1986

Functional Regions And Spatial Interaction: The Black/white Model

Valerian Titus Noronha

Recommended Citation

Noronha, Valerian Titus, "Functional Regions And Spatial Interaction: The Black/white Model" (1986). *Digitized Theses*. 1481.
https://ir.lib.uwo.ca/digitizedtheses/1481

National Library of Canada
Canadian Theses Service
Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada
Services des thèses canadiennes

CANADIAN THESES

THÈSES CANADIENNES

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

NL-339 (r 86/06)

Canada

Functional Regions and Spatial Interaction: The Black/White Model

by

Valerian T Noronha
Department of Geography

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Graduate Studies
The University of Western Ontario
London, Ontario
July 1985

ISBN 0-315-29469-8

Abstract

The functional region is usually conceptualized as a spatial partitioning