KOCH dargestellt. Diskutiert werden die Wahl des Schätzverfahrens, die Metrisierung der abhängigen Variablen und die Möglichkeiten kausalanalytischer Behandlung nominaler Daten mit Hilfe der neueren Verfahren. Die Anwendungsbeispiele sollen zeigen, daß nicht nur Befragungen, sondern auch andere in der Geographie verbreitete Daten sinnvoll mit den geschilderten Verfahren ausgewertet werden können.

1. INTRODUCTION

Nominal and other variables with categorical values frequently occur in geography. Typical examples are spatial categories such as agglomeration or rural space, the ethnical residential quarters of an African town, each type of quarter marked by a particular symbol, the distribution of valley forms or of different types of potential natural vegetation. Very often the research problem requires the measurement of associations between such variables. To analyse the relationships, it was usual to proceed from crosstables set up for two variables, and to compute bivariate measures of the strength of association. A multivariate perspective in which the effects and interactions of several variables are considered simultaneously has now been made possible by a set of related procedures, about which WRIGLEY (1979) has given an excellent review. In this paper it is not intended to give a systematic description of the various methods, but rather to present some remarks on the specific possibilities and problems of several important techniques and the interpretability of the results.

2. SYMMETRICAL APPROACHES

2.1. THE LOGLINEAR MODEL

It is usual to distinguish symmetrical and asymmetrical approaches. In asymmetrical approaches the variables are divided into dependent (response) and independent (explanatory) variables, whereas in symmetrical models there is no such distinction. In the latter case the analysis starts with a contingency table, each cell in this table containing a frequency, i.e. the number of observations for a particular combination of variable levels. The most widely used procedure for analysing such a multidimensional contingency table is the loglinear model (cf. BISHOP et al. 1975, FIENBERG 1977, GOODMAN 1978). In this approach the natural logarithm of the frequency of a combination is represented as a linear function of effect parameters of particular variables and variable groups. The choice of parameters follows from the hypotheses underlying the selected model. This can be demonstrated by an example of three variables. Let frequencies x_{ijk} be given for all combinations ijk, which are characterized by the level i of the first variable, level j of the second, and level k of the third variable. Each combination defines a subpopulation in the set of all individual cases. If we assume the independence of all variables, the following model results:

$$\log m_{ijk} = u + u_1(i) + u_2(j) + u_3(k) \quad (1)$$

Here m_{ijk} represents the expected frequency of combination ijk assuming the validity of the presupposition, and the u's are effect parameters. u is the mean of all log frequencies and the u_n are functions, which describe the deviations from the grand mean u for the value levels of variable n. There are two problems to be solved: first the estimation of the expected frequencies and second the computation of the parameters u_n. The first problem, which in most cases is solved with the aid of the maximum likelihood estimation technique, appears to be most crucial. In order to apply this method one must assume that the data are the outcome of a particular sampling scheme (Poisson, multinomial, or product-multinomial, see FIENBERG 1977). To check whether the empirical frequencies x_{ijk} are significantly different from the estimated expected frequencies $\hat{m}_{ijk}$, the well-known chi-square-test can be used. So it can be shown that under the assumption of the null hypothesis - no difference - the value

$$LRX^2 = 2 \sum_{i,j,k} x_{ijk} \log \frac{x_{ijk}}{\hat{m}_{ijk}} \quad (2)$$

which is called "Likelihood ratio chi-square", is distributed like a chi-square-distribution with a number of degrees of freedom corresponding to the number of all combinations minus the number of estimated parameters.

If model (1) is not appropriate one has to assume that there are associations or interactions between some variables. The most general model containing all possible interactions, is the following saturated model:

$$\begin{aligned}\log m_{ijk} &= u + u_1(i) + u_2(j) + u_3(k) \\ &+ u_{12}(i,j) + u_{13}(i,k) \\ &+ u_{23}(j,k) + u_{123}(i,j,k) \quad (3)\end{aligned}$$

This model fits the data exactly, indeed it presents a particular arrangement of the data. It is usual to consider hierarchical models only, because the maximum likelihood estimation for other models may be very difficult. In hierarchical models a higher order interaction may be included only if all lower order effects, which are generated by splitting, are also included. If e.g. a model comprehends the term u_{12}, it must include u_1 and u_2 as well. The model is represented by those effects through which all other effects can be derived by splitting. Therefore [123] denotes the saturated model.

2.2. A LOGLINEAR EXAMPLE

A first example is taken from LEWIS (1977). It demonstrates that results from mappings, something very common in geography, can be analysed by means of the new techniques. Within a sector of northeast London, 1407 buildings were mapped and classified by three variables: age, degree of decay, and use of building. The variables "age of building" and "decay" were measured by 5 categories each, and the use by a threefold subdivision into residential, manufacturing and warehousing, and offices. The most suitable model, which fitted the data of the 5x5x3 contingency table, was the model [12] [13] [23] with all second-order interactions. On the basis of substantial hypotheses these relationships could be expected. But beyond this, a geographical interpretation should consider what kind of relationships appear. For that purpose it is necessary to compute the effects u. Unfortunately, LEWIS did not carry out this part of analysis.

The missing results could be computed with the data of the contingency table published by LEWIS. As an example Table 1 shows the effects of the interaction use/degree of decay. It should be noticed that in the table the intervening influence of the third variable, i.e. age of building, is excluded. If one regards, as widely usual, only the crosstable use/decay from the original data (Table 2), one could infer that residential buildings tend to be in good condition. Especially the category "no decay" appears to be represented above average. But if we look at the values of the interaction effect, we

decay:	use: residential	manufacturing	offices
no decay	-0.276	-1.382	1.658
slight decay	0.177	-0.269	0.092
much decay	0.196	0.297	-0.493
substantial d.	-0.207	1.112	-0.905
severe decay	0.110	0.242	-0.352

Table 1: Loglinear parameters: interaction use/degree of decay

decay:	use: resid.	manuf.	offices	N
no decay	25.3	16.3	58.2	656
slight d.	16.7	30.0	53.4	365
much decay	17.6	64.3	18.1	182
subst. decay	4.9	90.2	4.9	163
severe decay	9.8	73.2	17.1	41
total	19.3	36.3	44.4	1407

Table 2: Crosstable of original data: row percentages

shall recognize that the category just mentioned is represented below average concerning the residential use. This discrepancy results from the influence of the third variable, i.e. age of building. In the study area the buildings with exclusively residential use are relatively new, and, not very surprisingly, new buildings do not show much decay. Therefore it can be expected that residential use in this case is connected with a good condition of the building. But the results of loglinear analysis demonstrate that this condition is not so good as could be expected from the age of the buildings. Beyond it nineteenth century buildings with offices were more often renovated than residential buildings of the same period.

2.3. ZEROS IN CONTINGENCY TABLES

The problem of the selection of a loglinear model will not be treated here. For that purpose several strategies have been developed to find an "optimal" model. Important criteria for the selection are statistical fit, parsimony, and substantial interpretability. It is an advantage of the loglinear model that relationships between categories of variables can adequately be analysed in detail. But if the number of variables and/or of categories increases the number of effects grows and the interpretation may be difficult, especially concerning higher order interaction effects. A further advantage of the loglinear approach is the possibility to handle a certain amount of zeros in the contingency table. Two types of zeros should be distinguished. Fixed or structural zeros occur if the appropriate cell of the table can a priori not contain an observation. For example, in a crosstabulation of European cities according to their function it is impossible to find a Swiss seaport. Such contingency tables are called incomplete. For a detailed discussion see BISHOP et al. (1975) (cf. also ARMINGER 1979). Frequently, zeros occur because of a relatively small size of the sample in comparison to the number of cells. The above mentioned contingency table of the buildings in London did not contain buildings from the age period 1939 or later, which had much, substantial, or severe decay, although such a combination is imaginable. Such sampling zeros put no problems as long as the expected frequencies are not zero either. This property is an important reason for using the expected frequencies instead of the empirical frequencies in the model. But even if some expected frequencies, or more precisely, estimated expected values are zero, it is possible to carry out a loglinear analysis, in which the degrees of freedom of the chi-square-test must be changed (cf. FIENBERG 1977). Difficulties only arise if certain marginal distributions of the fitted values contain zeros. Generally in loglinear analysis it is not necessary to collapse categories with few observations or to add a small constant to all cells as often proposed.

2.4. SOME PROBLEMS OF LOGLINEAR ANALYSIS AND POSSIBLE SOLUTIONS

To end this section, we shall discuss two further problems of loglinear analysis which are important in the case of many applications. The first problem concerns the limitation to hierarchical models. In practice this class of models usually offers sufficient possibilities for the selection of a suitable model, but sometimes it would be preferable to choose a nonhierarchical model for the sake of parsimonious modelling. The widely used computer programs like ECTA or BMDP can only handle hierarchical approaches. It can be demonstrated however that by modifying the data a nonhierarchical problem can be changed into a hierarchical one, and hence can be solved with standard techniques. For that purpose "indicator variables" are introduced and new combinations of variables are formed, where zero values occur. For details see MAGIDSON (1976) or LANGEHEINE (1979).

A second problem becomes relevant when many frequencies are small. The fit of a loglinear model is tested by the chi-square-statistic. As is well known, there are certain assumptions of the test concerning the size of the cell values, e.g. the values m. These assumptions must be checked for every fitted model (KÜCHLER 1979, 251, cf. FIENBERG 1977, 37). If they are not fulfilled the reliability of the tabulated significance limits is doubtful. This situation often occurs, for example if there are many zero cells. Accordingly the above mentioned possibility of the loglinear approach to process zeros has some clear limitations. For small samples it may be recommendable, instead of the loglinear model to use the related approach of the "configuration frequency analysis" (Konfigurationsfrequenzanalyse = KFA), which was developed by the German statisticians KRAUTH and LIENERT (1973). This approach avoids the chi-square-test and uses the binomial distribution for the comparison between empirical and expected frequencies of each combination of value levels. The results are combinations, or configurations in the terminology of KRAUTH and LIENERT, which occur significantly more or less often than expected. If it is required to obtain, instead of these "individual" statements, more generalizing results about the influence of one variable or of one variable interaction, the tests of significance become much more difficult, and therefore the KFA is not so flexible as the more elegant and more consistent loglinear approach. Nevertheless, for particular problems and for small samples KFA can be a useful alternative to the loglinear model.

3. ASYMMETRICAL APPROACHES

Asymmetrical approaches which involve a distinction between response and explanatory variables are especially important, because they can directly be related to substantial problems and theoretical arguments. By these approaches explicitly formulated hypotheses basing on causal interpretations are operationalized and tested. Using the symmetric loglinear procedure it can be dangerous to look for the statistically "best" model and, at the same time, to neglect such hypotheses about relationships. For the sake of a closer connexion between quantitative and theoretical geography it seems desirable for applications of loglinear models to concentrate on explanatory rather than on exploratory analysis.

3.1. THE LINEAR LOGIT MODEL

It is not very difficult to transform a loglinear model into an asymmetrical model. The transformation is particularly simple if the dependent variable is dichotomous. Suppose e.g. that a saturated loglinear model with three dichotomous variables is given and that variable 3 is defined as response variable. The equation of the loglinear model can be rewritten for the two levels of the dependent variable:

$$\log m_{ij1} = u + u_1(i) + u_2(j) + u_3(1) + u_{12}(i,j) + u_{13}(i,1) + u_{23}(j,1) + u_{123}(i,j,1) \quad (4)$$

$$\log m_{ij-1} = u + u_1(i) + u_2(j) + u_3(-1) + u_{12}(i,j) + u_{13}(i,-1) + u_{23}(j,-1) + u_{123}(i,j,-1) \quad (5)$$

Subtract (5) from (4) and notice that u-values from the same function add to zero because they are deviations from the grand mean, thus $u_3(1) + u_3(-1) = 0$ etc. It follows:

$$\log \frac{m_{ij1}}{m_{ij-1}} = 2u_3(1) + 2u_{13}(i,1) + 2u_{23}(j,1) + 2u_{123}(i,j,1) \quad (6)$$

Defining $a_o = 2u_3(1)$, $a_1 = 2u_{13}(1,1)$, $a_2 = 2u_{23}(1,1)$, $a_3 = 2u_{123}(1,1,1)$ and $p_{ijk} = m_{ijk}/(m_{ij1} + m_{ij-1})$ yields:

$$\log \frac{p_{ij1}}{1-p_{ij1}} = a_o + a_1X_1 + a_2X_2 + a_3X_3 \quad (7)$$

X_1, X_2, X_3 are dichotomous variables with the values 1 and -1 and p_{ij1} is the conditional probability that the response variable takes the value 1 assuming that the categories i and j appear. The term log (p/(1-p)) is called log-odds or logit of p, and (7) represents a linear logit model. The variables X_1, X_2 reflect the influence of the two explanatory variables onto the respond variable and X_3 the interaction effect of these two variables.

As an example we choose the building classification data. An asymmetrical model can be generated by defining the state of decay as response variable and age and use as explanatory variables. To dichotomize the variables, we collapsed variable 3 "state of decay" into the two categories "no decay" (1) and other (-1), variable 1 (age) into "new", i.e. built after 1914, (1) and other (-1), variable 2 (use) into "manufacturing" (1) and other (-1). A saturated log-linear model, which was the sole model fitting the data, yielded the following results: $a_1 = 1.954$, $a_2 = -1.304$, $a_3 = 0.464$.

Therefore the most important effect influencing the state of decay is produced by the age of the building (a_1), but function also provides an important effect in so far as with manufacturing land-use there is much decay. The interaction function/age can not be neglected. So the effect of use on state of decay depends on age, because the combination "old" and "manufacturing" is connected with more decay than expected, whereas the combination "new" and "manufacturing" shows less decay.

3.2. POSSIBILITIES AND LIMITS OF CAUSAL MODELLING

To facilitate a causal interpretation path analysis or structural equation analysis were developed for metric data. These extensions of the usual regression analysis are especially advantageous if there are many relationships, some variables are dependent as well as independent, and so direct and indirect effects are present in the model. If only effects without feedbacks occur, the model is called recursive, otherwise non-recursive. GOODMAN and others have proposed to build up a nonmetric equivalent to path analysis by suitable transformations of log-linear models in logit equations. The following example, which is taken from KÜCHLER (1979), can demonstrate this procedure. The problem was the explanation of the voting behaviour of the German population. Although this is a problem of political science or of sociology, and not directly one of geography, there are clear connexions to the geography of elections, and similar analyses can be performed with data from surveys in the scope of more explicit geographical inquiries. The data originate from about 1200 interviews which were taken after the election to the parliament (Bundestag) of West Germany in 1976. The following variables were used:

X_1 = party voted for in 1976 (-1= Christian Democrats, +1 = other parties)

X_2 = membership in the trade unions (-1 = no, + 1 = yes)

X_3 = church attendance (-1 = non-regular , (+1 = regular)

X_4 = religion (+1 = catholic, -1 = other)

X_5 = class (+1 = working class, -1=other)

It was postulated that the support of the conservative Christian Democrates would depend on the variables X_2 to X_5, secondly that the membership in trade unions would depend on class, and thirdly that catholics and the middle class would more frequently attend church than the other groups. For these expected relationships three structural equations were formulated as startingpoint for a recursive path model:

$$\log(p_1/q_1) = a_{10} + a_{12}X_2 + a_{13}X_3 + a_{14}X_4 + a_{15}X_5 \quad (8)$$

$$\log(p_2/q_2) = a_{20} + a_{25}X_5 \quad (9)$$

$$\log(p_3/q_3) = a_{30} + a_{34}X_4 + a_{35}X_5 \quad (10)$$

Here p_1 is the conditional probability p_{1jkln} that X_1 takes the value +1 assuming that the other variables take the constant values j,k,l,n and q_1 is defined as $1-p_1$ etc.

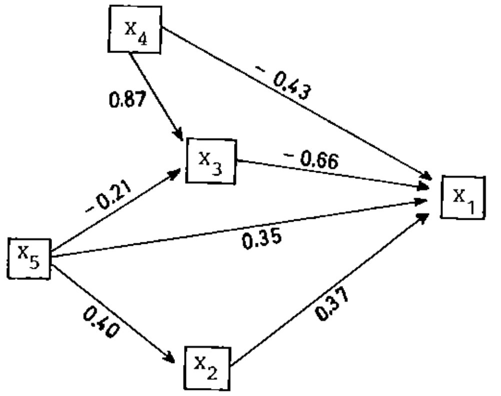


Figure 1 Nonmetric path-model: voting behaviour

By obvious transformations the three structural equations can be rewritten as a loglinear model. The hierarchical model which is characterized by the interactions [12] [13] [14] [15] [25] [34] [35] [45] could be accepted. Using the technique described in 3.1, the coefficients a could be derived from the values for the u-terms. Figure 1 shows these coefficients which confirm the suppositions about the relationships. The variable most strongly influencing party preferences was church attendance, whereas the socio-economic variables X_2 and X_5 were of minor importance. A diagrammatic representation like that of Figure 1 is certainly quite impressive, but one should notice several differences to the metric analysis which severely reduce the usefulness of path analysis in the case of categorical variables:

(i) The coefficients a yield information about the direction and size of relationships, but beyond that cannot be compared to metric path coefficients. So it is impossible to compute indirect effects by multiplication of coefficients. Therefore FIENBERG (1977,91) regards the assignment of numerical values to the arrows in the path diagram as problematic.

(ii) The representation of three-factor or higher-order interaction terms in the path diagrams is complicated and difficult to interpret.

(iii) For variables with more than two categories a split into dichotomous variables seems to be necessary, otherwise a relationship would have to be described by a multidimensional matrix rather than by a numerical value.

(iv) There are no residual or error terms.

(v) There is no possibility of testing the model by computing expected coefficients for missing arrows.

3.3. ESTIMATION METHODS: ML and WLS

The parameters in a linear logit model, which is a transformation of a loglinear approach, are computed by maximum likelihood (ML) estimation. However, there is another frequently used estimation procedure, the method of weighted least squares (WLS). Consider for example the model of equation (7). Here a metrix response variable is represented as a linear combination of dichotomous variables. The units of analysis are the subpopulations, which are defined by the combinations of the categories of independent variables. One can suggest to compute the parameters by multiple regression analysis. If this is done by ordinary regression (ordinary least squares = OLS), unbiased estimators result, but generally it is not possible to compute standard errors of the regression coefficients. To compute standard errors by OLS we must assume that the residuals have homogenous variances, i.e. that there is homoscedasticity. It is not difficult to see that this assumption does not hold with the type of response variable in question (cf. KOCHLER 1979). Hence the preferable estimation method will be WLS. By a technique of suitable weights standard errors can be computed for data characterized by heteroscedasticity (cf. THEIL 1971). This procedure was used by THEIL (1970) as an estimation method for a linear logit model. Now, what are the advantages and disadvantages of the two estimation techniques? WRIGLEY remarks: "As the WLS procedure does not involve the numerical optimization necessary in the ML procedure, it is less costly in terms of computation and thus has become the generally adopted method" (WRIGLEY 1979, 334). Besides the kind of computation there are some further differences between the two procedures, which are relevant for applications of linear logit models. On the one hand, both models are large sample approaches, but the WLS procedure is more susceptible of small samples. As KOCHLER (1979, 169) suggests, each subpopulation should contain at least 20-30 observations with the exception of a small number of cells. With ML estimation however, one can treat smaller cell values. On the other hand, the usual application of the loglinear approach and hence of the linear logit approach with ML estimation is limited to hierarchical models. Although, as we have seen, it is possible with ML estimation to handle non-hierarchical models, the WLS procedure can be done in the hierarchical as well as in the non-hierarchical case and it is very easy to fit a wide variety of models. Therefore estimation by WLS has some advantages, although with small samples one would prefer the ML method.

3.4. THE LINEAR MODEL OF GRIZZLE, STARMER AND KOCH

The linear logit model with WLS estimation can be extended by allowing transformations of the response variable different from the log-odds. A general linear model which contains such an enlargement was presented by GRIZZLE, STARMER and KOCH in an influential article (1969). We shall not describe this procedure, which is called GSK-approach, in its whole generality, but only compare two possibilities of metrification of the response variable by taking the logits or the untransformed percentages. The metrification of a categorical variable is based upon the computation of a percentage in each subpopulation with the same combination of explanatory variable levels. We shall only regard the most simple case, i.e. that all variables are dichotomous. Then it is possible to compute for each subpopulation i the percentage p_i of all units in the subpopulation that fulfill a particular, arbitrarily selected category of the response variable. Another metrification follows from taking the logits log $(p_i/(1-p_i))$. The advantages of the more complicated logits are based on statistical arguments. At first a regression with a percentage value as dependent variable y can yield predicted values y outside the range of 0 to 1. This is especially problematic if the percentages are interpreted as probabilities. Secondly, the statistical techniques of WLS assume that the range of values of the dependent variable is unlimited. By the logits the interval (0,1) is mapped into $(-\infty, +\infty)$, so that no statistical problems arise. On the other hand, there are also arguments in favour of the untransformed percentages (see KOCHLER 1978), especially because of the better interpretability of the results. The metrification by percentages together with a suitable coding of the categories allows the results to be appreciated by wide audiences without any background in statistical methods. As far as a model is not used for prognosis the range of the predicted values seems to be of minor importance. For many problems of application the interpretation is based upon the regression coefficients, not upon the predicted values. It should also be noticed that a model with logits will give more weight to the subpopulations with very high or very low percentages p_i than a model with untransformed metric values, whereas in the interval (0.3, 0.7) the weights are rather equal.

Hence, although the overwhelming majority of applications is carried out with the logits, the metrification by percentages can also be useful. An example shall demonstrate the transparent manner of interpretation. The data come from a survey concerning the leisure mobility of the population of Bonn, Germany (cf. KEMPER 1977). One of the objectives was the explanation of the participation in vacation journeys. It was assumed that the following dichotomous variables could be determinants:

X_1 social class (-1 = lower class/ +1 = middle class)

X_2 household income (-1 = low/ +1 = high)

X_3 size of household (-1 = 3 or more persons / +1 = 1 or 2 persons)

X_4 availability of a car (-1 = no/ +1 = yes)

The response variable was obtained by computing the percentage of all persons in a subpopulation, which had not undertaken a journey in the year before the interview. By the selection of the subpopulations as units of analysis, the relationships of the correlated independent variables are orthogonalized. Each variable was coded by the values +1 and -1. Table 2 shows the so-called design matrix of the saturated model, which contains the values of the main effects and of all interaction effects, together with a column of 1-values, which represents the constant term of the regression. From the design matrix it is easy to see that all independent variables including the interaction variables are not correlated, i.e. orthogonal, on the basis of all 2^4 = 16 subpopulations.

With the selected coding, the results of the saturated model (Table 3) can easily be interpreted. The constant term a_0 represents the overall mean of the response variable. On the average 33 % of a subpopulation did not undertake a vacation journey. The main effects, namely a_1,a_2, a_3,a_4, are differences of means, i.e. the differences of the means in those subpopulations containing a particular category of the explanatory variable and of the mean a_0. So the value a_1 = -8.36 shows that in the middle class (X_1 = 1) the

	OLS estimation effects	WLS estimation effects	standard errors
a_0	32.89	31.89	1.79
a_1	-8.36	-10.41	1.75
a_2	-5.75	- 6.35	1.85
a_3	-7.84	- 5.42	1.42
a_4	-9.63	- 8.95	1.86
a_{14}	4.70	4.84	1.71

Table 3: Effects of the saturated model (in part, OLS) and of the model with interaction 14 and all main effects (WLS)

i	p_i	1	X_1	X_2	X_3	X_4	X_1X_2	X_1X_3	X_1X_4	X_2X_3	X_2X_4	X_3X_4	$X_1X_2X_3$	$X_1X_2X_4$	$X_1X_3X_4$	$X_2X_3X_4$	$X_1X_2X_3X_4$
1	7,5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	19,0	1	1	1	1	-1	1	1	-1	1	-1	-1	1	-1	-1	-1	-1
3	14.6	1	1	1	-1	1	1	-1	1	-1	1	-1	-1	1	-1	-1	-1
4	27.8	1	1	1	-1	-1	1	-1	-1	-1	-1	1	-1	-1	1	1	1
5	20.6	1	1	-1	1	1	-1	1	1	-1	-1	1	-1	-1	1	-1	-1
6	17.9	1	1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	1	1
7	35.7	1	1	-1	-1	1	-1	-1	1	1	-1	-1	1	-1	-1	1	1
8	53.1	1	1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1
9	14.3	1	-1	1	1	1	-1	-1	-1	1	1	1	-1	-1	-1	1	-1
10	44.4	1	-1	1	1	-1	-1	-1	1	1	-1	-1	-1	1	1	-1	1
11	37.1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1
12	52.4	1	-1	1	-1	-1	-1	1	1	-1	-1	1	1	1	-1	1	-1
13	16.7	1	-1	-1	1	1	1	-1	-1	-1	-1	1	1	1	-1	-1	1
14	60.0	1	-1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	1	-1
15	39.6	1	-1	-1	-1	1	1	1	-1	1	-1	-1	-1	1	1	1	-1
16	65.5	1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	1

Table 2 : Design-matrix saturated model: vacation journeys

percentage of non-vacationists declines to an average of 32.89 - 8.36 = 24.53, whereas in the lower class (X_1 = -1) it rises to 32.89 + 8.36 = 41.25. The difference between the two classes is therefore 2 x 8.36 = 16.71 %. The most important effects are the main effects of the four independent variables, the first position being hold by the availability of a car. All socio-economic variables surpass the demographic variable "size of household". As for the interaction effects, the terms X_1X_4 and $X_1X_3X_4$ are the only worth mentioning. The more important term X_1X_4 shows that the simultaneous coincidence of middle class and car availability, resp. of lower class and no car, results in raising the percentage of non-vacationists by 4.7 % above the value expected from the main effects. In a saturated model OLS and WLS estimations are the same, and the sum of squares of the coefficients for the main effects and the interaction effects is equal to the variance of the response variable y. From that one can compute straightforwardly the share of each effect in the variance of y, which in a saturated model is explained by 100 %. The 4 main effects already explain 82.7 %. Looking for a model with few interaction terms, the following equation could be acceptable:

$$p_i = a_o + a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + a_{14}X_1X_4 \qquad (11)$$

The estimation of the effect parameters by WLS yields the values noted in Table 3. In principle, these values are similar to be interpreted as the coefficients of the saturated model, the difference being substitution of "average" by "weighted average". If a model fits the data well the sum of squares of the residuals is distributed like a chi-square-distribution with n-m degrees of freedom. Here n is the number of subpopulations and m the rank of the design matrix. With model (11) a value of 8.26 results, which is not significant for 16-6 = 10 degrees of freedom at the 0.05 level. Therefore this model is acceptable. On the whole, the substantial interpretation does not depart from the effect interpretation of the saturated model, but there are some changes in the order of the independent variables.

3.5. THE EXPLANATORY POWER OF A MODEL

Finally we refer to a problem that may demonstrate the differences between multivariate methods for categorical and for metric data. As is well known, the coefficient of determination in a metric regression analysis indicates, how much variance of the response variable is statistically explained by the regression. In contrast to that, in the non-metric approach we have described, a model is selected with no significant deviations from the data. But in such models the statistical power of explanation can be very different depending on the sample size, although mostly it will be high. With the asymmetrical analysis of categorical data a strange phenomenon may appear, namely that by adding new explanatory variables the power of explanation declines if only the main effects are considered. The example of vacation travel can be used to show this effect. Because in a WLS approach the usual coefficient of determination is not computed, we use the OLS estimation for the following models. This is possible, for we are only interested in the statistical power of explanation, not in the significance of the terms. On the basis of the main effects, the degrees of explained variance of the response variable p is as following, when different independent variables are used:

X_1	100.0 %
X_1X_2	98.2 %
$X_1X_2X_3$	91.3 %
$X_1X_2X_3X_4$	87.7 %

These results are surprising only at first sight. They become understandable by considering the differing number of units. If the number of independent variables increases, the number of combinations of variable levels will rise and consequently the number of subpopulations too. Furthermore, with the growing number of variables the number of possible interaction effects will increase in a higher proportion and therefore the percentage of the main effects will decline. The underlying cause of this situation seems to be that the GSK approach, as well as the loglinear approach, is not based on individuals, but on data aggregated to homogeneous subpopulations.

Since in the GSK approach only the fit of data is tested, it is advisable additionally to measure the explanatory power of the model. This will be particularly helpful, when many potential variables are at disposal, but not all of them can be integrated into a model, because the size of the subpopulations would become too small then. Suppose for instance that only three variables were possible with the example of vacationists. As KÜCHLER (1979) suggests, before testing a GSK model one should, with the aid of an association coefficient, select the combination of variables which shows the largest coefficient on the basis of a contingency table, in which the columns represent the levels of the response variable and the rows the levels of the combination. Concerning our example, we have computed the contingency coefficient C as well as GOODMAN's and KRUSKAL's λ_r that has been developed starting from a conception of proportional reduction in error. The following values are the result for different sets of variables:

Variables	C	λ_r
$X_1X_2X_3$	0.306	0.0387
$X_1X_2X_4$	0.317	0.0885
$X_1X_3X_4$	0.314	0.0000
$X_2X_3X_4$	0.307	0.0436

Both coefficients favour the combination of the three socio-economic variables X_1, X_2, X_4. Hence for this combination an appropriate GSK model should be selected. On the other hand, the differences of the C and λ_r-values are relatively small so that other combinations could be taken into consideration too, if theoretical arguments support them. Another proposal for measuring the explanatory power of the determining factors was suggested by THEIL (1970), who has used an information-theoretic entropy measure.

4. CONCLUSION

The aim of this paper was to show that among the various techniques of categorical data analysis there is no optimal procedure in general, but that a suitable procedure should be used depending on the problem of inquiry and the possibilities as well as the restrictions of the data used. The same arguments concern the selection of the kind of coding and of metrification, which can be different according to the purpose of the analysis, the importance of prognosis and the expected groups of readers or users of the results. In the years to come geographers will increasingly use the new methods for multivariate analysis of categorical data. But certainly it would not be desirable if a single procedure would dominate as an unreflected standard, as in the case of factor analysis the technique of PCA with orthogonal varimax rotation (cf. KEMPER 1975). The potential fields of application for the new procedures are extensive, and besides surveys, which have represented the most often used source of categorical data so far, geographers should also pay attention to the evaluation of other sources such as mapping or census results, the quantitative analysis of which has hitherto been restricted mostly to bivariate associations.

REFERENCES

ARMINGER,G.: Loglineare Modelle zur Analyse des Zusammenhangs zwischen nominalen Variablen. In: HOLM,K. (ed.): Die Befragung, vol. 6. München: Francke 1979, 218-261.

BISHOP,Y.M.M., FIENBERG,S.E., HOLLAND,P.W.: Discrete Multivariate Analysis: Theory and Practice. Cambridge, Mass.: MIT Press 1975.

FIENBERG,S.E.: The Analysis of Cross-Classified Categorical Data. Cambridge, Mass.: MIT Press 1977.

GOODMAN,L.A.: Causal Analysis of Data from Panel Studies and Other Kinds of Surveys. American Journal of Sociology, vol. 78, 1973, 1135-1191.

GOODMAN,L.A.: Analyzing Qualitative/Categorical Data: Log-linear Models and Latent Structure Analysis. Cambridge, Mass.: Abt Assoc. 1978.

GRIZZLE,J.E., STARMER,C.F., KOCH,G.G.: Analysis of Categorical Data by Linear Models. Biometrics, vol. 25, 1969, 489-504.

KEMPER,F.J.: Die Anwendung faktorenanalytischer Rotationsverfahren in der Geographie des Menschen. In: GIESE,E. (ed.): Symposium "Quantitative Geographie". Giessener Geogr. Schriften 32, 1975, 34-47.

KEMPER,F.J.: Inner- und außerstädtische Naherholung am Beispiel der Bonner Bevölkerung. Arbeiten zur Rheinischen Landeskunde 42, 1977.

KRAUTH,J. and LIENERT,G.A.: Die Konfigurationsfrequenzanalyse und ihre Anwendung in Psychologie und Medizin. Freiburg and München: K. Alber 1973.

KÜCHLER,M.: Alternativen in der Kreuztabellenanalyse. Zeitschrift für Soziologie, vol. 7, 1978, 347-365.

KÜCHLER,M.: Multivariate Analyseverfahren. Teubner Studienskripten zur Soziologie 35. Stuttgart: Teubner 1979.

LANGEHEINE,R.: Multivariate Analyse nominalskalierter Daten via Goodmans Modell: Sehr wohl eine Alternative. Zeitschrift für Soziologie, vol. 8, 1979, 380-390.

LEWIS,P.: Maps and Statistics. London: Methuen 1977.

MAGIDSON,J.: The Multivariate Analysis of Qualitative Variance: Analyzing the Probability of an Event as a Function of Observable Variables. Ph.D. Evanston, Northwestern Univ. 1976.

THEIL,H.: On the Estimation of Relationships Involving Qualitative Variables. American Journal of Sociology, vol. 76, 1970, 103-154.

THEIL,H.: Principles of Econometrics. New York: J. Wiley 1971.

UPTON,G.J.G. and FINGLETON,B.: Loglinear Models in Geography. Transactions Institute of British Geographers, N.S., vol. 4, 1979, 103-115.

WRIGLEY,N.: Developments in the Statistical Analysis of Categorical Data. Progress in Human Geography, vol. 3, 1979, 315-355.

R E I N E R   S C H W A R Z

REEXAMINATION OF THE GERMAN PHYSIOGRAPHICAL REGIONALISATION (NATURRÄUMLICHE GLIEDERUNG)

From:

BAHRENBERG, GERHARD und STREIT, ULRICH (Eds.):
German Quantitative Geography. Papers presented at the 2nd European Conference on 'Theoretical and Quantitative Geography' in CAMBRIDGE 1980. Münstersche Geographische Arbeiten 11. Schöningh, Paderborn 1981, p. 59 - 66

Author:

Or. Reiner Schwarz

Geographisches Institut
Universität Tübingen

Hölderlinstraße 12

7400 Tübingen

ZUSAMMENFASSUNG

Die Naturräumliche Gliederung Deutschlands (MEYNEN, 1953-1962) versuchte Regionen einheitlichen Gefügemusters physischer Merkmale auszugliedern. Da das Programm dieser Unternehmung keine exakt nachvollziehbare Anweisung enthält, kommt dem Ergebnis selbst die Bedeutung eines empirischen Forschungsgegenstandes zu. Es ist dabei zu fragen, mit welchem Gewicht die einzelnen im Programm dieses Unterfangens vorgesehenen Merkmale in das Gliederungsergebnis eingehen. Darüber hinaus soll der Anspruch überprüft werden, die naturräumlichen Einheiten 3. Ordnungsstufe stellten Großregionen dar, in denen 'Landesnatur und die ihr entsprechende kulturräumliche Ausstattung zur räumlichen Deckung' kommen (MÜLLER-MINY, 1962, S. 259 f).

Um diese Frage zu klären, werden elf thematische Karten physischen Inhalts über Baden-Württemberg ausgewählt, deren thematisches Spektrum (vgl. Tab. 1) in etwa dem Programm der Naturräumlichen Gliederung entspricht. Zu Vergleichszwecken wird zusätzlich die Karte der Naturräumlichen Einheiten 3. Ordnungsstufe in die Untersuchung einbezogen. Mit Hilfe eines arealinvarianten Distanzkoeffizienten wird eine Distanzmatrix dieser Merkmale erstellt und die Verwandtschaftsbeziehungen der Merkmale untereinander mit nicht-hierarchischer Clusteranalyse ermittelt.

Das Ergebnis identifiziert zwei deutlich getrennte Merkmalscluster, deren erstes die Merkmale 'Geologie', 'Lithologie', 'Hydrogeologie', 'Bodentypen', 'Reliefenergie' zusammen mit den 'Naturräumlichen Einheiten 3. Ordnungsstufe' enthält. Dieses Cluster beschreibt Merkmale, die eine enge Beziehung zum oberflächennahen Untergrund zeigen. Die Zugehörigkeit der Naturräumlichen Einheiten zu diesem Cluster bestätigt, daß das Hauptinteresse der Naturräumlichen Gliederung diesen weitgehend aus geologischen und topographischen Karten ableitbaren Merkmalen galt.

Das zweite Cluster enthält die Merkmale 'Reliefbasis', 'Reliefhülle', 'Wuchsklima', 'Potentielle Natürliche Vegetation', 'Leistungsfähigkeit des Waldes' und 'Bodengüte'. Es faßt Merkmale zusammen, welche die Wachstumsbedingungen der Vegetation in enger Anlehnung an die Höhenlage zum Ausdruck bringen. Da gerade diese Merkmale in naher Verbindung zur kulturräumlichen Ausstattung stehen, wird deutlich, daß die Naturräumlichen Einheiten 3. Ordnungsstufe nicht dem oben genannten Anspruch einer auch nur näherungsweisen Wiedergabe kulturräumlicher Grundmuster gerecht werden.

Diese Tatsachen werden zurückgeführt auf das Gliederungsprinzip 'von unten nach obem', das einem hierarchischen Weg folgt, in dessen Verlauf einmal aufgrund von kleinräumlich variierenden Merkmalen getroffene Abgrenzungen in höheren hierarchischen Fusionsstufen nicht mehr mit Blick auf die Gesamtstruktur revidiert werden. Um diesen entscheidenden Nachteil zu vermeiden, sollte bei derartigen Gliederungsversuchen ein nicht-hierarchisches Denken Anwendung finden. Selbstverständlich ist dies nur in Verbindung mit entsprechenden exakt nachvollziehbaren Methoden sinnvoll.

1. OBJECTS OF STUDY

The BRD Physiographical Regionalisation (MEYNEN, 1953-1962) tried to delimit uniformly patterned regions based on natural constituents: rocks, relief, climate, water budget, soils, natural vegetation. Combining mainly topographical and geological maps without rigorously defining this process the authors ended with a regionalisation that can be considered an object of empirical study itself.

Demonstrating the value of mathematical thinking and techniques can best be successful when dealing with problems that are common to traditional geographers. One such question might be to ask for the weight of all programmed constituents and the degree to which climate, soil types and natural vegetation are implicitly involved in the presented Physiographical Regionalisation. Another desideratum represents the statement brought forward by one main author of this programme (MÜLLER-MINY, 1962, p. 259f) the features of the 3rd order physiographical regions correspond with cultural features. On the basis of this statement it was demanded, successfully for a time, to refer statistical surveys on population and agriculture to the proposed physiographical units. Both problems will be revisited by means of rigorous analytical methods.

2. DATA BASIS

According to the programme eleven different physical maps on Baden-Württemberg were chosen, mostly from DEUTSCHER PLANUNGSATLAS BADEN-WÜRTTEMBERG (1969). Relief information was supplied by the topographical map 1 : 50 000 using computational aid to recombine rectangles overlapping the sheet margins. For the purpose of comparison the physiographical regions of 3rd order were added as a twelfth such map. The contents of these maps are taken as nominal variables attached to the cells of a regular grid pattern overlay, quantitative data being transformed into nominal data by means of one dimensional non-hierarchical clustering.

3. DISTANCE OF VARIABLES

3.1. CONCEPTS AND INVARIANCE PROPERTIES

For any two of these nominal variables their mutual similarity can be expressed by a distance coefficient, the selection of which will be discussed here. Basis for its derivation constitutes the contingency table which contains the distribution with regard to the product categories. At least three different concepts are available in this context, the PHI concept originating in inference statistics (HAYS, 1963), the elementary TAU concept (GOODMAN and KRUSKAL, 1954) and the entropy concept derived from information theory. A critical examination of these concepts can be based on the following invariance properties:

(1) size invariance, i.e. invariance with regard to size effects of the whole sample area (total number of grid data)
(2) class invariance, i.e. independence of the number of categories within each variable
(3) area invariance, i.e. invariance with respect to categorical distribution within each variable (marginal distribution of the contingency table).

Size invariance is met by all these concepts, whereas no one supplies class invariance.

3.2. AREA INVARIANCE BY CHANNEL CAPACITY

Critical property is area invariance, which is not met by any of the hitherto applied coeeficients. But it can be attained by using the entropy concept. Such coefficients, as they are usually employed, contain the quotient of synentropy and total entropy. This quotient can be interpreted as information content common to both compared variables in relation to the information that contain both variables together in combination. But the resulting coefficients are dependent on the single distribution of each participating nominal variable, i.e. on the accidential cut of the sample region. In order to get rid of this dependence, synentropy is replaced by channel capacity, which is the maximum synentropy for all possible marginal distributions.We do not choose channel capacity because of its maximizing property but as it is dependent only on the noise matrix.

Unfavourable for the intended usage as a metrical distance coefficient is the loss of symmetry. Exchange of source and sink variables produces different amounts of channel capacity. In order to make up for this deficiency the greater of the two channel capacities is taken as characteristic of the relation. This means marking out the direction in which the single variable distributions can be fitted to the relation so that their common information content can be expressed best.

To evaluate channel capacity is much more expending than to calculate synentropy. An elegant algorithm was published by BLAHUT (1972). It avoids the uncomfortable necessity of excluding negative probabilities that complicate the method based on LAGRANGE-multipliers ,offered in some textbooks on information theory (e.g. KÄMMERER, 1971 or SPATARU, 1973). The approximation algorithm (see SCHWARZ, 1977, p. 8) leads to a solution that depends on the applied accuracy limit. This has to be small enough to secure the metrics of the resulting distance measure which can be proved by simple enumerative computation and test of all possible triangular inequalities within the distance matrix, distance D (A,B) of two variables A and B being expressed by D (A,B)=1-S(A,B), where S(A,B) is the quotient of channel capacity and total entropy.

3.3. CORRECTION ON THE STRENGTH OF CLASS INVARIANCE

Unfortunately there seems to be no concept available to attain class invariance. The distance coefficients will always be dependent on the number of categories within each nominal variable. We can only try to improve coefficients with regard to some specified influences. In this respect it is assumed that distance coefficients depend on the number of elements in the contingency table and on its side proportions. Be (A) the number of categories of variable A, then we can try to predict the distance d of two variables A and B by a linear regression model

$$d = a + b * x_1 + c * x_2$$

with the independent variables

$x_1 = (A)*(B)$ and $x_2 = \max\left[(A)/(B),\ (B)/(A)\right]$

The product moment correlation coefficient $r(d,x_1) = -.264$ indicates a tendency to greater distance coefficients for smaller contingency tables, whereas $r(d,x_2) = .097$ shows a petty inclination to smaller distance coefficients for more quadratic contingency tables. In order to correct the distance coefficients with respect to these specified influences we computed residuals relative to the estimated hyperplane

$$d = .5211 + .0030\, x_1 - .1327\, x_2,$$

and added them to the arithmetic mean of all distance coefficients. It might be that the procedure produces corrected coefficients a little greater than 1 or less than 0. In this case we would have to transform all values so that they shrink into the interval (0,1). Here this final correction was not necessary. The resulting distance coefficients are presented in Tab. 1.

Tab. 1: Area invariant distance matrix after correction on the strength of class invariance

No.	Number of Categories	Variable	1	2	3	4	5	6	7	8	9	10	11	12
1	6	Hydrogeology	0	.655	.623	.512	.484	.476	.476	.336	.653	.511	.688	.348
2	7	Soil Quality	.655	0	.629	.569	.540	.576	.647	.607	.600	.570	.513	.541
3	9	Growth Climate	.623	.629	0	.509	.523	.504	.847	.709	.668	.601	.556	.635
4	7	Relief Envelope	.512	.569	.509	0	.142	.111	.647	.548	.613	.507	.501	.391
5	7	Relief Basis	.484	.540	.523	.142	0	.219	.665	.590	.570	.520	.496	.413
6	7	Relative Relief	.476	.576	.504	.111	.219	0	.625	.559	.584	.539	.472	.383
7	14	Geology	.476	.647	.847	.647	.665	.625	0	.264	.387	.640	.543	.393
8	7	Lithology	.336	.607	.709	.548	.590	.559	.264	0	.463	.616	.768	.313
9	6	Soiltypes	.653	.600	.668	.613	.570	.584	.387	.463	0	.727	.795	.452
10	7	Pot. Nat. Vegetation	.511	.570	.601	.507	.520	.539	.640	.616	.727	0	.545	.465
11	5	Forestry Efficiency	.688	.513	.556	.501	.496	.472	.543	.768	.795	.545	0	.421
12	10	Physiograph.Region.	.348	.541	.635	.391	.413	.383	.393	.313	.452	.465	.421	0

4. NON-HIERARCHICAL CLUSTERING

To clarify the relationships between the different variables R-mode application of cluster analysis provides a suitable tool. There are two fundamental ways of hierarchical and non-hierarchical clustering. Our desire for well separated homogenous clusters of variables cannot be met by hierarchical cluster algorithms at all because they only optimize the linking of two clusters at a time. A linkage once finished cannot be revised by considering the whole cluster structure. Even subsequent revision by discriminant analysis in order to provide well separated convex clusters, as some authors propose (e.g. KING, 1969), is not satisfactory because there exist many arbitrary solutions of convex clusters, which can easily be constructed by setting separating hyperplanes at will. Those that are fixed by the result of hierarchical analysis are by no means optimal in the sense of maximal internal homogenity.

Non-hierarchical cluster analysis on the contrary is able to provide clusters that are both, well separated and homogenous. HARTIGAN (1975) proves that a cluster structure with minimal internal variance is automatically convex. We therefore apply the k-MEANS principle to minimize variance internal.Indeed, this principle does not spontaneously lead to an overall optimal cluster structure, but the solution only provides local optimality that depends on the partition to begin with. The better the starting partition the better the result with respect to minimal internal variance.In order to find out plausible clusters of variables to start with, the similarity matrix $S = 1 - D$ is subjected to principal component analysis, D being the distance matrix (see Tab. 1). Extraction of two principal components and subsequent VARIMAX rotation hints at a reasonable heuristic starting partition. Application of the k-MEANS algorithm (program CLUSYMA published by SPÄTH, 1975) on the distance matrix D leads to the final solution.

5. RESULTS

The result identifies two clearly separated clusters of variables. The first contains variables describing the subsoil features 'Geology', 'Lithology', 'Hydrogeology', 'Soiltypes', 'Relative Relief', together with the '3rd Order Physiographical Regions'. The second cluster of greater 'within'- variance consists of the variables 'Altitude' (i.e. 'Relief Basis' and 'Relief Envelope'), 'Growth Climate', 'Potential Natural Vegetation', 'Forestry Efficiency', 'Soil Quality'. Obviously this second cluster which is not considered by the Physiographical Regionalisation summarizes growth conditions of vegetation that are closely tied to altitude. Especially the variables of this second cluster, which does not include the 3rd Order Physiographical Regions, are more related to cultural features of agriculture, forestry or population distribution than the variables of the first clusters are. Therefore the statement of the 3rd Order Physiographical Regions to correspond with cultural features can no longer be maintained.

This result shows the subsoil features to have been of main interest in the German Physiographical Regionalisation. Neglecting climatic variation and subsequently growth conditions of vegetation can be explained with the 'bottom to top' principle of hierarchically joining areas that could be considered similar to each other, starting with the smallest sites perceived as being homogenous internally. In this regard the traditional concept resembles hierarchical clustering with linkages once executed never to be changed in fusions of higher hierarchical order. In this process distinctive criterias for small areas being always based upon subsoil features are perpetuated. Patterns with extended spatial variation, such as climatic features, never get a chance to be involved. In order to overcome this shortage physical regionalisation should proceed to non-hierarchical thinking and, of course, to fully communicable corresponding methods.

SUMMARY

The German Physiographical Regionalisation tried to delimit regions with uniform regions based on natural constituents: rocks, relief, climate, water budget, soils, vegetation. As the authors mainly combined topographical and geological maps withouth defining the process, it is interesting to question the weight for all the programmed constitutents and the degree to which climate, soils and vegetation are implicitly involved.
Grid based nominal data were derived from various physical maps on Baden-Württemberg. By non-hierarchical cluster analysis based on a channel capacity correlation coefficient, which is supposed to be invariant to marginal values of the contingency table, two clusters of variables could be identified. The first contains variables describing the subsoil features 'Geology', 'Lithology', 'Hydrogeology', 'Soiltypes', 'Relative Relief', together with '3^rd Order Physiographical Regions'. The second cluster of greater 'within'-variance consists of the variables 'Altitude', 'Growth Climate', 'Potential Natural Vegetation', 'Forestry Efficiency', 'Soil Quality'. Obviously this cluster, which is not considered by the Physiographical Regionalisation, summarizes growth conditions of vegetation that are closely tied to altitude.

Especially the variables of this second cluster, which does not include the Physiographical Regionalisation, are more related to the cultural features of agriculture, forestry or population distribution than the variables of the first cluster are. Therefore the statement brought forward by one main author of the Physiographical Regionalisation (MÜLLER-MINY, 1962, p. 259f), the features of the 3^rd Order Physiographical Regions correspond with cultural features, can no longer be maintained.

This result is due to the 'bottom to top' principle of hierarchically joining areas that are similar to each other with respect to small scale varying features. In this process of hierarchical clustering patterns with extended spatial variation, such as climate or altitude, never get a chance to be involved. Only rigorous non-hiearchical methods could remove this deficiency.

REFERENCES

ANDERBERG,M.R.: Cluster Analysis for Applications. New York 1973.

BLAHUT,R.E.: Computation of Channel Capacity and Rate-Distortion Functions. In: IEEE-Transactions on Information Theory, Vol. II-18, No. 4, 1971, p. 460-473.

DEUTSCHER PLANUNGSATLAS 6, Baden-Württemberg. Ed. by Akademie für Raumforschung und Landesplanung, and Innenministerium Baden-Württemberg. Hannover 1969.

FISCHER,M.M.: Nicht-hierarchische Clusteranalyse in der Regionalforschung. Österreichische Studiengesellschaft für Kybernetik, Berichte, 1976.

GOODMAN,L.A. and W.H.KRUSKAL: Measures of Association for Cross Classifications. In: Journal of the American Statistical Association 49, 1954, p. 732-764.

HARTIGAN,J.A.: Clustering Algorithms. New York 1975.

HAYS,W.L.: Statistics. London 1963.

KÄMMERER,W.: Einführung in mathematische Methoden der Kybernetik. Berlin 1971.

KILCHENMANN,A.: Statistisch-analytische Landschaftsforschung. In: Geoforum 7, 1971, p. 39-53.

KILCHENMANN,A.: Die Merkmalanalyse für Nominaldaten - eine Methode zur Analyse von qualitativen geographischen Daten basierend auf einem informationstheoretischen Modell. In: Geoforum 15, 1973, p. 33-45.

KING,L.: Statistical Analysis in Geography. Englewood Cliffs 1969.

MEYER-EPPLER,W.: Grundlagen und Anwendungen der Informationstheorie. Berlin 1959.

MEYNEN,E. (ed.): Handbuch der naturräumlichen Gliederung Deutschlands. 9 Lieferungen. Remagen, Bad Godesberg 1953-1963.

MÜLLER-MINY,H.: Betrachtungen zur Naturräumlichen Gliederung. In: Berichte zur deutschen Landeskunde 28, 1962, p. 258-279.

OTREMBA,E.: Die Grundsätze der naturräumlichen Gliederung Deutschlands. In: Erdkunde 2, 1948, p. 156-167.

SCHMITHÜSEN,J.: Grundsätze für die Untersuchung und Darstellung der naturräumlichen Gliederung von Deutschland. In: Berichte zur deutschen Landeskunde 6, 1949, p. 8-19.

SCHWARZ,R.: Arealinvarianter Zusammenhang von Nominalmerkmalen. Karlsruher Manuskripte zur Mathematischen und Theoretischen Wirtschafts- und Sozialgeographie 21, 1977.

SPÄTH,H.: Cluster-Analyse-Algorithmen zur Objektklassifikation und Datenreduktion. München 1975.

SPATARU,A.: Theorie der Informationsübertragung. Signale und Störungen. Braunschweig 1973.

OBERLA,K.: Faktorenanalyse. Eine systematische Einführung für Psychologen, Mediziner, Wirtschafts- und Sozialwissenschaftler. 2. ed., Berlin etc. 1971.

W O L F H A R D   S Y M A D E R

THE APPLICATION OF CANONICAL CORRELATION ANALYSIS IN GEOECOLOGY

From:

BAHRENBERG, GERHARD und STREIT, ULRICH (Eds.):

German Quantitative Geography. Papers presented at the 2nd European Conference on 'Theoretical and Quantitative Geography' in CAMBRIDGE 1980. Münstersche Geographische Arbeiten 11. Schöningh, Paderborn 1981, p. 67 - 76

Author:

Dr. Wolfhard Symader

Geographisches Institut
Universität Köln

Albertus-Magnus-Platz
5000 Köln 41

ZUSAMMENFASSUNG

Wegen des großen Meßaufwandes können die wesentlichen Prozesse in einem Ökosystem immer nur exemplarisch erfaßt werden. Die Planung braucht aber Entscheidungshilfen für weitaus größere Räume, so daß die Ergebnisse von Punktmessungen auf diese Räume übertragen werden müssen. Das ist nur möglich, wenn bekannt ist, welchen Einfluß die herrschenden Umweltbedingungen auf diese Prozesse haben. Mit ihrem Ansatz, die Abhängigkeit zwischen zwei Variablensätzen zu erfassen (Prozeßablauf und Umweltbedingungen), scheint die kanonische Korrelationsanalyse einen Lösungsansatz zu diesem Problem zu bieten.

Trotzdem wird sie kaum angewandt, da sie weder wie die ihr verwandte multiple Regression zu Vorhersagen eingesetzt werden kann, noch bei einer konventionellen Anwendung wesentliche Informationen zur Datenstruktur liefert.

Bei einem Verzicht auf die Aussagen des kanonischen Korrelationskoeffizienten und der Gewichte zur Berechnung der kanonischen Variablen können die Korrelationen zwischen Ursprungs- und kanonischen Variablen wie die Ladungen einer Hauptkomponentenmatrix interpretiert werden. Auch in diesem Fall hilft eine Varimaxrotation, die Datenstruktur zu verdeutlichen.

Leider stellte sich heraus, daß bei einer zu hohen Zahl von Interkorrelationen die ersten kanonischen Korrelationen den gleichen Betrag aufweisen können, wodurch sich keine eindeutige mathematische Lösung ergibt. In der Praxis antwortet der Rechner mit Programmabbruch, aber es können durch Rundungsfehler auch mathematische Artefakte auftreten, die nicht ohne weiteres als solche erkannt werden können. Von dem Gebrauch der kanonischen Korrelationsanalyse wird daher abgeraten.

INTRODUCTION

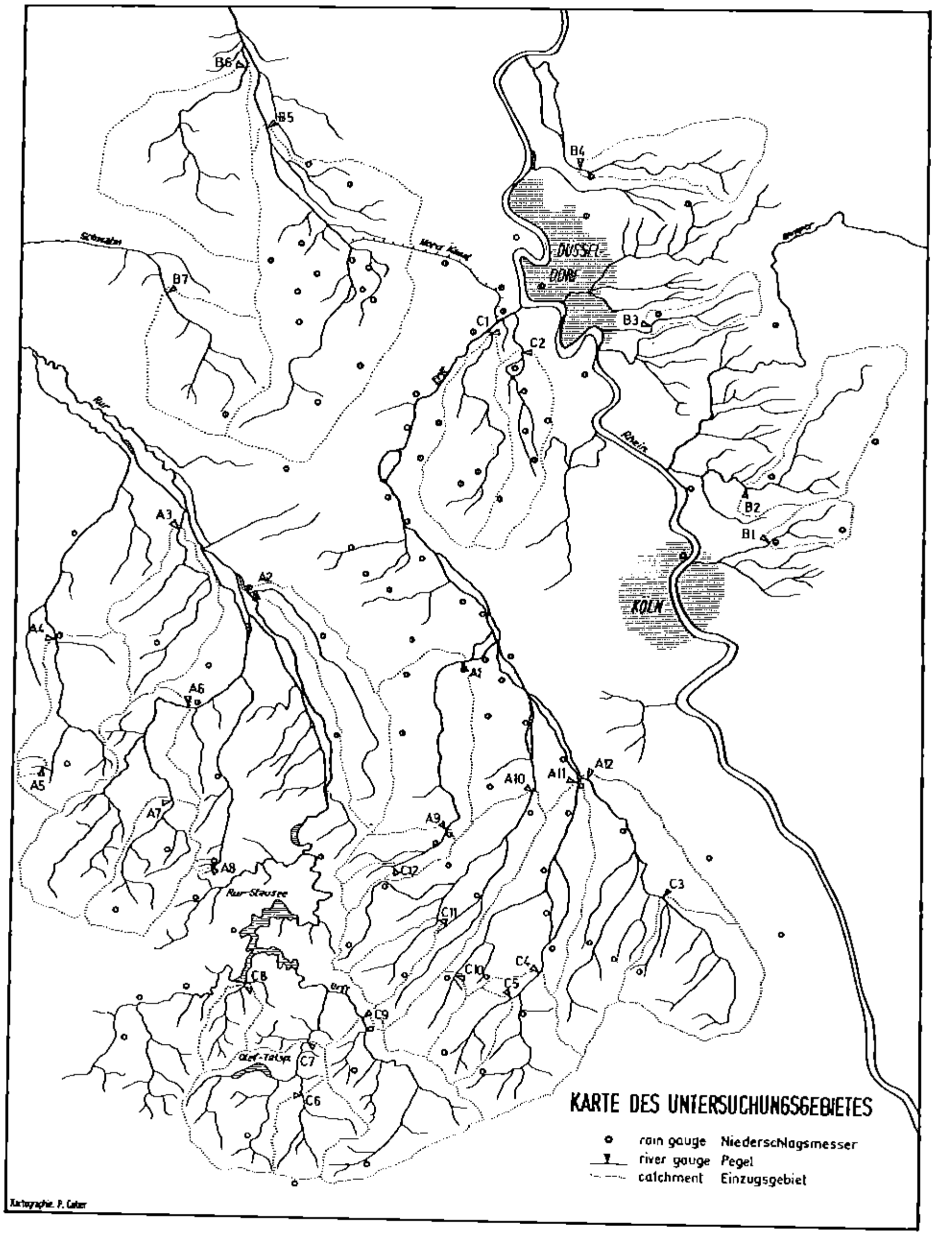
By dealing with two sets of variables canonical correlation analysis shows a very interesting approach to solve the most urgent problem of geoecology. Measurements of time dependent processes of energy, water, or solids are only possible at great expense. Therefore they are mostly restricted to small areas. However problems of planning do not stop at those areas but cover large regions. If geoecology considers its results as useful for practical purposes, it must be able to extend them after certain modifications to a large area or even to transfer them to other regions where no measurements have been made.

As it was shown by W. SYMADER (1980a, B3-B6) transfer of measurements is possible if the influence of the environment on the process is known. Both processes and environment are described by a set of variables so that canonical correlation analysis seems to be the appropriate statistical procedure.

But as it is difficult to estimate its efficiency, canonical correlation analysis is not very often used in physical geography. An example of a conventional application exhibites some of these reasons.

From May 1976 until May 1977 water samples were taken biweekly in 31 catchments of the Northern Eifel mountains and the adjacent Lower Rhine area (W. SYMADER & W.THOMAS 1978; W.SYMADER & R.HERRMANN 1979; W.SYMADER 1980b) and analysed for dissolved and suspended heavy metals and nutrients. Besides, data of heavy metal content in fluvial sediments were available (W.THOMAS 1977). The characteristics of the catchments were described by a set of variables listed in Table 1.

As canonical correlation analysis can be regarded as the general form of multiple regression, it must fulfill the same premises such as normal distribution and, what is more important, independency of the predictor variables. If the variables of a set are intercorrelated among each other, canonical coefficients are not



KARTE DES UNTERSUCHUNGSGEBIETES
rain gauge Niederschlagsmesser
river gauge Pegel
catchment Einzugsgebiet
DÜSSELDORF
KÖLN
Rhein
Rur
Rur-Stausee
B1
B2
B3
B4
B5
B6
B7
A1
A2
A3
A4
A5
A6
A7
A8
A9
A10
A11
A12
C1
C2
C3
C4
C5
C6
C7
C8
C9
C10
C11
C12

TABLE 1: Spatial variables of the characteristics of the catchments

Code	
FN	drainage area
FLDI	density of rivers (km/km^2)
FWALD	area of a catchment covered by forests,
FGRUEN	by grassland,
FACK	by farmland,
FSIED	by settlement,
FOED	and by non usable land (km^2).
PCWALD	percentage of a catchment covered by forests,
PCGRN	grassland,
PCACK	farmland,
PCSIED	and settlement.
GEF	mean slope of the river basin
FORM	L^2/FN, L=axial length of the catchment.
FSAND	area of a catchment with sandstone,
FKALK	calcareous rocks,
FGRD	shales and greywackes,
FQRT	quaternary rocks.
PCSAND	percentage of a catchment with sandstone,
PCKALK	calcareous rocks,
PCGRD	shales and greywackes,
PCQRT	quaternary rocks.
EGW	indicator of waste water pollution. It gives the amount of waste water compared to the waste water production of an average person.
BETR	number of industrial plants
BESCH	number of employees within a catchment
FEIND	number of plants working up iron
NFEIND	number of plants working up metals except for iron.
STAHL	number of steel mills,
CHEM	chemical industries,
TEXT	textile industries,
BEKL	clothing industries,
LEBMI	food industries,
GALV	galvanic industries.
EINW	number of inhabitants of a catchment
SIGEW	area of settlements, weighted according to their distance to the sample point.

steady and the significance of the canonical coefficient is overestimated. To get variables independent of each other a principal component analysis is used first (Tables 2 and 3).

The matrix in table 2 shows that only a few variables are intercorrelated e.g. the nutrients Na, K and PO_4 coming from domestic sewage or Ca and Mg as indicators of water hardness.

Zn is a heavy metal which can be found in nearly every type of waste water. Thus it shows good correlations with many other elements leading to medium loadings in several components. As a set of independent variables Mn, Cu, Pb, Cd, Ni, Cr, Co, Mg and K are selected. Note that such interesting elements as Zn, Na and PO_4 are not taken into consideration any more.

A great deal of the intercorrelations in Table 3 is due to the heterogenity of the whole sample. The catchments of the Northern Eifel mountains with high percentages of shales, greywackes and quartzites

Table 2: Varimaxrotated principal component matrix of spatial means of dissolved heavy metals and nutrients

	1	2	3	4	5	6	7	8	9	
Variation %	22,7	12,5	11,1	10,3	9,8	8,8	8,4	6,4	6,3	sum=96.3%
Zn	48		66				-41			
Fe					-47					
Mn				-95						
Cu						83				
Pb								-79		
Cd			97							
Ni									79	
Cr					-82					
Co	48						-87			
Mg		96								
Ca		72								
Na	87									
K	87									
PO_4	87									

All variables are log transformed. All loadings are multiplied by 100. Loadings less than |40| are omitted.

Table 3: Varimaxrotated principal component matrix of spatial characteristics of the catchments

		1	2	3	4	5	6	7	8	9	10	11	
	Variation %	26.9	15.1	11.2	10.2	5.3	4.2	3.9	3.5	3.4	2.7	2.6	sum=89 %
X	FN				92								
X	FLDI		86										
	FWALD		44		82								
	FGRUEN				76								
	FACK		-57		52								
	FSIED	94											
X	FOED									-95			
	PCWALD		86										
X	PCGRN							-91					
	PCACK		-80					42					
	PCSIEO		45		71								
X	GEF										81		
X	FORM											83	
X	FSAND								93				
	FKALK					81							
	FGRD		60		70								
	FQRT	59	-61										
X	PCSAND								42				
X	PCKALK					91							
	PCGRD		81										
	PCQRT		-79										
X	EGW	97											
	BETR	87											
	BESCH	89											
	FEIND	57					-56						
X	NFEIND						-88						
	STAHL			89									
	CHEM	70		63									
	LEDER	72		47									
	TEXT	96											
	BEKL	95											
	LEBMI	84		41									
X	GALV			96									
	EINW	91											
X	SIGEW	45		44									

X indicates that this variable is chosen for canonical correlation analysis. All loadings are multiplied by 100. Loadings less than |40| are omitted.

covered by forests and grassland differ significantly from the Lower Rhine area with its arable loess-soils. Note the different signs of PCGRD and PCQRT or PCWALD and PCACK in the second component. The first and the third components discriminate between two different aspects of industrialization. As many catchments of the densely populated Lower Rhine area show a high density of industry, variables of settlement (FSIED,EINW) and industry occur together in the first component. On the other hand the third component reflects branches which developed from the special conditions in the ancient mining districts. This industry is mainly situated near the mountains or in the broad valleys.

The fourth component is a dimension of size.

As it is the purpose of this principal component analysis to create a set of independent variables this short interpretation must do. A detailed description of the area under investigation is given by H.KRUTZ (1979, 22-26 and appendix).

A comparison between the two sets of variables by canonical correlation analysis shows that only the first pair of canonical variables is significant on the 0.1% level, the second one only on the 5% level. In Table 4 the matrix of canonical loadings is given in order to understand the structure of the relation between the dissolved elements and the spatial characteristics of the catchments.

Although only the first two dimensions are statistically significant, four dimensions can be usefully interpreted. R.E.BLACKITH & R.A.REYMENT (1971, 90-91) point out that a clear distinction is to be drawn between the statistical and the biological importance of an axis of variation. A lack of significance does not mean that a pair of canonical variables is not important. If only a few variables are measured or a small sample is taken, the last canonical correlations will hardly be significant. But they still can be used in understanding the information of the data.

The first dimension shows a strong relationship between dissolved Ni and the galvanic industry. The second dimension displays a more general fact. Rivers highly polluted by waste water (EGW) carry high concentrations of nearly all measured elements. On the other hand the degree of settlement is very low in those parts of the mountains which lie far away from the broad valleys. In these areas the river density is high, because all the little tributaries have not yet become connected to important rivers. That is why FLDI shows an opposite sign to EGW. The little rivers of the Eifel are hardly polluted by waste water. That means that the second dimension describes the contrast between the highlands of the Eifel and the high concentrations of dissolved solids in polluted water. This relationship is not so clear for Cu and Cd compared to the other elements, because these heavy metals can also be found in unpolluted waters of the western part of the Eifel caused by extended ore deposits.

From this example it follows that a conventional application of canonical correlation analysis takes a lot of effort, because it must be carried out in two steps. A principal component analysis or at least a correlation analysis must be calculated, before a comparison of two sets of variables is possible. If you use variables which are independent of each other then only few dimensions are significant and the percentage of common variance is small. Therefore the percentage of explained variance in the individual dimensions is also small. The first dimension e.g. explains only 7.8% of the total variance of the first set of variables and 12.8% of the second set. This result is not only confined to the example presented above, but can also be proved by other investigations (R.HERRMANN et al. 1979, 655-659).

The interpretation of the matrix of canonical loadings gives interesting results about some aspects of intercorrelation. But a principal component analysis with all variables (not published in this paper) does nearly the same and can be interpreted in a more convenient way, because of a high number of high loadings. As a conclusion from this there seems to

Table 4: Matrix of canonical loadings for dissolved solids and catchment characteristics

	1	2	3	4	
Variation %	7.8	31.0	16.6	8.7	sum = 64 %
Mn	18	51	42	54	
Cu	33	38	58	-1	
Pb	-11	50	62	-2	
Cd	21	34	37	27	
Ni	60	51	34	-4	
Cr	-6	79	34	-25	
Co	-1	67	3	11	
Mg	-37	10	43	-56	
Variation %	12.8	18.0	13.3	5.7	sum = 50 %
FN	16	7	56	-1	
FLDI	26	-73	8	19	
FOED	13	-10	8	7	
PCGRN	5	-16	14	43	
GEF	-9	-64	21	-21	
FORM	12	48	10	-57	
FSAND	-50	-21	44	-1	
PCSAND	-51	-28	35	-12	
PCKALK	-32	-40	57	-30	
EGW	-12	73	40	-11	
NFEIND	53	13	55	16	
GALV	71	20	29	-11	
SIGEW	36	53	33	-18	
r^2	0.98	0.95	0.92	0.74	
p	0.00	0.03	0.44	0.94	

be no need for a canonical correlation analysis for this problem.

Using variables with a medium degree of intercorrelation the results differ considerably from those gained by a conventional application.

P.VINCENT & V.CLARKE (1980) compare two sets of variables describing morphological properties of terracettes and soil properties. They show a close relationship between both sets and interpret this relationship by means of canonical weights. Using intercorrelated variables the statistical significance between two sets will be overestimated just as in the case of multiple regression when intercorrelated predictor variables are used (H.GAENSSLEN & W.SCHUBÖ, 1973, 125). In addition to that the canonical weights are not steady. They may vary widely between different samples and even insignificant variables may show high weights. But this last difficulty can be avoided, if canonical loadings instead of weights are used for interpretation.

The second example of this paper shows typical results of this approach. In Table 5 the average concentrations of dissolved

Table 5: Matrix of canonical loadings of heavy metals in flowing water and sediment

		1	2	3	4	5
	Variation %	15.9	12.1	16.7	6.6	11.9
flowing water	Zn	-51	68	-20	-12	29
	Fe	-51	32	30	27	23
	Mn	-30	45	-24	7	25
	Cu	-49	30	11	11	53
	Pb	- 6	33	42	1	63
	Cd	-21	66	-12	-22	2
	Ni	-48	25	-22	- 4	27
	Cr	-39	20	55	18	50
	Co	-51	36	- 3	-50	36
	Mg	38	16	28	-57	22
	Ca	- 5	1	55	-31	52
	Na	-55	12	64	- 7	1
	K	-39	3	62	-24	13
	PO_4	-32	13	67	- 7	13
	Variation %	5.7	6.6	12.4	4.7	4.2
sediment	Zn	-11	53	-25	-22	1
	Fe	7	- 6	-51	33	-35
	Mn	46	- 6	-60	21	-20
	Cu	-16	15	-31	8	20
	Pb	38	27	-26	20	30
	Cd	0	45	-41	- 6	-15
	Ni	4	35	-50	13	- 2
	Cr	-19	17	- 9	38	8
	Co	- 8	36	-59	-16	- 6
	Mg	44	-19	-12	-29	4
	Ca	28	- 8	12	-27	31
	Na	10	- 9	9	18	-38
	K	24	-11	-12	-12	- 9
	PO_4	14	11	24	-18	-16
	r^2	1.00	0.98	0.93	0.90	0.89
	p	0.000	0.000	0.000	0.001	0.009

elements are compared to the concentrations in fluvial sediments. All variables were taken into consideration.

Five dimensions are statistically significant on the 1 % level but this significance is partly due to the sampling error whose effect increases excessively with the number of variables. As no principal component analysis was calculated in beforehand, information about the intercorrelations is only given by the canonical loadings.

The first dimension of Table 5 displays that the low concentrations of many dissolved heavy metals coincide with high concentrations of Mn and Mg in the sediments. This coincidence is a combined effect of a low degree of pollution and a high capacity of self purification in many flowing waters of the Northern Eifel mountains.

The second dimension shows high concentrations of Zn and Cd both in flowing water and sediment which are due to ore deposits and industrial pollution. The high loadings of dissolved nutrients in the third dimension coincide with low concentrations of Fe, Mn, Ni and Co in the sediment. The explanation of these high loadings is similar to the first dimension. The Lower Rhine loess area has a smaller background level for heavy metals than the sediments of the Eifel mountains. Besides that, the high concentrations of dissolved nutrients indicate that the rivers are polluted by domestic sewage rich in organic carbon and poor in oxygen. Under these circumstances the sediments often consist of black mud with a foul odour and low redox- and pH-values. Therefore remobilization dominates and immobilization is blocked. Thus many sediments contain amounts of heavy metals even smaller than their low background level. It is worth noticing that the highest loadings in the sediment set are given by Fe and Mn, both very sensitive to changes in redox-potential and Ni and Co which cannot easily be immobilized.

Results of additional multivariate procedures clarify that this interpretation is the only meaningful possibility. The results of the canonical correlation analysis are not very convincing. The interpretation is based on loadings from 0.4 to 0.6. The most important variables have 20 % to 30 % of common variance with the canonical dimensions, which is the reason for a difficult interpretation. The fourth and fifth dimension cannot be interpreted at all.

Therefore the matrix of canonical loadings was rotated according to the varimax procedure. By doing so, the loadings increased, but nearly all of the original structure was lost. The result of this experiment was similar to the result which was achieved with the principal component analysis (Table 2).

Using variables with medium intercorrelations, a canonical correlation analysis followed by a varimaxrotation of the canonical loadings offers a good alternative to principal component analysis and gives more information about the relationship between two sets of variables. But some experience in interpreting matrices of loadings is needed because of many medium loadings.

In a third example a calculation of canonical correlation analysis using highly intercorrelated variables was attempted. This attempt failed completely because the program repeatedly stopped with an "indefinite condition". This happens, when two or more eigenvectors are equal, so that the canonical weigths cannot be computed.

But what is worse, errors by rounding off may produce slight differences between the eigenvectors which will lead to mathmatical artefacts. These fictional results cannot be checked without further calculations. A useful test is the sum of squares of the loadings for each variable which cannot exceed the value 1.00.

As it is not always possible to decide if variables show high or medium intercorrelation, before you start analysing the data, a principal component analysis is recommended as a first step in analysing a bulk of data. But in that case a canonical correlation analysis would not give much additional information, so that this procedure cannot be regarded as a powerful tool in physical geography.

REFERENCES

BLACKITH,R.E. & REYMENT,R.A.: Multivariate morphometrics. Academic press, London - New York, 412 p. 1971.

GAENSSLEN,H. & SCHUBÖ,W.: Einfache und komplexe statistische Analyse UTB, 326 p., 1973.

HERRMANN,R., BOLZ,U., SYMADER,W., RUMP,H.H.: Interpretation and prediction of spatial variation in trace metals in small rivers by canonical and discriminate analysis. In: Morel-Seytoux,H. et al. (eds.). Surface and Subsurface Hydrology, Proc. 3rd Int. Hydr. Symp. on Theoretical and Applied Hydrology at Fort Collins 1977. Water Res. Publ., 636-660, 1979.

KRUTZ,H.: Cyanverbindungen in Fließgewässern. Eine mathematische Beschreibung ihres raum-zeitlichen Verhaltens und ihrer Wechselwirkungen zwischen Wasserkörper und Sediment. Veröffentlichung des Instituts für Wasserforschung G.m.b.H. Dortmund und der Hydrologischen Abteilung der Dortmunder Stadtwerke AG, Nr. 28, Dortmund, 139 p., 1979.

SYMADER,W.: Zur Problematik landschafts-ökologischer Raumgliederungen. Landschaft und Stadt 12, 81-89. 1980a.

SYMADER,W.: Suspended heavy metals. An investigation of their temporal behaviour in flowing waters. Catena 7, 1-26, 1980 b.

SYMADER,W. & HERRMANN,R.: Schwermetalle in Lösung. Eine Untersuchung über ihr zeitliches Verhalten in Fließ-gewässern. Catena 6, 1-21, 1979.

SYMADER,W. & THOMAS,W.: Interpretation of average heavy metal pollution in flowing waters and sediment by means of hierarchical grouping analysis using two different error indices. Catena 5, 131-144, 1978.

THOMAS,W. : Schwermetalle in Flußsedimenten und ihre mathematische Vorhersage. Diplomarbeit, Köln 51 p., 1977.

VINCENT,P. & CLARKE,V.: Terracette morphology and soil properties: a note on a canonical correlation study. Earth Surface Processes 5, 291-295, 1980.

A R I B E R T   P E T E R S

THE DECOMPOSITION OF SPATIAL VARIATIONS INTO STRUCTURAL AND REGIONAL COMPONENTS

From:

BAHRENBERG, GERHARD und STREIT, ULRICH (Eds.):
German Quantitative Geography. Papers presented at the 2nd European Conference on 'Theoretical and Quantitative Geography' in CAMBRIDGE 1980. Münstersche Geographische Arbeiten 11. Schöningh, Paderborn 1981, p. 77 - 90

Author:

Aribert Peters

Internationales Institut für Management und Verwaltung
Wissenschaftszentrum Berlin

Platz der Luftbrücke 1-3

1000 Berlin 42

ZUSAMMENFASSUNG

Die Analyse interregionaler Disparitäten erfordert in den meisten Fällen eine Bereinigung um Einflüsse, die auf unterschiedliche strukturelle Zusammensetzungen der Regionen zurückgehen. Beispiele dafür sind der Einfluß regional unterschiedlicher Siedlungsstrukturen (d.h. ungleicher Mix verschiedener Siedlungstypen in den Regionen) oder unterschiedliche Branchenstruktur auf die Einkommens- oder Beschäftigungsungleichheiten zwischen Regionen. Ein duales Problem tritt beim Strukturvergleich von Regionen auf, wobei von strukturunabhängigen Regionsunterschieden, die im obigen Beispiel errechnet wurden, abgesehen werden muß.
Es läßt sich also eine Symmetrie zwischen strukturellen und regionalen Einflußfaktoren erkennen, die auf eine im ersten Teil des Beitrags entwickelte varianzanalytische Lösung des Problems hinführt.

Anschließend wird diese Lösung mit der Shift-Share-Analyse verglichen. Daraus ergibt sich ein alternativer Shift-Share-Ansatz, der im Gegensatz zur traditionellen Formulierung Regions- und Struktureffekte äquivalent behandelt. Erst weitere Untersuchungen werden zeigen, ob der hier formulierte Ansatz die Lösung für ein sehr allgemeines Problem räumlicher Analysen darstellt.

1. INTRODUCTION

This article does not deal with a highly sophisticated new method, but rather it introduces a very simple and common sense way of analysis that is useful in studies of urbanization effects and of related spatial disparities. The justification for its simplicity is based on the belief that we suffer today more from a lack of simple communicable and easily understandable methods than from a lack of highly developed quantitative techniques.

Disparities within a country are in most cases reduced either to disparities among specific types of settlements (e.g. urbanized or rural settlements) or disparities among major regions. This usually pertains, for example, to investigations of population growth, per capita incomes and especially for the consideration of urbanization processes. The aim of the paper is to unify these two approaches into a single framework of analysis. The common schemes of analysis, the ANOVA and the shift-share approach are compared with respect to this problem.

The selection of the regions (e.g. their delineation and size) is of crucial importance, not only for the investigation of disparities. The choice of regions is guided by the research interest, by the scale of operation of the processes and by the theoretical framework which is in the researcher's mind. Among others, there are two main paradigms applied in the choice of regions:

1) Subdivision of a country into several major nodal regions, where each of the regions is comprised of many different settlements. These regions often have "an obvious geographical integrity related also to cultural and quasi-national factors" (HUGHES 1979, p. 174). They are more or less autonomous administrative units (federal states, provinces, economic planning regions) and their number ranges in most cases between five and fifty. "The flow of statistics that followed upon the acceptance of the Standard Regions increased their use as the obvious units for spatial analysis" (HUGHES 1979, p. 174).

2. Subdivision at a local level into regions, which are comprised of mainly a single settlement (e.g. communities). The number of such small regions is usually very high in a country (mostly several hundreds of regions). Their great number calls for grouping prior to analysis and this is often performed according to functional characteristics such as economic and social structure, size or relative location via major centers. This grouping results in a set of contingent regions, later reffered to as structural regions. Studies of urban hierarchies are often performed in this way.

Much analysis is also based on regions which are in the range between the two extremes mentioned above.

The relevance of both major nodal regions and structural regions stems from the fact that they have a special political and theoretical relevance:

1) Theoretical reasons:
The theoretical importance of "major nodal" or "standard" regions results from their use as a base for regional, economic or development models.
On the one hand, this results simply from the availability of data, as well as from the structure of the current econometric models. Spatial explanatory models (e.g. city hierarchy models), on the other hand, often use structural classifications of towns.

2) Policy reasons:
Regional economic policies are often based on comparison between the major nodal regions, whereas important land-use decisions are carried out by the local authorities.

It can be concluded that both major regional and structural regional considerations complement each other, both covering important but distinct political, theoretical and practical topics. Further, there is a gap between the two levels, and this missing link is not brought about by investigation at an intermediate regional scale: "Emerging economic problems of urban areas has become a more dominant source of spatial imbalance and inequality; but it has tended to fall into a policy 'gap' caused by the separation of 'industrial' (economic) and physical planning policies" (HUGHES 1979, p. 173). This article provides an analytical framework for bridging the gap. The theoretical implications, as well as the empirical applications, will be published later.

2. THE ANALYTICAL PROBLEMS

The starting point for the present consideration is a subdivision of the whole investigated surface into very small areas ("settlements). These settlements can be classified, on the one hand, according to their gross location in space and to other things which they have in common with their neighbours. This would result in subdivision of the whole surface in a few "major" regions" (index j). Alternatively, the settlements can be classified according to their relative location, according to the type or structure of the settlement (index i), i.e. urban, rural, commuting etc. Each settlement is labelled by the region and by settlement-type it belongs to: the "region" carries the gross spatial and the "type" the structural characteristic of each settlement. This way of indexing will be referred to by the term "structural-regional-taxonomic system".

The settlements in the same region i and of the same type j are numbered by a third index k.

When investigating a deliberately chosen variable (e.g. per capita income) there are four different questions of interest which are to be answered:

1) To which degree can the value of an individual settlement V_{ijk} be explained or expressed by the fact that it belongs to a certain region i and to a certain settlement type j; what are the individual contributions of the two factors and how do they interact? The contribution of the region j to the value v_{ijk} will be abbreviated by β_j, the settlement type contribution by $(\alpha\beta)_{ij}$.

2) What are the differences among the major regions standardized for their different settlement patterns (i.e. distribution of the types of settlements within the region)? An example will stress the problem. The north of the Netherlands is a rural peripheral area. To what extent can the lower per capita income in the north be explained by the fact that it possesses a different settlement structure? If it is assumed that the north has the same settlement pattern as the other parts of the country, then would it still have a lower income, and if so, how great would the difference be?

3) What are the differences among different settlement types, considering the fact that they are unevenly distributed over regions and that they provide distinct socio-economic environments?

4) What will be the influence of the different settlement sizes in terms of population, families, housing etc., based on the results of the three previous questions? The settlement sizes will be measured by a variable g_{ijk}. If the different settlement sizes are neglected, then every settlement would enter into the computation with the same weight. In this case, the result would be biased towards small settlements, whereas in comparison to them, the major cities would be almost ignored.

An approach to the solution of these questions will be given in the next section by means of an analysis of the variance-approach. Subsequently, the relationship to the shift-share-analysis will be discussed.

3. THE ANOVA-SOLUTION

The analysis of variance (ANOVA) is an appropriate solution to the problems raised in the last section if a linear relationship between the variable v_{ijk}'s values and the effects of the two explaining independent categorical variables α_1 and β_j and their interaction $(\alpha\beta)_{ij}$ is assumed:

$$v_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk} \quad (1)$$

e_{ijk} is a normally independent distributed error with zero mean and known variance. The squared sum of errors will be minimized in the process of parameter estimation. Conventionally, the formula (1) is written with

$$\sum_{i,k} \alpha_i \cdot g_{ijk} = \sum_{j,k} \beta_j \cdot g_{ijk} = 0 \quad (2)$$

If the weighted sum of effects does not add up to zero, then the actual value of the sum can be subtracted from the effects and added to μ. This will not change the fit of the model, but the formula thus modified will fulfil the requirements of (2).

The standard ANOVA-approach will compute results for the following three kinds of questions:

1. Significance testing by computing statistics for the sum of squares;
2. Computation of μ and the effects of α_i and β_j;
3. Testing the adequacy of the model (1) as compared to the alternative models, in which one or more terms of the full model (1) are dropped.

It turns out that significance-testing is not necessary in the present case, because the full population (i.e. all settlements) are used in the computation and it is not intended to infer to any unknown other cases. The procedure mainly serves as a tool for statistical description and analysis of the situation. It follows therefore, that as a first step of computation, all settlements with the same indices ij can be aggregated into one case with an appropriate weight. In empirical investigations it should be made clear, however, that the variances lost in this step are of a smaller order of magnitude than the differences that are to be explained in the analysis after the first step.

The adequacy-of-model-questions will be raised next. It is assumed first of all that the model is as simple as the theoretical consideration will permit. If a model of this most simple kind fits the data well, then the model will be justified. Many other multivariate procedures confirm their assumptions in this way.

What are the essential terms in equation (1)? Whereas the two effects α_i and β_j are central to the problem, an interaction between the effects cannot be expected a priori. A good fit without interaction would be an important result, and would mean that both effects operate independently from each other. Therefore an interaction-free-model of the type

$$v_{ij} = \mu + \alpha_i + \beta_j + e_{ij} \quad (1a)$$

is tested first.

The meaning of equation (1a) and the structure of the problem can be easily visualized by means of a contingency-table. In the following example, only two regions, R1 and R2, are thought to exist and there are only two types of settlements K1 and K2.

In the major double frame of Figure 1, the mean values of all settlements of a type and region are shown as well as the value of the residuals e_{ij} (in % of the original value) and the relative weights of each cell (in %). The single frames above and beside the double frame contain the means, the effects and the sum of μ and its effects. This kind of arrangement, as well as the methods of the effects computation, can be found in TUKEY (1977).

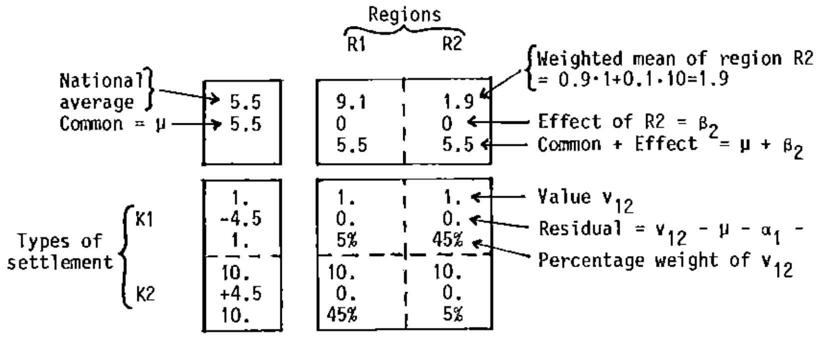


Figure 1: Estimation of region and settlementtype effects in a nonorthogonal case and demonstration of the influence of weights.

Problems in the computation of the effects arise because of the unequal cell frequencies and respective weights encountered in (1a). It follows that the values of α_i cannot be computed independently of the values of β_j. In our example, the low average of region two can either be interpreted as an effect of the region or as a consequence of the high share of low-valued type-one-settlements in region two. In the latter case, the "region-two-effect" would be minor because of the major effects of settlement types. This is more probable because of the small variance between the settlement-one-type values.

More generally, a nonorthogonal ANOVA-problem is posed here. This problem is by no means statistically simple and there is ongoing research on this question (APPELBAUM 1974): "Due to the complexities in analyzing nonorthogonal designs generally, the best solution to the problem is not to have unequal N_{ij}'s" (TIMM 1975, p. 526), or to avoid such designs at all costs (APPELBAUM 1974, p. 335).

For the reasons mentioned above, statistical testing will not be considered here, and therefore the problem reduces to the computation of the main effects. The main effects can be fairly easily computed by subtracting iteratively the weighted means of the rows and the columns in the table (TUKEY 1977). The procedure converges rapidly after the first few steps. In an empirical study, the fit of the model (1a) was found to be so good (average error around 3%), that there was no need for testing the interaction hypothesis (1). This result will shortly be discussed in section 7.

4. COMPARISON WITH THE SHIFT-SHARE APPROACH

There is a strong similarity between the shift-share-technique and the above presented approach. Broadly speaking, the shift-share-analysis computes to which extent changes in employment, income, outgoings etc. in a region can be attributed to the industrial composition of the region (structural effect), and what part of its change is due to the fact that the industries in these regions grow faster or slower than their national counterparts (differential shift). The latter contribution is often interpreted as a measure of the locational advantages or disadvantages of the region (RICHARDSON 1978, p. 18). Some investigations have even traced growth and decline of specific regions back to specific sectors. For most problems appertaining to the shift-share-approach, it is irrelevant whether any growth rates or other variables are considered. If, instead of the industrial composition, the settlement structure of the region is assumed to have major influence, then the goals of the shift-share-approach will equal the problems that have been raised in the previous sections.

The most common formulation of the shift-share-approach reads.

$$E_{ij}^{1} - E_{ij}^{\circ} \equiv E_{ij}^{\circ}\, r_{++} + E_{ij}^{\circ}\,(r_{i+} - r_{++}) + E_{ij}\,(r_{ij} - r_{i+}) \qquad (3)$$

The super scripts 1 and 0 refer to the terminal period and the beginning period, respectively. E_{ij} is the employment in industry i and region j whereas

$$r_{++} = \frac{E_{++}^{1} - E_{++}^{\circ}}{E_{++}^{\circ}}$$

$$r_{i+} = \frac{E_{i+}^{1} - E_{i+}^{\circ}}{E_{i+}^{\circ}} \qquad (4)$$

$$r_{ij} = \frac{E_{ij}^{1} - E_{ij}^{\circ}}{E_{ij}^{\circ}}$$

are the respective growth rates

and

$$E_{++} = \sum_{ij} E_{ij} \quad , \quad E_{i+} = \sum_{j} E_{ij} \qquad (5)$$

is used for the initial as well as for the terminal period (HERZOG 1977, p. 442).

Equation (3) decomposes the increase in employment in industry i and region j into a component that would equal $E_{ij}^{1} - E_{ij}^{\circ}$ if the industry i in region j would grow at the same rate r_{++} as all industries in all regions grow on average (national component). The second term in (3) expresses the consequences of the fact that industry i might grow nationally at a different rate than the average of all industries and respects, at the same time, the different regional shares of that industry (structural component).

The first and second terms express the expected regional employment change in industry i, and take into account the share of that industry in the region. It is also assumed that the regional growth rate equals the national growth rate. The third term accounts for the fact that the regional growth rate of an industry might differ from its national counterparts (differential or competitive component).

The third residual term claims to be free of influences from the regional industrial structure and therefore expresses, in our former terminology, the "effect" of the region. ESTABAN-MARQUILLAS argues that the differential component $E_{ij}\,(r_{ij} - r_{i+})$ will still be biased by the industrial structure, because the differences $r_{ij} - r_{i+}$ are weighted by the industrial employment structure E_{ij}. He therefore splits the term into a standardized 'homothetic'-term and a remaining 'allocation-effect' (ESTABAN-MARQUILLAS 1972, p. 249; HERZOG 1977).

To compare (3) with our previous approach, the equation will be modified into a form where it does not matter whether we analyze growth rates or other variables:

$$\frac{E_{ij}^{1} - E_{ij}^{\circ}}{E_{ij}^{\circ}} = r_{ij} = r_{++} + (r_{i+} - r_{++}) + (r_{ij} - r_{i+}) \qquad (6)$$

Equations (1c) and (6) are formally equivalent, if the last term in (6) is interpreted as a residual. A weakness of this comparison is that the statistical properties of a residual are not necessarily met by the differential component in (6).

In (6) we can interchange r_{ij} by any other regional indicator that depends on industry i and region j. Equation (4) is changed into

$$r_{++} = \sum_{ij} r_{ij} \qquad\qquad (7)$$
$$r_{i+} = \sum_{j} r_{ij}$$

Equation (7) is written without weights. However, if necessary the weights can be reintroduced.

In equation (6) a major weakness of the shift-share approach becomes obvious: Industrial and regional structure are treated in an asymmetrical manner, e.g. the regional influence is not explicitly considered at all. A regional shift is only explained by the industrial structure of the region as compared to the national industrial structure. The remainder is treated as residual. However, the performance of the region as a whole also influences the growth rate, and this effect should be treated separately from the residual.

The example shown in Figure 1 can be used in a slightly modified way to illustrate this point:

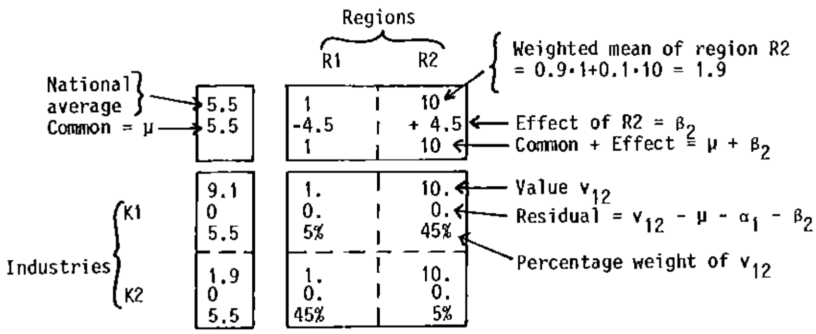


Figure 2: Estimation of regional and sectoral effects in a nonorthogonal case

The growth in region 1 will be expressed by (6) as:

Region 1 Sector 1

$$1 = 5.5 + (9.1 - 5.5) + (1 - 9.1)$$
$$= \underset{\text{national}}{5.5} + \underset{\text{structural}}{3.6} - \underset{\text{differential}}{8.1}$$

Region 1 Sector 2

$$1 = 5.5 + (1.9 - 5.5) + (1 - 1.9)$$
$$= \underset{\text{national}}{5.5} \quad \underset{\text{structural}}{-3.6} \quad \underset{\text{differential}}{-0.9}$$

From the structure of the data it is much more probable that region 1, as an economic system, tends to let all industries grow by just 1. This is a common feature of the industries in region 1. Equation (6) does not account for this kind of reasoning, whereas equation (1a) treats the influence of industrial and regional structure in a symmetrical way.

As soon as this becomes clear, either the shift-share formula could be adapted to (1a) or the settlement structure could be treated in a similar way, as are the industries in the shift-share-approach. Both possibilities will be discussed shortly in the next two sections. However, a decision between these possibilities can only be made after more extensive theoretical discussions and empirical applications have been attempted.

5. AN EXTENSION OF THE SHIFT-SHARE-CONCEPT

There are two main arguments that call for an extension of formula (6):

- The unequal formal treatment of the effects of the industries as a whole on the national level, on the one hand; and the region as a whole, comprised of all industries, on the other hand. Why should a specific industry in a specific region depend more on that industry at the national level, than on other industries in the same region?

- The residual in (6) ($r_{ij} - r_{i+}$) explains how much a region's growth differs from that industry's national average growth. However, it does not explain how much of this difference may reasonably be attributed to the growth of the region as a whole.

A straightforward expansion of (6) would read

$$r_{ij} = r_{++} + (r_{i+}-r_{++}) + (r_{+j}-r_{++}) + ((r_{ij}-r_{i+}) - (r_{+j}-r_{++})) \qquad (7)$$

where r_{i+} and r_{+j} are symmetrically employed (HOPPEN 1978, p. 181).

The first two terms on the right-hand side are identical to (6). The third term expresses the effect of the region, and the fourth residual term is similar to (6). The fourth term equals the difference between the actual growth of industry i in region j and the national average of that industry minus the regional effect. Equation (7) equals precisely the solution of an analysis-of-variance approach in the orthogonal case.

The discussion of nonorthogonal analysis of variance problems does show, however, that equation (7) is not an appropriate formulation. The regional and industrial effects are interrelated because of unequal 'cell'-frequencies. The nonoptimality of (7) becomes more clear if (7) is interpreted in terms of the general linear model, where the last term equals the residual and the first three terms are parameters to be estimated. The example shown in Figure 2 illustrates this points:

$$\begin{aligned} r_{11} &= 5.5+(9.1 - 5.5) - (1 - 5.5) + ((1 - 9.1) - (5.5 - 1)) \\ &= 5.5 + 3.6 - 4.5 - 8.1 + 4.5 \end{aligned}$$

The optimal use of the data would result in a zero industrial structure effect and a -4.5 regional effect.

Equation (7) should therefore be generalized as

$$r_{ij} = \mu + \alpha_i + \beta_j + c_{ij} \qquad (8)$$

where r_{++} has been replaced by μ, $r_{i+} - r_{++}$ by α_i, $r_{+j} - r_{++}$ by β_j and the residual $((r_{ij} - r_{i+}) + (r_{++} - r_{+j}))$ by e_{ij}. Equation (8) should be completed by expecting

$$\sum_i \alpha_i = \sum_j \beta_j = 0 \qquad (9)$$

Equation (8) can be estimated as in equation (1a) above.

The extended shift-share (7) and the ANOVA-approach (8) differ in the extend of variance that is attributed to the independent variables. In (8) only the differences with respect to the mean values of regions and industries are considered, whereas the influence of industrial structure on the mean of regions i.e. the interaction of regions and industries is neglected. This may lead to distorted results, as the example proves. The ANOVA-approach accounts for that interaction and explains thereby as much variance as possible. There is a difference in both philosophies, but the shift-share approach in its extended traditional version (7) is unable to combine structural and regional aspects in a sensible way. Empirically this has been proved by HOPPEN (1978, p. 179).

6. SETTLEMENT VERSUS REGIONAL EFFECTS BY A SHIFT-SHARE-APPROACH

The shift-share-approach is now used to correct the value of a specific type of settlement in a specific region, from the influences of the settlement-structure of the region as a whole. It is a reasonable assumption to identify types of settlements with industrial sectors:

$$r_{ij} = r_{++} + (r_{i+} - r_{++}) + (r_{ij} - r_{i+}) \qquad (6)$$

Equation (6) in this context has the following interpretation: The value of settlement type i in region j can be decomposed into three terms, one of which can be deduced from the overall national mean and one that is caused by the regional settlement structure itself, e.g. because settlements of type i differ even at the national level from the national average of all settlement types. The last term gives the difference between the value of settlement type i in region j from the national average of settlements of this type. The argument of using (6) in this context is the same as in the former case; whereas the structural component can be interpreted and computed very straightforwardly. The price one has to pay for this is that regional influences cannot be dealt with using this approach. In a regional policy context, equation (1a) should be preferred because of its broader explanatory power.

7. EMPIRICAL RESULTS

In an empirical study of the urbanization of the Netherlands, many different socioeconomic variables have been investigated by means of the structural-regional-taxonomic system and a subsequent analysis of variance has been made applying model (1a). The study identified six settlement types (A = rural communities, B12 = urbanized rural communities, B3 = commuter communities, C12 = small towns, C34 = medium-sized towns, C5 = big towns) in four regions of the country (north, east, west, south). For most variables investigated, model (1a) gave a good fit with a relative mean percentage error of around 3%. As an example, Figure 3 shows the original values in the large double-frame, and below the relative deviation of the fit in percent, whereas the rest of the table is structured like Figure 1.

In Figure 4, a graphical representation of the values are given by means of a three-dimensional plot, where the ground-plan represents the independent variables and the height shows the respective values of the variable. As an example, the income per head of population in 1971 is shown in relation to the national average.

8. GENERALIZATION OF THE STRUCTURAL/ REGIONAL-APPROACH AND OTHER APPLICATIONS

The previously discussed problems show the following common features: The units of observation (i.e. regions) consist of subunits, which can be classified into groups that have different properties. In the settlement-case, the subunits have been the individual communities, whereas in the shift-share-case they have been the individual firms. Many other scientific fields are concerned with the very same problem:

the geologist with different tectonic structures,
the biologist with different crops,
the epidemologist with different diseases,
the sales-manager with different household-characteristics,
the labour-market researcher with different sex-and-age structures.

Almost no field of research can be imagined, where in the process of regional disaggregation, a problem of this type does not arise and challenge the intellect of the researcher.

Several distinct situations can be distinguished:

1) All values v_{ijk} are known. The questions are then: How great is the influence (effect) of the region as compared to the influence of the structure, and what model describes the relationship between region and structure in an appropriate way? (Compare with page 8).

2) All values v_{ij} are known. The question is the same as in 1) above, but the model adequacy can hardly be tested.

3) Only the marginals v_{i+} and v_{+j} are known. Given the relative weights g_{ij}, what is the best estimation of v_{ij}?

4) Only the regional averages v_{+j} and the weights g_{ij} are known, and structural influences are suspected or expected. Can the structural effect be confirmed and its size be estimated?

5) Same as 4), only structure averages v_{i+} are known and regional effects are suspected. Is there a regional effect and what is its size?

if no structure properties are considered (GRIFFITH 1978; GREER-WOTTEN 1972).

Obvious extensions of the basic question arise, if the basic subunits are classified by two interfering structural properties. Another extension could refine the spatial dimension by introducing a multi-level-spatial hierarchy. MOLLE compared the EC-regional values with their national average,

Average percentage error = 2.17%

		North Region	East Region	West Region	South Region	
National Average	1.00	0.872	0.903	1.120	0.876	Mean values of the regions
μ	1.00	0.076	-0.067	0.085	-0.097	β_j
		0.923	0.932	1.086	0.902	$\mu + \beta_j$
Mean	0.86	0.808	0.808	0.975	0.804	A-type-settlements
α_1	-0.11	-0.200	-0.323	0.352	1.882	
$\mu + \alpha_1$	0.886					
Mean	0.844	0.825	0.814	0.955	0.810	B12-type-settlements
α_2	-0.108	1.180	-1.247	-2.354	1.967	
$\mu + \alpha_2$	0.891					
Mean	1.155	1.135	1.046	1.203	0.951	B3-type-settlements
α_3	0.111	8.813	0.164	0.445	-6.585	
$\mu + \alpha_3$	1.111					
Mean	0.927	0.888	0.918	0.978	0.868	C12-type-settlements
α_4	-0.056	2.363	4.595	-5.154	2.536	
$\mu + \alpha_4$	0.943					
Mean	0.996	0.901	0.969	1.062	0.923	C34-type-settlements
α_5	-0.003	-2.070	4.148	-1.939	2.615	
$\mu + \alpha_5$	0.996					
Mean	1.121	0.970	0.991	1.198	0.950	C5-type-settlements
α_6	0.091	-4.646	-3.297	1.694	-4.664	
$\mu + \alpha_6$	1.092					

Figure 3: Per Capita Income of groups of Dutch communities in 1971, the logical structure compares to figure 1.

Although much has been said on this problem in individual cases, regional science or geography has not yet picked up the general problems. The current spatial ANOVA-applications are mainly addressed to the question of the appropriate scale of analysis (in quadrat-analysis the GREIG-SMITH Scheme: THOMAS 1980, p. 253; for hierarchically nested regions: MOELLERING and TOBLER 1972), or to spatial sources of variation, as well as with the communities mean (MOLLE 1980, p. 260).

The general solution is certainly beyond the scope of this paper. However it is felt that a lot of questions still remain to be answered.

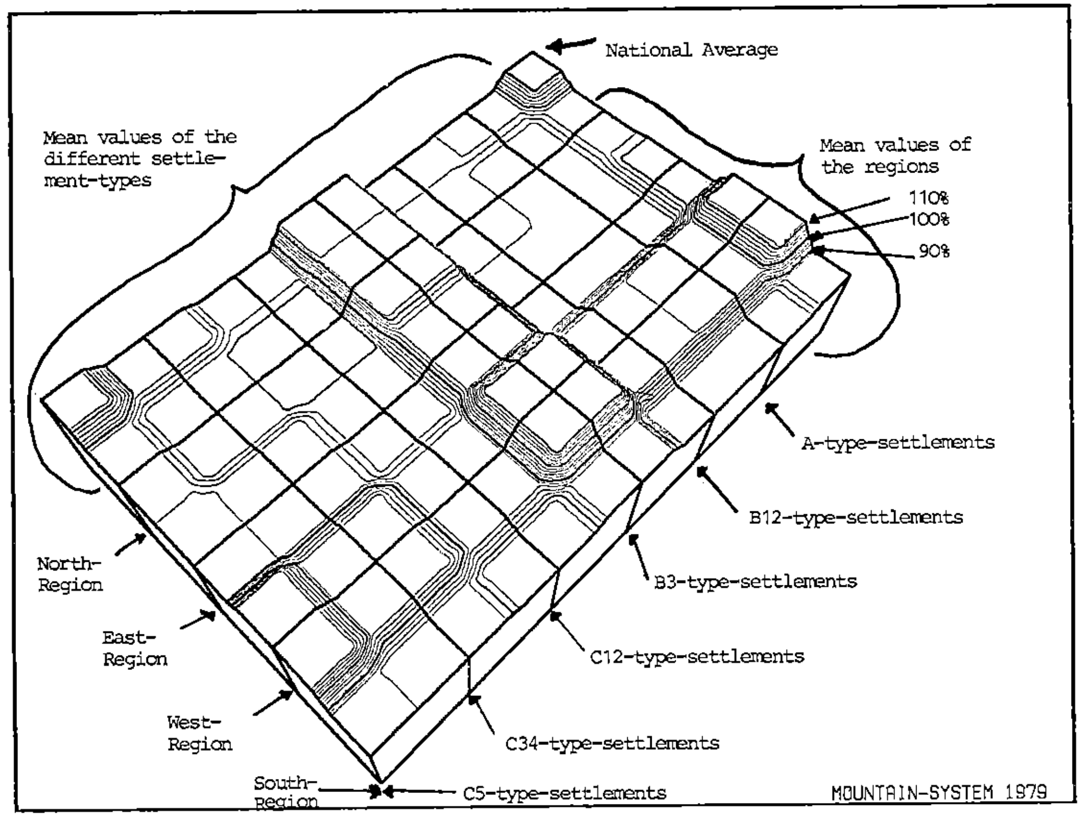


Figure 4: Graphical representation of the per capita incomes of Dutch communities in 1971 as analysed in Fig. 3.

REFERENCES

1 APPELBAUM, M.I.; CRAMER, E.M.: Some Problems in the Nonorthogonal Analysis of Variance, Psychological Bulletin, 81, 1974, pp. 335-343.

2 ESTABAN-MARQUILLAS, J.M.: A Reinterpretation of Shift-Share Analysis, Regional and Urban Economics, 2, 1972, pp. 249-261.

3 GREER-WOTTEN, B.A.: A Bibliography of Statistical Applications in Geography, Washington D.C., A.A.G. Commission on College Geography, Technical Paper No. 9, 1972.

4 GRIFFITH, D.A.: A Spatially Adjusted ANOVA-Model, Geographical Analysis, 10, 1978, pp. 296-301.

5 HERZOG, H.W.; OLSEN, R.J.: Shift-Share Analysis Revisited, Journal of Regional Science, 17, 1977, pp. 441-454.

6 HOPPEN, H.D.: Regionale Sektorprognosen, Raumforschung und Raumordnung, 1978, pp. 179-185.

7 HUGHES, J.T.: An Urban Approach to Regional Problems. From: Regional Policy, MACLENNAN, D. (Ed.), Glasgow Social and Economic Research Studies, 6, Oxford, 1979.

8 MOELLERING, H.; TOBLER, W.: Geographical Variances, Geographical Analysis, 4, 1972, pp. 34-50.

9 MOLLE, V.: Regional Disparities and Economic Development in European Community, Saxon House, 1980.

10 RICHARDSON, H.W.: The State of Regional Economics : A Survey Article, International Regional Science Review, 3, 1978, pp. 1-48.

11 THOMAS, R.W.; HUGGETT, R.: Modelling in Geography, Harper & Row, London, 1980.

12 TIMM, N.H.: Multivariate Analysis, Brooks/Cole, Monterey, Cal., 1975.

13 TUKEY, J.W.: Exploratory Data Analysis, Addison-Wesley, 1977.

G O N T E R   L Ö F F L E R

A MULTIVARIATE ANALYSIS OF THE RELATIONSHIP BETWEEN DIFFERENT SPATIAL STRUCTURES - SOME ASPECTS OF DEVELOPING REGIONAL SPACE-(TIME-)MODELS OF SWEDISH AGRICULTURE

From:

BAHRENBERG, GERHARD und STREIT, ULRICH (Eds.):
German Quantitative Geography. Papers presented at the 2nd European Conference on 'Theoretical and Quantitative Geography' in CAMBRIDGE 1980. Münstersche Geographische Arbeiten 11. Schöningh, Paderborn 1981, p. 91 - 104

Author:

Dr. Günter Löffler
Universität Trier
FB III
Postfach 3825
5500 Trier

ZUSAMMENFASSUNG

Wenn das Ergebnis einer Regionalisierung oder einer Clusteranalyse von Raumeinheiten mit bestimmten Merkmalen als generalisierte Beschreibung der Raumstruktur aufzufassen ist, dann existieren für verschiedene Merkmale oder Merkmalsgruppen unterschiedliche Raumstrukturen. Unter der Annahme, daß die Existenz kongruenter Raumstrukturen, abgeleitet aus verschiedenen Merkmalsgruppen, den Schluß auf "kausale" oder "funktionale" Zusammenhänge zulassen, kann ihre Analyse das räumliche Beziehungsgefüge der verwendeten Merkmalsgruppen aufzeigen. Die einzelnen Raumstrukturen lassen sich quantitativ durch Distanz-/Ähnlichkeitsmatrizen beschreiben, wie sie aus den unterschiedlichen Merkmalsausprägungen der Raumeinheiten zu berechnen sind. Der Grad der Kongruenz zweier Raumstrukturen kann unter Verwendung aller Distanzen mittels des "cophenetic correlation coefficient" nach SOKAL und ROHLF (1962, 33-40) bestimmt werden, der von ihnen zum Vergleich allgemeiner Clusterstrukturen verwendet wurde. Dieser Korrelationskoeffizient für Distanzen oder Ähnlichkeiten entspricht dem Produktmoment-Korrelationskoeffizienten nach Pearson und erlaubt es daher, alle auf diesem aufbauenden statistischen Verfahren ebenso anzuwenden.

Die so verstandene Raumstrukturanalyse wird im ersten Teil des Beitrages herangezogen, um das Beziehungsgefüge verschiedener Variablengruppen aus schwedischen Agrarerhebungen und Volkszählungen sowie solchen über die naturräumliche Ausstattung zweier Untersuchungsgebiete in Mittel- und Südschweden in Form von Pfadmodellen zu quantifizieren.

Am Beispiel der Pfadmodelle für die Erhebungsjahre 1960/61 und 1975/76 des südschwedischen Untersuchungsraumes werden verschiedene Entwicklungstendenzen und Strukturveränderungen aufgezeigt. Anschließend erfolgt ein quantitativer intra- und interregionaler Vergleich der für jeweils 4 Erhebungszeitpunkte in Mittel- und Südschweden berechneten Pfadmodelle.

Im zweiten Teil des Beitrages werden einzelne Aspekte der weiteren Anwendung der Raumstrukturanalyse anhand kurzer Beispiele erläutert. So erlaubt die gleichzeitige Verwendung von Distanzmatrizen (Unterschiede zwischen den Raumeinheiten) und kilometrischen Distanzmatrizen oder Nachbarschaftsmatrizen - alle drei sind in ihrem Aufbau identisch - reichweitenorientierte und gerichtete Raumanalysen durchzuführen. Dabei werden Fragen nach den unterschiedlichen Reichweiten räumlicher Erhaltensneigung von Variablengruppen, nach dem Auftreten räumlicher Kreuzkorrelationen zwischen ihnen und nach gerichteten und ungerichteten Reichweiten von Einflüssen und Zusammenhängen tangiert, um einen Einblick in die weiteren Einsatzfelder des als Raumstrukturanalyse bezeichneten Prinzips der Verwendung von Distanzen zu geben.

1. INTRODUCTION

If we characterize the result of a regionalization or cluster-analysis of choristically defined criteria as a generalized description of the spatial structure, there is a number of spatial structures of different significance corresponding to the various theoretical concepts which are operationalized by groups of variables. Distance-matrices used for a lot of classification strategies, that take into account the differences between spatial units, can be regarded as quantitative descriptions of these spatial structures.

Similar to BERRY's field theory (1969, 419-430) in which the relationship between such spatial structure and an interaction structure, operationalized by a distance-matrix and an interaction-matrix, is tested, we are able to measure the relationship between two different spatial structures by computing the "cophenetic-correlation-coefficient" r_c (SOKAL and ROHLF, 1962, 33-40) and making use of distance-matrices (LÖFFLER, 1981).[1] When measuring the degree of spatial congruence of two spatial structures in this way we may ascertain whether a correlation exists between them or not. As the "cophenetic-correlation-coefficient" can be handled in the same way as the product-moment-correlation-coefficient we are allowed to use it for the same statistical techniques. The application use of differences or distances between objects or spatial units is not new. Besides BERRY who - as far as I know - made use of it in Geography for the first time this strategy is normally used in the social sciences in order to compute correlation-regression- and pathcoefficients (HOLM, 1977/ LEITNER and WOHLSCHLÄGL, 1980, 81-106), with regard to non-metric data. In geography the application of differences between variables of spatial units means that we do not only compare the variability of two phenomena in all units arranged like the pearls on a chain, but we also compare both spatial structures of the phenomena as they are described in the distance matrices.

2. THE ANALYSIS OF SPATIAL STRUCTURE

2.1 DATAS AND SAMPLE AREAS

The objective of the study described here is an analysis of the influence of urban agglomerations on the agrarian structure of their "umland". On the basis of the official statistics, an attempt was made to explain the influence of urban development on agrarian structure. The data used for this purpose were those available at the communal level for the years 1950/51, 1960/61, 1965/66, 1970/71 and 1975/76 for two regions of Sweden, the Mälar region with Stockholm and Southern Sweden with the agglomeration Malmö/Lund.

The first step was to make all available variables comparable over time and space, with regard to both their definitions and to the frequently changing administrative boundaries. Then the variables were tested as to their applicability to the operationalization of the theoretical concepts agrarian structure and urban agglomeration to data with the same definition for all of the years, the number of variables was still 68. Therefore, in a further step they were combined into groups with the same or complementary substance. In addition to the groups of variables that describe the natural environment, such as climate, quaternary deposits and soil associations, which were included in the analysis, in order to quantify these influences, which can be regarded as constant in their spatial variation between 1950 and 1976 there are nine groups of variables. Since the definition and quantitative description of agrarian structure and urban agglomeration are not objectively clear, no attempt was made to operationalize these two concepts. Instead, subconstructs were chosen on the basis of the definitions used in the official statistics. Fig. 1 shows the hierarchy of theoretical concepts and the operationalization at the third level.

Before the actual analysis could begin, the data had to be standardized, because in the chosen spatial reference system (administrative divisions as of 1.1.76) the individual communes vary in size and form. After standardization 12 comparable sets of data were available for the further analysis, multiplied by the number of census years.

2.2 THE METHOD

In the search for an appropriate method of statistical analysis, which will eventually lead to insights into the influences of urban agglomerations on the agrarian structure, all known multivariate approaches had to be rejected,because they turned out to be of little value for this complex problem. Multiple regression and the computation of path models would have been suitable for solving the problems. However, they could not be used, because the dependent variables are defined multivariately, and the large number of individual variables (all are metric) would have led to an overproduction of information, which cannot easily be analyzed. The familiar procedures for variable reduction (cf. JOHNSTON, 1980), such as principal component analysis, factor analysis or canonical correlation analysis, were purposely not used, because we did not want to replace the groups of variables, which were derived on the basis of theoretical considerations, with mathematically derived factor scores, which are often difficult to interpret.

I) AGRARIAN STRUCTURE

II)	FARM SIZE	OWNERSHIP OF ARABLE LAND	USE OF AGRI-CULTURAL LAND	CROP PRODUCTION	LIVESTOCK	STRUCTURE OF EMPLOYMENT IN AGRICULTURE
III)	NUMBER OF FARMS WITH	ha ARABLE LAND HELD IN LEASE	CULTIVATED NATURAL PASTURES	USE OF ARABLE LAND WITH CROPS OF DIFFERENT KINDS	NUMBER OF LIVESTOCK	NUMBER PER ha ARABLE LAND
	1) 2 - 5 ha 2) 5 -10 ha 3) 10 -20 ha 4) 20 -30 ha 5) 30 -50 ha 6) 50-100 ha 7) > 100 ha arable land in % of all farms >2 ha	1) in % of all arable land	1) in % of arable land	1) winter wheat 2) spring wheat 3) winter rye 4) spring rye 5) barley 6) oats 7) mixed grain 8) oleiferous plants 9) potatoes 10) sugar beets 11) pasture on arable land 12) plants for silage 13) ley 14) fallow land crops in % of arable land	1) horses 2) cows 3) heifers, steers 4) calves (younger than one year) 5) pigs 6) sheep per 100 ha arable land	1) male, 2) female: employers 20-w hours per week 3) male, 4) female: employees 20-w hours per week 5) male, female: family workers

I) URBAN AGGLOMERATION

II)	STRUCTURE OF GENERAL LAND USE	STRUCTURE OF GENERAL ECONOMIC ACTIVITIES	SPATIAL POPULATION STRUCTURE
III)	LAND USE	EMPLOYEES, EMPLOYERS 20-W HOURS PER WEEK	POPULATION IN TÄTORTER (SETTLEMENTS)
	1) arable land 2) cultivated natural pasture 3) settlements (tätorter) in % of all land	1) male, 2) female: in % of the whole population 3) male in agriculture, 4) female in agriculture: in % of all economically active people	1) % of population in tätorter 2) number of people per km² in tätorter

I) NATURAL ENVIRONMENT

II)	CLIMATIC STRUCTURE	STRUCTURE OF SOIL QUALITY	
III)	rainfall and temperature	quaternary deposits in % of the commune	soil associations in % of the commune (according to FAO classification)
	A) average rainfall 1) April 2) May 3) June 4) July 5) August 6) September 7) year B) temperature 8) min. (annual) 9) max. (annual) 10) number of days with more then 5°C per year	1) Bogs and swamps 2) Clay, silt and very fine sand 3) Sand, coarse fine sand and fluvial sediments 4) Glacifluvial sediments 5) Moraine clay 6) Moraine 7) Exposed bed rock	1) organic soil (O) 2) podzolized and organic soils (P/O) 3) podzolized soils (P) 4) Gray-brown podzolic soils and podzolized soils (Gb/p) 5) Gray-brown podzolic soils (Gb) 6) Lithosols, rankers and pod-solized soils (Li/Rh/P) 7) podzolized soils, rocky, stony or shallow (P(R))

I) theoretical constructs   II) theoretical subconstructs

III) variables for operationalizing the theoretical subconstructs for all years at the communeal level (1.1.1976)

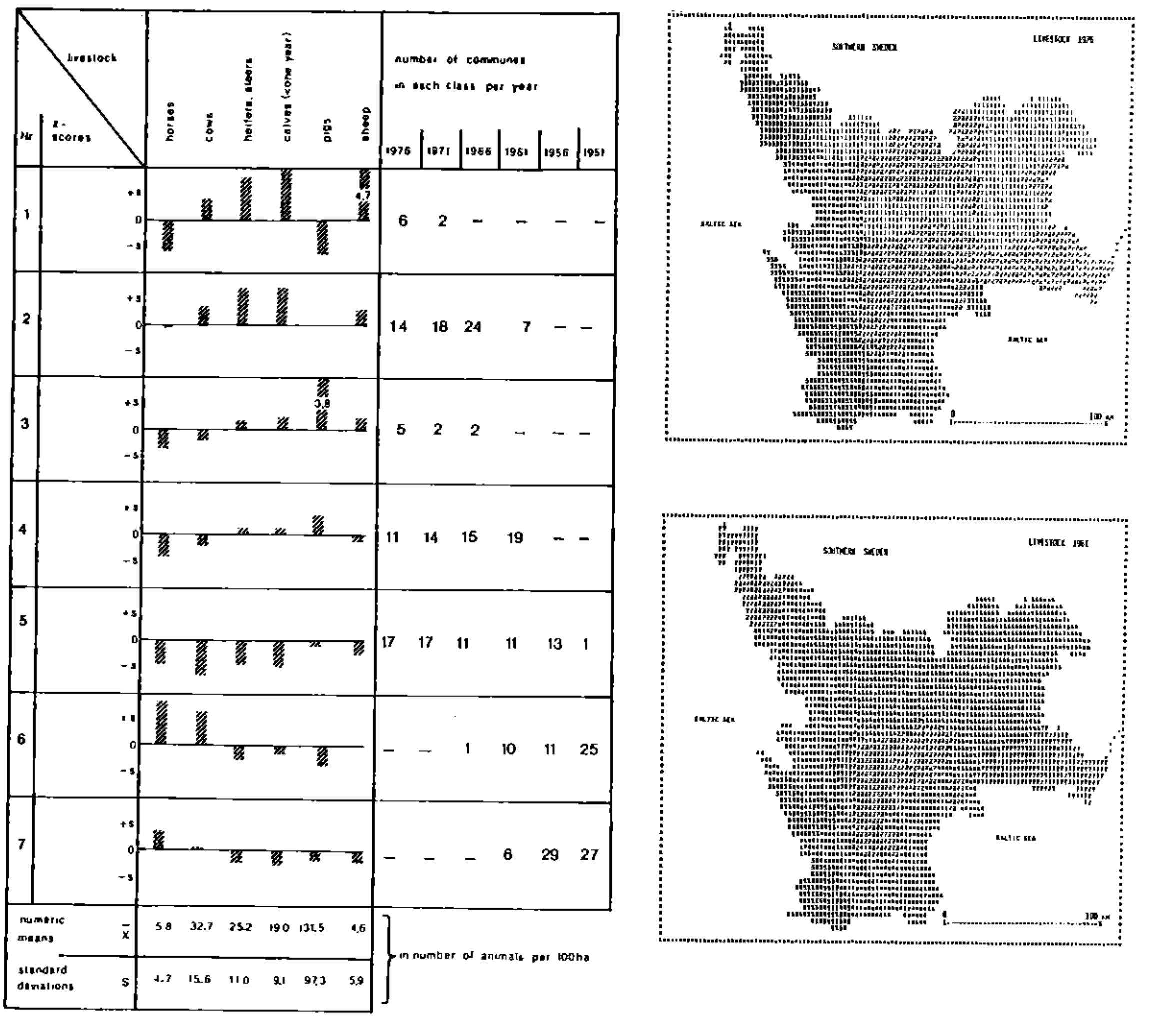
Fig. 1 The hierarchy of theoretical constructs and their operationalisation levels

As described in the introduction, the objective was to demonstrate the spatial structure of the groups of variables and then to reason on the basis of their spatial congruence. The first step in this direction was to classify the communes multivariately with respect to each group of variables.

In order to make the changes in the spatial structure of the individual groups of variables "over time" visible, the cluster analysis [2] was computed not only for one year, but for all years. This procedure led to different spatial structures for each year. The substance of each spatial class (region), however, remained comparable. This can be seen in Fig. 2 for the spatial distribution of livestock, using 1961 and 1976 as examples. Fig. 2 clearly shows the differences in the distribution of the spatial classes (regions) formed on the basis of the type of livestock. In all three figures the numbers of the classes are identical. We used the z-scores, which show the positive and negative deviations form the numeric means, to describe the content of each class.

The spatial structure of the groups of variables is derived from the distance matrix D and is generalized by the classification procedure. Thus, the use of this distance matrix $\underset{\sim}{D}$ for the quantitative description of the original spatial struc-

Fig. 2 : The use of Cluster Analysis over space and time



ture suggests itself, whereas the generalized spatial structure can be described by a generalized distance matrix. The generalization of the distance matrix is carried out by means of the "expanded ultrametric transformation" as it is designated by the author (cf. LÖFFLER, 1979/1981) [3]. The purpose of the generalization is the detailed substantive interpretability of the computed results, as will be explained in detail later on.

Because after the classification is concluded, all years are contained in a distance matrix D of the size

$$Tn \cdot 1/2 \cdot (Tn-1), \text{ with } T= \text{no. of years}, \; n= \text{no. of cases}$$

before the further analysis, T matrices of the size $n \cdot 1/2 \cdot (n-1)$ are extracted. These distance matrices describe the spatial structure of the 12 groups of variables at different times [4].

The distance matrices of a group of variables for each year are perceived as a quantitative description of the spatial structure. Accordingly, they are used for computing equations of multiple regression from which a totally recursive path model is computed according to a prepostulated theoretically based causal order. Although it is disputable from a methodological point of view (cf. HOLM, 1977, p. 57), the theoretically excludable paths were eliminated from this regional path model of the relationships between urban agglomeration, agrarian structure and natural environment. A total of 33 of 66 possible paths were considered to be potentially existent. Because a test of normal distribution of the distances was only positive in 50 % of all cases, and thus no statistical inferences can be made, the decision which path coefficients could be considered relevant involved a compromise between the computed significance and the absolute value of the coefficients (cf. HOLM 1977, p. 57).

Fig. 3 shows the results for the years 75/76 and 60/61 in the form of a diagram. Before the two models are explained and the differences between the two years are discussed, we must investigate the interpretative value of this model. Thus a standardized path coefficient of +0,592 between the spatial structures of the groups of variables VG_1 and VG_2 only tells us that the difference (distance of VG_2 between commune a and b) increases by + 0,592 when the distance of VG_1 increases by 1. We do not, however, know anything about the situation in commune a and b unless we look at the original data. If we have 58 (53) communes, we would have to analyze 1653 (1378) distances in a tedious procedure in order to come to direct substantive insights into the correlation whereas a general correlation can be inferred directly from Fig. 3.

The cluster analysis that we performed in order to generalize the spatial structures and the already mentioned expanded ultrametric transformation for the quantitative description of this simpler spatial structure now permit a renewed computation of path models.

These generalized path models are no longer as informative as those that were computed over the original distances. They can, however, be interpreted substantively much more easily, because the substance of the spatial classes upon which they are based is known (cf. Fig. 2). Since when we compute path models by means of the expanded ultrametric distances, we expect an extremely poor congruence of the spatial structures, this model by no means overestimates the relationships (cf. LÖFFLER, 1979/1981).

In the framework of this paper only a few bivariate relationships of our path model of the original distances are interpreted in somewhat more detail [5].

2.3 A BRIEF INTERPRETATION OF THE PATH DIAGRAMS 1976/75, 1961/60 IN SOUTHERN SWEDEN

Both diagrams are based on the same prepostulated orders of the groups of variables that were tested against alternative orders on the basis of the data for 1975/76. Since the results scarcely differed from those of the order chosen on the basis of theoretical considerations, it was retained for all years.

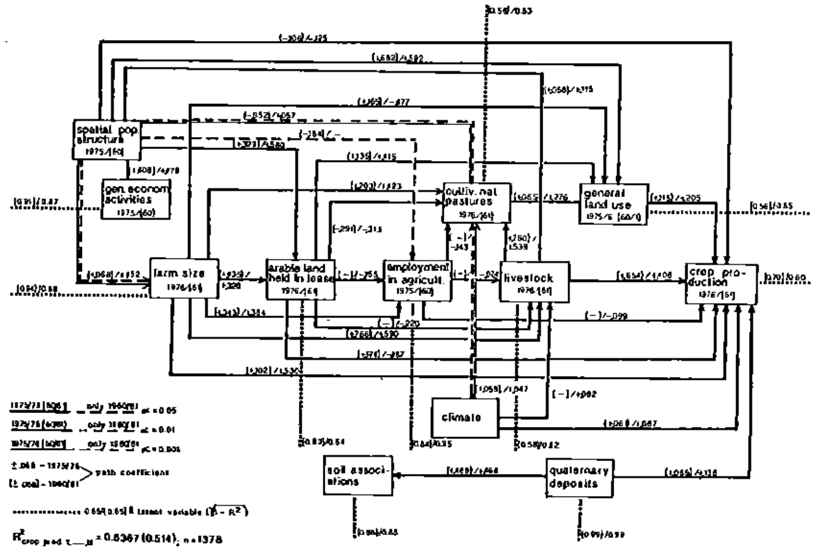


Fig. 3: Path diagram Southern Sweden 1975/76 and 1960/61

The correlation between the variables "population distribution and density" and "general land use" is + 0.682, i.e., a change of 1 unit in the intercommunal differences in"population distribution and density" causes a change of + 0.682 units in "land use". The decrease in the coefficient up to 1975/76 (0.592) can be explained by the increasing concentration of the population in urban areas along with an only slowly increasing or even stagnating demand for settlement area. In addition to a number of coefficients that remained constant between 1960/61 and 1975/76, changes can be determined in the area of agrarian structural components. Whereas in 1960/61 regional differences in "population distribution" hardly allow inferences as to "farm size" (+ 0.068), the coefficient of 0.452 15 years later shows an increase in significance. A decrease in the correlations between the regional differences in "farm size" and the proportion of "land held in lease" (from 0.435 to 0.320) and "livestock" (from 0.766 to 0.590, compared with 1960/61) is opposed to an increase in the correlation with "crop production" (from 0.102 to 0.530). This means that the diversity of crops produced depends increasingly on the size of the farms (specialization related to inputs of land, machinery and labour). On the other hand, "livestock production" is no longer based on "farm size", because of the increasing purchase of fodder (decrease of the influence of "livestock production" on "crop production" from 0.654 to 0.406). The proportion of "land held in lease" is no longer influenced as strongly by "farm size". A general tendency towards an increase in medium-sized and large farms can be deduced from the raw data, indicating that enlargement of farm size by leasing land depends increasingly on other factors. The regional differences in the proportion of "land held in lease" are influenced in 1976 much more strongly by those in the "distribution of population" (increase from 0.322 to 0.580). Whether a stronger tendency to give up farming and lease out land is a result of the "migration" from the country to towns or whether it sets in the immediate vicinity of the metropolitan areas can not be determined at this point.

2.4 A COMPARISON BETWEEN THE SAMPLE AREAS

On the basis of all path models of both sample areas, we can show that the differences between the two regions are grea-

Tab. 1 : An inter- and intraregional comparison of the sample areas

S.S./M.R.U.	Southern Sweden 75/76	70/71	65/66	60/61
75/76	R = .366 α = .018 CC = .684	R = .869 α = .001 CC = .880		
70/71	R = .628 α = .001 CC = .862	R = .395 α = .012 CC = .818	R = .786 α = .001 CC = .896	
65/66		R = .845 α = .001 CC = .885	R = .425 α = .007 CC = .719	R = .893 α = .001 CC = .918
60/61			R = .405 α = .010 CC = .739	R = .525 α = .001 CC = .847

Mälar region/Uppland

ter in each year than the differences between two consecutive years in the individual regions. This situation can be surveyed by a quantitative comparison of the path coefficients.

If we calculate the rank correlation coefficient (R) and, in addition, indicate the corrected contingency coefficient (CC) for the correspondence between the levels of significance of the path coefficients, certain intra- and interregional structural changes can be demonstrated (Table 1).

Although, as explained above, it is not completely permissible, the levels of significance were referred to in this comparison, in order to show the differences between the path models.

Although in general all correlations are significant their decrease from 1960/61 to 1975/76 in the interregional comparison points to two "causal models" for the variables used, which are continually developing apart. It appears that the influence of the urban agglomerations and the influences among the variables of agrarian structure in the sample area in middle Sweden increasingly differ from those in Southern Sweden. This tendency is not so obvious from the corrected contingency coefficients, because only the significance, but not the absolute value, of the path coefficients was used.

In the innerregional comparison of the individual years, the relatively low coefficient (R = .405) between 1960/61 and 1965/66 for the sample area in middle Sweden is striking. This "break" in the development of this area shows up in many analytical steps, which we cannot go into in detail here. Although it does not occur in the sample area in Southern Sweden, it can be conjectured that the cause is related to the change in Swedish agrarian policy during the sixties, which shows up in this clarity only here.

In general, both the intraregional and the innerregional differences can be shown only by a substantive interpretation of all path models [5], because a quantitative comparison with the procedures used here is too crude, and can only supply preliminary reference points.

3. THE RANGE-ORIENTED AND DIRECTED ANALYSIS OF SPATIAL STRUCTURE

All regional sciences are lacking in suitable methods of statistical analysis for computing the range and direction of influence of more or less located phenomena, which operate directed or undirected in space with varying ranges, such as the influence of urban agglomeration on agriculture in the "umland" in this study (cf. v. Thünen's model). Particularly in the last decade, research on problems of spatial "persistence" and the spatial autocorrelation of phenomena has increased considerably (e.g. CLIFF and ORD 1973/ NIPPER and STREIT 1977/78). The approach proposed here can be employed in this direction - and this is its advantage and strength.

As we showed in the first section, the concept is based on the use of distance matrices, which describe the differences between the communes with respect to the chosen groups of variables. These matrices of the form n · 1/2 · (n-1) are now in their structure directly comparable with matrices that contain the kilometric distances between all possible pairs of communes or the neighbourhood situations, i.e., the number of communes located between two given communes. Because of the identical structure of the two matrices, a number of analyses can be carried out, which approach the objective of quantifying ranges and directions of phenomena and their spatial persistence.

We can compute for a phenomenon the cophenetic correlation coefficients (r_c) between its distances (to be understood here as a description of spatial variability and the complete neighbourhood matrix $\underset{\sim}{W}$)[6]. A step by step reduction of both matrices beginning with the largest range or order of neighbourhoods permits to compute for n orders of neighbourhood (n-1) correlation coefficients, because the neighbour of the first order cannot be controlled by this coefficient. These coefficients enable us to reject or conform a hypothesis on the existence of spatial persistence for special ranges.

If, e.g., in the sample area in Southern Sweden there is k = 12 (k = order of neighbourhood) we obtain 11 coefficients, which are each based on the number of remaining pairs of cases.

It can be clearly seen from this table that the group of variables "spatial population structure" has the maximum spatial persistence with respect to the range of order <3 and <4, because there is no correlation with the neighbourhood matrices used.

For the group of variables "arable land held in lease" there is no indication of a spatial persistence. In general the strength of the correlation between substantive distances and kilometric distances - here neighbourhood relationships - increases with increasing range.

In our study the question of the spatial persistence of a group of variables was

Order of Neighbourhood (k)		<3	<4	<5	<6	<7	<8	<9	<10	<11	<12	<13
N. of cases		322	579	817	1003	1152	1245	1308	1348	1369	1376	1387
A	r	,033	,047	,091	,083	,157	,138	,154	,153	,163	,170	,167
	$\hat{z}$	0,55	1,25	2,57	2,50	5,52	5,00	5,36	5,56	5,93	6,30	6,30
B	r	,156	,135	,143	,148	,206	,181	,205	,217	,239	,239	,235
	$\hat{z}$	2,91	3,50	4,00	4,69	7,24	6,43	7,50	8,15	9,26	9,26	8,89

A = spatial population structure/B = arable land held in lease

$-1{,}96 < \hat{z} > +1{,}96 = \alpha < 0{,}05$

$\hat{z}$ ist independent of the number of cases[7] (SACHS 1974, p. 331)

Table 2: The range of spatial persistence

of less interest than the question of the correlations and the changes in them between 2 groups of variables at given ranges.

Our approach to the analysis of spatial structure shows the dependence of the correlations on the chosen scale of steps very well. Thus Table 3 shows the correlation between groups of variables "spatial population structure" and "structure of livestock" in Southern Sweden (1960/61) and (1975/76).

are based on the range of influence of chosen communes, as in the last example.

Before we go into the possibility of choosing fixed directions and ranges derived from the problem, the application of our approach to problems connected with spatial cross-correlation will be explained briefly. If our hypothesis is that a phenomenon at location A influences, causes, etc. a phenomenon at location B, a spatial cross-correlation can verify or disprove this hypothesis. If the hypothesis is that phenomenon a

ranges (k)		<2	<3	<4	<5	<6	<7
no. of cases		121	332	579	817	1003	1152
1975/76	r	-,177	-,105	-,089	-,028	-,000	+,090
	$\hat{z}$	-1,96	-2,00	-2,25	-0,86	0,00	+ 3,1
1960/61	r	-,090	-,100	-,010	+,070	+,070	+,140
	$\hat{z}$	-0,97	-1,82	-0,25	+2,00	+2,19	+ 4,83

Table 3: The dependence of the correlation between "spatial population structure" and "livestock" on the range chosen (Southern Sweden 1975/76 and 1960/61)

The dependence of the correlation coefficients on the chosen scale of spatial steps can be clearly recognized. This can be explained by the varying spatial "persistence of the variables" (cf. REYNOLDS 1974 and GEARY 1954) [8].

Since the spatial persistence of "livestock" is greater than that of the "spatial population structure" the correlation of the two leads to the change from negative to positive coefficients, which can be recognized in Table 3. The differrence between 1975/76 and 1960/61 is now based on the changes in the spatial persistence of the groups of variables in these 15 years.

The value of this computation is, however, mainly theoretical and its contribution to the understanding of the effects of spatial persistence is only a general one. From the substantive point of view, this application of our approach is only meaningful if particular hypotheses

correlates with phenomenon b with respect to the range k ($k = 1, \dots n$), the following correlation coefficients are computed: phenomenon a with b for all cases of the range or order $\leq k$, and each phenomenon with the neighbourhood relationships of the range/order $\leq k$. If the partial correlation coefficient between a and b is computed, a high partial correlation coefficient confirms the hypothesis that there is a spatial cross-correlation with respect to the chosen range or order $\leq k$, as the following example shows.

In the sample area in Southern Sweden the correlation coefficients between the distances of the groups of variables "crop production (1)", "spatial population structure (2)" and the elements of the neighbourhood matrix (3) are $r_{1,2}=0.14$, $r_{1,3} = + 0.37$ and $r_{2,3} = + 0,17$, when the complete matrices are used. The partial correlation coefficient $r_{1,2.3}$ has the much higher value of + 0.29 (n=1378).

This increase in the coefficient confirms the hypothesis that there is a spatial cross-correlation between the groups of variables. At which range a maximum can be found can be easily determined by testing all neighbourhood orders.

In our last example the hypothesis on the influence of urban agglomerations on the "umland" is tested by using the variables "spatial population structure" (V_1), "arable land held in lease 1976" (V_2), "farm size 1976" (V_3), "crop production 1976" (V_4) and "livestock 1976" (V_5). For this purpose a neighbourhood matrix was generated that includes the neighbourhood relationships of urban agglomerations with more than 30.000 inhabitants to those with fewer inhabitants. With the aid of these selected distances we can now test to what extent changes occur in the correlation with increasing distance from the urban agglomerations. For special problems, ranges of influence can also be tested along individual axes, etc.

Table 4 shows the results for the groups of variables mentioned above in Southern Sweden. Whereas the influence of the urban communes on the groups of variables "arable land held in lease" and "farm size" increases greatly with increased range, a definite influence on crop [9] and livestock production in all zones near cities with more than 30.000 inhabitants does not begin until the 7th, or respectively, 6th step of spatial range is included.

The results presented here cannot be compard directly with models such as v. THÜNEN's in this form, because ranges are not tested individually. Instead, changes resulting from an increase in the "radius" are sought. Since, however, the absolute strength of the correlation (z-scores) can change when the succeeding neighbours are included only if the correlation between the pairs of cases themselves has changed, inferences can be made indirectly as to the influences in individual zones around the urban agglomerations.

Thus we can assume that the differences (distances) in the proportion of "arable land held in lease" correlate highly with those of the spatial population structure", and that the strength of correlation increases with increasing distance. The substantive interpretation shows that a relatively high proportion is found in the immediate vicinity of these cities, and with increasing distance from the cities the proportion of arable land held in lease decreases. The increasing correlation with increasing distance between the differences in "farm size" and the "spatial population structure" means that

Order of Neighbourhood		<2	<3	<4	<5	<6	<7	<8	<9	<13
N. of cases		35	79	118	185	213	254	280	299	322
$V_{1/2}$	r	.471	.618	.745	.789	.800	.823	.838	.846	.835
	ẑ	2,91	6,26	10,27	14,45	16,06	19,96	20,30	21,34	22,64
$V_{1/3}$	r	.539	.508	.356	.340	.383	.422	.461	.427	.395
	ẑ	3,42	4,87	3,96	4,73	5,84	7,13	8,32	7,92	7,86
$V_{1/4}$	r	-.320	-.180	-.209	-.128	-.062	.069	.124	.158	.169
	ẑ	-1,88	-1,57	-2,25	-1,76	-0,88	1,10	2,00	2,75	3,18
$V_{1/5}$	r	-.272	-.070	-.056	.036	.089	.225	.279	.304	.313
	ẑ	-1,60	-0,61	-0,64	0,54	1,31	3,49	4,41	5,16	5,99

Table 4: Ranges of influence of cities with more than 30.000 inhabitants

with increasing distance from the cities the proportion of medium-sized and small farms increases.

The weak and, in part, very weak correlations between the 4th group of variables and the group of variables "spatial population structure" result from the varying spatial persistence. It is relatively high in the 1st group of variables; the communes surrounding the metropolitan areas are themselves fairly large settlements with relatively high population densities. However, the correlation with crop and livestock production, particularly in the vicinity of the chosen communes, is nonexistent. Thus large differences in group 4 occur simultaneously with smaller ones in group 1. Only at a certain distance are the differences in all groups approximately equal.

Unfortunately, a more detailed interpretation of all of these aspects would go beyond the limits of this paper. It remains to be mentioned that the strength of the correlation and the differences, which are in part only very slight, should not be interpreted absolutely. Instead, they must be seen relatively in the interregional and intraregional comparison for the entire 25 year period of study.

4. OUTLOOK

All of the examples given here demonstrate the usefulness of our approach as a descriptive statistical procedure for developing multivariate regional space-(time-) models. They also show possibilities for demonstrating ranges and directions of influence, spatial persistence and spatial cross correlations. Still, the analytical procedures outlined here certainly do not yet represent the optimum.

In the entire field of correlation analysis and multiple regression, we need to examine whether it would be better to replace Gauß's method of least squares for estimating parameters with the maximum likelihood method, and whether in the area of path analysis it would not be better to work with nonrecursive models. More clearcut criteria and investigation procedures need to be found for the comparison and selection of path coefficients. In addition, a large number of different case studies are needed to test the range of application of this method in regional sciences.

The reason for presenting our approach towards the analysis of spatial structure at this point is to stimulate a larger number of comparisons, to stimulate and assimilate criticism and impulses that may arise, and to present an alternative to the constant adaption of analytical techniques that were not actually developed for spatial problems.

FOOTNOTES

1)

$$r_c = \frac{\sum(d_{i,j} - \bar{d}_{i,j})\cdot(d^*_{i,j} - \bar{d}^*_{i,j})}{\left[\sum(d_{i,j} - \bar{d}_{i,j})^2\cdot\sum(d^*_{i,j} - \bar{d}^*_{i,j})^2\right]^{\frac{1}{2}}}$$

$$d_{i,j} \in \underset{\sim}{D} \; ; \; d^*_{i,j} \in \underset{\sim}{D}^*$$

2) A hierarchical procedure, according to WARD, with iterative relocation of the last 20 fusing steps was chosen. (Clustan 1^C, WISHART, Procedure WARD and Relocate).

3) In contrast to the conventional ultrametric transformation (cf. BOCK 1974), the expanded ultrametric transformation can also be used for nonhierarchical classification procedures.

4) The preceding analysis already took into account the changes over time. In the further course of the analysis, the factor time is thus disregarded, and all models are quantified only according to years in the form of spatial models.

5) An interpretation of all path models and a description of the changes in the relationships between the groups of variables of urban agglomeration, agrarian structure and natural environment in both regions will be published later on in full length.

6) The neighbourhood matrix $\underset{\sim}{W}$ contains all unweighted relationships $w^{(k)}_{i,j}$; with $k \geq 1$ and $i,j = 1,\dots,n$ (n = no. of spatial units)

(k= order of neighbourhood; number of boundaries which must be crossed when taking the direct (shortest) path between the spatial unit i and j).

7) $\hat{z} = \frac{\dot{z}}{s_{\dot{z}}} = \dot{z} \cdot \sqrt{n-3}$; with $\dot{z} = \frac{1}{2} \ln \frac{1+r}{1-r}$

8) Reynolds tested the existence of spatial autocorrelation for 2 variables and then for the residuals resulting from the regression. The reason he gave for the fact that both variables, but not the residuals, were spatially autocorrelated was an equivalent spatial autocorrelation of the variables.

9) It can be seen already in Fig. 1 that special crops are not included. Unfortunately, there were no figures available at the communal level.

REFERENCES

BERRY,B.J.L.: A syntesis of formal and functional regions using a general field theory of spatial behaviour. In: B.J.L. BERRY and O.F. MARBLE (ed.);Spatial Analysis, Prentice Hall, Englewood Cliffs, New Jersey, 1968, 419-430

BOCK,H.H.: Automatische Klassifikation. Göttingen 1974

CLIFF,A.D./ORD,J.K.: Spatial autocorrelation. London 1973

GEARY,R.C.: The contiguity ratio and statistical mapping. The Incorporated Statistician, 5,1954, 115-141.

HOLM,K.(Hrsg.): Die Befragung. Bd. 5, UTB 435. München 1977

JOHNSTON,R.J.: Multivariate Statistical Analysis in Geography, Longman, London,1980.

LEITNER,H. and WOHLSCHLÄGL,H.: Metrische und ordinale Pfadanalyse: Ein Verfahren zur Testung komplexer Kausalmodelle in der Geographie. Geographische Zeitschrift, 68. Jg., 1980, Heft 2, 81-106

LÖFFLER,G.: Ultrametrische Raumstrukturanalyse - ein methodischer Beitrag zur Quantifizierung des raumzeitlichen Beziehungsgefüges multivariat gebildeter agrar-geographischer Raumtypen. Kieler Geographische Schriften 50: 149-163

LÖFFLER,G.: Cluster-Strukturvergleiche - Wahl der Klassenzahl und "natürliche" Klassifikation. Theorie und quantitative Methodik in der Geographie, Tagungsband Zürich 1981, 26.-28.3.1980 (im Druck)

NIPPER,J./STREIT,U.: Zum Problem der räumlichen Erhaltensneigung in räumlichen Strukturen und raumvariaten Prozessen. Geograph. Zeitschrift, 65. Jg., 1977, 241-263

NIPPER,J./STREIT,U.: Modellkonzepte zur Analyse, Simulation und Prognose raum-zeit-varianter stochastischer Prozesse. Bremer Beiträge zur Geographie und Raumplanung, Heft 1, 1978, 1-17

REYNOLDS,D.R.: Spatial contagion in political influence processes. In:K.R. Cos, D.R. Reynolds and S. Rokkan (eds.), Locational Approaches to Power and Conflict, John Wiley, New York, 233-274

SACHS,L.: Angewandte Statistik (4th ed.). Berlin 1974

SOKAL,R.R. and ROHLF,F.J.: The comparison of dendrograms by objective methods. Taxon 11: 33-40, 1962

WISHART,D.: Clustan 1^C USER Manual. London 1975

K U R T E. K L E I N

ASPECTS OF TRAFFIC-ZONING IN INVESTIGATING INNER-URBAN CONSUMER TRAFFIC WITH THE AID OF INTERACTION MODELS

From:

BAHRENBERG, GERHARD und STREIT, ULRICH (Eds.):
German Quantitative Geography. Papers presented at the 2nd European Conference on 'Theoretical and Quantitative Geography' in CAMBRIDGE 1980. Münstersche Geographische Arbeiten 11. Schöningh, Paderborn 1981, p. 105 - 114

Author:

Dr. Kurt E. Klein

Institut für Geographie
Universität Regensburg

Universitätsstraße 31

8400 Regensburg

ZUSAMMENFASSUNG

Für einen Teil des innerstädtischen Einkaufsverkehrs gibt das Entropie-Maximierungsmodell von Wilson eine geeignete Erklärungsbasis ab. Die Verkehrsverteilungsmatrix wird durch die übertragbare Präferenzmatrix ersetzt: an ihr kann beispielhaft der Einfluß des Punktmusters auf die Modellergebnisse gezeigt werden. Um diesen zu minimieren, sind regelmäßige Punktmuster mit geringer Variation der Aufkommensgrößen Versand/Empfang gesucht. Ausgehend vom Hexagonalnetz läßt sich eine Transformation angeben, welche die Variation der Bevölkerungsdichte im innerstädtischen Bereich in den auszuwählenden Verkehrszellen minimiert. Regensburg dient als Anwendungsbeispiel für die vorgetragenen Überlegungen.

1. INTRODUCTION

For the succeeding considerations it is supposed that number and shape of traffic zones are to be determinated without restriction. In case of optimizing some basic pattern of traffic zones the works of Openshaw [14,15] could be referred.

Two main thoughts determine the following investigation:

(i) traffic zoning as regionalisation should square with the basic trip distribution hypothesis;

(ii) transfer of the results of model application should be granted; to put it in other words: minimizing the impact of point pattern which is formed by the centres of traffic zones and taken as a basis for model calibration.

ad(i). As there is no universal regionalisation for the purpose of traffic investigations important hints of how to choose and shape traffic zones can be drawn by careful investigation of the underlying trip distribution hypothesis or interaction model. As to those models with explicitly expressed hypothesis (e.g. route/time minimization) solutions like Thiessen-polygons constructed around sinks are easily offered. Difficulties are growing if one has to use models with implicitly given hypothesis like entropy-maximizing methods.

ad (ii). Close relations of trip distribution hypothesis with regionalisation are also desirable to minimize the - in its amount contested - impact of point pattern on the results of model application. At least three ways could be followed to filter out the influence of point pattern:

- construct a reference trip distribution without using spatial variables
- compare model results with those being derived by use of an ideal point pattern
- compute two trip distribution matrices, one with a global, the second with spatially differenciated impedance function(s).

Taking up the first two thoughts, section 2 deals with a suggestion how to construct a trip distribution without relying on areal variables. Results are used to derive some general properties of subdivision of space if the entropy-maximizing hypothesis is thought of being suited best for the problem under investigation. Section 3 focuses the general results of section 2 on the regionalisation problem of the multinodal inner-urban space, while section 4 contains an application of the ideas put forward.

Before going further some remarks should be made upon inner-urban consumer traffic. In view of consumer-behaviour it can be subdivided into at least two subgroups (SG):

SG1 : Time/Distance/Effort Minimization in satisfying high-frequency demand (mainly food). General interaction-theory offers 'Linear Optimization Models' in connection with the 'Theory of Central Places' and its shaping of market areas.

SG2 : With decreasing frequency of demand of goods traffic considerations take into account 'distance' in various measures, 'relative location' in form of accessibility or intervening opportunity and size/facilities/multiple choice of possible destinations. Apart from special behavioural hypothesis and dependent on available information intervening opportunity or entropy-maximizing models could be used.

In case of poor and rather global information - that is: application of entropy-maximizing models - the weight of underlying point pattern in analysing consumer traffic increases: explanation of trip distribution is based solely on spatial variables like 'distance', 'relative location' or derived measures like accessibility', 'potential' or 'areal competition'. The representation of the investigation area within the model especially the zoning completely determines the success of further interpretation. Therefore besides questions of measuring distance (real, ordinal, perceptive) or effort, expenditure etc. one has to raise the question about a neutral point pattern or zoning-system, too. Regular point patterns are well suited because they inherit the lowest information, overall as well as regionally differentiated.

2. IMPACT OF POINT PATTERN ON MODEL RESULTS

Interaction models generate a trip distribution matrix under use of

- information of local structure at sources/sinks
- characteristics of the investigation area (distance matrix, route network, barries, etc.)
- global indicator about traffic (total expenditure, etc)
- the trip distribution hypothesis.

Since these characteristics are dependent on space it is suggested to substitute the preference matrix G for the computed matrix.

Given:

$V: = \{V_1,\dots,V_n\}$ Set of traffic amount generated by origin zones $\{1,..,n\}$

(1) $E: = \{E_1,\dots,E_n\}$ ... destination zones $\{1,..,n\}$

$T: = \sum_{j=1}^{n} E_j$ Total amount of traffic

$= \sum_{i=1}^{n} V_i$

The matrix

$$(2) \quad F = (F_{ij}) = \left(\frac{V_i \; E_j}{T}\right)$$

has the following properties:

- up to a multiplicative constant $\left(\frac{1}{T}\right)$ this is the matrix of the interaction probabilities under assumption that V and E are independent in sense of probability theory
- under given row and column sums this is the matrix of maximal entropy
- the following constraints are satisfied:

$$(3) \quad \sum_{i=1}^{n} \frac{V_i \; E_j}{T} = E_j \quad \forall j \in \{1,..,n\}$$

$$(4) \quad \sum_{j=1}^{n} \frac{V_i E_j}{T} = V_i \quad \forall_i \in \{1,\dots,n\}$$

The accompanying preference matrix $G=(G_{ij})$ of a real (or computed) trip distribution matrix $T= (T_{ij})$ will be defined as

$$(5) \quad (G_{ij}) = \left(\frac{T_{ij}}{F_{ij}}\right)$$

Applied to the entropy-maximizing model one gets

$$(6) \quad G_{ij} = \frac{A_i B_j V_i E_j d_{ij}^{-\alpha}}{\frac{V_i E_j}{T}} = \frac{1}{T} \; A_i B_j d_{ij}^{-\alpha},$$

($d_{ij}^{-\alpha}$: logarithmic perception of distance d between orig. i and dest.j)

whereas the balancing factors A, B are defined like

$$(7) \quad A_i: = \frac{1}{\sum_{j=1}^{n} B_j E_j d_{ij}^{-\alpha}}, \quad B_j: = \frac{1}{\sum_{i=1}^{n} A_i V_i d_{ij}^{-\alpha}}$$

As to the above described Wilson model the impact of point pattern on the computed trip distribution could be estimated by examining the balancing factors A, B. Since latter are computed with the aid of iterative methods [7] the initial step of the Furness procedure can serve as approximation. So we get for example:

$$(8) \quad B_j \approx \frac{1}{\sum_{i=1}^{n} V_i d_{ij}^{-\alpha}}$$

A Taylor-expansion of the denominator holds:

$$(9) \quad \sum_{i=1}^{n} V_i d_{ij}^{-\alpha} \approx n \left\{ \bar{V} \, d_{.j}^{-\alpha} + (-\alpha) d_{.j}^{-\alpha-1} \cdot \mathrm{cov}(V,d_{.j}) + \frac{1}{2} \bar{V} \, \sigma^2(d_{.j}) \cdot (-\alpha)(-\alpha-1) d_{.j}^{-\alpha-2} \right\}$$

and

$$(10) \quad \mathrm{cov}(V,d_{.j}) := \frac{1}{n} \sum_{i=1}^{n} (V_i - \bar{V}) \, (d_{ij} - d_{.j})$$

$d_{.j}$:= average distance of trips/ traffic to destination zone j.

Hence follows that the areal competition of senders at sink j : $\sum_{i=1}^{n} V_i d_j^{-\alpha}$ can be expressed in a first approximation as the product of average amount of traffic sent by sources and α-averaged distance of traffic with zone j as destination. From that can be concluded that the areal variation of B-values is dependent on to what extent the distribution of receiver-distances differs from sink to sink. Therefore pattern with homogeneous point-density will minimize this areal variation.

It still remains to discuss the second term of the Taylor expansion (10). While the first approximation is only regionalized by the relative location of destination zone j (expressed by $d_{.j}^{-\alpha}$), its corrector term $\mathrm{cov}(V, d_{.j})$ takes into account the co-variation of amount of traffic at sources, V_i, and distance to sink j. Examples for the relation of point pattern and cov. cf. Klein [11, pp.191-196]. This by no means ignorable contribution to the areal competition can be minimized by

- probability theoretical independence of V and $d_{.j}$
- or low areal variation of V.

At the point of entropy-maximizing method that much can be said that the impact of point pattern on model results can be delimited by

- use of a regular pattern as well as
- minimizing the areal variation of V and E.

3. DERIVATION OF A PATTERN TRANSFORMATION

This section deals with a transformation how a regular pattern can be adjusted to a special problem under consideration of the ideas developed so far.
The investigation area is overlaid with a hexagonal network. Latter seems to be particularly suited for traffic investigation within the multinodal inner-urban region. The centres of hexagons form also the point pattern required for model application. Within the hexagons persons act as potential tripmakers. Their areal distribution is shown in Fig. 1. It is evident that the number of persons in traffic zones (hexagons) varies considerably.

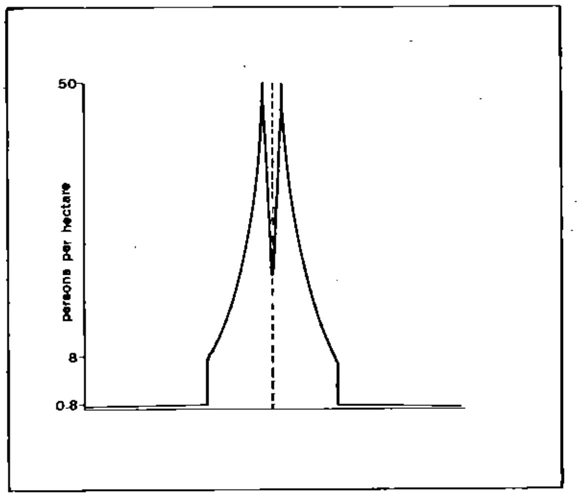


Fig. 1. Theoretical distribution of population density across a city centre. (Adapted from Dixon [6,p. 20])

So traffic zones are to be transformed to minimize this areal variation. Choose a transformation centre (x_m, y_m). Let $r_1, \ldots, r_k$ denote the distances from (x_m, y_m) to those hexagon centres which will be touched by the half-profile shown in Fig. 1. F be the area of a hexagon.

$$(11)\quad V_l = \int_{F(r_l)} p(F)\,dF \qquad \forall l \in \{1, \ldots, k\}$$

p: function of probability density, shown in Fig. 1.

If $r_1 < r_2 < \ldots < r_k$ then - outside the city centre -

$V_1 > V_2 > \ldots > V_k$ and the decline will correspond to the one of p.

Transformation t should be chosen to achieve

$$\{r_1, \ldots, r_k\} \rightarrow \{r_1^{(t)}, \ldots, r_k^{(t)}\} \quad \text{with}$$

$$(12)\quad \bar{V} = V_l = \int_{F(r_l^{(t)})} p(F)\,dF \qquad \forall l \in \{1, \ldots, k\}$$

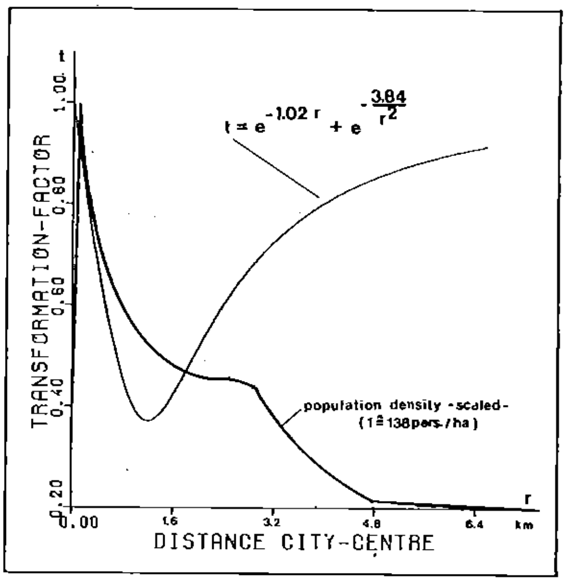


Fig. 2. Graph of transformation function for the case of neglecting barriers (cf. Fig. 4). Empirically derived graph of population density for Regensburg has been opposed.

To characterize the transformation function in words:

- expansion in the range of low population density
- contraction in the range of high

In the present case the following transformation function has been used:

(13) $t = e^{-ar} + e^{-\frac{b}{r^2}}$ (cf. Fig. 2)

$r := d((x,y), (x_m,y_m))$

(x,y) : coordinates of a hexagon centre

$d(,)$: distance to transformation centre (x_m,y_m)

$0 < a,b$ a : dependent on the extension of the CBD and the mesh-width of hexagonal network

b : dependent on the total urban and especially densely populated area (a first guess could be the square of the standard distance of population)

Each hexagon centre (x,y) will be transformed into $(x^{(t)}, y^{(t)})$ like:

$$x^{(t)} = (x-x_m) \cdot t_{(x,y)} + x_m$$

(14)
$$y^{(t)} = (y-y_m) \cdot t_{(x,y)} + y_m$$

The numerical estimation of the transformation parameters a, b uses the object function :

(15) $$ZF = \frac{1}{n} \sum_{i=1}^{n} (V_i - \bar{V})^2 = \min .$$

4. APPLICATION

Handling the information of a two years period study of inner-urban consumer traffic at Regensburg a zoning system for analysing SG 2(cf. 1.) is needed. With the method outlined above a first design will be worked out which has to be modified in small scale to take into consideration road network, distribution of shops and interaction barriers.

As in section 3 the areal distribution of sources will be thought as coincident with population distribution. As a consequence of its historical development and physical (river Danube, relief) as well as anthropogeneous (railway, highway) barriers the graph of population density differs from the ideal one (cf. Fig. 2.). There is no exponential decline from centre to fringe and the density gap of the CBD is only slightly marked.

As to possible sinks on the basis of a total survey about 70 inner-urban clusters of shops or shopping centres of differing size and facilities have been established. Only a small part of these is chosen as direct destination under the aspect of SG 2. Latter are located within the inner city, along arterial roads - mainly in the area of pre-1914 expansion and incorporated village centres - and at the urban fringe. However inquiries on consumer behaviour show that multi-purpose shopping trips occur frequently. So the whole 70 centres can be regarded as indirect destinations of SG2.

Regarding the shopping behaviour there is much evidence that the entropy-maximizing method with its interpretation of areal competition will be appropriate to reconstruct it. For there are not only centripetal traffic flows but also flows along the margin and CBD-crossing flows. Latter is due to the described pattern of sinks and the tendency of retail decentralisation.

A hexagonal zoning system of 69 cells

has been chosen which covers the whole settled area of Regensburg. With regard to population density Regensburg seems to be partly overbounded (cf. Fig. 3). Two calibration runs have been made which differ in the consideration of barriers. Results may be seen from Table 1.

Tab. 1 Calibration results of transformation parameters

	Without barr. Hexag.	Without barr. Transf.	With barr. Hexag.	With barr. Transf.
$\bar{V}$	2029	2029	2029	2029
ZF	2265	1738	2458	1858
V_{min}	0	0	0	0
V_{max}	7871	7593	10121	7541
$a[km^{-1}]$	-	1.02	-	1.02
$b[km^{2}]$	-	3.84	-	3.20

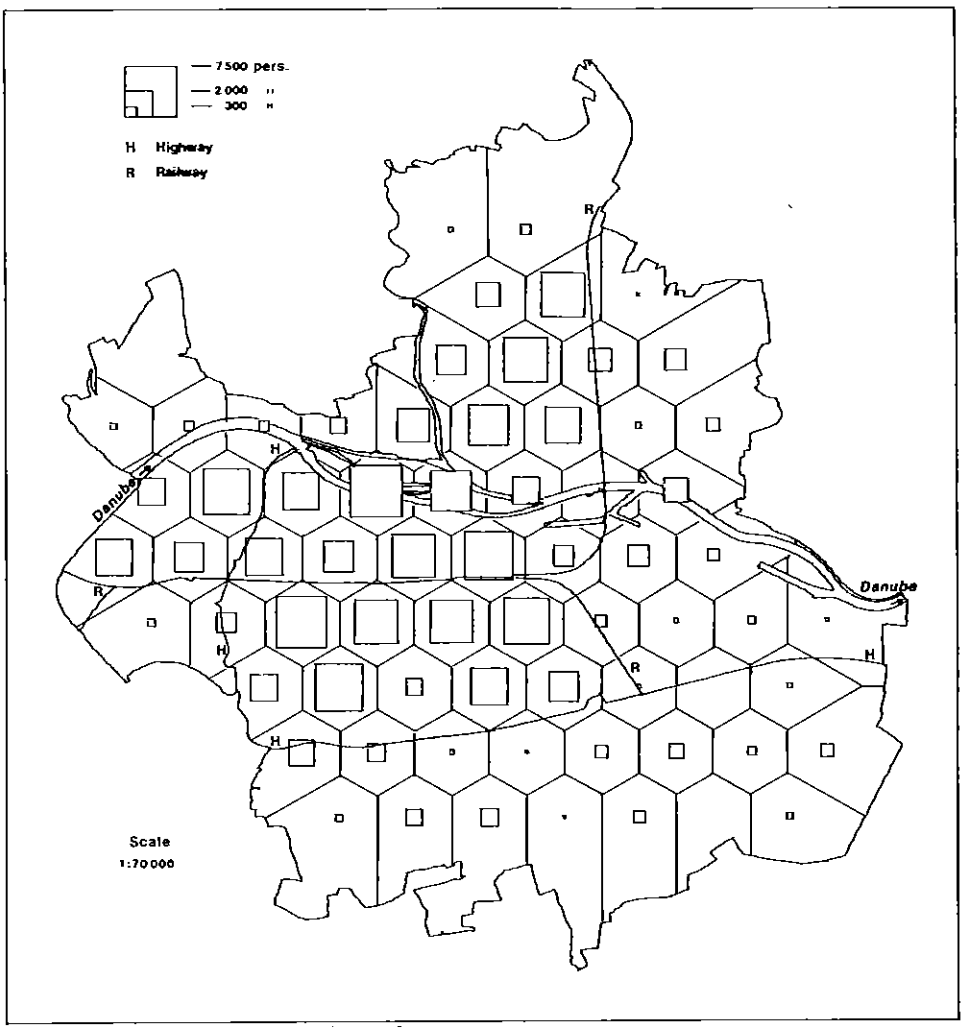


Fig. 3. Regensburg: Aggregation of population for a hexagonal zoning system on the basis of blocks. (Source of data : City Regensburg, Department of Statistics, 1980)

Fig. 2 shows that in the immediate neighbourhood of the chosen centre (x_m, y_m) points will be moved only slightly to avoid a splitting up of the CBD. With increasing distance from city centre and exceeding the densely populated area points will be contracted to even out population in zones. With decreasing population density the original hexagon point pattern undergoes only a slight distortion.
Comparison of Fig. 3 and Fig. 4 indicates that redistribution of population has been successful in city centre and margin, not always at the fringe. The areal variation has been restricted (standard deviation declines from 2265 pers. to 1738 pers.) but no regular pattern is achieved.

What about the impact of point pattern ? Table 2 provides the figures to estimate success of failure of the transformation.

Tab. 2 Measures for the impact of point pattern

	Without barr. Hexag.	Transf.	With barr. Hexag.	Transf.
	$\cdot 10^4$	$\cdot 10^4$	$\cdot 10^4$	$\cdot 10^4$
$\overline{cov}$	0.673	0.425	0.665	0.460
s_{cov}	1.02	0.581	1.01	0.635
v_{cov}	1.52	1.37	1.52	1.38
$(\alpha=-1.)$	$\cdot 10^{-4}$	$\cdot 10^{-4}$	$\cdot 10^{-4}$	$\cdot 10^{-4}$
$\bar{B}$	1.059	0.832	1.060	0.870
s_B	1.552	1.135	1.552	1.129
v_B	1.47	1.31	1.47	1.30

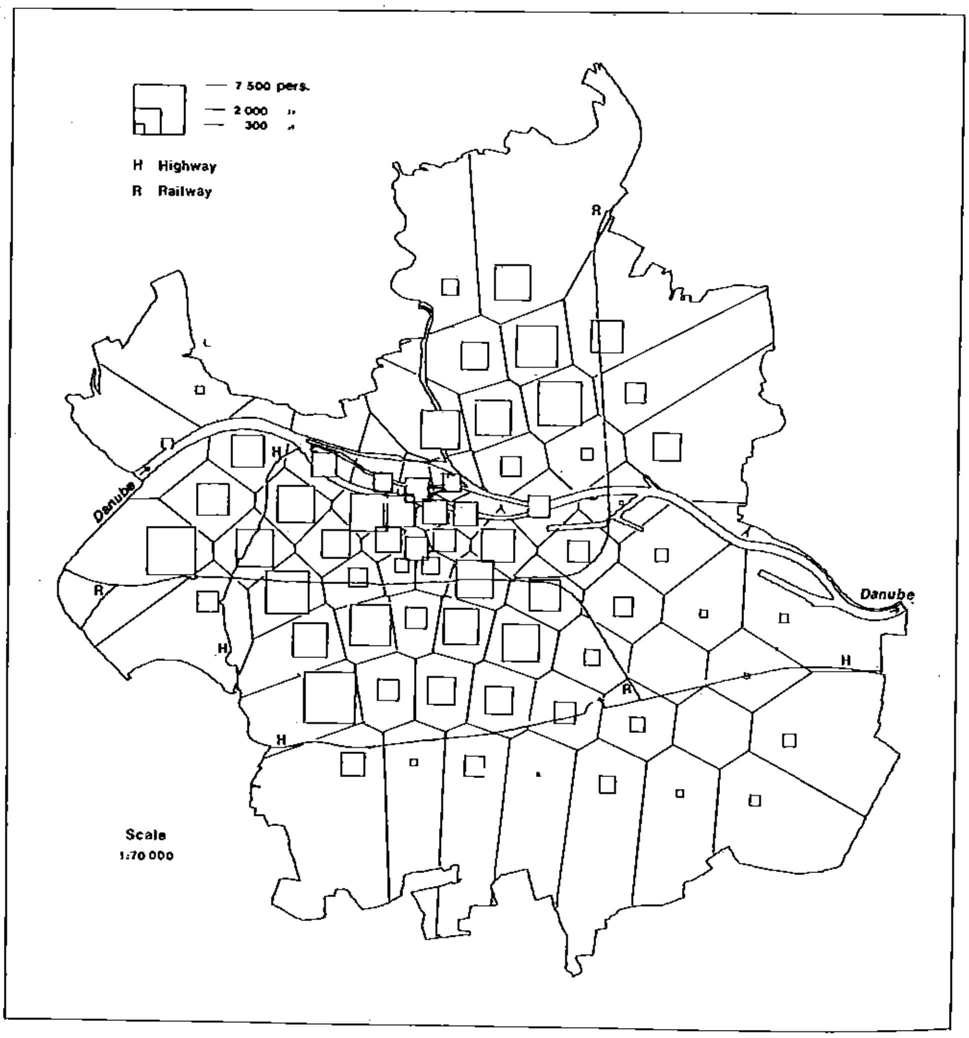


Fig. 4. Regensburg: Aggregation of population for the transformed zoning system on the basis of block data - without consideration of barriers. (Source of data : City Regensburg, Department of Statistics, 1980)

The distortion of the hexagonal zoning system with the consequence of varying distance distribution between sinks will be compensated by the lower variation of V. Therefore cov- as well as 8-values fall short of the correspondent values of the hexagonal point pattern. This is valid absolute as well as relative (coefficient of variation $v_{..}$). Note: Completely regular distribution of V could be obtained but at the cost of the still limited variation of point density and in consequence an unbalanced zone shape and area - not only at the fringe like in Fig. 4.

Closer to reality is the assumption that intersection barriers have to be considered. Indeed the rivers Danube and Regen as well as highways and railways affect and hinder the overall interaction. A second computation has been made and the results remain the same (cf. Tab. 1/2). Also the relating of blocks to zone centres within the 11 areas bordered by barriers shows extreme settings only at the fringe (cf. Fig. 5). In most cases the transformed points correspond with the population centre of the zone.

A few zones have been empty before and remain empty after transformation. This is due to the fact that population density is not symmetrical relative to (x_m, y_m). To avoid this and to bring zoning in close accordance with the local conditions it is suggested to depend the transformation function not only on distance but also on direction

$$t = f(r, \varphi) .$$

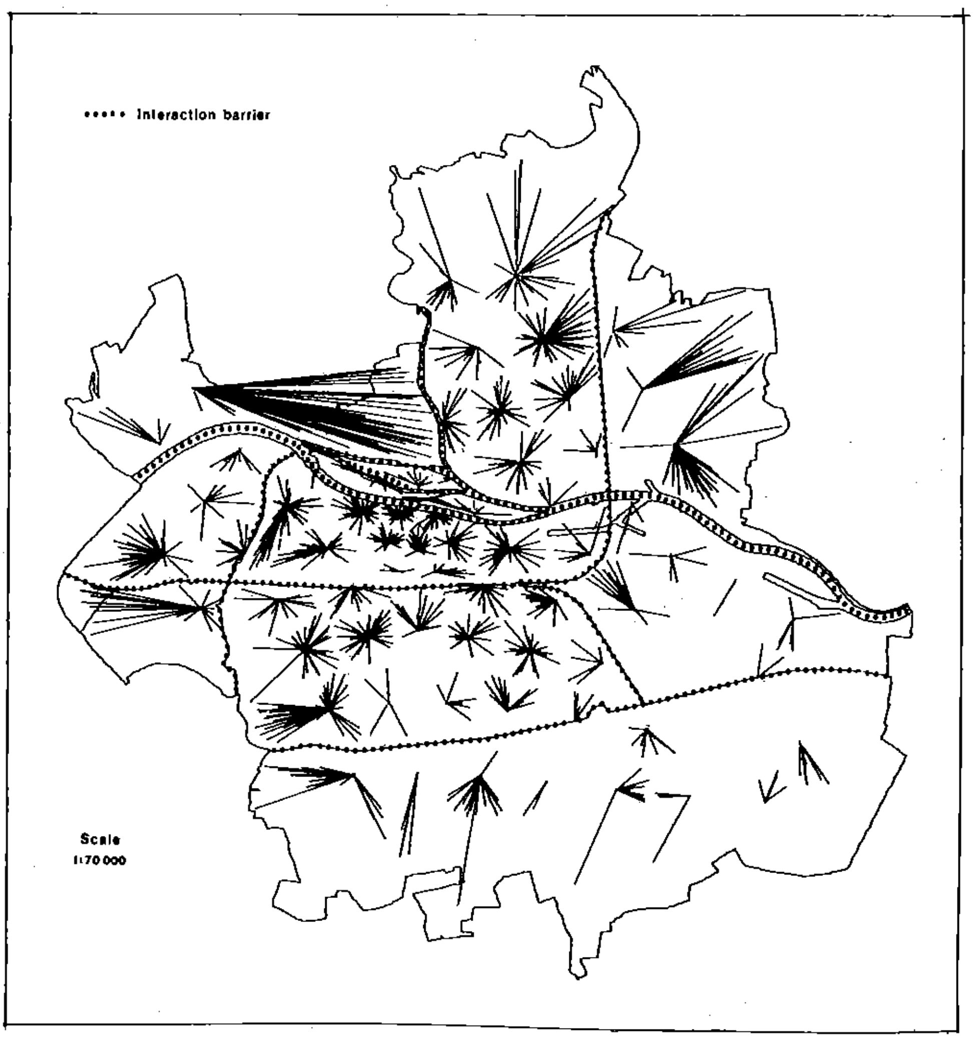


Fig. 5 Regensburg: Aggregation of population for the transformed zoning system on the basis of blocks - with consideration of barriers. Centres of stars are identical with the transformed points, the endpoint of stars are the centres of blocks. (Source of data : City Regensburg, Department of Statistics, 1980)

REFERENCES

1 ANGEL, S., HYMAN,G.M.: Urban spatial interaction. In: Environment and Planning 4,1972,pp.99-118

2 BERRY,B.J.:: Geography of market centres and retail distribution. Englewood Cliffs,N.J., 1967

3 BOOTS,B.N.: Some Models of the Random Subdivision of Space. In: Geografiska Annaler 55 B, 1973, pp.34-47

4 CLIFF,A.D.,MARTIN,R.L.,ORD,J.K.: Evaluating the friction of distance parameters in gravity models. In: Regional Studies 8,1974,pp.281-286

5 CURRY,L.: A spatial analysis of gravity flows. In: Regional Studies 6, 1972, pp. 131-147

6 DIXON,O.m.: Methods and progress in choropleth mapping of population density. In: Cartographic Journal 9, 1972, pp.19-29

7 EVANS,A.W.: Some properties of trip distribution methods. In: Transportation Research 4, 1970, pp. 19-36

8 HARRIES,B.: A note on the probability of interaction at a distance. In: Journal of Regional Science 5, 1964, pp. 31-35

9 JOHNSTON,R.J.: Map pattern and friction of distance parameters: a comment. In: Regional Studies 9, 1975, pp. 281-283

10 KIRBY,H.R.: Normalizing factors of the gravity model - an interpretation. In: Transportation Research 4, 1970, pp. 37-50

11 KLEIN,K.E.: Theoretische Untersuchung eines räumlichen Konkurrenzmodells zur regionalen Verkehrsverteilung. Mit Süddeutschland als Anwendungsbeispiel (Regensburger Geographische Schriften Heft 15). Regensburg, 1980

12 LOWE,J.C., MORYADAS,S.: The Geography of Movement. Boston 1975

13 MARQUARDT,P.W.: An algorithm for least squares estimation of nonlinear parameters. In: J.Soc. Indus.Ap.Math. (SIAM) 11, 1963, pp. 431-441

14 OPENSHAW,S.: Optimal zoning system for spatial interaction models. In: Environment and Planning A, 9, 1977, pp. 169-184

15 OPENSHAW,S.: An empirical study of some zone- design criteria. In: Environment and Planning A, 10, 1978, pp. 781-794

16 TIMMERMANS,H.: Consumer spatial choice strategies: A comparative study of some alternative behaviour spatial shopping models. In: Geoforum 11, 1980, pp. 123-131

17 TOBLER,W.R.: Geographic area and map projections. In: Geographical Review 53, 1963, pp. 59-78

18 WARNES,A.M., DANIELS,P.W.: Urban retail distribution: an appraisal of the empirical foundations of retail geography. In: Geoforum 11, 1980, pp. 133-146

19 WILSON,A.G.: Advances and problems in distribution modelling. In: Transportation Research 4, 1970, pp. 1-18

G E R H A R D   B A H R E N B E R G

A LOCATION-ALLOCATION MODEL FOR MAXIMIZING SUPPLY DISPERSION OF CENTRAL FACILITIES

From:
BAHRENBERG, GERHARD und STREIT, ULRICH (Eds.):
German Quantitative Geography. Papers presented at the 2nd European Conference on 'Theoretical and Quantitative Geography' in CAMBRIDGE 1980. Münstersche Geographische Arbeiten 11. Schöningh, Paderborn 1981, p. 115 - 124

Author:

Prof. Dr. Gerhard Bahrenberg

Schwerpunkt Geographie
Universität Bremen

Bibliothekstraße

2800 Bremen 33

ZUSAMMENFASSUNG

Es wird ein diskretes Standort-Zuordnungsmodell zur Diskussion gestellt, dessen Aufgabe es ist, für eine vorgegebene Mindestgröße einer zentralen Einrichtung die maximal mögliche Anzahl von Standorten dieser Einrichtung in einem Gebiet mit gegebener räumlicher Nachfrage- bzw. Benutzerverteilung zu bestimmen. Dieses "Modell maximaler Angebotsdispersion" operationalisiert somit eins der Axiome der Theorie zentraler Orte, nämlich die Existenz einer "Marktschwelle" zentraler Einrichtungen. Zusammen mit dem "vollständigen Überdeckungsmodell" von TOREGAS, das dem Prinzip der (oberen) Reichweite zentraler Güter entspricht, kann es sinnvoll für die Standortplanung zentraler Einrichtungen eingesetzt werden.
Am Beispiel der Standortplanung für Grundschulen im ländlichen Raum, bei der die notwendige Mindestjahrgangsstärke der entscheidende Parameter für die Frage der Erhaltung bzw. des Neubaus einer Grundschule ist, wird eine wesentliche Anwendungsmöglichkeit des Modells maximaler Angebotsdispersion aufgezeigt.

INTRODUCTION

During the last 15 years location-allocation modelling has become a field of growing interest in Geography, Regional Science, Planning, and other related disciplines. This is not surprising since these models are intended to deal with what ABLER, ADAMS and GOULD (1971) called "the geographer's unsolved problems". It is the question of the best location which can be formulated as follows: Given a population distributed in two-dimensional space. Where are to be located facilities of a central good/service for an optimal (as good as possible) provision of the population? Closely connected and by no means separable are the questions

- how many facilities are to be established,
- how are the catchment areas or hinterlands to be delimitated.

From an applied or planning point of view which I want to take here location-allocation models are of special importance for locational decisions of public facilities. Only in public facility planning it is possible to set objectives for a whole locational system whereas private firms decide under criteria of profitableness of the single location in question.

Hitherto most location-allocation models developed in geography and regional science for public facilities assume the number of locations to be within externally determined boundaries derived from crude cost and/or capacity considerations. The objective function usually represents a suitable measure of accessibility or transport costs that has to be maximized or minimized[1]. Examples are among others

- the classical transportation model which marks the beginning of location-allocation modelling.

It's purpose is to optimize spatial interactions resp. assignments by minimizing total transport costs between a set of potential facility locations (each with a specific capacity) and a set of population locations (each with a specific demand). Although formulated as an allocation model[2] it can be used for finding the facility locations as well[3]. Regarding that purpose it is extremely restrictive since minimizing total transport costs is the only criterium for the selection of the facility locations.

- the so called p-median model which was introduced by REVELLE and SWAIN (1970).

It was a starting point for many other similar discrete location-allocation models[4]. It is aimed at finding optimal central facility locations by minimizing the total travel costs of users when the number of facility locations is given.

Since the number of facility locations determines the accessibility (regardless of the accessibility measure actually applied) to a large extent it is necessary to develop models that answer the question for reasonable upper and lower limits of the number of facility locations.
The first step in this direction was made by TOREGAS with his "location set-covering model"[5]. This model is intended to find the minimum number of facility locations necessary to serve every demand location within a given time or distance

limit. Therefore it corresponds perfectly to the axiom of the (upper) range of a good/service in classical central place theory. Even if used alone, i.e. not in combination with other location-allocation models, it is of great help in the locational planning of emergency service facilities. While the location set-covering model leads to a lower limit of the number of facility locations on the basis of minimum accessibility requirements I would like now to introduce a model that yields an upper limit of the number of facility locations. In general the critical parameter this upper limits depends on is the minimum (threshold) size of the facility at each location. The threshold size can be the result of economic, organizational and/or service quality requirements. It is especially important for locational strategies in rural areas with a low population resp. demand density.

THE MAXIMUM SUPPLY DISPERSION MODEL

The model which I call "maximum supply dispersion model" has the aim to find the greatest possible number of locations (and simultaneously the locations themselves) of a central facility, given the constraint that the facility at each location be of minimum (threshold) size. It corresponds well to the second axiom of central place theory, namely the existence of a threshold (or lower range) of a good/service. Together with the "location set-covering model" it makes it possible to apply basic principles of central place theory to the locational planning of one central good/service but - in contrast to the central place model - also for an unevenly distributed population resp. demand.

The basic structure of the model is very simple and is described here in a form suitable for the application demonstrated below. Let us assume that there is an area divided into n basic spatial units, $i=1,\dots,n$ and that the demand or the population using the facility is b_i in unit i. Let us also assume that every unit is assigned to exactly one facility location, i.e. the hinterland or district of each facility location is a set of whole units. In that case we can introduce in the usual way a binary variable x_{ij}, defined as follows:

$$x_{ij} = \begin{cases} 1, & \text{when unit } j \text{ is a facility location and the population from unit } i \text{ is assigned to the facility at } j; \\ 0, & \text{otherwise} \end{cases}$$

in the particular case

$$x_{ii} = \begin{cases} 1, & \text{when unit } j \text{ is a facility location,} \\ 0, & \text{otherwise} \end{cases}$$

The locations of the facility and their hinterlands are therefore defined by the x_{ij}'s. It follows that the size of the facility at location j is $\sum_{i=1}^{n} b_i x_{ij}$. Let the minimum (threshold) size be c_1.

If the planning constraints that every unit is being served by one facility location, and that the districts shall be connected in a topological sense are included, then the model may be formally defined as follows:

(B 1) $\sum_{j=1}^{n} x_{ij} = 1$ (for all $i=1,\dots,n$)

(B 2) $x_{ij}=1$, thus $\exists\, i_1=i, i_2, \dots, i_{p-1}, i_p=j$, so that i_l and i_{l+1} are neighbouring units and $x_{i_l j}=1$ (for all $l=1,\dots,p$)

(B 3) $\sum_{i=1}^{n} b_i x_{ij} \geq c_l$ (for all j with $x_{jj}=1$)

(Z 1) Maximize $Z_1(x_{ij}) = \sum_{i=1}^{n} x_{ii}$

AN APPLICATION OF THE MODEL

In this section I want to demonstrate the application of the maximum supply dispersion model to elementary school planning in the former Landkreis (county) Rotenburg/Wümme which is located in northern Germany roughly midway between the cities of Bremen and Hamburg.

During the late 60s and early 70s a reform in West Germany's educational system took place that led to a spatial concentration of schools in rural areas. Large, multi-stream entry schools were preferred since they seemed to provide better education because

- they allowed for more powerful internal differentiation,
- there would be more specialized and therefore better qualifical teachers, and
- there would be better technical back-up (teaching equipment).

As a result numerous small schools were closed, because they were no longer viable in educational terms and instead one stream and, later, two stream entry schools were created in locations where there were catchment areas with a sufficiently large pupil potential. This policy was pursued right through the system down as far as the elementary schools and the 1974 Education Act in Lower Saxony (the state in which country Rotenburg/Wümme is located) prescribed that two stream entry elementary schools should be the rule.

This policy produced a rapid decline in the number of elementary schools which fell down in the twenty years between 1956 and 1977 to 28 % of its original level.

Year	Number of elementary schools in the former Landkreis Rotenburg/Wümme
1956	61
1968	46
1972	33
1976	21
1977	17

Source: W. STEINGRUBE (1979)

The closure of so many elementary and other schools increased the distances pupils had to travel in order to get to school and necessitated an alarming expansion of the school bus service. Travel costs are now the second largest element in the education budget (coming after salaries, but before running costs and learning materials). In Lower Saxony alone the total allocation for school transport rose from 13 million DM in 1971 to 200 million DM in 1977.

It is however not simply the cost of school transport that gives rise to concern, but also the length of time involved in journeys. According to a report in a Bremen daily paper ("Weser Kurier") on 18.7. 1978 journeys in excess of one hour are not uncommon for pupils and to this can be added the time spent waiting at the bus stop.

If this school policy would be retained, then inevitably the problem would become even more serious. The medium-term projections for migration losses in rural areas primarily affect the age groups with children and there are likely to be further reductions in the number of pupils, due to the narrowing of the gap between fertility levels in rural and urban areas. Thus in Lower Saxony it is estimated that the number of elementary school children in 1990 will be about 60 % of the 1974 level[6)].

Even without the steadily growing opposition from parents, paediatricians, and psychologists to the actual closure of elementary schools, the demographic changes currently taking place would make it imperative that there be a change in school policy. The first steps have already been initiated. In Lower Saxony the requirement that there be two form entry elementary schools has already been dropped in favour of one form entry, and a new official minimum has been set for class size, which is 15 % lower than the previous norm. It means that in elementary schools the minimum class size is now 24 instead of 28. The Minister for Culture has also let it be known publicily that the possibility of elementary schools with a minimum class size of 18 should not be ruled out. Even this figure may not be the final lower limit and it now seems conceivable that in the future there will once again be elementary schools with less than one complete class entering each year.

In order to be able to define the potential scope for future action in school policy, it is necessary to know the realtionship between the maximum possible number of elementary schools in a region and the number of pupils per year required to retain or establish a school. For that purpose the above introduced maximum supply dispersion model is applied.

The former county of Rotenburg/Wümme was divided into 75 basic spatial units each being represented by the mind-point of its settled area.
The b_i's were defined as the expected number of elementary school children in unit i entering school in 1990. This number was found by a prognosis of the number of children in the 5-9 year age group for 1990. The prognosis was based on an age-structured population projection by the Federal Research Institute for Regional Geography and Spatial Planning (Bundesforschungsanstalt für Landeskunde und Raumordnung) for the Bremen region (in which county Rotenburg lies) for 1990. It was assumed that population development in every unit i would be the same in relative terms and that each would have the same age structure in 1990 as the Bremen region as whole[7]. In other words the estimate of b_i for 1990 assumed that the following definition equations held true:

$$b_i = BAG_i^{(1990)}/5$$

$$\frac{BAG_i^{(1990)}}{BEV_i^{(1990)}} = \frac{BAG_{HB}^{(1990)}}{BEV_{HB}^{(1990)}}$$

$$\frac{BEV_i^{(1990)}}{BEV_i^{(1974)}} = \frac{BEV_{HB}^{(1990)}}{BEV_{HB}^{(1974)}}$$

with
$BAG_i^{(t)}$ = Number of 5-9 years old children in unit i at time t.
$BAG_{HB}^{(t)}$ = Number of 5-9 years old children in the Bremen region at time t.
$BEV_i^{(t)}$ = Total population in unit i at time t.
$BEV_{HB}^{(t)}$ = Total population in the Bremen region at time t.

If one is actually involved in the forward planning for elementary schools in any given area, then the starting point must of course be the existing distribution of schools. That means we have not considered how many elementary schools there ought to be in the former county Rotenburg in the light of the expected decline in the number of pupils of elementary school age. Instead we have looked at how many and which ones of the existing elementary schools can be retained in relation to the required minimum class size.

In 1977 there were still elementary schools in 17 units (cf. Fig. 1). In addition to constraints (B 1)-(B 3) of the basic model a further constraint was therefore adopted, to the effect that at most there should be elementary schools at 17 locations in the future[8].

(B 4) $\sum_{i=1}^{n} x_{ii} \leq 17$

By the same token, units which at present have no elementary school were excluded as potential elementary school locations, i.e.

(B 5) $x_{ii} = 0$ for the 58 units without elementary schools in 1977.

Finally it was also assumed, that those units with an elementary school, where there were likely to be enough children living in 1990 (at least 20 newcomers per year), would automatically retain their elementary school. It is assumed that it would not be possible politically to close these elementary schools. This produced the following further constraint:

(B 6) $x_{ii} = 1$ for the 4 units with $b_i \geq 20$.

This means that of the present 17 elementary school locations, 13 are possible candidates for change.

So far there has been no discussion of accessibility which is the dominant variable in most location-allocation models. It is however implicit in (Z 1), for the larger the number of possible locations, the greater will be the general level of accessibility. Nevertheless accessibility has also been considered explicitly as an independent constraint. It was assumed that no unit was further than 10 kilometres by road from the elementary school to which it was assigned.

(B 7) $\max_{i,j} x_{ij} \, d_{ij} = 10$

with d_{ij} = the shortest distance by road between units i and j.

The 10 km-limit should be seen as an initial trial maximum distance. School buses travel at an average speed of about 40 km/h and, if 5 minutes waiting time is included, then 10 km represents a travel time of 20 minutes. If it is further assumed that each pupil has a walk of up to 10 minutes to reach the bus stop, then the total travel time would amount at most to 30 minutes which doesn't seem to be unreasonably high.

The task therefore is to maximize the number of elementary school locations (Z 1) given constraints (B 1)-(B 7). The objective (Z 1) corresponds to the goal to safeguard as far as possible the elementary education through a local supply of schools. It contains implicitly criteria such as the extent to which a population is able to identify with the state, its organizations and facilities, and the extent to which the school is integrated into the life of the community. Both are difficult to define in quantitative terms and can best be summarized by the term "psychological proximity". This important psychological proximity does not decline continuously with distance but with a discrete jump when a facility vanishes from the daily experienced environment.

In most instances the model (Z 1), (B 1)-(B 7) will have several solutions all of equal value for the objective function $Z_1(x_{ij})$. When this happens, that solution ought to be preferred - on grounds of the 'psychological proximity' and the implications for school transport - which enables the largest possible number of pupils to reach their school on foot. The Lower Saxon School Act expects pupils living within 2 km of their school to reach their school on foot whereas pupils living farther away can use the school bus gratis. Therefore a second objective function was incorporated into the model:

(Z 2) Maximize

$$Z_2(x_{ij}) = \sum_{\substack{i=1 \\ d_{ij}\leq 2}}^{n} \sum_{j=1}^{n} b_i x_{ij} \Big/ \sum_{i=1}^{n} b_i$$

The model (B 1)-(B 7), (Z 1), (Z 2) was solved for 6 different minimum class sizes, i.e. with c_1 = 22, 20, 18, 16, 14 and 12 respectively in constraint (B 3). The results are given in Table 1, which also includes some further indicators for evaluating the spatial pattern. These show the proportion of pupils in each year, who have to travel more than 10 km, 5 km and 2 km respectively to their elementary school.

Table 1 Solutions of the location-allocation problem for elementary schools in the former county Rotenburg/Wümme

1	2	3	4	5	6	7	8	9
No.	Minimum class size	No. of spatial units with an elementary school	No. of pup. with a≤2km journey to the elementary school (in per cent)	No. of pupils (per academic year) with journeys of more than .... to the elementary school			Total journey distance of pupils (per academic year) with individual journeys of more than 2km	Maximum journey length (in km)
				2km	5km	10km		
1	22	13*	67,47 %	179,60	50,16	5,1	798,75	13,2
2	20	14	70,09 %	165,15	31,25	0	665,89	7,0
3	20	15*	71,83 %	155,50	43,03	0	691,15	9,6
4	18	15	71,76 %	155,90	36,83	0	653,56	9,6
5	18	16*	73,70 %	145,18	35,20	0	620,14	9,6
6	16	16*	74,25 %	142,15	15,51	0	544,74	9,6
7	14	16	74,25 %	142,15	11,26	0	528,565	7,5
8	14	17*	75,57 %	134,88	13,01	0	489,315	7,5
9	12	17*	75,57 %	134,88	8,67	0	473,796	6,5
10 §	12 (14)	15	72,26 %	153,16	38,47	2,55	670,17	10,8

* optimal solutions

§ proposal in the school development plan

Incidentally the last figure also shows how many pupils in any elementary school year use the school bus system. Of great interest is the column headed 'The total journey length of pupils (per entry year) with individual journeys of more than 2 km'. This provides a first approximation of the costsof the school bus system in relation to the number of elementary school locations. Finally the maximum length of journey is also included (last column).

For purposes of comparison Table 1 also includes solutions for several minimum class sizes, but with one elementary school respectively location less than the optimal one. Finally solution 10 is that proposed in the current school development plan for the county.

Table 1 shows clearly the connections between minimum class size, the number of elementary school locations, the degree of accessibility, and pupil transport costs. For c_I = 22 there is no feasible solution with a maximum journey as long as 10 km. Furthermore the solution makes necessary the closure of four schools (see first row in Table 1).

To maintain the present system of elementary schools, the highest the minimum class size can be is 14, ten less than that presently laid down by the School Act. Nevertheless, it is a figure well within the bounds of practical possibility, if the statements of intent by Lower Saxony's Minister for Culture are to be believed.

In general a larger number of elementary schools does not raise the number of children with a journey of less than 2 km (i.e. the number of children who can go to school on foot) as much as it does reduce overall school transport costs (compare columns 4 and 8). The main effect of additional school locations is to reduce the number of children with journeys of more than 5 km and to increase the number of children with medium length journeys (2-5 km). Even if these children cannot get to school on foot, at least they can get there by bicycle, which not only for cost reasons is preferable to a journey in a school bus. A prerequisite for this would however be a safe network of cycle routes.

Another interesting lesson to be drawn from Table 1 is that once the maximal possible number of school locations for any given minimum class size has been derived, a check should be made to see whether or not it would be possible to reduce further class sizes with this number of school locations. In certain circumstances such a rearrangement could improve accessibility and lower the costs of school transport, without increasing the salary bill and the running costs. That such improvements are possible can be seen in our example by comparing solutions 5, 6 and 7. The improvements can be made in this instance by a better choice of locations and some alterations in the delimitation of the catchment areas. This example also demonstrates the senselessness of adopting inflexible norms for class sizes.

Interesting are further the instances in Table 1 where it would be feasible to have one additional elementary school (see the transitions from solutions 1 to 3, 3 to 5, 5 to 8). The question that has to be answered here is the extent to which the increase in the salary bill (1 teacher per class, making a total of 4 per school) and in running costs can be offset by savings in school transport costs. The above analysis provides no clear answer, for column 8 'Total travel costs ...' only provides a rough first approximation of real costs for busing. Nevertheless column 8 suggests that there are considerable savings to be made.

Let us finally look briefly at the proposed solution in the school development plan, which is currently being debated in county Rotenburg/Wümme. The proposal is that two schools be closed, but, despite the fact that it will create very small classes in some locations[9], it is actually no more effective than either of solutions 2 or 3, which are based on miniuum class sizes of 20 and 18 respectively (see Table 1). The unsatisfactory measures for the school development plan have been brought about at least to some extent by

- the requirement that elementary school catchment areas coincide with community boundaries, and

- the merging of many small communities into bigger administrative units (the so called 'large communities') during the local government reform in 1974.

CONCLUDING REMARKS

In conclusion I should like to make some tentative suggestions about ways in which this study might be extended and expanded in the future.

1. So far this model has only been used to study the possibility of stopping the process of elementary school liquidation in rural areas. In the long-term, however, the goal must be to reverse this process and the model can be used equally well for this purpose - and that by omitting constraint (B 4). Trials ought to be made with minimum class sizes of less than. 12. This may perhaps seem a somewhat utopian prospect, but I believe that school development planning generally should be based on variable minimum class sizes depending on population, or more precisely pupil density. It ought to be accepted that class sizes in urban and rural areas are different, and in that case small classes in rural areas won't lead necessarily to astronomical increases in the salary bill.

2. The complex problem of school transport has been widely ignored. But it is certain that the structure and organization of the school bus system, the choice of bus routes, the size of the busses and the degree of its usage all have important implications for school development planning. Their inclusion in a location-allocation model seems to be too difficult to realize at the moment.

Despite these omissions, the study might have given some indication of how location-allocation models may be used positively to help in the planning of elementary schools in rural areas. I firmly believe that it is not for academic researchers to produce 'best' solutions to planning problems, but they can fulfil a most useful role by demonstrating a range of possibilities of action and their associated problems and consequences - say along lines I have tried to demonstrate by means of Table 1. In this way they can provide a basis for rational (in the sense that the options are derived from explicit criteria and objective evaluations) consideration of questions that must, in the last analysis, be politically decided.

Generally the model could be applied to studies of all kinds of service provision, so long as the size of the facility at any given location is the critical parameter.

FOOTNOTES

1 See the early reviews by SCOTT (1970, 1971) or the more recent ones by BAHRENBERG (1978), HODGART (1978).

2 See YEATES (1963) for an early application to a school districting problem.

3 HAGGETT, CLIFF and FREY summarize the applications of the transportation model (1975, 479-485 and 491-516) and HAY(1977) gives an elementary introduction with further references.

4 The development during the last 10 years has been covered widely by "Geographical Analysis". See also BAHRENBERG (1978), HODGART (1978).

5 See TOREGAS (1971), TOREGAS and REVELLE (1972,1973) and TOREGAS, SWAIN, REVELLE and BERGMAN (1971). The links to other location-allocation models are discussed in CHURCH (1974) and CHURCH and REVELLE (1976).

6 See KOCH (1976).

7 These assumptions were necessary, as there were no small-scale population projections for county Rotenburg available.

8 If the problem is formulated differently of course then this constraint can be omitted.

9 Our estimates forsee a minimum class size of 12 if the School Development Plan is put into practice. However the School Development Plan itself lays down a minimum class size of 14, based on the author's somewhat unrealistic projections for population and pupils.

REFERENCES

ABLER,R.;J.S.ADAMS and P.GOULD: Spatial organization. The geographer's view of the world. Englewood Cliffs,N.J.: Prentice-Hall 1971.

BAHRENBERG,G.: Ein allgemeines statisch-diskretes Optimierungsmodell für Standort-Zuordnungsprobleme. Karlsruhe: Geographisches Institut der Universität 1978 (= Karlsruher Manuskripte zur Mathematischen und Theoretischen Wirtschafts- und Sozialgeographie 31).

CHURCH,R.L.: Synthesis of a class of public facilities location models. (PhD thesis, The John Hopkins University. Baltimore 1974.

CHURCH,R.L. and C.S.REVELLE: Theoretical and computational links between the p-Median, location set-covering, and the maximal covering location problem. Geographical Analysis 8, 1976, 406-415.

HAGGETT,P.;A.D.CLIFF and A.FREY: Locational analysis in human geography. Volume 2: Locational methods. 2nd edition London: Edward Arnold 1975.

HAY,A.: Linear programming: Elementary geographical applications of the transportation problem. Norwich: Geo Abstracts 1977 (=CATMOG 11).

HODGART,R.L.: Optimizing access to public services: A review of problems, models and methods of locating central facilities. Progress in Human Geography 2, 1978, 17-48.

KOCH,R.: Prognose der Arbeitsplatzzahl und der Bevölkerung in den 38 Gebietseinheiten für das Bundesraumordnungsprogramm 1974-1990. Teil 2: Erweiterung der Bevölkerungsprognose von 1973. Bonn-Bad Godesberg: Bundesforschungsanstalt für Landeskunde und Raumordnung 1973.

REVELLE,C.S. and R.W.SWAIN: Central facilities location. Geographical Analysis 2, 1970, 30-42.

SCOTT,A.J.: Location-allocation systems; A review. Geographical Analysis 2, 1970, 95-119.

SCOTT,A.J.:Combinatorial programming, spatial analysis and planning. London:Methuen 1971.

STEINGRUBE,W.: Die Entwicklung der Schulen im Landkreis Rotenburg. Bremen 1979 (unpublished manuscript).

TOREGAS,C.: Location under maximal travel time constraints. PhD thesis, Cornell University. Ithaka 1971.

TOREGAS,C. and C.S.REVELLE: Optimal location under time or distance constraints. Papers and Proceedings of the Regional Science Association 28, 1972, 133-143.

TOREGAS,C.and C.S.REVELLE: Binary logic solutions to a class of location problems. Geographical Analysis 5, 1973, 145-155.

TOREGAS,C.;R.W.SWAIN;C.S.REVELLE and L. BERGMAN: The location of emergency service facilities. Operations Research 19, 1971, 1363-1373.

YEATES,M.H.: Hinterland delimitation: A distance minimizing approach. Professional Geographer 16, 1963, 7-10.

W O L F   G A E B E

IMPORTANCE OF INFRASTRUCTURE, SERVICES AND CONTACTS FOR INDUSTRIAL LOCATION DECISIONS

From:

BAHRENBERG, GERHARD und STREIT, ULRICH (Ed.):
German Quantitative Geography. Papers presented at the 2nd European Conference on 'Theoretical and Quantitative Geography' in CAMBRIDGE 1980. Münstersche Geographische Arbeiten 11. Schöningh, Paderborn 1981, p. 125-134

Author:

Prof. Dr. Wolf Gaebe

Geographisches Institut
Universität Mannheim

Schloß

6800 Mannheim 1

ZUSAMMENFASSUNG

Bericht über den Versuch, die Bedeutung der Infrastruktur, Dienstleistungen und Kontakte für industrielle Standortentscheidungen herauszufinden und die staatlichen Möglichkeiten eines Einflusses auf die räumliche und sektorale Entwicklung abzuschätzen. Untersuchungsraum sind 25 Gemeinden des westlichen Ruhrgebietes und linken Niederrheins. Makroanalytische Datengrundlage sind Statistiken über die Industriemobilität und Überprüfungen der Entwicklung neuerrichteter und verlagerter Betriebe, mikroanalytische Datengrundlage Befragungen von Unternehmern. Die Auswertung erfolgt nach den Variablen: Branche, interne Unternehmensorganisation, Betriebsgröße, Standortanpassung und Ortsgröße.

The aims of a current research project are:

1. Research into the meaning of infrastructure, services and personal contacts for industrial settlement and mobility decisions (newly founded firms, branch establishments, relocations), presumedly most important variables of external economies, and
2. Estimation of state possibilities of influencing the spatial development of industrial acitivity. The willingness of an employer to change the location of his enterprise if necessary is the prerequisite for an effective regional policy, intended to create new and safe jobs in underdeveloped areas.

There is not much usable information on the evaluation of infrastructure, services, contacts. There are some of the variables which describe the locational factors or location demands of the employer . Any attempt to collect all the variables would necessarily fail since dense reciprocal relationships exist which have not even vagely been clarified. It is therefore necessary, although rather frustrating, to choose some variables from the large amount of possibilities of which one can assume after some investigation that they influence the employer's decisions. Locational desires, including for example the size and price of the site, the qualification of labour, the viability of the ways of transport and the investment climate, are, according to behavioural investigations and considerations only one decision component. Two further components are the expectations of the employer and his willingness to take action. All three components, location desires including experience and perception, expectations and the willingness to take action, are connected with each other and form the cognitive and emotional attitude towards settlement and mobility decisions.

1. Development of industrial activity

In 1970 some 8.6 m. people were employed in industry in the Federal Republic,Tab. 1. Since this post-war record employment in industry decreased not only absolutely by 1.25 m to 7.3 m in 1977, but also relatively:

	total of industrial employees	difference (yearly average)	employees of newly founded and relocated firms	employees of closed fimrs (date of closure)	internal employment changes (recruitment/ dismissals)
			1000		
	I	II-IV	II	III	IV
1955-1957 (1)	7342	.	53	- 7	.
1958-1960	7800	153	47	- 15	121
1961-1963	8306	169	53	- 24	140
1964-1965	8381	38	48	- 20	10
1966-1967	8114	- 134	30	- 78	- 86
1968-1969	8104	- 5	45	- 37	- 13
1970-1971	8571	234	38	- 45	241
1972-1973	8383	- 94	23	- 47	- 70
1974-1975	7826	- 279	13	- 66	-226
1976-1977	7393	- 217	10	- 38	-189

(1) till 1962 without Saarland

Sources: Statistisches Bundesamt, Bundesanstalt für Arbeit.

Tab. 1: Industrial employment in the Federal Republic of Germany 1955-1977

	1960	1977
Primary Sector	16 %	8 %
Secondary Sector	46 %	43 %
Tertiary Sector	38 %	49 %
	100 %	100 %

The loss of jobs was due not so much to closures as to internal employment decrease. According to investigations of the Federal Ministry of Labour some 390 000 jobs were lost through closures between 1970 and 1977 and some 170 000 won through the foundations and relocations of firms (as yearly average the figures in Tab. 1 have to be doubled). Thus only a balance of some 220 000 jobs or about 20 % of 1.25 m. derived from settlement and closure decisions. And so, taking into consideration dismissals which took place before the closure, the total will be about 25 % of the reduction of employees, whereas 75 % came from internal locational adaptations and investment decisions.

The post-war wave of firm-founding has passed. Today the probability of the settlement of industry and of improving the supply of jobs is small for most of the 8 500 parishes in the Federal Republic as long as only about 200 firms are founded in the last years with a total of some 10 000 jobs. At the end of the 60s and the beginning of the 70s some firms were being founded annually. In fact new firms are being established in less than one per cent of the parishes in the Federal Republic. Quantitatively more important than new investments are substitute, enlargement and rationalisation investments. Investments aimed at creating new jobs are becoming less important. The need for jobs becomes greater the more enterprises are closed down.

Some of the possible reasons for the decrease in the number of people employed in industry are:

- a structural change in the economy (including changed world economic division of labour, increased importance of private and public services),
- innovation gaps (employment decrease in branches of industry based on research was 8 % from 1970-77; in other branches 21 %).

2. INVESTIGATION AREA

The macro and micro analysis was carried out in an area of 25 parishes in North Rhine Westfalia including the cities of Duisburg and Krefeld and the rural countries of Viersen, Kleve and Wesel beyound the edge of the congested Ruhr area. The density gradient decreases towards the north-west and west.
Development axes of second order of the Development Plan I/II of North Rhine Westfalia divide up the investigation area. Population and industrial decreases are higher than average, development chances are very unfavourable.

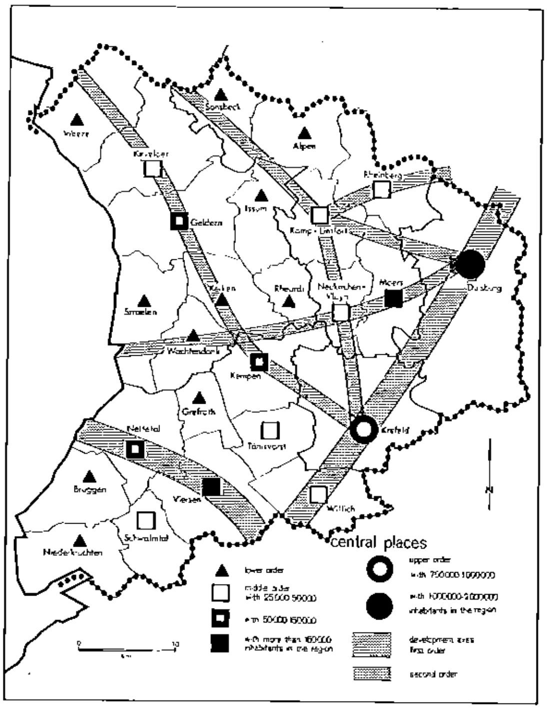


Fig. 1: Investigation area

A steady population decrease expected for the whole area, especially Duisburg and Krefeld is due to a birth deficit plus migrational losses, e.g. in Duisburg:

	(1000)	(%)
1961	663.1	100
1979	559.1	84
2000	455.1	69

In the cities of Duisburg and Krefeld the GNP per head of the population is noticeably higher than in the rural areas, yet lower than in Cologne or Düsseldorf, other central places of the Rhine-Ruhr-Region.

Levels of productivity and of mobility are not identical. A more favourable migration balance might be expected where there is a higher GNP per head. The share of the ter ary sector is relatively low in all regions; the differences in secondary sector are in Duisburg larger due to the large industrial fringe.

Data on industrial activity are based on

1. official statistics on industrial mobility,
2. questioning of 322 enterprises with 5o and more employees in 1979; more than 90 % of all enterprises of this size could be included.

2.1 EMPLOYMENT CHANDES IN THE RHINE-RUHR-REGION 1968-1975

In the Rhine-Ruhr-Region industrial activity is decreasing more rapidly than the federal average. The founding of new establishments and of branches is decreasing from year to year.
According to the Federal Office of Labour there were in the 8 years from 1968-1975
93 foundations (1968: 29, ...,1975:4)
109 branches (1968: 20, ...,1975:2)
11 relocations within the Rhine-Ruhr-Region, 3 moved into the area and 2 left it. Positive is only the fact that more than half of the new industrial establishments belong to the investment goods industry that is as much as two thirds of those employed. The investment goods industry has the best chances of all industrial groups.

		Establish-ments	Employees
1. Locational changes			
- new founded firms, branches, relocations within the Rhine-Ruhr-Region, movements to the Rhine-Ruhr-Region	1968 - 1975	216	22.000
- closures	1968 - 1977	- 73	- 3.000
2. Changes within the firms	1968 - 1977		
(Number employed minus number dismissed)		-	9.000
		+ 143	+ 28.000

Tab. 2: Balance of mobility in the Rhine-Ruhr-Region

Of all the industrial establishments founded in the Rhine-Ruhr-Region between 1968 and 1975, a third had already closed down again by 1977. It was above all establishments of the consumer industry which closed down though none of the establishments which had changed their location. There were about as many closures among newly founded establishments as among branch establishments.

Despite closures, establishments founded between 1968 and 1975 had more employees in 1977 than at the time of settlement (+ 28 %). The share of the investment good industry increased according to the number of establishments and numbers of people employed as a result of a low closure quota and greater internal growth.

The closure quota among branch establishments founded within the Rhine-Ruhr-Region was little different from that of establishments founded outside the area but whith the main establishment in the Rhine-Ruhr-Region.
The number of employed increased by 28.000. Some 100 branches with 4.750 employees were founded between 1968-1975 outside the Rhine-Ruhr-Region by employers within the region. The structural development indicates general unwillingness of the firms and branches and the increasing negative balance of the internal employment changes. The result of locational deficits is more often a closure than a change of location. Relocation means that working places are reshuffled only and that scarcely any new ones are created. Changes of location are the expression of a positive valuation of the new site.

The number of new establishments which can be expected to be spread into the regions becomes smaller. That impedes regional promotion based on industrial incentives and mobility. Public aid is welcome towards extension investments, it diminishes the investment risk. Yet the share of these extension investments susceptible ot the market is decreasing, whereas the share of replacement and rationalisation investments which enable the creation of new jobs are becoming less and less important.

2.1 LOCATIONAL DECISIONS

Services, contacts and infrastrukture are not isolated variables of settlement, mobility and internal investment decisions of the establishment, but are related to other variables. For the following variables, the connection with service demands was investigated statistically:

a) Branch,
b) Internal organization,
c) Size of the unit (employment),
d) Locational adaptation (flexibility) and
e) Size of the location (inhabitants).

ad a: BRANCH
With different priority connectivity, skilled workers, large and cheap sites are most important in all branches. Ways and costs of transport were of primary importance for employers of the basic and production goods industry. In the investment goods industry skilled workers were most important, more than connectivity and large sites. In the consumer industry ways of transport, skilled labour and site had equal importance. In the food and drink industry the sales market is the most important locational choice variable.

A large sales area and a high degree of export business increase the importance of public and private services, e.g. for machinery and engineering industries. For firms in the consumer industry with primarily regional markets and little input of manufactured goods traffic infrastructure within the area is of more importance than outside the area. They need services and contacts above all on the spot. For firms of the textile and clothing industry with a high sales radius and little input of manufactured goods the interregional traffic network is of greater importance.
Banking, legal- and tax advice are the services most needed by the enterprises that were questioned. The need for advice increases, in general, with the size of the location. It decreases, however, with the increase in size of the enterprise in accordance with the proportion of self-produced services. In larger enterprises with very differentiated demand, services and advice, there is a further increase. It is, relatively, the lowest in the basic and production goods industry, and highest in the investment goods industry.

ad b: INTERNAL ORGANIZATION
In all branches, the most important functions are management, stock-keeping, marketing, central buying, and endproduction; these are strongest in the food and drink industry, and weakest in the basic and production goods industry. With increase in size of the enterprise, the service and trade functions decrease whilst the production functions become more important and diversify, just as research and development.
Branches and the organization of enterprises have in general a greater influence on the need of infrastructure, services and contacts than size of enterprise and formal position of the enterprise within the firm (branch,main,firm). The internal exchange of goods or services within the enterprise create relatively autonomous spatial relationships.

In general, with firms with many units and branches the local and regional demand for private and public services is less developed than with firms which have only a single unit. The regional quality of a branch for the security of jobs is greater when the enterprise produces goods which are further manufactured within the firm than if it simply enlarges its capacity. Such enterprises which are not integrated are more likely to be closed down when there is a drop in the sales than enterprises which manufacture parts for other enterprises in firm.

The proportion of half- and parts-production, as well as processing in the basic and production goods industry; the proportion of assembling, research and development in the investment goods industry; and the proportion of manufacture for the final consumption in the consumption-orientated industrial units are as expected. The unit functions correspond to the structure of the enterprise: in one-unit enterprises the management and trade functions are most important, in branch units it is the final-, half- and parts-production, as well as further processing, and in main units of multi-unit enterprises, assemblage, research and development are most important.

ad c: SIZE OF THE UNIT

With increased enterprise size, the proportion of management and trade functions decrease, and the proportion of product differentiation, research and development increase.

In general, larger firms need more public and private services, contacts and specialized consultations because of their broad activity spectrums. Demands on level of education, knowledge and experience of employees increase the larger the firm. The same holds good for transport and reloading facilities and the usefulness of the site, larger firms in general spend more money on research and development. They also make more use of public research promotion.

ad d: LOCATIONAL ADAPTATION

If no internal adaptation is achieved a locational decision (branch, take over) is preferred to a change of location (shift). In principle a slight an tendencial decreasing willingness of the employers to move can be discerned. Locational persistence is generally high, it is highest in the basic and production goods industry. The behaviour of the employer is more greatly influenced by the present locational situation than by information on alternative locations. The more complex the decision situation, the less probable a change of location. Environmental protection costs, shortness of skilled workers, lack of extension possibilities, high prices of site speak in favour of locational division (branch establishment) or a change of location (relocation). On the other hand, money spent on investment, capital accumulation, contacts, spatial relations to banks, economic organizations, authorities, deliverers, customers, friends, acquaintances speak against this. Risk estimation is determined by the estimation of the economic development. For a branch establishment the decision situation is less complex and risky than for a relocation. Locational alternatives are seldom moved in depth. If there are bottlenecks or if there is a lack of location it is more likely that a firm will close down rather than change its location. The quota of closures in the consumer industry is above average, especially in the leather, textile and clothing industry. These branches are competitivly both structually and especially susceptible and are dependant on external services. With the structural shifting of industrial activity due to the changing economic division of labour between high and less developed countries, the share of industries needing services, informations and contacts, will increase, primarily offered in metropolitan areas and high ranking central places.

ad e: SIZE OF THE LOCATION

The relationships between the communications- and contact-intensity and unitand location-size, and between service- and advice needs and unit- and location-size, are not clear. Telephone calls are, in general, the most important form of business contact. Personal visit contact, trips of leading employees, and information per teleprinter increase with size of the location and size of the enterprise. Contact intensity is lowest in the consumer industry, and highest in the food and drink industry.

The local supply of goods and services increases as population figures rise. All the parishes in the investigation area have been classified in the central place hierarchy of the regional development plan (Fig. 1) because of the real or planned supply capacity. The supply of goods and services is a characteristic and an important explanation variable of locationally tied urbanization economies. Further characteristics are: an extensive labour market with employees of varying education and experience, a concentration of purchasing power and demand, extended and elaborate infrastructure, konsultation, maintenance, repair, market research, and publicity services, organizations to arouse interest, to collect, to process and to distribute informations.

Supply weaknesses of the investigated area, above all in the private economic and services supply of the larger central places correspond

- to a below-average representation of economic and administrative decision centres (headquarters of national and international industrial concerns),
- to a below-average representation of research and development institutions,

- to the lesser amount of publically sponsered research and development projects,
- to lack of research and development intense activities (e.g. in the pharmaceutical and electrotechnical industry),
- to the relatively small intensity of contacts with other firms (by telephone, by post, personally),
- to the very low share of employees in contact and information intensive activities (i.e. in publishing houses, association, consultation firms, industrial fairs, data processing, communication and reproduction technology).

Although in the investigation area an increase of supply and use of services, contacts and infrastructural establishments in relation to the size and rank of the central place can be proved (contacts within the firm, visitors, official journeys) yet this relation is considerably weaker than might have been expected considering the population and working place density on the periphery of the Rhine-Ruhr-Region.

The weakness of infrastructure facilities and the scanty staffing with contact and information intensive activities would explain the extraordinarily great difficulties of the parishes, to create new industrial jobs and to secure the existing ones. The largely negative location decisions of the households and enterprises are not only due to the perceived environmental burden (including microclimatical changes) and to the intense land use, especially in the core areas of the metropolitan area but above all to the lack of external economies, of urbanization economies. The obviously negative location preference: the above average emigration surplus of the population and the losses of jobs in industry is not only the balance of individual experiences, perceptions, expectations and intentions but also due to the lack of private economic supply. The public supply of services and infrastructure is considered as favourable by the enterprises; the deficit of political decision centres is not estimated very high.

According to the requiry the characteristic location for a newly built industrial establishment is a middle-sized near a large town. The characteristic location for a relocated firm is the periphery of metropolitan areas. This can be explained by the fact that the further the distance from the old location in the center of the congested area is, the more the information flow and external economies decrease. This distance - in accordance with the attainability of other spatial connections - is some 25 - 40 km . The areas of industrial inflow, however differentiate more than administrative areas or the other functional areas of goods supply or labour markets. The size of the site, the price and development cost for investment are reckoned as higher than the intensity of spatial relations or external economies by the firms which have moved site. Their decision to remain in the metropolitan areas is obviously so much a matter of fact that only characteristics influencing the micro-location are mentioned. Even the result that the proximity of suppliers and customers plays only a small part in the choice of location does not contradict the agglomeration hypothesis, if the decision for the macro-location is not taken into consideration. According to the enquiry one can deduce that choices are made between one metropolitan area and another, and not between one metropolitan area and a rural area.

Other reasons are more important in the location decision for a branch. These establishments are usually so-called 'prolongued workshops', that is small, one-sided firms, dependant both for organization and for production technology on the mother firm, without research and sales department. Such production and assembly plants are the first to be closed when demand decreases.

The results of the enquiries show:

1. that single characteristics of supply do not influence the macro-location for newly founded firms and relocations but the concentrated supply of infrastructure and services together with the supply of qualified labour, especially individual and ex-

pert knowledge and long experience at the job.
The probability of settlement in a metropolitan area and in a more important central place increases (or settlement in rural areas decreases)
a. with capital accumulation, since the settlement risk increases the higher the investment, yet it decreases with external economies,

b. with the need for more sophisticated services, since for most of the firms certain supply levels can only be attained in the metropolitan areas, e.g. the traffic supply, although this spatial variable for location decision is becoming less important the more technical and social infrastructure is increased,

c. with the share of qualified labour, as only here expert labour reserves for varying abilities and knowledge are to be found. Preference for living demands rise with the social status (educational, health, recreational facilities, etc.),

d. with the amount of international contact, of foreign visitors and journeys abroad because of the national connectivity and a representative address.

2. that direct personal contacts are probably of no great importance. Reliability and stability of business relations would seem more important than technical and economic relations. Here innovation perception and experience play a part together. Proximity to services is less important the more seldom it is made use of and the more specialised the use is. Solutions to research and consultation problems are sought there, where suitable partners are presumed. The probability of settlement in a metropolitan area and an important central place rises (or the settlement in a rural area sinks)
a. with the complexity of the firm's organization patterns and production processes, that is with the amount of forward and backward relations and the need of goods, services, knowledge and contacts,

b. with the number of location and investment decisions and the size of the enterprise. Personal variables can be levelled out and decision processes rationalised more easily,

c. with dependence on technical progress, on innovation and cost of research and progress. There is a transfer of technology between the large, stateowned research establishments and industry.

3. that the choice of site (micro-location) is influenced by its size, costs, extension possibilities, location advantages (e.g. connection to water ways), living and recreational opportunities (environmental quality, attractiveness of the city).

3. CONSEQUENCES FOR REGIONAL POLICY

Regional economic policy has been unable to realize its aims of diminishing the regional disparities in job opportunities and gross national product. Subvention in rural and structurally weaker areas is generally more difficult as a result of the decreasing mobility of private investment. The total increase in jobs in industry has become smaller. If one takes into consideration jobs which have been lost by the closing down of establishments, regional economic policy will be reduced to a nonentity with strong preferences for attractive metropolitan areas. The dependence of the rural areas on the metropolitan areas strengthens the unintended demand of unskilled labour and comparatively badly equipped firms and jobs mainly for female workers. In the metropolitan areas, on the other hand, most of the suppliers, customers, administrative and development departments of the firms are to be found. The foundation of research and development capacities in rural areas has scarcely been promoted directly. A common deglomeration policy in the metropolitan areas is hardly realisable politically.

There is only a weak connection between

promotion and investment in spite of the large amount of public aid received. Most research comes to the conclusion that the influence of subsidies on local decisions is small. Even the employment policy of public spending programmes hardly touches on the peripheral rural areas.

In order to achieve the regional development aims the following factors must be taken into consideration:

- the rising international competitive pressure in industry,
- the minimum level of the supply patterns,
- the rising demands on basic communal provisions,
- the more conscious perception and estimation of the environmental quality, in particular the residential environment.

A concentrated expansion of infrastructure and a wide economic promotion in rural and less developed areas must be limited to medium-size and large-size central places, connected with a network of roads and means of communication. As research into changes of location has showed, site advantages, economic and social chances through a wide activity potential are in general prerequisits for an employer's location decision. In the economic and social system of the Federal Republic the employer himself has to decide whether he makes use of these possibilities. The state contributes in deciding basically on the general pattern of land use, infrastructure investment, renewal of residential areas, and further means of improving the urban environment. In order to improve activity conditions of industry, regional policy must systematically care for existing and for the supply of goods and services. The quality of housing and recreation is becoming more and more important in employers location decisions.

REFERENCES

BUNDESMINISTER FOR ARBEIT UNO SOZIALORDNUNG(Hrsg.): Die Standortwahl der Industriebetriebe in der Bundesrepublik Deutschland und Berlin (West). Bonn 1979= Neuerrichtete, verlagerte und stillgelegte Industriebetriebe in den Jahren 1976 und 1977.
- Die Standortwahl der Industriebetriebe in der Bundesrepublik Deutschland und Berlin (West). Bonn 1977 = Neuerrichtete, verlagerte und stillgelegte Industriebetriebe in den Jahren 1972 bis 1975.
- Die Standortwahl der Industriebetriebe in der Bundesrepublik Deutschland und Berlin (West). Bonn 1975 = Neuerrichtete, verlagerte und stillgelegte Industriebetriebe in den Jahren 1970 und 1971

FÜRST,D.und K.ZIMMERMANN unter Leitung von K.W.HANSMEYER: Standortwahl industrieller Unternehmen. Ergebnisse einer Unternehmerbefragung. Bonn 1973= Schriftenreihe der Gesellschaft für regionale Strukturentwicklung, Bd. 1

GAEBE,W.und K.HOTTES(Hrsg.): Methoden und Feldforschung in der Industriegeographie. Mannheim 1980= Mannheimer Geographische Arbeiten,Heft 7

GROTZ,R.: Räumliche Beziehungen industrieller Mehrbetriebsunternehmen. Ein Beitrag zum Verständnis von Verdichtungsprozessen. Stuttgart 1978= Stuttgarter Geographische Studien,Bd. 93,S. 225-243

HAMILTON,F.E.I.(Hrsg.): Contemporary industrialization. Spatial analysis and regional development. London 1978.

HAMILTON,F.E.I.und G.J.R.LINGE(Hrsg.): Spatial analysis,industry and industrial environment. Progress in research and application.Vol.1: Industrial systems. Chichester 1979

W I L F R I E D   W I T T E N B E R G

STUDIES ON THE MOVEMENT PROCESS OF INDUSTRIAL PLANTS*)

From;

BAHRENBERG, GERHARD und STREIT, ULRICH (Eds.):
German Quantitative Geography. Papers presented at the 2nd European Conference on 'Theoretical and Quantitative Geography' in CAMBRIDGE 1980. Münstersche Geographische Arbeiten 11. Schöningh, Paderborn 1981, p. 135-144

*) The author is indebted to Mr. P.A. Longley (Bristol) for help in the translation of this article.

Author:

Dr. Wilfried Wittenberg

Geographisches Institut
Universität Karlsruhe

Kaiserstraße 12

7500 Karlsruhe 1

ZUSAMMENFASSUNG

Die mobilitätsorientierte Regionalpolitik in der Bundesrepublik Deutschland ist bisher von der unausgesprochenen Annahme ausgegangen, daß ein nach Qualität und Quantität ausreichendes Potential an mobilen Industriebetrieben zur Verfügung steht. Während sich schon in den 60er Jahren zeigte, daß die in den Fördergebieten errichteten Industriebetriebe nicht die von der Regionalpolitik erwünschten Qualitätsmerkmale hatten, zeigt sich seit dem Beginn der 70er Jahre, daß sich die Zahl der mobilen Industriebetriebe ständig verringert. Deshalb ist es erforderlich, sich in stärkerem Maße als bisher mit der Mobilitätsbereitschaft von Industriebetrieben zu beschäftigen. Im Rahmen dieser, sich als Pilotstudie verstehenden Untersuchung, wird aufgrund der bisherigen Ergebnisse von theoretischen und empirischen Arbeiten postuliert, daß die Mobilitätsbereitschaft im wesentlichen von der konjunkturellen Entwicklung und den am alten Standort bestehenden Standortengpässen abhängig ist. Die empirische Überprüfung der zuerst genannten Abhängigkeit zeigte besonders die unterschiedliche Mobilitätsneigung von verschiedenen Industriegruppen und den beiden Errichtungsarten (Betriebsverlagerung und Zweigbetriebsneugründungen) auf. Demgegenüber wirken die Standortengpässe, der zweite Mobilitätsverursachungskomplex, auf alle Industriegruppen gleichmäßig. Bisher ist es jedoch nicht gelungen, die Standortengpässe im Rahmen eines Regressionsmodells zu quantifizieren. Die bisherigen Ergebnisse, die auf einem kurzen Beobachtungszeitraum und auf einem kleinen räumlichen Stichprobenumfang beruhen, legen die Vermutung nahe, daß die Mobilitätsbereitschaft nicht nur in Abhängigkeit von den oben genannten Verursachungskomplexen betrachtet werden muß. Denn der räumliche und sektorale Strukturwandel der Industrie kann auch als ein raum-zeitlicher Prozeß aufgefaßt werden.

1. REGIONAL POLICY AND MOBILE PLANTS

Regional Policy in the Federal Republic of Germany attempts to advance the economic standard of the less developed regions by improving their industrial structure and by increasing the number of industrial jobs in these areas. As the public and administrative authorities do not wish to regulate the regional development process directly, the main emphasis of Regional Policy has been upon the improvement of physical infrastructure and the subsidisation of industrial investment. In order to increase the number of jobs in the depressed regions there are two alternative strategies: first, the expansion of already existing industrial plants and, second, establishment of new industrial plants in depressed areas. The prevailing opinion is that the second strategy is much more effective, because diversification of the industrial structure is realized in a shorter period of time. Given that newly-established plants are few in number and exhibit and even spatial distribution, Regional Policy is forced to concentrate upon transfers of mobile plants.

In order to achieve the maximum policy effect in the depressed regions, much information was required concerning the factors deemed to be important by industrialists in the search- and decision-making processes. Thus many empirical studies have been focussed upon the availability of, and demand for, various locational factors and the scaling of their relative importance. With this information Regional Policy was able to improve the location factors in the depressed regions in order to make them attractive for industrial location. However, because most such studies only consider the destination region, it is assumed (as in traditional location theory) that the decision to move is made before the start of the search process for a new location. Furthermore, there was considered to be an abundance of mobile industrial plants which wanted to leave their present sites in the urban areas, in order to relocate or to found a branch in the depressed regions.

SPATIAL CATEGORIES

Kinds of Industrial Establishment	Assisted Areas				Non-assisted Areas				Total	
	Rural		Agglomeration		Rural		Agglomeration			
	abs.	rel. in %	abs.	rel. in %	abs.	rel. in %	abs.	rel. in %	abs.	rel. in %
NUMBER OF PLANTS										
New Founded Plants	639	38.4	151	9.1	431	25.9	443	26.6	1 664	100.0
Relocated Plants	449	23.8	231	12.2	498	26.3	712	37.7	1 890	100.0
Establishment of Branch Plant	2 027	48.1	349	8.3	1 176	27.9	661	15.7	4 213	100.0
Total	3 115	40.1	731	9.4	2 105	27.1	1 816	23.4	7 767	100.0
NUMBER OF JOBS										
New Founded Plants	31 453	32.9	10 023	10.5	18 417	19.3	35 740	37.4	95 633	100.1
Relocated Plants	28 065	19.6	17 565	12.3	32 014	22.3	65 711	45.8	143 355	100.0
Establishment of Branch Plants	146 225	39.8	44 700	12.2	80 889	22.0	96 013	26.1	367 827	100.1
Total	205 743	33.9	72 288	11.9	131 320	21.6	197 464	32.6	606 815	100.0

Table 1: Number of newly-established industrial plants and jobs with regard to their location into different areas in the Federal Republic of Germany (excluding West-Berlin) 1955-1971
(Source: WITTENBERG 1978)

The assumptions and the strategy of Regional Policy seemed to be valid until 1971. Despite some anomalies during recession of the economic business cycle, about 50 % of the newly-established industrial firms and about 46 % of the new jobs were founded in the assisted areas between 1955 and 1971 (Table 1). However, with regard to qualitative aspects, Regional Policy did not achieve its aims of diversifying the industrial structure and generating self-sustaining economic development in the depressed regions. Most of the newly-established plants were branches of labour-intensive industries (such as the clothing and leather industry), employing less than 100 people with a predominance of female workers.

Furthermore, time has shown that the number of new plants and new jobs created has been very small: indeed, the number of newly-established plants amounts to an annual average of just 0,1 % of the total of existing plants. It is therefore evident that the potential redistributive effect of newly-established plants was seriously overestimated by Regional Policy. A further problem has become important from 1971, since when the number of newly-established plants has been declining rapidly (Table 2). Data since 1975 suggest that rationalisation programmes are now more common than simple transfers in the industrial relocation process. This new development calls into question the emphasis upon industrial movement by Regional Policy.

More information is therefore required about the mobility potential of industrial plants. Regional Policy incentives alone will not influence the location decision if the industrial decision-maker does not wish to invest in an assisted area. We must therefore consider the main issue: namely, under what circumstances and for how long will an industrialist adapt to changes in the internal conditions of his plant and in the external conditions of the environment without changing location. We should consider not only the decision to move, but also the decision to remain in the same location.

This paper considers whether it is possible to estimate the number of mobile plants. Few theoretical works and empirical studies have considered the industrial decision-making process. So there is little information about the reason to move in such studies investigating the relocation process.

2. DETERMINING FACTORS FOR INDUSTRIAL MOVEMENT

In addition to the information by empirical studies (for example FOERTSCH 1973, FORST/ZIMMERMANN 1973, GAEBE 1977) the investigation by BADE (1976) has analysed the problem of industrial mobility from different aspects, such as investment-, behavior- and decision-theory. It has been postulated that there are three factors influencing the willingness of the industrial decision-maker to move.

The first ist the presence of inertia in industrial decision making. This results when conditions at an existing plant location have been satisfactory for a long period of time: although both the internal conditions of the firm and the external conditions of the environment are changing, the industrialist is either ignorant of or oblivious to these changes. If the scale of negative location factors is very small, the industrialist can justify the decision to remain at the existing location by stressing the positive factors associated with the location. Thus the industrialist justifies the present plant location by ignoring its disadvantages. Another reason for industrial inertia is the fact that the decision to invest at the existing location is more simple than an investment decision coupled with changing location. In bounding the complex relocation problem, the industri-

Time period	Newly Established plants	New Enterprise abs.	New Enterprise rel. in %	Mobile Plants: Total abs.	Total rel. in %	Relocated Plants abs.	Relocated Plants rel. in %	Branches abs.	Branches rel. in %
				Number of Plants					
1955-1958	983	185	18,8	798	81,2	265	33,2	533	66,8
1959-1961	1 149	120	10,4	1 029	89,6	207	20,1	822	79,9
1962-1963	699	116	16,6	583	83,4	159	27,3	424	72,2
1964-1965	1 663	334	20,1	1 329	79,9	391	29,4	938	70,6
1966-1967	863	231	26,8	632	73,2	268	42,4	364	57,6
1968-1969	1 301	327	25,1	974	74,9	285	29,3	689	70,7
1970-1971	1 288	426	33,1	862	66,9	331	38,4	531	61,6
1972-1973	839	251	29,9	588	70,1	276	46,9	312	53,1
1974-1975	453	164	36,2	289	63,8	141	48,8	148	51,2
				Number of Jobs					
1955-1958	165 908	23 063	13,9	142 845	86,1	40 978	28,7	101 867	71,3
1959-1961	121 723	10 604	8,7	111 119	91,3	20 491	18,4	90 628	81,6
1962-1963	59 153	9 277	15,7	49 876	84,3	11 891	23,8	37 985	76,2
1964-1965	78 652	11 643	14,8	67 009	85,2	20 949	31,3	46 060	68,7
1966-1967	41 254	7 928	19,2	33 326	80,8	13 339	40,0	19 987	60,0
1968-1969	75 041	16 000	21,3	59 041	78,7	18 254	30,9	40 787	69,1
1970-1971	78 307	21 267	27,2	57 040	72,8	19 498	34,2	37 542	65,8
1972-1973	46 570	11 442	24,6	35 128	75,4	19 769	56,3	15 359	43,7
1974-1975	26 039	8 175	31,4	17 864	68,6	10 845	60,7	7 019	39,3

Tab. 2: Number of newly-established industrial plants and jobs with regard to kinds of industrial establishment in die Federal Republic of Germany (with West-Berlin) 1955-1975.
(Source: WITTENBERG 1978, DER BUNDESMINISTER FÜR ARBEIT UND SOZIALORDNUNG 1977)

alist reduces the amount of relevant information by considering only a small number of potential locations and essential prerequisites at the beginning of the search- and decision-making processes. A further aspect of the decision to stay can be illustrated by analysing the distances moved: Figure 1 suggests that the industrialist searching for a possible new location does not wish to break his connection with the local environment. This is shown by the different distributions of relocated and branch plants.

gional disincentives. The most important internal conditions are the manufacture of a new product, expansion of production, obsolescence of production facilities, and wear of machines.

The exact threshold beyond which difficulties at the old location are perceived to be insurmountable cannot be directly ascertained, since the importance and effect of the disadvantages of agglomeration are related to the economic business cycle, the third factor influencing the willingness to move, in a complex way.

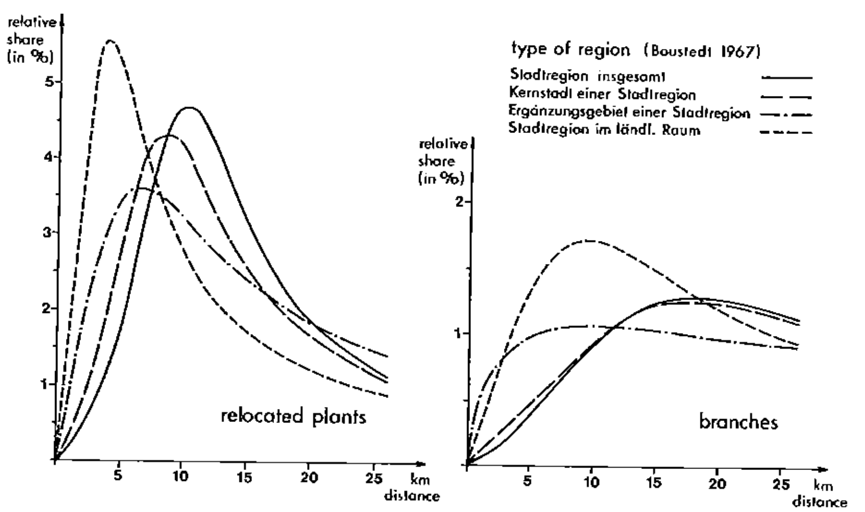


Fig. 1: Frequency distribution of distances involved in the industrial transfer by relocated and branch plants with regard to different areas of density in the Federal Republic of Germany 1955-1971
(Source: WITTENBERG 1978)

The industrialist establishing a branch plant is able to maintain close contact with the existing headquarters plant, and is thus able to move much further afield than the industrialist who transfers his entire production from one location to another.

The decision to move is precipitated by a crisis situation within the firm which must be so strong that the existence of the firm is threatened. In considering this second factor of willingness to move we have to distinguish between external and internal conditions. Interview surveys of industrialists suggest that the external conditions are the disadvantages of agglomeration, such as changes in quality and quantity of labour and space, changes in the market linkages, and re-

If the economic business cycle is on the downturn the industrialist tries to avoid the decision to move and tries to solve the problem in another way. If the economic business cycle is going up he tends to be more willing to accept the need to move than would otherwise be the case.

From this short discussion of the three factors influencing the willingness of the industrialist to move a plant the following point should be emphasized: given that the first factor (the behaviour of the industrialist) does not favourably influence the decision to migrate, we need only to consider the second and third factors. Thus the number of mobile plants and the spatial distribution of mobile plants with regard to the old location or the location of a main plant must be a function of the changing of internal and external conditions, and of the economic business cycle.

3. EMPIRICAL RESULTS

Contrary to the above theoretical framework, the mobility rate MR (defined as percentage of total industrial plants at time t considered mobile plants between time periods t and t+1) was used as the dependent variable. It was therefore necessary to verify the relationship between mobility rate and the changing of internal and external conditions of the plants, and between the mobility rate and stage of the economic business cycle. As data relating to plant mobility were only available on an annual basis for the period 1964 to 1971 the use of detailed time series analysis was not feasible. The results of this preliminary study using correlation-, regression-, cluster- and discriminant analysis should be interpreted with caution.

3.1. MOBILITY RATE AND THE ECONOMIC BUSINESS CYCLE

The empirical analysis relating mobility rate to several variables measuring stage of the economic business cycle shows that the best connections (accounting for about 60 % of total variance) exist between mobility rate and the annual relative changes in the "Bruttosozialprodukt" (Gross National Product), the "Index der industriellen Bruttoproduktion für Investitions- und Verbrauchsgüter", and the relative changes in the total number of industrial plants and jobs. The actual situation is shown in Fig. 2 in which the great differences between the six groups of industries may be seen. The leather, textile and clothing industry (group 5) is very susceptible to fluctuations in the business cycle while the paper and timber and stone and ceramic industry (group 1 and 4) respond only slightly to changes in the cycle. These differences are explained by the different abilities

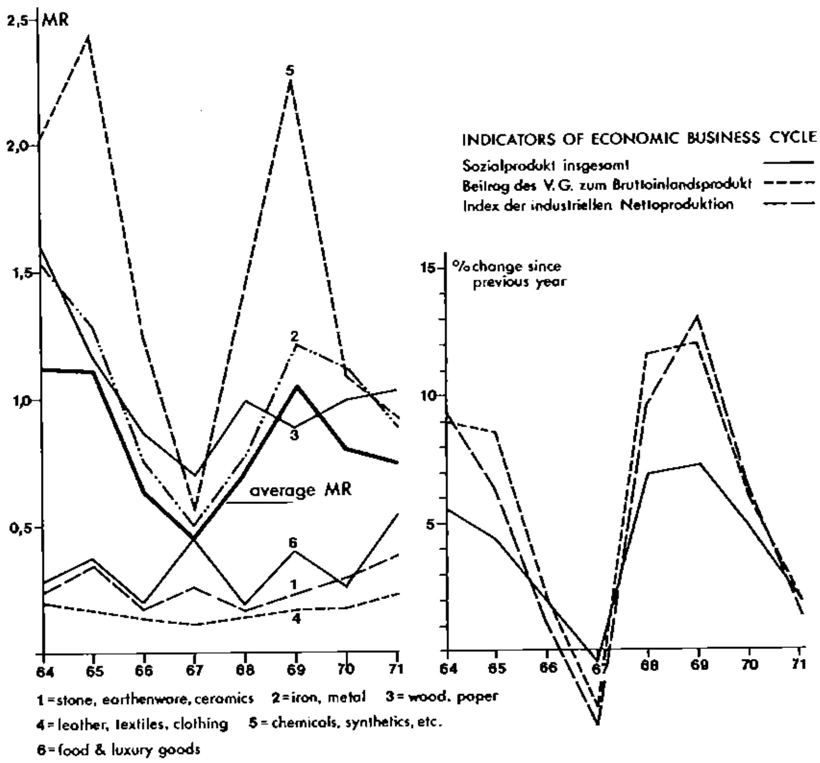


Fig. 2: Mobility rate of industrial sectors and indicators of economic business cycle in the Federal Republic of Germany 1964-1971

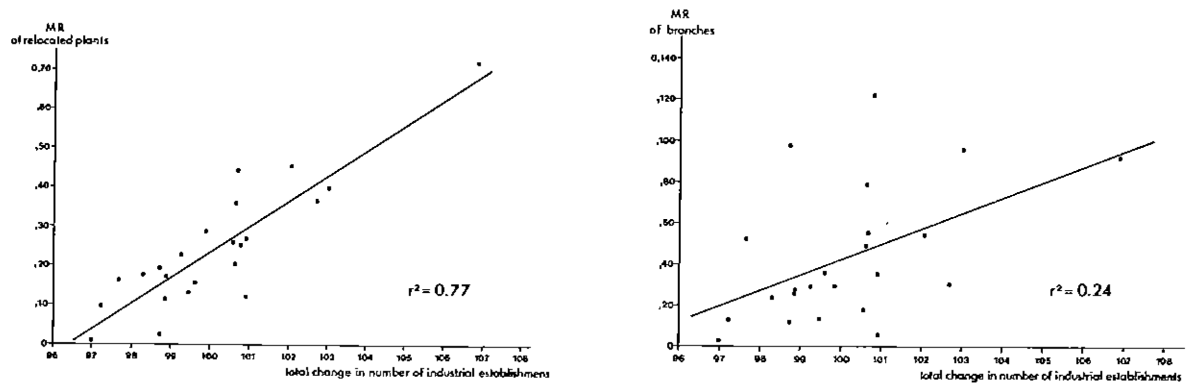


Fig. 3: Mobility rate of relocated plants and branches and total change in the number of industrial establishments for different industrial sectors in the Federal Republic of Germany 1964-1971
(Source: BADE 1978)

of these industries to establish branch plants. Distinguishing between branches and relocations in the same analysis, the total variance explained increases for relocations to approximately 70 % while falling to about 50 % for branches. This effect is also illustrated in Fig. 3 by the analysis of BADE (1978), using a different classification of industries and the relative changes in the total number of industrial establishments as an independent variable.

Both studies suggest that there are several groups of industries, such as the clothing industry, which exhibit a tendency to establish branch plants. Because of their production process they are able to react very quickly to short term economic change by establishing branches in boom periods, and closing them when the economic business cycle is on the downturn.

Given that the branch plants react quickly to changes in the business cycle, it is to be expected that the relationship between movement rates and earlier stages in the economic business cycle will be looser for branch plants than for relocations. Similar correlations were therefore made between mobility rate and the same variables of the business cycle in the year before the decision to move. The empirical results were as expected: whereas the total variance explained declined only about 10 % to about 60 % for relocations, in the case of branches it drops from 50 % to about 30 %.

In a further analysis the variation in time of mobility rate was investigated at different spatial scales (e.g. national, rural and urban scales and special "Gemeinden"). The results show the same trend in the variation of mobility rate at every spatial scale. The observed deviations from the general trend and the different levels of mobility rate must therefore be explained by the special conditions of the different regions, i.e. the overriding disadvantages of agglomeration.

3.2. MOBILITY RATE AND DISADVANTAGES OF AGGLOMERATION

Given that the above analysis suggests the economic business cycle influences the mobility rate of each of the industrial groups in a different way, a comparison was made between the national scale and the scale of the "Verdichtungsräume" (agglomerations). In Fig. 4 the "Verdichtungsraum Stuttgart" is compared with the national scale. It suggests that the diseconomies of agglomeration affect every sector or group of industry. It is therefore apparent that declining and rising

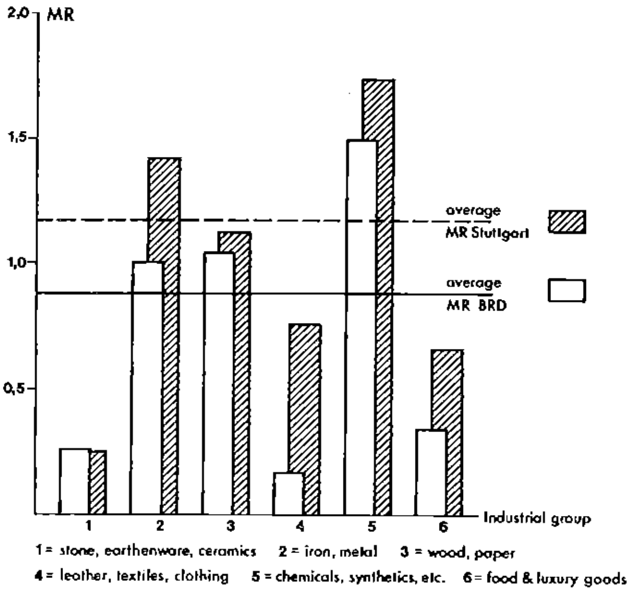


Fig. 4: Mobility rate of industrial groups at national scale and the scale of "Verdichtungsraum Stuttgart" 1964 to 1971

industries are affected in the same way by the disadvantages of agglomeration. As the mobility rate at the "Verdichtungs-räume"- scale is higher than at any other spatial scale it must primarily be explained by the external conditions of the environment. Industry in the agglomerations is not able to substitute for changing of external conditions by changing the internal conditions, and therefore the mobility rate of every industrial group is higher at the agglomeration scale.

However, the results of cluster- and discriminant analysis, designed to classify the "Kreise" within Baden-Württemberg with respect to annual mobility rate between 1964 and 1971 (Fig. 5), suggest that there are rural regions (such as that south of Stuttgart) which have the same level of mobility rate as the agglomeration regions (such as Stuttgart or Mannheim). Similar results were obtained by analysing the problem for different groups of industry. The existence of rural regions with an equal level of mobility to the agglomerations can be explained by their special situation with regard to the existing number of plants and the sectoral structure of their industry. This would seem to suggest that industrial mobility should be viewed within a dynamic space-time framework, a factor not considered in the present analysis.

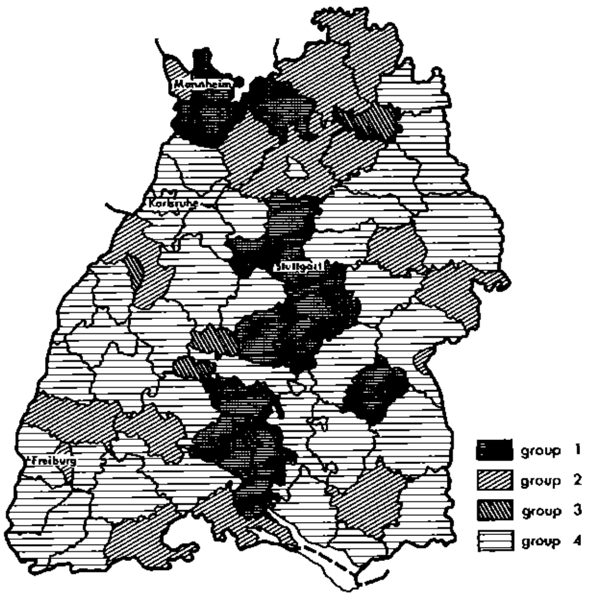


Fig. 5: Classification of the "Kreis" of Baden-Württemberg with regard to annual mobility rate 1964-1971

This suggestion is a further result of the above analysis, because there are different trends of mobility rate within the groups one, two and four (group three exhibited deviations from every general trend).

Group four comprises those spatial units, whose mobility rate is at the lowest level of the three groups and is sensitive to the variations in the economic business cycle as discussed above. Group one and two have, on average, a higher level of mobility rate and higher deviations than expected by the changes of the business cycle. The differences between these two groups are that the deviations of group one are during the first three years (1964-1966) of the observed time period, while group two has its greatest deviations in the years 1969 and 1970. Although the time period is very short, it would appear that mobility rate is declining for group one over time, and it is rising for group two. This suggests that the structural change of industries in the areas of group one is by and large completed, but is is becoming important for the areas of group two. Further empirical study over a longer time period is necessary to verify the above suggestion.

Although the above results are essentially descriptive in nature, a number of interesting questions arise which suggest directions for further empirical research. The attempt to establish connections between mobility rate and characteristics of their location with respect to external conditions such as labour and space by multiple regression analysis were not successful, since they only illustrate the effect of competition between industry and the tertiary sector within labour markets.

REFERENCES

BADE,F.-J.: Die Mobilität von Industriebetrieben. Diss., Karlsruhe 1978.

BOUSTEDT,O.: Die Stadtregionen in der Bundesrepublik Deutschland 1961. Hannover 1967.

BUNDESMINISTER FÜR ARBEIT UND SOZIALORDNUNG (Hrsg.): Die Standortwahl der Industriebetriebe in der Bundesrepublik Deutschland und Berlin (West) in den Jahren 1972 bis 1975. Bonn 1977.

BUTTLER,F.,GERLACH,K.,LIEPMANN,P.: Grundlagen der Regionalökonomie. Hamburg 1977.

ESENWEIN-ROTHE,I.: Die Persistenz von Industriebetrieben in strukturschwachen Wirtschaftsgebieten. In: Industrialisierung ländlicher Räume, 65-93. Hannover 1961.

FOERTSCH,H.-J.: Industriestandorttheorie als Verhaltenstheorie. Diss., Köln 1973.

FREUND,U.,ZABEL,G.: Zur Effizienz der regionalpolitischen Industrieförderung in der Bundesrepublik Deutschland. Raumforschung und Raumordnung 36: 99-106, 1978.

FORST,D.: Die Standortwahl industrieller Unternehmer: ein Überblick über empirische Erhebungen. Jahrbuch für Sozialwissenschaft 22: 189-220, 1971.

FURST,D.,ZIMMERMANN,K.: Standortwahl industrieller Unternehmer. Bonn 1973.

GAEBE,W.: Erklärungsversuche industrieller Standortentscheidungen. Seminarberichte der Gesellschaft für Regionalforschung 13: 161-184, 1977.

GANSER,K.: Großräumige und kleinräumige Konflikte in der Verteilung von Arbeitsplätzen. Bauwelt/Stadtbauwelt 57: 436-440, 1978.

GRITZKA,C.: Die Anwendung der heuristischen Systemanalyse zur Vorbereitung industrieller Standortentscheidungen. Diss., Mannheim 1974.

HANSMANN,K.-W.: Entscheidungsmodelle zur Standortplanung der Industrieunternehmen. Diss., Hamburg 1972.

KEINATH,K.: Regionale Konjunkturschwankungen - eine empirische Analyse der Bundesrepublik Deutschland 1950-1974. Tübingen 1978.

ROSCHENPÖHLER,H.: Der Standort industrieller Unternehmungen als betriebswirtschaftliches Problem. Köln 1958.

STATISTISCHES BUNDESAMT (Hrsg.): Lange Reihen zur Wirtschaftsentwicklung. Stuttgart 1976.

STRUFF,R.: Regionale Wirtschaftspolitik auf dem Prüfstand. Structur 5: 105-111, 1975.

TUCHTFELDT,E.: Infrastrukturinvestitionen als Mittel der Strukturpolitik. In: R.JOCHIMSEN,U.E.SIMONIS (Hrsg.): Theorie und Praxis der Infrastrukturpolitik, 125-151. Berlin 1970.

WIETING,R.: Regionale Strukturpolitik auf dem Prüfstand. WSI-Informationen 29: 611-617, 1976.

WITTENBERG,R.: Neuerrichtete Industriebetriebe in der Bundesrepublik Deutschland 1955-1971. Giessen 1978.

WOLF,F.: Wie effizient ist die regionale Wirtschaftsförderung. Informationen zur Raumentwicklung 9: 431-437, 1975.

W O L F G A N G   T A U B M A N N

REGIONAL DIVISION OF LABOUR AND MINIMUM STANDARDS OF LIVING CONDITIONS IN THE FEDERAL REPUBLIC OF GERMANY

From:

BAHRENBERG, GERHARD und STREIT, ULRICH (Eds.):
German Quantitative Geography. Papers presented at the 2nd European Conference on 'Theoretical and Quantitative Geography' in CAMBRIDGE 1980. Münstersche Geographische Arbeiten 11. Schöningh, Paderborn 1981, p. 145 - 153

Author:

Prof. Dr. W. Taubmann

Schwerpunkt Geographie
Universität Bremen

Bibliothekstraße

2800 Bremen 33

ZUSAMMENFASSUNG

Die Ordnungsvorstellungen der Raumplanung orientieren sich nach wie vor an dem Ziel der strukturellen Ausgewogenheit der Teilräume des Bundesgebietes. Neuerdings wird jedoch diskutiert, ob diese Konzeption ergänzt oder gar abgelöst werden soll durch den Ansatz der sog. funktionsräumlichen Arbeitsteilung, welche die Optimierung der unterschiedlichen Eignung der jeweiligen Teilregionen stärker in den Mittelpunkt raumordnerischer Überlegungen rückt.

Die Begründungen für beide Konzeptionen werden untersucht. Insbesondere wird auch analysiert, inwieweit Übereinstimmungen oder Gegensätze zwischen beiden Modellen bestehen, u.a. unter dem Aspekt der intra- und interregionalen Arbeitsteilung.

Bei Zielkonflikten wird vorgeschlagen, den Indikatorenkranz für "gleichwertige Lebensbedingungen" auf zentrale Meßgrößen zu reduzieren und andere Indikatoren regional zu modifizieren bzw. substitutiv zu verwenden.

According to the Regional Planning Report (Raumordnungsbericht) 1978 it still is the principal objective of the Federal Government's regional planning policy to attain equal social well-being in all regions of the Federal Republic. This planning goal is laid down in the Regional Planning Law (Raumordnungsgesetz) of 1965 as well as in the Federal Regional Planning Program (Bundesraumordnungsprogramm) of 1975, and it is also emphasized in the Recommendations of the Council for Regional Planning (Beirat für Raumordnung) of 1976 [1].

During the last few years an intensive discussion has begun whether the concept of structural balance or balanced functional regions should be supplemented by another concept which focuses on the optimization of the differential potentialities of the various regions and which accepts resulting additional disparities (cf. WEYL 1981).

This new line of argument refers to one particular passage in the Recommendations of the Council for Regional Planning of 1976 where it is postulated as a further task for regional planning policy "to make full use of the remaining economic, ecological, and other chances of all regions and to develop in an optimal way each region according to their functionally set or desired possibilities".

Whether both target functions - i.e. equal living conditions and inter-regional functional division of labour - conflict with each other or whether the first function is "only" the corrective of the second is still a matter of controversial debate. In the following chapters the two positions will be analyzed, with particular emphasis on the main differences and conflicting views between them.

The regional division of labour approach is mainly based upon partial models developed in regional economic theory, but these cannot be dealt with in any detail within the framework of this paper. In their basic assumptions these explanatory models consist of a combination of areal and sectoral structures (for further details cf. HAMPE 1981).

Within the economic division of labour and specialization those locations are chosen which offer an optimal utilization of comparative cost benefits for specialized production due to the presence of natural (e.g. mineral resources) and man-made resources (e.g. capital equipment). If there is a demand for those produced goods and services from outside the respective areas, inter-regional exchange relations will develop. The regional division of labour is also conceived as resulting from the general economic or social division of labour (a view similar to HAMPE's is held by BRÖSSE 1977, p. 100).

---

[1] "Die sozialstaatliche Forderung nach gleichwertigen Lebensbedingungen ist nach wie vor ein grundlegendes Ziel jeder Raumordnungspolitik. Denn nicht nur das verfassungsmäßig verbürgte, sondern auch politisch unabdingbare Maß von Gerechtigkeit, zu dem der Sozialstaat seinen Bürgern gegenüber verpflichtet ist, verlangt die Erfüllung der Mindeststandards für die Bevölkerung aller Teilräume".

The concept of "funktionsräumliche Arbeitsteilung" (division of labour between functional regions) used in regional planning policy is primarily substantiated by ongoing sectoral structural changes, national economic and demographic conditions and the international competitive situation (cf. e.g. SÄTTLER 1980a, 1980b). I.e. sectoral structural changes lead to an increase in the differences between the locational qualities of the regions of the Federal Territory and agglomerations become more and more important as locations for growth- or innovation-orientated industries. Because of the slackening growth of the national economy and the division of labour within the world economy the structurally weak rural areas have by now lost all attraction as locations for labour-intensive branches of internationally operating big firms. Wage-intensive phases of production are increasingly moved to Third World countries (cf. also BODENSCHATZ/HARLANDER 1974). Considering the lack of funds it must be an urgent goal for regional planning policy to safeguard those industries which are competitive on the world market and viable. This finds expression in a "concerted policy of providing locations" (gezielte Standortvorsorgepolitik) for existing firms especially in agglomerations.

Apart from the economic justification of the concept of regional division of labour there is one which questions the legitimation of balanced functional regions resp. the concept of minimum standards (cf. WEYL 1979 and 1980, HOBLER et al. 1980).

HOBLER et al. (1980, p. 31ff.) maintain that the objective "attainment of equal living conditions" cannot - as it is frequently done - be derived from the passage "guarantee of the uniformity of living conditions" (Basic Law Art. 72, Para. 2, No. 3). Even the "welfare state principle" (Basic Law Art. 20, Para. 1) is - according to the same authors - inappropriate for substantiating this objective. "The 'performance state' (Leistungsstaat) is obliged to create better living conditions only as far as it is possible in consideration of its remaining tasks" (op. cit., p. 43).

WEYL's argument is similar to that of the group around HOBLER. He extends the meaning of the term "equivalent living conditions", now designating the whole spectrum of chances for the free development of personality. But while for WEYL the aim of removing disparities is directed towards the immobile sections of the population staying in their home regions, the modern industrial and 'service' society is characterized by a high degree of spatial and social mobility. A mobile citizen, however, can take advantage of the chance for self-realization anywhere in the Federal Republic (WEYL 1980, p. 818).

It is well-known and therefore calls for no further explanation that theoretical approaches developed in the field of regional economy can be used for explaining spatial processes of differentiation / processes of spatial differentiation (cf. e.g. BUTTLER/GERLACH/LIEPMANN 1977 or SCHÄTZL 1978).

But viewing the problem for the opposite side I doubt whether an insufficient legitimation through the Basic Law is reason enough to call the objective of removing disparities into question. We can agree with DAVID (1981) who declares that regional planning policy should be defined by the legal and constitutional system only to a very limited extent, but we may just as well ask whether the request for balanced functional regions needs a legitimation through the Basic Law at all. Normative guide-lines, like e.g. regional minimum standards, at best presuppose a consensus within a society which regards itself as a solidarity. Therefore it seems that the dispute in favour of or against the request for balanced functional regions cannot be settled by reference to the Basic Law.

WEYL's opinion that the request for a removal of regional disparities is directed towards a "pre-industrial" or "immobile" type of person and supports the leveling of regional diversity, in my view neglects that a substantial population group especially in structurally weak areas is dependent upon their home regions because of age or lack of professional qualification.

Within migration research it is an established fact that age, level of eduction and position in the family cycle constitute the most important situational variables for decisions to migrate. This means that with decreasing degree of training and advancing age it becomes more difficult to start a career outside the previous profession and normally also outside the home region, and therefore people rather prefer to stay in their native place. In addition to this MIETH (1978, p. 216f.) has pointed out that in Germany readiness for regional mobility has been declining for a long time and that income differentials induce less and less decisions to migrate, since secondary factors like being tied to small social sub-systems or low housing costs in the countryside have gained in importance. Therefore in contrast to WEYL's assumption mobility especially in the advanced industrial and service society is decreasing with growing selectivity of migrating groups. This probably explains why the programs of the Federal Labour Office have failed, which tried to increase regional and professional mobility by means of removal allowances in order to create a balance between the labour markets. Once again, it was those people with the lowest qualifications who benefited the least. Regional planning policy must take note of this situation, if a one-sided commitment to the group of the highly mobile persons is to be avoided.

In the discussion about the equivalence of living conditions or the inter-regional division of labour another two complexes of problems are of considerable importance: Firstly, the definition of the objective of inter-regional functional division of labour, and secondly, the problem of choosing indicators or establishing minimum standards which are to mark a reasonable low level of regional living standards in the Federal Republic.

For THOSS, one of the leading advocates of the concept of balanced functional regions resp. regional minimum standards, there is no contradiction between balanced functional regions and the regional division of labour approach, since a region without preferential areas cannot be a balanced functional region (1980, p. 180). In his opinion only preferential areas or locations for natural parks, harbours or traffic routes establish a balanced functional region. The spatial demand in connection with the location of such preferential functions could be met within the respective functional region without any problems.

If we compare THOSS' short definition of a balanced functional region (op. cit., p. 179) with the original version of the concept, it becomes obvious that he ignores the aspect of intra-regional division of labour. A functional region namely is established above all through the fact that "potentiality-dependent regional division of labour only enables the balance of a functional region". "Signs of urbanity, economic strength and infrastructure" in the poorly developed centres of structurally weak regions of the Federal Territory should therefore be supported (cf. also MARX 1975, p. 7ff.). It follows that intra-regional division of labour is a prerequisite for a balanced functional region, i.e. quasi-ubiquitous functions should be established, if possible, in all functional regions.

The so-called "region concept" (Regionenkonzept) does not concern functions whose range is beyond the own region. But on this very aspect focuses the concept of inter-regional functional division of labour. Of course, it is true that the locations or areas of supra-regional functions (e.g. sea- and airports, fairs, research institutions, banks and insurance companies, etc.) "can easily be accommodated within the various balanced functional regions" (THOSS, op. cit., p. 180). But this does not reveal anything about the importance, range or eventual further support of those functions.

The analysis of inter-regional division of labour contains one trivial and one discussible aspect.

It is trivial that ports are situated in coastal regions or that winter holiday resorts are hardly to be found in the lowland of northern Germany. But this argument of triviality - quite popular with the critics of the regional concept - does not grasp the subject matter under discussion.

There is no point in describing obvious distributions of locations. Crucial is the contribution of a preferential function to the efficiency of multi-functional key areas of the Federal Territory. For the time being we shall leave aside the special case of large ecologically valuable preferential areas.

An airport, for example, not only increases the demand for regional inputs, but also improves the factor combination of its locational area because of the "air traffic sensitivity" of service and producing industries. The Untermain/Frankfurt Region depends on fast connections with world air traffic because of the dominance of the tertiary sector (banks, communications, insurance companies, lobbies, etc.) and because of its industrial structure (chemistry, electrical industry, engineering industry, light engineering and optical industry). Taking into account the multiplication effect, Frankfurt Airport had an estimated share of 15 percent in the gross domestic product of Hesse in 1977 (cf. TAUBMANN 1981).

It is by no means trivial to work out what kind of consequences an insufficient capacity of the airport or, for instance, a shift of partial functions to Munich could have for the whole region or what kind of locational preservation policy (Standortsicherungspolitik) should be pursued. The concept of balanced functional regions says nothing about these problems.

In my view the inter-regional functional division of labour approach is complementary to the concept of balanced functional regions. As far as the spatial demand of supra-regional functions is concerned there can be no clash with the basic ideas of the concept of balanced functional regions, since the locations must be conceived as points and not as areas regarding the viewing scale to be chosen. Their spatial demand can therefore be ignored.

An important exception are large ecologically valuable preferential areas which aim at an - even one-sided - exchange with other regions (e.g. water supply in large water protection areas). However, it remains an open question how an exchange between ecologial "compensatory areas" (Ausgleichsräume) and egological "stress areas" (Lasträume) can be accomplished (cf. FINKE 1981).

It is evident that without any case studies or model computations no exact propositions can be advanced, only some general ideas. Thus it should be argued whether regional specializations resp. clusterings of functions are to bring about regionally modified minimum standards or whether it is appropriate at all to determine minimum standards for all spheres of life.

Those social indicators suggested by the Council for Regional Planning in 1976 which describe the real living conditions by means of measurable criteria resp. which are meant to serve as a yardstick for the allocation of financial means and finally as a measure of success are still accepted without reservation. Recently THOSS (1981) has pointed out that the observance of uniform minimum standards for all regions of the Federal Territory is a pre-condition for balanced functional regions and that a substitution of indicators should not be permitted. For this reason it may be worthwhile to analyze the set of indicators suggested by the Council at least in some detail.

Certainly the minimum standards of the Council are a step forward compared with the indicators given in the Federal Regional Planning Program, especially because they also try to evaluate the non-economic spheres of life. Still there remain those points of criticism which refer to the significance of the indicators in the sense of their representativeness for those spheres of life to be measured or even to the normative set points and their substantiation.

Some indicators seem to be dispensable, because they contain hardly any information about the quality of life in the regions. For example, the indicator which stipulates that the length of motorways - relative to the population size of a region - should not fall below a certain value is inappropriate for determining the quality of life.

Other indicators must probably be modified regionally, because they conflict with the concept of regional specialization or the inter-regional functional division of labour. To give an example, specialization coefficients are formed on the basis of the number of employees in the 10 sectors of the economy, and the values of these coefficients are to remain below a certain level. Another indicator gives certain percentages for the employees in the various sectors of the economy (agriculture and forestry, power-producing industry, mining industry, etc.) relative to the total number of employed persons. These percentages should not be exceeded in order to avoid mono-structured regions or, because a lopsided economic structure affects the economic stability and the free choice of an occupation. These and other indicators quite definitely interfere with the sectoral specialization of certain regions. Consequently, modifications or substitutions must be admitted already on the level of the minimum standards.

The same applies to the indicator "number of foreign workers and their family ..." which should not exceed 12 p.c. in the districts (Kreis). Apart from the fact that there are urban districts showing higher values, "the present concentration of foreign workers in the big cities arises from a genuine demand" (MIETH, op. cit., p. 217). Even in the case of a decline in economic activity it is likely that industrial branches in agglomerations with a high percentage of lowly rated jobs will be dependent on foreign workers to a high degree. This particular job situation must be taken as an argument for regionally differentiated indicators.

Some other indicators are too vague in their denotations. The seemingly simple indicator 'employment deficit', measured as the rate of unemployed persons relative to the numer of employees on the first of September or the first of January, is problematical in many ways (cf. e.g. DEITERS 1980, p. 57-58): 1. It does not grasp "hidden" unemployment, 2. it includes no information about employees who have left the region after an unsuccessful search for a job, and 3. it does not embrace problem groups among the jobless, like e.g. young persons and women in structurally weak rural areas or old unemployed persons and those with "wrong qualifications" in the agglomerations.

Criticism of the indicator approach must not result in abandoning the use of indicators altogether. Rather regionally nonvariable and substitutable minimum standards should be limited to a data set which is indisputable under normative aspects of requirements. This particularly applies to the material and manpower infrastructure and to institutions in the field of education, culture and health. On the other hand, there is quite a range of indicators which can be regionally modified and substituted. It is true that the indicators 'income' or 'economic strength' which are only weakly developed in rural areas are confronted with distinctly more favourable values for non-utilized space, recreation, housing supply, etc. Especially with regard to the "equivalence of living conditions in the various regions" it is quite important to decide which indicators should be used in a substitutive or complementary way.

The cardinal problem of structurally weak areas cannot be solved by the provision of public infrastructure, namely providing a sufficient number of jobs, whose level of qualification would have to be higher than that of the existing jobs. Even the Regional Planning Report of 1978 must admit: "In the structurally weak and thinly populated areas the increase in the number of jouth entering into working-life is extremely big, whereas the job losses there are also extremely high". As already mentioned above, the peripheral rural areas can hardly expect any additional industrial jobs in the future. I doubt it very much that this problem can be solved by changing the methods and tools used in regional planning policy. This also applies to the proposal which favours a restriction of the funds for industrial promotion to growth industries and a few central places (appr. 30, instead of the previous 300) in order to attract firms to certain areas and to achieve within a fairly short period a self-reinforcing growth (cf. HÜSCH 1979) as well as to the concept of the urbanization of rural regions which is similar in principle. There is a possibility that inter-regional disparities could be reduced thereby, but on the other hand this policy will certainly aggravate regional disparities.

Of some interest is the proposal by GANSER (1977) to shift within a long-term development strategy those institutions with supra-regional functions to peripheral structurally weak areas, thereby utilizing the advantages of the regional division of labour for peripheral rural regions. This strategy implies above all the creation of less capital-intensive jobs in administration, management, research and development, the location of private services (wholesale trade, business consulting firms, planning, etc.), and the enlargement of institutions of higher education and of government boards. As such defined jobs are predominantly found within the Public Service, even this proposal is hardly realizable in view of the financial situation of the state budgets.

Apart from the securing of infrastructural minimum standards it will probably become necessary to make financial transfers to those areas which are no longer attractive as locational areas for jobs in the secondary or even tertiary sector (cf. SÄTTLER, op. cit., p. 175). However, such transferable incomes can only be produced by the productive agglomerations, if these remain internationally competitive.

BIBLIOGRAPHY

Beirat für Raumordnung. Empfehlungen vom 16. Juni 1976.

BODENSCHATZ, H./R. HARLANDER: Industrieansiedlung im ländlichen Raum. In: Stadtbauwelt 43, 1974, p. 197-201.

BRÜSSE, U.: Zur begrifflichen Klärung und Abgrenzung des Konzepts funktionsräumlicher Arbeitsteilung. In: Raumforschung und Raumordnung, 35 (1977), p. 97-102.

DAVID, C.-H.: Stellungnahme zu den Konzepten "großräumige Zuweisung von Funktionen" und "funktionsräumliche Arbeitsteilung im Bundesgebiet" aus verfassungsrechtlicher Sicht. Typescript, Celle 1981.

DEITERS, J.: Industrialisierung ländlicher Räume gescheitert? In: Geographie heute, No. 2, 1980, p. 57-58.

GANSER, K.: Die Bedeutung von Einrichtungen mit überregionalen Funktionen für strukturschwache Räume. In: Landwirtschaft und Landentwicklung in peripheren Gebieten. Hannover 1977, p. 49-66 (= Schriftenreihe für ländliche Sozialfragen, No. 78).

HAMPE, J.: Funktionsräumliche Arbeitsteilung aus ökonomischer Sicht (Typescr.) (To be published in: Funktionsräumliche Arbeitsteilung im Bundesgebiet. Vol. 1).

HÜSCH, F.: Industrielle Entwicklungsmöglichkeiten für strukturgefährdete ländliche Räume. In: Strukturgefährdete ländliche Räume. Hannover 1979, p. 89-104 (= Veröff. der Akademie für Raumforschung und Landesplanung, Forschungs- und Sitzungsberichte, Vol. 128).

HÜBLER, K.-H.: Raumordnungskonzepte und ihr Realitätsbezug. Zu einigen Ursachen des Vollzugsdefizites beim Bund. In: Innere Kolonisation, 29 (1980), p. 167-170.

HÜBLER, K.-H./E. SCHARMER/K. WEICHTMANN/S. WIRZ: Zur Problematik der Herstellung gleichwertiger Lebensverhältnisse. Hannover 1980 (= Veröff. der Akademie für Raumforschung und Landesplanung, Abhandlungen Vol. 80).

MARX, D.: Zur Konzeption ausgeglichener Funktionsräume als Grundlage einer Regionalpolitik des mittleren Weges. In: Ausgeglichene Funktionsräume. Hannover 1975, p. 1-18. (= Veröff. d. Akademie für Raumforschung und Landesplanung, Forschungs- und Sitzungsberichte, Vol. 94).

MIETH, W.: Zur Bedeutung regionaler Arbeitsmärkte für die Raumordnungspolitik. In: Raumforschung und Raumordnung, 36 (1978), p. 215-219.

Raumordnungsbericht 1978. Deutscher Bundestag. Bonn 1978. Drucksache 8/2378.

Raumordnungsprogramm für die großräumige Entwicklung des Bundesgebietes (Bundesraumordnungsprogramm). Bonn 1975 (= Schriftenreihe "Raumordnung" des BM für Raumordnung, Bauwesen und Städtebau).

SÄTTLER, M.: Ökonomische Perspektiven und ihre Folgen für die Raumordnungs- und Regionalpolitik. In: Innere Kolonisation, 29 (1980), p. 173-175.

SÄTTLER, M.: Standortpotential des Rhein-Neckar-Raumes und Entwicklungsmöglichkeiten ausgewählter Industrien dieses Verdichtungsraumes. Hannover 1980 (= ARL-Arbeitsmaterialien, No. 41).

TAUBMANN, W.: Funktionsgruppe Umschlagzentren. See- und Flughäfen. Typescript. (To be published in: Funktionsräumliche Arbeitsteilung im Bundesgebiet, Vol. 2).

THOSS, R.: Zurück zur passiven Sanierung? In: Innere Kolonisation, 29 (1980), p. 178-182.

THOSS, R.: Großräumige Funktionszuweisungen und Ausgeglichene Funktionsräume. Typescript. Celle 1981.

WEYL, H.: Funktion und Wirkungspotential der Raumordnung. Hannover 1979 (= Veröff. der Akademie für Raumforschung und Landesplanung, Abh. Vol. 79).

WEYL, H.: Ist das raumordnungspolitische Ziel der "wertgleichen Lebensbedingungen" überholt? In: Die Öffentliche Verwaltung, 33 (1980), p. 813-820.

WEYL, H.: Sozialstaatliche Komponenten raumordnungspolitischer Zielvorstellungen. In: Innere Kolonisation, 29 (1980), p. 165-167.

WEYL, H.: Das Konzept "Funktionsräumliche Arbeitsteilung im Bundesgebiet". Typescript, Celle 1981.

WITTE, G.: Möglichkeiten und Grenzen der Verwendung von gesellschaftlichen Indikatoren in der raumordnungspolitischen Planung der BRD. Gesellschaft für wirtschafts- und verkehrswissenschaftl. Forschung e.V., Bonn 1976).

H A N S   K E R N

REPRESENTATION OF SOFT AND HARD DATA FOR URBAN AREAS-CASE STUDIES IN BERLIN AND KARLSRUHE*)

From:

BAHRENBERG, GERHARD und STREIT, ULRICH (Eds.):
German Quantitative Geography. Papers presented at the 2nd European Conference on 'Theoretical and Quantitative Geography' in CAMBRIDGE 1980. Münstersche Geographische Arbeiten 11. Schöningh, Paderborn 1981, p. 155 - 169

*) With one map in the cover

Author:

Hans Kern

Geographisches Institut
Universität Karlsruhe

Kaiserstr. 12

7500 Karlsruhe 1

ZUSAMMENFASSUNG

Das Gegensatzpaar "weak data" und "hard data" bezieht sich in diesem Bericht auf planungsrelevante, räumliche Informationen, die in Stadtgebieten gewonnen werden.
Ob dabei Daten als "weak" oder "hard" zu verstehen sind, entscheidet sich auf zwei unterschiedlichen Ebenen:

- Art der Datengewinnung
  "hard data" als Ergebnis von standardisierten Erhebungsmethoden, die zeitlich entweder in kurzen Abständen anfallen (z.B. Daten aus dem Einwohnermeldewesen) oder unregelmäßig bzw. mit größerem zeitlichen Abstand gewonnen werden, dann aber durch den Umfang des erhobenen Materials herausragen (z.B. Volkszählungen). Die Erhebung dieser Daten ist zumeist gesetzlich geregelt.

  "weak data" als Ergebnis von Sondererhebungen.
  Sie werden zu aktuellen Fragen durchgeführt, beziehen sich teilweise nur auf bestimmte Stadtgebiete oder sind zumindest nicht landesweit koordiniert.

- Typ der Information
  "hard data" sind danach Informationen, die mit objektiven Maßstäben gewonnen werden (z.B. Alter einer Person, Einkommen eines Haushalts, überwiegende Nutzung einer Fläche, Zustand eines Gebäudes).

  "weak data" beinhalten Einstellungen, Motive, Wünsche von Personen.

Hinsichtlich der kartographischen Informationsdarstellung werden zwei Projekte besprochen, die wegen des Umfangs des Erhebungsmaterials oder wegen der Komplexität der Untersuchung den Einsatz der automatischen Datenverarbeitung erfordern.
Damit ergeben sich folgende Fragen:

- Vorteile/Nachteile der computer-unterstützten Kartenerstellung
- Anschluß der automatisierten Kartenerstellung an die Aufbereitungs- und Analyseverfahren
- Einfluß der Analysemethoden auf das Kartenbild
- derzeitige Möglichkeiten der Kartengestaltung mit automatisierten Verfahren.

Die gezeigten thematischen Karten entstanden im Zusammenhang mit den Projekten "Sozialplanbefragung in Berliner Sanierungsgebieten" und "Funktionale Abgrenzung der Karlsruher Innenstadt".

INTRODUCTION

Computer cartography - in the following, mainly thematical computer cartography is addressed - has increasingly opened up new application areas for itself over the past years (ref. Laboratory for Computer Graphics and Spatial Analysis [11]). This process has captured less the traditionally non-cartographical areas[1], but it has rather been the case that step by step, manual procedures have been replaced by automated ones. Step by step in as far as equipment and programs become available and efficient, step by step also in as far as skilled personnel can be employed. The DP revolution in cartography has not taken place and will probably not do so (see Arnberger[1]). At least this applies to the graphical means of representation used. It can rather be observed how the development of cartography programs orientates itself on previous traditional standards and strives to attain their qualities.

The phase of printer cartography was therefore not a new beginning, but rather a step towards automation which was accompanied by loss of quality. Even if printer cartography is now obsolete, its principles and programming techniques - as demonstrated by the example of the ink-jet plotter from Lund- have not lost in topicality. There is a widespread belief that the matrix principle as opposed to the vector principle is going to be the method of the future. In the following, a few aspects are to be highlighted that deal with the relation between type of data respectively type of data acquisition and automated cartographical representation. This point is of significance, because thematical computer cartography can be effective only in those cases where DP analyses and cartography can be combined. This contribution therefore deals with aspects of the interfaces between DP analysis and cartography.

SOFT AND HARD OATA, DEFINITION

The contrasting set of 'soft data' and 'hard data' relates to spatial and planning information which is acquired in municipal areas. Whether in this connection data are to be understood as 'soft' or 'hard', is determined on two different levels:

- Type of Data Acquisition
  'hard data' as a result of standardised survey methods which arise either at short time intervals (e.g. data from the field of residents' registration) or are collected at larger time intervals, but would then stand out by the volume of the acquired material (e.g. population census). The gathering of this data is in most cases legally regularized.
  'soft data' as a result of special surveys. They are carried out on topical questions and in part are related only to selected municipal areas. Usually they are not co-ordinated nationwide.
- Type of Information
  'hard data' is thereby information gathered on the basis of objective criteria (e.g.age of a person, income of a household, predominant utilisation of an area, condition of a building).
  'soft data' comprise attitudes, motives, wishes of persons.

Regarding the cartographical representation of information, two surveys are discussed which require the use of automatic data processing because of their sheer volume or because of the complexity of the survey. The following questions will thus arise:

- advantages/disadvantages of computer-assisted map production
- linking of automated map production to the editing and analysing process
- influence of the analysis methods on the cartographic representation
- present possibilities of map representation with automated procedures.

The thematic maps presented were created in connection with the two projects 'Social Plan Survey within the Redevelopment Areas of Berlin' and 'Functional Boundaries of the City of Karlsruhe'.

INFLUENCE OF THE DATA SURVEY ON CARTOGRAPHICAL PROGRAMMING SYSTEMS

In the previous section, a distinction was made between soft and hard data depending on the type of their acquisition, the data survey. The question arises now whether in fact any influence is exerted by the data survey on the computer programmes for cartographical representation. In my opinion, the nature of the data survey is of considerable significance for the cartographical representation. In the case of hard data surveys, the cartographical programmes must fit smoothly into the total concept of data collection, data storage, data retrieval, data analysis, and data representation. With the cartography programs, this becomes noticeable in rather pronounced standardising restrictions, e.g. the placement of headers may be confined or the structure and placement of the legend may be pre-determined.

This would indicate then - to stay with the example - that headers and legends serve merely as internal information, that is to rule out possible errors, and that the maps must subsequently be manually retouched for later printing by a cartographer.

There is possibly one further deciding factor for standardisation: on integration of cartography into an information system for spatial observation (in Berlin, this system is called structure and planning data base [19]), it must be ensured also that the user can obtain adequately expressive cartographical representations despite little previous cartographical knowledge. In that case the maps fulfill the function of working maps. As a rule, they are single-layer maps, the spatial units are fixed, or there are only few different spatial levels of aggregation. Multi-layer effects are obtained by multi-colour overprinting.

This coincides with the observation that at statistical offices plotters are in general no precision instruments; whereas for example the land registry offices of a town already have precision drafting equipment at their disposal, the statistical office which is charged with the task of spatial observation, merely possesses a fast drum plotter.

We can therefore observe that the standardizing restrictions imposed on cartographical representations of hard data surveys are stipulated on the one hand by the integration of computer cartography into information systems for spatial observation, and on the other by the insufficient cartographical knowledge of the user of the information systems.

For the soft data survey as mentioned earlier, e.g. special surveys and inofficial surveys, the circumstances are slightly different. Here, data analysis and graphical representation are as a rule not so closely interlinked. Frequently, computer cartography is even carried out separately in time and organisation. The maps obtain thus a more individual appearance. Though in most cases a disadvantage is that a feedback of a working map, and the spatial structures or processes it displays, to another new data analysis only can be effected with considerable effort.

As a consequence of these considerations, two points ought to be emphasised:

- a good cartography system on the one hand must be integrable into a spatial information system, but should, on the other hand, be usuable as an independent unit, too.
- the users should possess an elementary cartographical knowledge beside the ability to cope with regional statistical evaluation processes.

SOFT RESPECTIVELY HARD INFORMATION

As the influence of the data survey on the possibilities of cartographical representation has already been discussed, the question to be asked now is: what importance does the type of information have for the thematic map?

Whereas hard information can be objectified, e.g. contains intersubjectively verifiable details, soft information in particular concerns attitudes and motives of persons.
Official statistics contain as yet almost exclusively hard information (an exception are election statistics).

Time-dependent changes of hard, person-related data are more likely to be caused by long-term trends, whereas short-term variations are more important for soft data.

Hard data is measured on the nominal (e.g. sex), ordinal (e.g. education), and metric (e.g.age) level. For soft data, the metric level is reached at best via a scale derived by multivariate statistics (e.g. scalogram analysis). For soft data, one is interested, in the main, in small-scale representations. Whereas the natural population development for example still can be estimated fairly well even on a regional basis from hard data, one is trying to explain the migration changes, which are important for municipal areas, by means of motivation research.

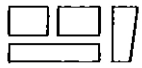
Soft information is gaining an ever increasing importance in the social sciences and in the economic sciences, too. The demand for small-scale representations is also on the increase. As the tasks are carried out with OP assistance in the majority of cases, we have here the ideal conditions for the use of computer cartography. Mainly required are representations of circles divided into sectors and representations with in part 30 and more easily distinguishable symbols (the need arises therefore for an automatic symbol generator).

TYPICAL CARTOGRAPHICAL TASKS OF TOWN AND REGIONAL PLANNING AND SUGGESTIONS FOR THEIR AUTOMATED SOLUTION

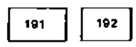
Typical cartographical tasks from the field of town and regional planning will be presented in this section and suggestions will be made regarding their automated solution. This is in fact an excerpt from the 'Extended Requisition Catalogue of Plotting Programs' of the Statistisches Landesamt Berlin [18,5]. The individual tasks have to be considered as cartographical elements. By varying the combination of these elements, planning maps are obtained which differ by their method of representation used and their content. The part of the catalogue of interest here differentiates between topographical bases, cartograms as representation of a numerical value or several numerical values per reference unit.

1. Topographical Basis

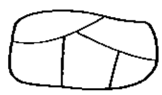
1.1 Block lay-out with street interspace (net blocks) → blocks defined as closed polygons or processed using segments.



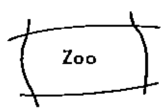
1.2 Same as 1.1, but with block numbers → block numbers have been entered as polygon names, which are placed at the 'optical midpoint'. The polygon name may be an arbitrary text (see map 2).



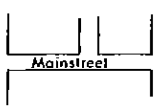
1.3 Lay-out for statistical areas, districts, traffic units → statistical areas, etc. defined as closed polygons. Alternative: segments defined as open polygons, attributes allocated to the segments identify area affiliation and line type. Automatic concatenation of segments to closed polygons.



1.4 Same as 1.2, but with numbers respectively names of reference areas → solution as 1.2



1.5 Labelling of main roads and similar within block lay-out → labelling with symbol-directive, possibly with blanking option



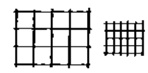
1.6 Variable boundary lines within hierarchically structured lay-out, e.g. for statistical areas and districts → boundary lines are entered as segments, attributes allocated to segments define hierarchical position of the boundary and thus its graphical realisation.

1.7 Aggregation of networks → aggregation and selection can be realised in case of networks provided the network has been defined by segments.

1.8 Partitioning of lay-out by means of straight lines → this function is described as windowing or scissoring.

1.9 Precise drafting of map content on to topographical basis → drafting of register marks. Additionally, precise fit is a demand made on the hardware and material used.

1.10 Grid square lay-out with different mesh densities → network construction with LINE directive.

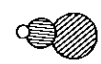


2. Cartograms

2.1 Circles of proportional values → CIRC directive.

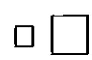
2.2 Smaller circles overlapping larger ones → hidden line algorithm.

2.3 Shading of circles with different hatchings → hatchings are defined by a hatching parameter,which is processed by the CIRC directive.

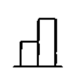


2.4 Overlapping circles with halo effect → border width of halo effect defined by HALO directive

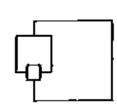
2.5 Squares of proportional value → own directive. The directive should be so prepared, that further symmetrical symbols are obtainable.



2.6 Bar-diagrams of proportional length-value → see 2.5.



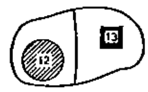
2.7 Overlapping squares → the hidden line algorithm handles all definable, graphical elements, e.g. regular geometric symbols (circles, squares), circles divided into sectors, any closed polygon (areas), arrow symbols, any open polygon (lines).



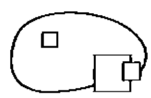
2.8 Shading of squares with different hatchings → the algorithm for hatching relates to all areal graphical elements, e.g. circles, squares, sectors, arrows, any closed polygon.

2.9 Shading of bars with hatchings → same as 2.8.

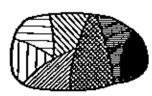
2.10 Leaving of blank spaces within hatched areas for key-numbers and names → when writing text or numbers, blanking rectangles are automatically created. The hidden line algorithm handles the blanking.



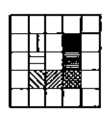
2.11 Leaving blank spaces within lay-out borders for overlapping symbols → same as 2.7.



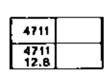
2.12 Shading of polygons with up to seven hatchings of gradated darkness → same as 2.8. Variation of distance and angle of hatching parallels, cross hatching, break-down of hatching parallels into regular dash or dot sequences, displacement of dashed lines, instead of dots plotting of 'meadow' and 'forest' symbols.



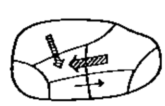
2.13 Shading of grid squares with up to seven hatchings of gradated darkness → same as 2.12.



2.14 Referencing of grid square lay-out with values or double values → symbol-or number directive. Values can also be defined as names of grid squares.

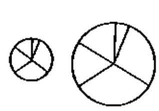


2.15 Arrow-cartograms on top of polygons: directed arrows of variable width → arrows are defined by initial and final points as well as by arrow width and hatching parameter.

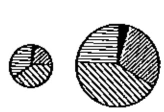


3. Cartodiagrams

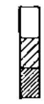
3.1 Proportional circles divided into sectors → CIRC directive with up to 10 sectors.



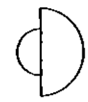
3.2 Same as 3.1, but with variable hatchings → each individual sector is defined by its initial angle and the hatching parameter.



3.3 Subdivided bars with hatchings → HIST directive.

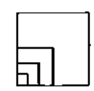


3.4 Proportional half circles facing each other → 'gateaux' of any chosen initial angle, each one with up to 10 sectors.



3.5 Same as 3.4 with subdivision into sectors with different hatchings → same as 3.4.

3.6 Nested squares with the same lower left-hand corner point → The definition is facilitated with the aid of a special directive in the fashion of the CIRC-directive.



3.7 Bar diagrams on a topographical basis → HIST directive.

Further desirable results of computer cartography systems could be:

- 3-d representation of spatial distribution [12]
- representations using isolines [4]
- continuous varying area-hatchings [3]
- representations on the basis of block-edges
- transportnets [8]
- automated color separation [9]

LINKAGE OF DP ANALYSIS WITH COMPUTER ASSISTED REPRESENTATION

According to the current situation, there are - separate from one another - efficient statistical analysis packages [7,15] and advanced cartography packages [1,11] . Occasionally, the statistical packages allow for rudimentary graphical representation on a high-speed printer, e.g. histograms and scattergrams. Equally, some of the cartography systems permit simple statistical calculations. Linking of such packages is necessary to obtain an efficient process from the initial data survey to the graphical finishing.

Point of departure for statistical packages are the attributes gathered from the inquiry objects (e.g. inquiry objects ≙ households; attributes ≙ age of head of household, number of household members). With the aid of logical and arithmetic operations, new attributes (variables) can be obtained (data manipulation), correlations between attributes can be calculated, and the deviance of the individual inquiry object to the general context can be defined (correlation and regression analysis) or factors can be extracted and the factorscores for the individual inquiry object can be stated (factor analysis). A large range of multivariate methods is available.

These examples intend to point out that statistical packages make on the one hand statements about the overall inquiry (e.g. in the form of correlation coefficients) and provide on the other hand new variables (residues, factorscores, etc.) for the individual inquiry objects.

In addition, new inquiry objects of a higher level can be defined by means of aggregation of inquiry objects (e.g. transition from inquiry level 'household' to inquiry level 'block' by aggregation of all households of a given block). If a spatial location is attached to the inquiry objects of an aggregation level, it suggests itself to cartographically represent the starting variables, the aggregated or the derived variables. To this end, the desired variables are written to a scratch file by the statistical package, the necessary spatial-geometric information is added (merging) and the cartography program then drafts the map. The linkage consists therefore of: output of spatial information by the statistical system, allocation of geometric information (merging function of an auxiliary program), processing for input to a cartography program.

POSSIBILITIES AND PROBLEMS OF LINKAGE

Initially, some types of maps will be explained which can be the end-result of the linkage described in the last section. Subsequently, details are pointed out that can limit the effectiveness of such a linkage respectively can require further development of computer cartography programs, particularly in the direction of interactiveness.

Types of Map

1 Diagrammatical Map with Uniform Symbols (e.g. circles) → The symbol size varies continuously , the symbol area is hatched. The hatching[2)] corresponds to a nominal or ordinal attribute. The realisation of the linkage is easy to effect as the 'spatial-geometric information' corresponds to the centre of the symbols. If these centres are for example the centre of a block, they are obtained either by direct manual digitizing or via a program 'optical midpoint' by the block vertices.
The demands on the cartography program are: drafting of simple symbols, hatching of symbols, overlaps, transformation of co-ordinates, parameters for the control of the relative symbol size.

2 As 1 above, but with different symbols (e.g. circles, squares, and hexagons) and possibly continuous varying hatchings [3] → The linkage does not create any new demands. The problem of overlaps must be solved in a more universally applicable way. The possibilities of hatching must be more comprehensive.

3 Diagrammatical Maps with Circles Divided into Sectors → Linkage and demands on the cartography program as above.

4 Diagrammatical Maps with Left and Right Halfcircles Divided into Sectors → This type is suitable for example for the representation of phases of a development at two separate points in time or to compare structures in two partial entities or to contrast actual against potential conditions. Linkage is simple. On the overlap problem, any type of 'gateaux' must be feasible.

5 Choropleth Maps with Fixed Areas within the Map Series → The areas are provided with hatchings. Hatchings correspond as a general rule to a nominal or ordinal attribute which could also have been ob-.

tained by classification (e.g. density of population). Continuous varying hatchings are possible. Linkage is more difficult as 'spatial-geometric information' consists of areas being described in the form of closed polygons. The areas can in this case be simple connected areas, but also multiple connected areas.
Appropriately, the closed polygons are built up from segments (i.e. vertices connecting two nodes and separating areas with different attributes). For this purpose special programs are used [10,13,17].

The cartography program must be able to hatch any, even multiple connected areas. Hatching of multiple connected areas can be viewed as hatching of simple connected areas under consideration of overlaps by 'islands'. Efficient algorithms have been developped [2,6,14,16].

6 Choropleth Maps with Variable Areas within a Map series → Whereas for the previous type of map, the DP analysis merely provided for each area a hatching parameter, in this case for each map new areas are built up from the smallest spatial units (basic areas) respectively from the segments defining them. Then a hatching is allocated for these new achieved areas.
For example applications could be the boundary definition for functional or homogeneous areas or - more generally - regionalisations. In these cases the segments bordering the areas are no longer fixed, but are altered by the analysis process.
Whereas the possibilities for aggregation with the aid of statistical packages already has been pointed out, here there is a necessity for spatial aggregation. The procedure is as follows: with the statistical package, the areas to be aggregated are defined as well as the hatching parameter which the aggregated area will receive. Subsequently, those segments are eliminated which separate aggregated basic areas. From the remaining segments, the aggregated areas are extracted[3]. The difficulty here is that it cannot be completely excluded that the aggregated areas are multiple connected. Therefore, the closed polygons must be examined regarding inclusion (see map 1).
With regard to the linkage, the program already mentioned under item 5) as helpful for the processing of segments has now become vitally necessary. It must be able to determine an interior point for any closed simple connected polygon and must be able to examine simple connected polygons regarding inclusion. The performance of the cartography program corresponds with that mentioned in item 5).

7 Maps for the Representation of Interactions → 'Inquiry objects' within the framework of the statistical packages are in this case spatial interactions (e.g. moves between different parts of the town). Linkage is comparatively simple. The cartography program must provide arrow symbols and handle overlaps (see map 5).

The different types of map, classified in this context by the various data processing requirements, can also be combined with one another. Problems would then arise though, because of complicated overlaps that can lead to large core requirements, long processing times, and priority conflicts (see maps 3 and 4).

The maps described under items 1 - 7 contain the thematical statement. Still missing for a complete map however are for example the map margin, legend, scale, captions, and partly (1 - 4, 7) the topographical orientation. Without these items, automation of map production (analysis - linkage - drafting) easily can be effected as far as time and effort are concerned.

The items mentioned above can be taken into consideration by:

- subsequent manual finishing
  Extensive retouching may become necessary which tends to nullify the efficiency of the automated draft.

- standardising specifications
  Title and legend of the map are for example always entered in a column on the right of the map. Strong standards make the map appear too stereotype. Compositional and esthetical aspects are neglected.
- integration into the automated draft
  Preparations are extensive and expensive. For example for the placement of texts, the co-ordinates of the starting points (lower left-hand corner of the first character) must be taken up. For the characters themselves there are as a rule no suitable type faces available. The advantage over manual processing for example is that the characters can be placed in blank surrounds. An advantage compared to the standardisation is the possible individual ductus of maps. This is the application area for interactive cartography programs. It is their aim, beside the definition of an optimal relative symbol size and the prevention of unfavourable overlaps, to complete the thematical draft to an expressive, individual map (respectively series of maps.

CASE STUDY IN BERLIN

Point of departure for the cartographical representations were polls in various re-development areas in Berlin. Apart from the demographical attributes, the questionnaires contained details about social status, net income of a household, present housing situation, and the desired housing situation.

The questionnaires were subdivided into a household questionnaire for each household and a residents' questionnaire for each person over 16 years of age (the assessment of combinations of residents' attributes with household attributes caused particular difficulties). As the re-development measures are orientated on the actual premises, it was necessary to align the tables, and wherever possible the cartographical representations as well, with this smallest spatial reference unit. Each premise examined was therefore represented by a circle divided into sectors. The size of the circle was computed in proportion to the number of households residing on the premises, whereas the size of the sectors reflect the percental distribution of a single attribute.

It was impossible to select the mid-points of the 'actual' premises as centre of the circle, because of the excessive overlaps of circles. Therefore, the following aspects were decisive for the determination of the centre of the circles:

- the circles ought to fall totally within the block
- only slight overlaps were permitted
- deviation from the 'actual' location of the premises should be kept to a minimum.

Free play in the placement of circle symbols was assured in that not all premises of a block vertice had to be investigated.

As we had no algorithms at our disposal for the solution of this localization problem - quite probably they do not exist yet - the definition of the centre of the circles had to be effected manually. It is quite evident that

- firstly, the expense is only acceptable for soft data surveys, and that
- secondly, these tasks can be dealt with more efficiently with interactive map processing (interactive map processing in those cases where formalised algorithms are not available!).

Representation with simple circle symbols divided into sectors was not sufficient in the following two cases:

- as far as significant differences were to be expected between the native German population and the foreign population (for example with regard to the age structure), these differences should be directly readable from the map.
- comparison of the actual housing situation with the desired housing situation should be possible 'at a glance' (Map 4).

In the first case, the left half circle represented the foreign population and the right one the native German population.

The size of the circle is directly proportional to the appropriate populations. In the second case, half circles of identical size were selected: the left half circle portrayed the present, the right half circle the desired situation.

The map program contained per re-development area (respectively per re-development phase) 17 maps. The smallest poll encompassed 151 households, the largest 2161 households. In total, the poll results of approximately 5300 households and 16000 persons were evaluated. Despite the great preparatory expense, it can safely be stated that the linkage of evaluation (with SPSS) and cartographical representation was an efficient one.

CASE STUDY OF KARLSRUHE

Whereas with the Berlin project, it had been possible to gain further experience concerning statistical analysis and its subsequent cartographical representation in the field of soft data, the possibilities of hard data representation were to be assessed with the Karlsruhe project. This study is also useful, because as yet only few reports are available for the automated cartographical processing of mass statistics for municipal areas.

With regard to town planning main emphasis is placed on the one hand on cartographical processing of the results of population censi on a small-scale basis, and on the other hand on the analysis of processes taking place over a certain period of time.

The most important steps are:

- creation of an open data base. Stored in it are the population census results on a block basis. The surveys are spatially classifiable via the block number. The data base contains data from the population census 1970/71 and it is to be enhanced with the results of the next population census.
- production of a digital town map as of 1970: The town map contains approximately 1800 blocks which were each digitized as closed polygons.
- digital town map 1980: The differences between the maps 1970 and 1980 consist of the extension of the municipal area due to incorporations, changes in the systematics of block numbering, changes in the block arrangement due to large-scale road construction projects, municipal re-development, and the creation of new residential districts.
- linkage of statistical analysis and cartographical representation: It was possible to simplify this item, which will lead finally to the thematic maps, to such an extent that participants of lectures and the computer cartography course of the Institute in October 1979 were able to draft thematic maps.

The tasks necessary for the map producer are to be highlighted with the aid of the attached example:

- the subject of the map is defined. In this case, direct public services are to be considered.
- the variable which is suitable for the representation is defined. This is therefore the operationalisation of the subject. The variable is defined here as an addition of variables contained in the data base. It might also be a relativation of a variable, with regard to population size or block size, or a variable derived with a multivariate process.
- a frequency distribution with fine classification (e.g. 30 classes) is calculated. Thus an overview is obtained of the distribution of an attribute to be represented. In the example, it is a strongly left-slanted distribution (blocks with few people employed in the direct public services predominate). An exponential classification would therefore be suitable in the case of this example.
- the final classification is defined.
- a provisional definition of the hatching parameters is effected. A predefined hatching may be used here.
- a first working map is drafted in black and white. This is done on the fastest possible drafting instrument, e.g. a

drum plotter. A microfilm plotter could even be used for small areas (e.g. the town centre), but for bigger areas the re-enlargement would not suffice to examine the quality of the graphical results.

- the working map is checked regarding the expressiveness of the represented variable, adequate classification, and correctly scaled hatchings.
- appropriate modifications are effected in the proceeding steps according to the outcome of the examination of the working-map.
- finally, a fair copy of the map is drafted. The colour plates were engraved automatically in this case, whereas the captions were entered manually.

The steps mentioned naturally are carried out in this or a similar form in the case of conventional map production. The difference lies in the considerable preparatory efforts (digitizing), which are paid for when a larger number of maps needs to be produced. The manual effort for the drafting of maps is however eliminated. The cartographer can therefore concentrate totally on the most suitable means of representation, because re-drafting is hardly limited at all by technical restrictions.

I would therefore like to close this report with the request to cartographers to utilise the techniques of automated map production - otherwise these tasks will in future be handled by the DP centres and this probably in an unsatisfactory rather than a satisfactory way.

FOOTNOTES

1) It is surprising though that large DP centres with their service functions awake a demand for cartographical representation in areas that so far have been alien to cartography. By nature, this demand is generated by people who are very open towards automated processes. It would therefore be desirable that the program documentation of DP centres contains a separate keyword for 'cartography'.

2) Hatchings are made up out of lines, dashes, or dots.

3) The term 'extraction' is used, because the process progresses in such a way that each segment must be extracted twice from the table of sorted segments corresponding to the two areas separated by the segment.

REFERENCES

1 ARNBERGER, E. (1979): Die Bedeutung der Computerkartographie für Geographie und Kartographie. Mitteilungen der Österreichischen Geographischen Gesellschaft, Vol. 121, Wien.

2 BRASSEL, K.E. and R. FEGEAS (1979): An algorithm for shading of regions on vector desplay devices, ACM-Computer Graphics (SIGGRAPH), Vol. 13, New York.

3 BRASSEL, K.E. and J.J. UTANO (1979): Design strategies for continuous-tone area mapping, The American Cartographer, Vol. 6, No. 1.

4 EGLOFF, P. und G. MAISS (1979): ISOFIX - Ein System zur grafischen Darstellung von Höhenlinien, Hahn-Meitner-Institut für Kernforschung, Berlin.

5 EVERS, K. (1979): Erweiterte Darstellungsmöglichkeiten bei thematischen Kartierungen statistischer Daten durch den Einsatz des computergesteuerten Zeichengeräts, Berliner Statistik-Monatsschrift, Berlin, Heft 8, 1979.

6 ENCARNACAO, J.L. (1975): Computer Graphics, München.

7 HELWIG, J.T. and K.A. COUNCIL (ed.) (1979): SAS user's guide, Raleigh.

8 HORN, K. (1976): EVA-regional, Programmsystem zur Erreichbarkeits- und Versorgungsgrad-Analyse auf der Grundlage regionaler Verkehrsnetze, DATUM e.V., Bonn.

9 KERN, H. (1978): ADP - Supported Color Separation, Nachrichten aus dem Karten- und Vermessungswesen, Reihe II, No. 35, Frankfurt a.M..

10 KERN, H. (1979): THEGIT, Computer-Kartographie-Kurs an der Universität Karlsruhe und an der ETH Zürich, Karlsruhe.

11 Laboratory for Computer Graphics and Spatial Analysis (1979): Mapping Collection, Vol. 1 to 11, Harvard University.

12 Laboratory for Computer Graphics and Spatial Analysis: PRISM, Harvard University.

13 MENNE, J. (1979): ANCER - A system for creation and management of geographic base files in local authorities, Proceedings of the 5th International Colloquium on Segment Oriented Referencing Systems, Bonn.

14 NAGY, G. and S. WAGLE (1979): Geographic data processing, Computing Surveys, Vol. 11, No. 2.

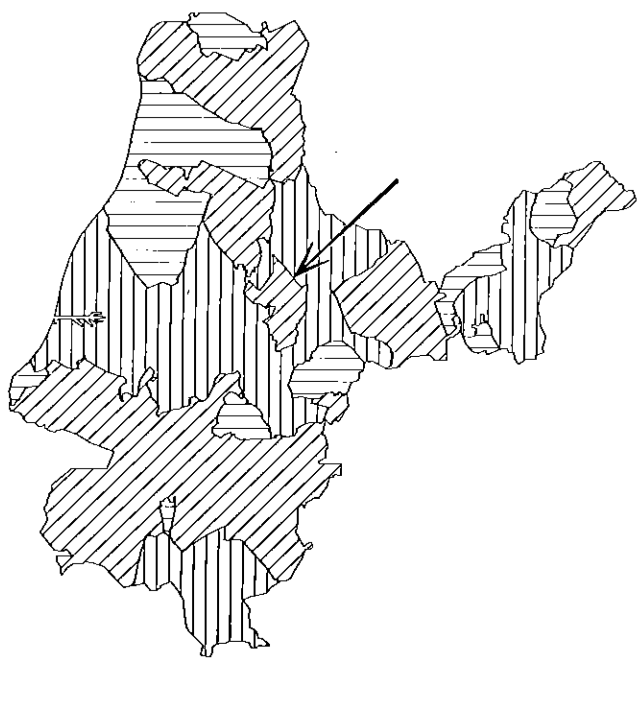
15 NIE, N.H. e.a. (1975): SPSS - Statistical Package for the Social Sciences, New York.

16 SHAMOS, J.J. and D. HOEY (1976): Geometric intersection problems, Proceedings of the Annual Symposium on Foundations of Computer Science, New York.

17 SILVER, J. (1977): The GBF/DIME system: Development, design and use, U.S. Bureau of the Census, Washington D.C..

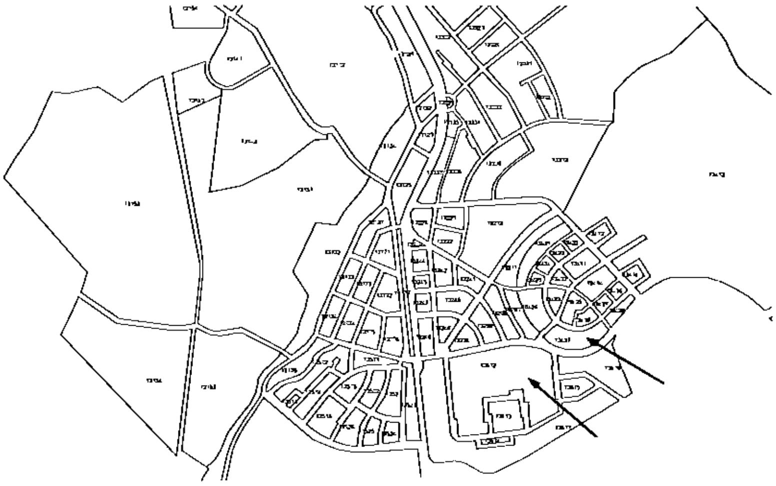
18 Statistisches Landesamt Berlin (1977): Erweiterter Anforderungskatalog für Plot-Programme, unveröffentliches Manuskript, Berlin.

19 Senator für Inneres (Hg.): Konzept für den Aufbau einer Struktur- und Planungsdatenbank im Rahmen eines Verwaltungsinformationssystems für Berlin (West), Berlin, 1974 (= Verwaltung in der Reform).



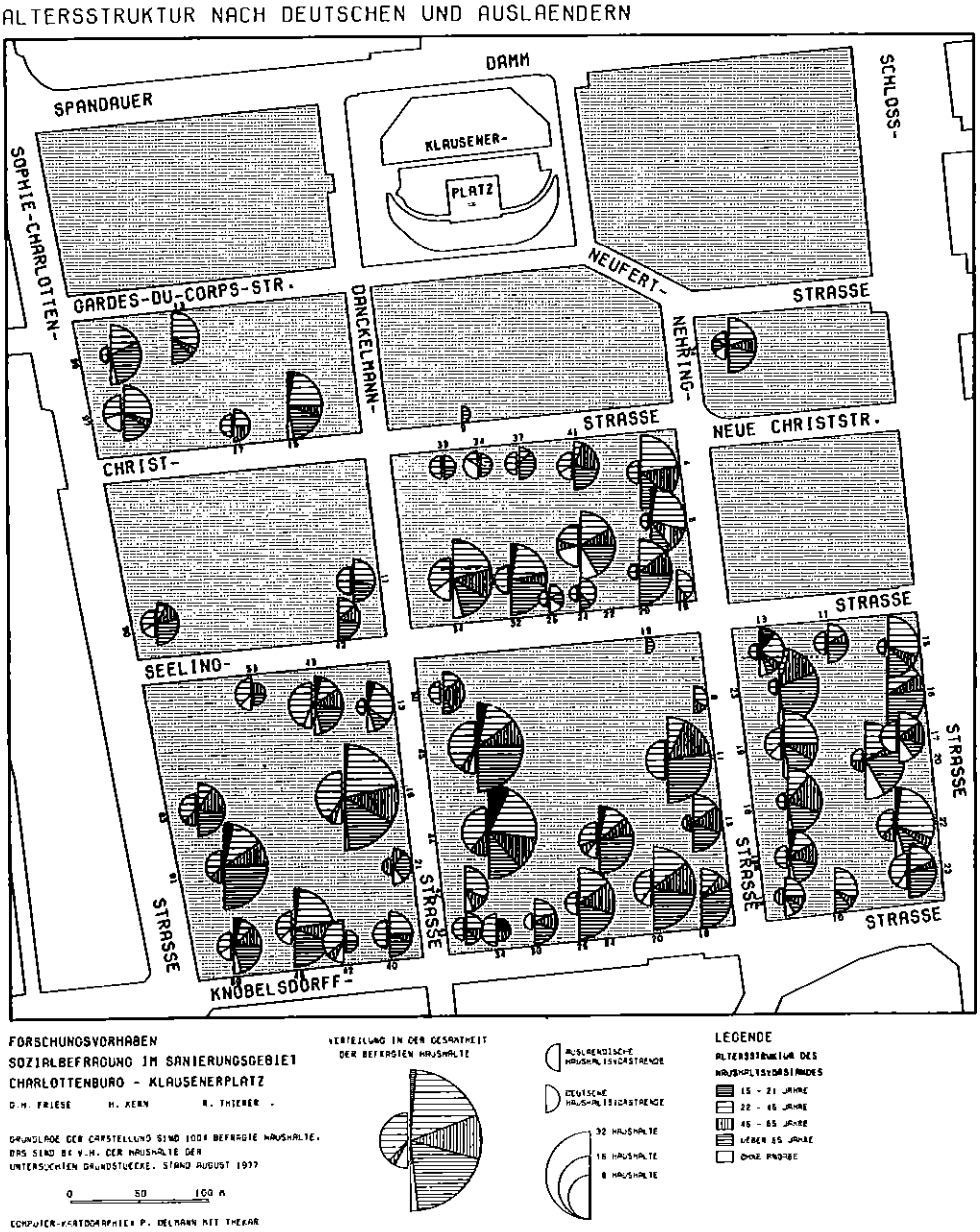
Map 1: AUTOMATED AGGREGATION, PROBLEM OF EMERGING ISLES, ENGRAVING

Map 2: KARLSRUHE, BLOCKS FOR CENSUS DATA, BLOCKNUMBERS ARE PLOTTED AT THE "OPTICAL MIDPOINT", ENGRAVING

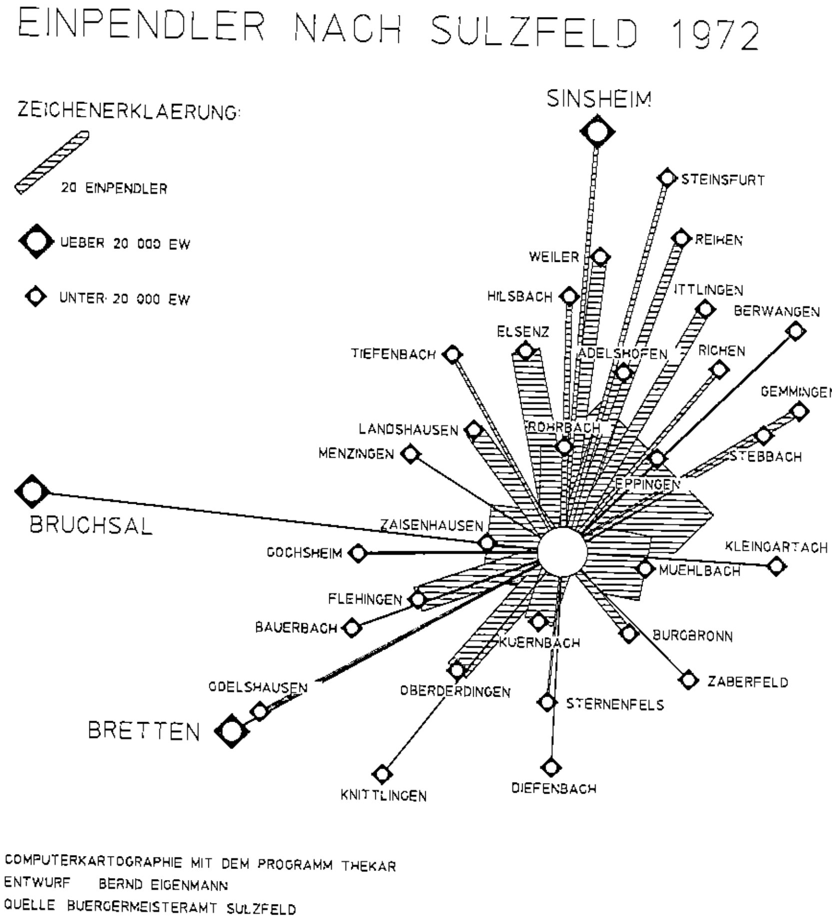


Map 3: AGE STRUCTURE OF GERMAN AND FOREIGN INHABITANTS, URBAN RENEWAL AREA Berlin-Klausener-Platz, Microplot

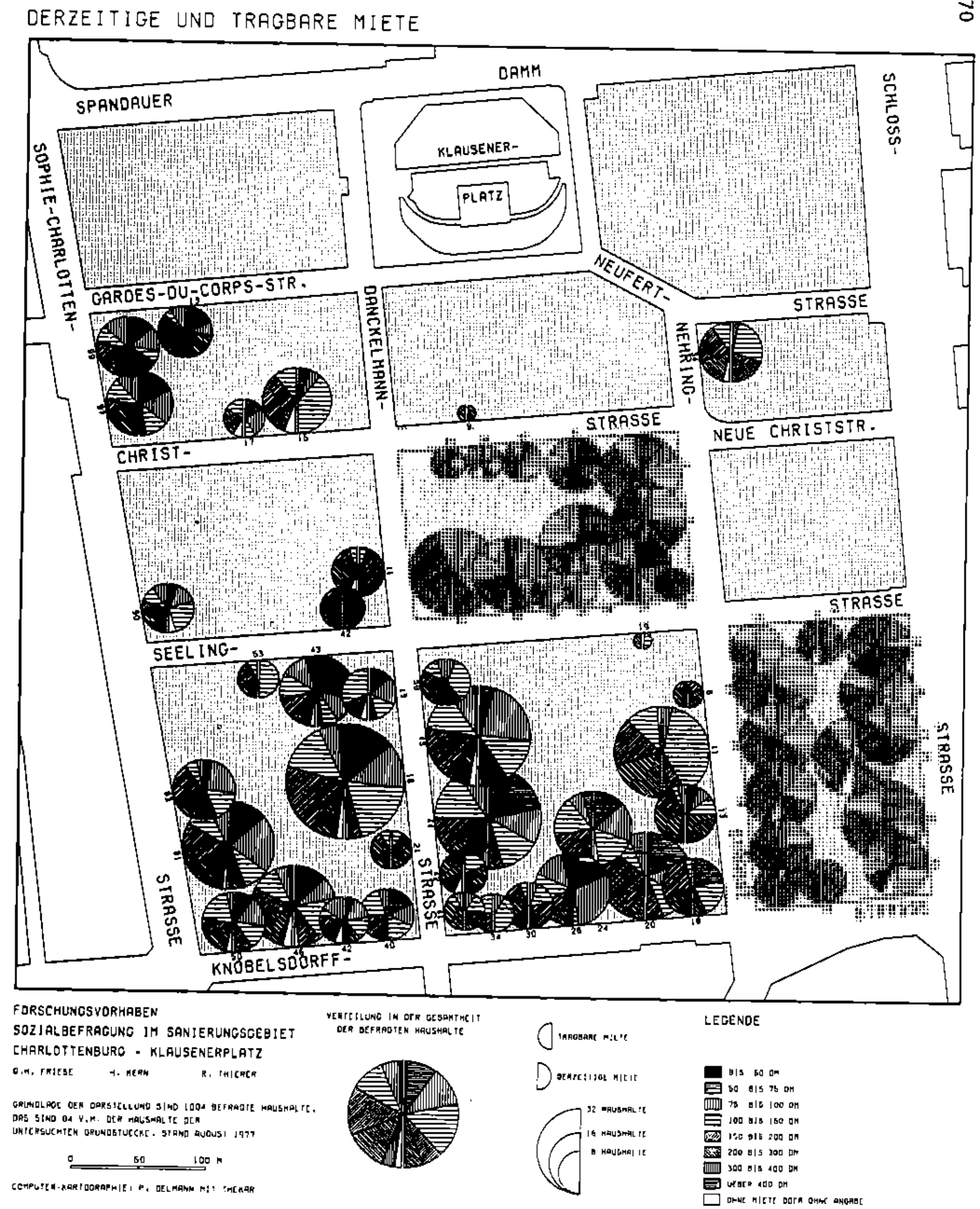
ALTERSSTRUKTUR NACH DEUTSCHEN UND AUSLAENDERN



Map 5: INTERACTION MAP, OVERLAPPING ARROWS, ARROWS TAKE CARE FOR TEXTS, ENGRAVING



Map 4: PRESENT AND DESIRED RENTS, URBAN RENEWAL AREA
Berlin – Klausener-Platz, Microplot



H U B E R T   B I S C H O F F und G E O R G   H E Y G S T E R

AUTOMATIC DIGITIZATION OF AREA-ORIENTED MAPS

From:

BAHRENBERG, GERHARD und STREIT, ULRICH (Eds.):
German Quantitative Geography. Papers presented at the 2nd European Conference on 'Theoretical and Quantitative Geography' in CAMBRIDGE 1980. Münstersche Geographische Arbeiten 11. Schöningh. Paderborn 1981, p. 171 - 180

Authors:

Hubert Bischoff
Georg Heygster

Rechenzentrum der
Universität Bremen

Bibliothekstraße

2800 Bremen 33

ZUSAMMENFASSUNG

Für die raumbezogene Datenerfassung und Datenanalyse eröffnen die Verfahren der elektronischen Datenverarbeitung in qualitativer und quantitativer Hinsicht neue Dimensionen. Große Mengen von Daten können gesammelt, gefiltert, gespeichert, verknüpft und als Informationen in Form von Tabellen und Karten verfügbar gemacht werden.

Als ein besonders langwieriger Prozeß erweist sich der Aufbau eines geographischen Informationssystems, da insbesondere die raumbezogene Datenerfassung, d.h. Koordination der im Informationssystem abzulegenden Objekte weitgehend manuell und damit zeitraubend und fehlerbehaftet durchgeführt werden muß.
In diesem Aufsatz soll in Form eines einführenden Überblicks ein Weg aufgezeigt werden, der den Flaschenhals der Datenerfassung im Bereich der Kartendigitalisierung überwindet.
Die automatische Digitalisierung von Grenzlinienkarten geschieht mit Hilfe von Raster Scannern unter Anwendung von Methoden der digitalen Bildverarbeitung. Es werden anhand eines Flußdiagramms die dafür notwendigen Schritte erarbeitet und erläutert.

1. INTRODUCTION

The use of computers and plotting devices in cartography increases the possibilities of map production during the last ten years. The advantage of computer cartography lies in its ability to store and retrieve information of large data sets.

Until recently maps were drawn or printed on paper, and the paper was both the storage medium and the display medium of the data (see BISCHOFF a.o. 1979). Beginning about twenty years ago, interest developed in the storage of maps in digital form, on magnetic disk or tape. In this form the paper copy was used only when a hardcopy output of the digitally stored data was desired.

The advantages of storing a map in digital form are impressive. First of all, it is a simple matter to copy the data and transmit them rapidly to remote locations without any loss of precision. Secondly, alterations of the data can be made easily, whether to correct errors or to incorporate additional information. Since the data are already in machine readable form, computer analysis or manipulation of map data is easily accomplished. Finally, maps for different purpose can be readily derived.

Before one may use a large data base for digital maps, one has to fill it with appropriate data. Any cartographic system needs the input of x-y-coordinates in digital form for processing procedures. That makes it necessary to collect data from an original or base map. There are several ways to do this; one can perform the digitization by hand which is a very long and errorprone work, especially if the map contains many details. Another way is the automatic digitization using a rasterscanner.

This article describes some of the techniques which are applicable to perform an automatic digitization of original maps. The literature on this problem is growing rapidly. Our primary aim is to present an overview because for most geographers the field is important but new and difficult to grasp.

2. MAP DIGITIZATION

The kind of data that may appear on a map can take different forms. The data can be classified as belonging to five major categories and any map is simply a set of overlays of data drawn from these categories (FREEMAN 1980).
These categories are

1) regions (which may overlap)
2) contour lines (without annotations), never intersecting
3) drainage lines (which tend to form tree structures terminating on the map boundary)
4) area-subdivision lines (which are used to represent political boundaries, road systems etc.).

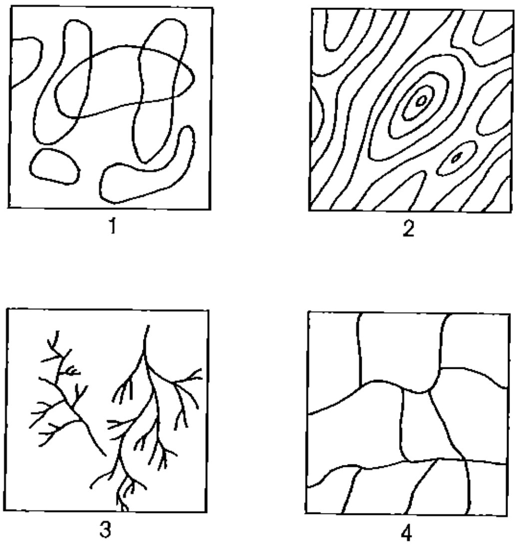


Fig. 1
Four categories of map data

The fifth category consists of annotation data, which may be combined with any of the other four categories. We will restrict our consideration to the fourth• category, that means area oriented maps.

The problem of translating the original map into a digitized set of data can be divided into five steps. Fig. 2 illustrates this process by means of a flow diagram.

2.1 QUANTIZATION

The first step in the translation process is the quantization of the original map data. Because the original map consists of analog data we have to perform an analog-to-digital conversion. The precision by which a map can be represented in digital form is limited by the degree of differentiation of the original analog medium and the ability to perform measurements on it (see FREEMAN 1980).
The quantization process depends on three decisions concerning the

(i) form of the quantization, i.e. the rules prescribing the quantization,

(ii) size of the quanta

(iii) approximation used to represent the quanta (FREEMAN 1980, p. 253).

There are at least two ways of performing the quantization:

1. the raster approach

2. the line - following approach

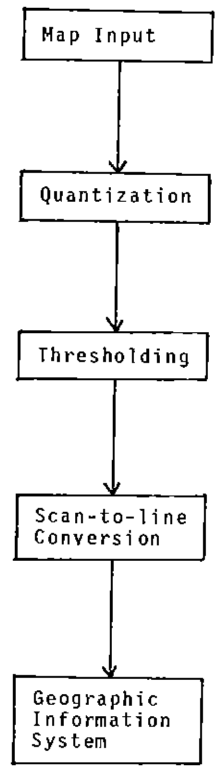


Fig. 2
Process of translation an original map into a digitized set of data.

2.1.1 RASTER APPROACH

In the raster approach the map is scanned over in a TV-line-by-line fashion and the density or luminance of each spot or pixel on the map is recorded by a photo detector. The analog-value of the luminance of each

pixel is converted into an integer number representing the grey-value of the corresponding element of the area. The raster representation preserves the spatial shape of a map and is a simple and easily handled form for communicating and redisplaying the map.
There exists a complete industry manufacturing raster scanners of various size and quality. The basic operation principle of a raster scanner needs a configuration consisting of a x-y-plotter, a photo-detector, an analog-digital-converter which digitizes the analog signal measured by the photo-detector and a projector which serves the map in form of a diapositive. Fig. 3 illustrates the configuration of a simple raster scanner.

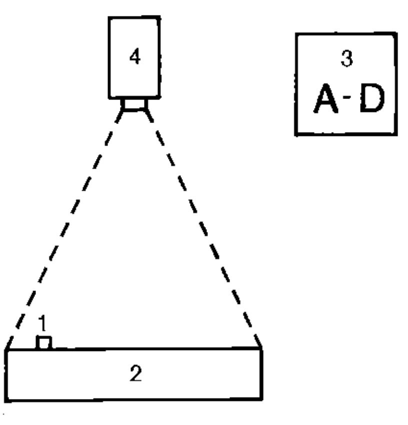


Fig. 3
Configuration of a simple raster scanner:
1 photo detector 2 x-y-plotter
3 analog-digital-converter 4 projector

The photo detector is moved by the x-y-plotter line-by-line taking up each pixel of the map-picture. The x-y-plotter is divided into an array of, for example, 512 * 512 pixels. Each spot of this array receives a grey value and is stored in a file for further processing. It is possible to digitize several maps of the same area and each map treats another thematical topic.

Some structuring of these map data can be realized by separating the data into distinct raster arrays, one for each topic and then creating raster overlays to obtain particular data combinations.
For example, one may create separate rasters for contour lines, for borders, highways, for land use etc.and then combine some subsets of these to obtain a desired digital map (FREEMAN 1980, p. 254).

2.1.2 LINE FOLLOWING APPROACH

The raster approach is not very efficient in terms of the computer memory used since it stores the grey values of all pixels. In area oriented maps one needs to retain only those (about) 10 % of the pixels which represent the area boundaries. Thus it is possible to save about 90 % of the memory space.

An alternative way of digitization is to encode the lines directly. A variety of such "lineal" representation schemes has been developed (BRADSHAW 1963). They can be obtained by so called automatic following algorithms, which encode the lines in the map in form of a series of (x,y) pairs in a cartesian coordinate system. The precision will be limited by the smallest differences in x and y that can be reliably detected in defining adjacent points on the curve.

Automatic line followers are usually flatbed plotters with an optical attachment (eye) that allows following the line under computer control up to its end. The finding of a new line and the marking of digitized lines is done manually. Therefore these systems offer rather an improvement of manual digitization than an automatic method (KANSY 1977).

In the following we will concentrate on the processing of data obtained by the raster approach.

2.2 THRESHOLDING

The result of the raster scan process (quantization step) is an array of grey-

levels (or greylevel matrix) in which each pixel is represented by a value between 1 and 256, for example. The next step to perform is the classification of each pixel into two categories. The first one contains those pixels which belong to the boundaries (lines) we are interested in; the second category contains the background pixels (representing the "white" area). The classification is done by thresholding the greylevel matrix. Usually the descision for the threshold-value is made on the basis of a greylevel histogram, which shows the absolute and relative frequency of each grey value of a particular map picture. If the histogram pattern is bimodal with two distinct peaks it is relatively easy to determine the threshold (see fig. 4) leading to a reliable classification. In other cases (e.g. monomodal histograms) it is necessary to fix the threshold heuristically.
After thresholding we get a binary matrix, in which each line (border) is described by a sequence of 1's and the background (area) by a cluster of 0's.

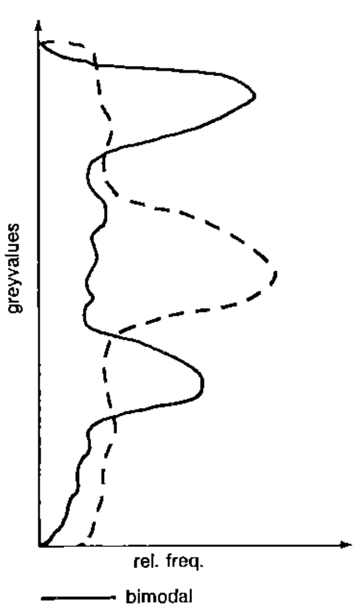


Fig. 4
Histogram types of greylevels

Thus all lines are detected, but some of them are broader than necessary.

2.3 LINE - THINNING

Broad lines in the binary matrix (described by 1's) have to be thinned out preserving their location and their topological properties. This line - thinning process is done by skeletonization of the objects in the binary matrix, i.e. reducing them to one pixel width (see fig. 5). This means that some 1-values in the binary matrix become 0-values. Numerous thinning algorithms are described in the literature (see PAVLIDIS 1980, STEFANELLI / ROSENFELD 1971, KREIFELTS a.o. 1974 and 1976, WOETZEL 1977, RIEGER 1978). Here we will concentrate on the principles of line - thinning by describing an algorithm from WOETZEL (1977).

```
ooo111ooooo1111oo        oooo1oooooooo1ooo
oooo111ooo111oooo        ooooo1ooooo1ooooo
oooo11ooo111ooooo        ooooo1oooo1oooooo
ooooo111o11oooooo        oooooo1ooo1oooooo
ooooo111111oooooo        ooooooo1o1ooooooo
ooooo111111oooooo        ooooooo11oooooooo
ooooo111111oooooo        ooooooo1ooooooooo
ooooo11oo11oooooo        ooooooo11oooooooo
oooo11ooo111ooooo        oooooo1oo1ooooooo

       (a)                      (b)
```

Fig. 5
Binary matrices with broad lines (a) and one-pixel-width lines (b)

A skeletonization process should have the following properties:

(i) Preservation of connectivity. The connectivity of a line structure represented in a raster matrix can be defined in terms of neighbourhood relations for the pixels.

(ii) Generation of a marked skeleton. The skeletonization should produce a raster matrix with four distinct types of pixel:

1) pixel belonging to the background
2) pixel having less than two neighbours which belong to the line (end point of a line)

3) pixel having exactly two neighbours, which are not neighbours of each other and which belong to the line (inner line point)

4) pixel having more than two adjacent neighbours (point of intersection)

(iii) Recognition of end points of lines: The end points of lines in a marked skeleton should correspond to the end points in the original map. There should be no "spikes" (additional short lines) nor should the lines in the original maps be shortened by skeletonization.

(iv) Lines in the skeleton should correspond to the centre of the lines in the original map (WOETZEL 1977, pp. 5-6).

The line thinning algorithms which have been developed for binary matrices apply usually 3x3 local neighbourhood transformations.

a	b	c
010	000	000
0X1	0X1	1X0
001	000	000

Fig. 6
3x3 local neighbourhood of X; a, b and c represent three of 256 possible types of neighbourhood. Fig. 6b and 6c are logically equivalent.

This means the particular transformation respectively the value of a pixel depends only on the values of its neighbouring pixels. The purpose of the transformation is to decide whether a pixel is to be removed (getting a value of 0 classified as belonging to the area) or has to be retained for preserving the connectivity of the line (getting a value of 1). In a 3x3 local neighbourhood a pixel has 8 neighbours, so one can distinguish 256 different types of neighbourhoods. A procedure developed by KREIFELTS (1974) uses 51 types of neighbourhood which remain from the 2**8 possible types in consideration of symmetry properties. In Fig. 6b and c the neighbourhood types are logically equivalent and only one type is used in the procedure.

An improvement is decribed by RIEGER (1978). His procedure works in three successive steps:

(i) preleminary step, which performs the contour smoothing of the lines

(ii) main step, performing the line thinning by "peeling" the broad lines in the binary matrix

(iii) final step, hitherto not grasped pixel are removed.

The resulting matrix consists of a set of one-pixel-width elements containing a value of 1 and representing the borderlines and pixel with value 0 representing the background.

2.4 SCAN-TO-LINE CONVERSION

The border lines are now represented as a sequence of one-pixel-width elements in the binary matrix. It is not possible to use these data in geographical information systems or for cartographical purposes. Therefore it is necessary to convert the lines from the binary matrix into a sequence of x-y-coordinates. In addition we have to determine also those nodes, which represent points of intersection.
WOETZEL (1977) describes an algorithm that performs the line-thinning and scan-to-line conversion simultaneously. KREIFELTS (1977) operates with a two-step-procedure.

Our starting point is the marked skeleton resulting from the line-thinning step.

If we use a cartesian coordinate system all points describing a border must be nodes of a uniform square grid with spacing equal to the minimum detectable difference in the coordinate values. The coordinate system is given by our binary matrix which contains 512*512 nodes. The task is to approximate geometric figures like curves and straight lines in the grid.

It is well known that if a curve is approximated by a sequence of straight lines these should be short where the curvature is great and long where the curvature is small. This suggests the representation of curves in a map by means of lines of varying length. The result will be what is commonly known as a polygonal approximation.

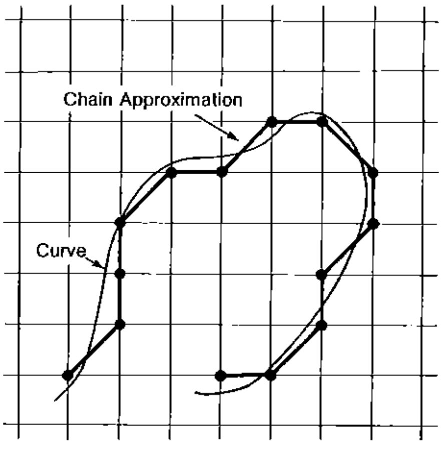


Fig. 7
The chain approximation scheme for a curve

The approximation scheme is called "chain approximation": a curve is approximated by linking each node to one of its eight neighbouring nodes (see fig. 7). Other approximation techniques are described by FREEMAN (1980).

If a border line is transformed into a sequence of nodes in a cartesian coordinate system, it is divided into polygons. A polygon is defined by the enclosed area since it represents the border of the area.

A polygon consists of several segments. Each segment is defined by a set of coordinates with the following properties:

(1) there exist two points in the segment, which have more than two neighbours (points of intersection with other segments; endpoints of the segment). End points of lines may have only one neighbour.

(ii) All other points in the segment have exactly two neighbours.

For cartographical purposes it is important to decide whether the borderlines are stored in a polygonal mode or in a segment mode (see fig. 8).

A segment consequently consists of nodes (end points of the segment) and so called intermediate points which are necessary if the segment is not running in a straight line. Each set of elements with a value of 0 which is bounded by segments from all sides without being intersected by another segment is called a region.

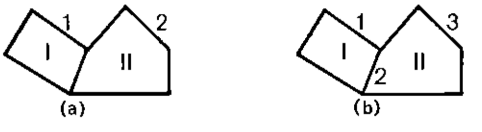


Fig. 8
Storage methods of borderlines:
a: polygon mode b: segment mode
In a the area I is described by the polygon 1; in b I is described by the segments 1 and 2.

Region boundaries consists of one or several segments. Bringing the different segments together, the starting point and the end point of the polygon are identical. Each segment separates exactly two regions. Each segment is stored only once, thereby decreasing the required storage capacity. Furthermore it secures higher flexibility in case of different types of representation. For many cartographical application criteria for example classification of land use it is thus easily possible to eliminate dividing segments. This is often required if a distinction between different land use types is no longer requested and two or more former land use types are combined into one class.

2.5 GEOGRAPHICAL INFORMATION SYSTEM

The processing of the segment data depends on the structure of the information system which is intended. For example, if we wish to store information concerning the streams, lakes and swamps and the roads, land use etc. of a region, we use a key which allows to retrieve and merge the data with the cartographic information stored under the same key.
The basic kinds of data found in maps are points, lines and areas.
A geographic entity consists of a set of attributes and an associated set of values. The geographic reference is established by the coordinates attributes. Both attributes and values may be integers or character strings.
The entity might be as simple as a point or as complex as a whole map. An entity has global properties, component parts and related geographic entities. For example, we may represent the city of Bremen by geographic entities. In this case global attributes are population, area, boundary, types of industries and so on. The numerical values of these attributes are stored together with the values of the boundary coordinates. One division of the particular city of Bremen is into Stadtteile (quarters); Stadtteile would also be represented by geographic entities.

The geographic entities that are related to a Stadtteil are its highways, railroads, rivers, lakes and so on. Some of these entities will be wholly contained in the Stadtteil and others will cross its boundaries. One way to represent this phenomenon is to use a binary relation where the first element of each pair denotes the geographic entity and the second element is a code indicating whether the entity is wholly contained in the Stadtteil. The geographic entities themselves would contain more specific information about their locations. Fig. 9 illustrates a simplified geographic information system containing an attribute value table, a Stadtteil adjacency relation, and a lakes relation for the city of Bremen (SHAPIRO 1979).

The collection of all necessary data can be done in most cases in an automatic way as described above.

In some cases one requires other facilities and procedures, e.g. for processing aerial photographs (see TOST 1977), but it is no longer necessary to collect spatial data by hand.

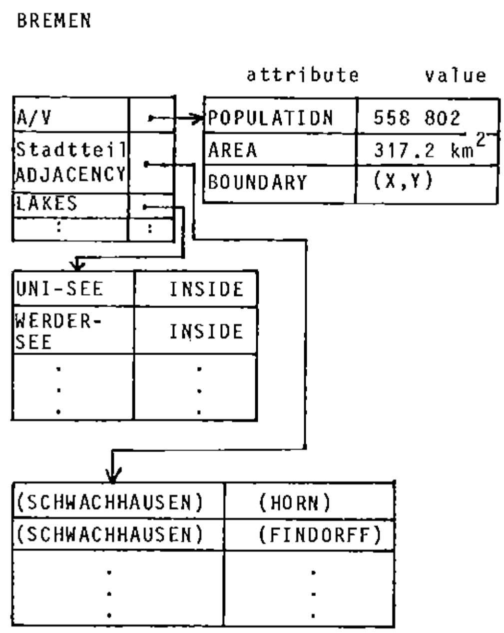


Fig. 9
A simple geographic information system

CONCLUSION

The comments above try to explain an automative way for digitizing area oriented maps. This way is faster and less errorprone than digitizing spatial data by hand. Additionally digitized maps are easier to update. The hardware and software requirements are available to most academic institutions like universities and others. Different methods for automatic digitization of area oriented maps are described. Automatic digitization of maps can be done economically today and is already used in some geographical and administration institutions as well as in universities for education.

REFERENCES

BISCHOFF, H.; H. NUHN; M. STUTZBACH (1979): Bearbeitung thematischer Karten mit dem Printer, Hamburg.

BISCHOFF, H., a.o. (1980): Regionale Informationssysteme für geographische Forschung und Lehre, Hamburg.

BRADSHAW, J.A. (1963): Letter recognition using a captive scan, IEEE Trans. Electron. Comp. EC - 12.

FREEMAN, H. (1980): Map and Line-drawing Processing, NATO Advanced Study Institute in Bonas.

HEYGSTER, G. (1979): Rangordnungsoperatoren in der digitalen Bildverarbeitung, Diss., Göttingen.

KANSY, K. (1977): Capture and Encoding of area - oriented geographic maps. Submitted for the 3rd Jerusalem Conference on Information Technology (SCIT).

KREIFELTS a.o. (1974 and 1976): Erfahrungen mit der Digitalisierung von rastermässig erfaßten Linienstrukturen I und II, GMD Mitteilungen Nr. 30 und Nr. 37.

PAVLIDIS, Th. (1980): Shape description. NATO Advanced Study Institute in Bonas.

RIEGER, B. (1979): Skelettierungsverfahren für die automatische Schreiberkennung, in: FOITH, J.P. (ed.): Angewandte Szenenanalyse, DAGM Symposium, Karlsruhe; Springer, Berlin.

SHAPIRO, L.G. (1979): Design of a spatial information system. NATO Advanced Summer Institute on Map Data Processing in Maratea.

STEFANELLI, R.; A. ROSENFELD (1971): Some Parallel Thinning Algorithms for Digital Pictures. Journal of the ACM, Vol. 18, No 2, pp 255-264.

TOST, R. (1977): Interactive Data Processing for Local and Regional Planning. Gesellschaft für Mathematik und Datenverarbeitung, Bonn.

WOETZEL, G. (1977): A fast and economic scan-to-line conversion algorithm. Gesellschaft für Mathematik und Datenverarbeitung, Bonn.

MÜNSTERSCHE GEOGRAPHISCHE ARBEITEN

Herausgegeben von den Hochschullehrern des Instituts für Geographie der Westfälischen Wilhelms-Universität

Wilfrid Bach - Hermann Hambloch - Heinz Heineberg - Ludwig Hempel - Friedrich-Karl Holtmeier - Cay Lienau - Alois Mayr - Karl-Friedrich Schreiber - Ulrich Streit - Dietbert Thannheiser - Peter Weber - Julius Werner

Schriftleitung : Alois Mayr

H.1 Alois Mayr: Universität und Stadt. Ein stadt-, wirtschafts- und sozialgeographischer Vergleich alter und neuer Hochschulstandorte in der Bundesrepublik Deutschland.
1979. 375 Seiten mit 43 Tab., 28 Abb., 24 Bildern, B Farbkarten im Anhang, DM 60,-. ISBN 3-506-73201-3.

H.2 Hermann Mattes: Der Tannenhäher im Engadin. Studien zu seiner Ökologie und Funktion im Arvenwald.
197B. B7 Seiten mit 52 Abb., DM 20,-. ISBN 3-506-73202-1.

H.3 Friedrich-Karl Holtmeier: Die bodennahen Winde in den Hochlagen der Indian Peaks Section (Colorado Front Range).
Ludwig Hempel: Physiogeographische Studien auf der Insel Fuerteventura (Kanarische Inseln).
197B. 103 Seiten mit 21 Abb., 52 Fotos und 1 Kartenbeilage, DM 20,-. ISBN 3-506-73203-x.

H.4 Peter Weber (Hrsg.): Periphere Räume - Strukturen und Entwicklungen in europäischen Problemgebieten. Mit Beiträgen von Bernhard Butzin, Hans Elsasser, Wilfried Heller, Cay Lienau/Hartmut Hermanns, Rolf Lindemann, Hartmut Lücke, Frank Nagel, Wolfgang Taubmann und Peter Weber (Berichtband einer Arbeitssitzung des Geographentags in Göttingen 1979).
1979. 183 Seiten mit insg. 23 Abb. und 32 Karten, DM 30,-. ISBN 3-506-73204-B.

H.5 Heinz Heineberg (Hrsg.): Einkaufszentren in Deutschland. Entwicklung, Forschungsstand und -probleme mit einer annotierten Auswahlbibliographie. Mit Beiträgen von Bernhard Butzin, Bernd R. Falk, Marianne Grewe, Heinz Heineberg, Alois Mayr und Winfried Meschede.
1980. 165 Seiten mit insg. 22 Tab., 20 Abb. im Text und 12 Abb. im Anhang, DM 30,-. ISBN 3-506-73205-6.

H.6 Wilfrid Bach/Ulrich Hampicke: Klima und Umwelt.
Wilfrid Bach: Untersuchung der Beeinflussung des Klimas durch anthropogene Faktoren.
Ulrich Hampicke/Wilfrid Bach: Die Rolle terrestrischer Ökosysteme im globalen Kohlenstoff-Kreislauf.
1980. 104 Seiten mit insg. 12 Tab., 17 Obersichten und 24 Abb., DM 31,50. ISBN 3-506-73206-4.

H.7 Peter Schnell/Peter Weber (Hrsg.): Agglomeration und Freizeitraum. Vorträge eines Symposiums der Arbeitsgruppe "Geography of Tourism and Recreation" der Internationalen Geographischen Union (IGU/UGI) in Münster 1979.
27 Beiträge von Autoren aus Bulgarien, Frankreich, Italien, der Tschechoslowakei, den USA und der Bundesrepublik Deutschland.
1980. 238 Seiten mit insg. 94 Abb., DM 40,-. ISBN 3-506-73207-2.

H.8 Norbert de Lange: Städtetypisierung in Nordrhein-Westfalen im raum-zeitlichen Vergleich 1961 und 1970 mit Hilfe multivariater Methoden - eine empirische Städtesystemanalyse.
1980. 17B Seiten mit 56 Tab. und 37 Abb. (davon 4 im Anhang), DM 34,-. ISBN 3-506-73208-0.

H.9 Ludwig Hempel/Hartmut Brettschneider: Beiträge zur "Energetischen Geomorphologie" in Trockengebieten.
Ludwig Hempel: Studien über rezente und fossile Verwitterungsvorgänge im Vulkangestein der Insel Fuerteventura (Islas Canarias, Spanien) sowie Folgerungen für die quartäre Klimageschichte.
Hartmut Brettschneider: Mikroklima und Verwitterung an Beispielen aus der Sierra Nevada Spaniens und aus Nordafrika mit Grundlagenstudien zur Glatthanggenese.
1980. 142 Seiten mit insg. 27 Fotos (davon 4 farbig), 43 Abb., 4 röntgenographischen Diagrammen und 3 Karten, DM 37,-. ISBN 3-5D6-73209-0.

H.10 Dietbert Thannheiser: Die Küstenvegetation Ostkanadas. 1981. 204 Seiten mit 41 Tab. und 166 Abb. (davon 16 Fotos), DM 41,50. ISBN 3-506-73210-2.

H.11 Gerhard Bahrenberg / Ulrich Streit (Hrsg.): German Quantitative Geography. Papers presented at the 2nd European Conference on 'Theoretical and Quantitative Geography' in Cambridge 1980. Mit Beiträgen von Gerhard Bahrenberg, Hubert Bischoff, Wolf Gaebe, Ernst Giese, Georg Heygster, Franz-Josef Kemper, Hans Kern, Kurt Klein, Günter Löffler, Josef Nipper, Aribert Peters, Reiner Schwarz, Ulrich Streit, Wolfhard Symader, Wolfgang Taubmann, Wilfried Wittenberg.
1981. 182 Seiten mit 22 Tab., 47 Abb. und 1 Farbkarte im Anhang. ISBN 3-506-73211-0.

In Vorbereitung

H.12 Bernhard Butzin (Hrsg.): Entwicklungs- und Planungsprobleme in Nordeuropa. Beiträge zu ausgewählten Problemen der Raumentwicklung in Dänemark, Schweden, Norwegen, Finnland und Island von Annette Bergmann, Ruth Bünning, Bernhard Butzin, Staffan Helmfrid, Rolf Lindemann, Friedhelm Pelzer, Peter Sedlacek, Dietbert Thannheiser, Erhard Treude und Unno Varjo. 1981.

H.13 Volker Rönick: Das nordöstliche Rio Grande do Sul/Brasilien - Naturräumliche Gliederung und wirtschaftliche Bewertung.

VERLAG FERDINAND SCHÖNINGH, POSTFACH 2540, D-4790 PADERBORN

Anfragen bezüglich Schriftentausch werden erbeten an die Schriftleitung, Institut für Geographie der Westfälischen Wilhelms-Universität, Robert-Koch-Straße 26, D-4400 Münster.

Druck: Henning Wocke, Offsetdruckerei, Karlsruhe, 1981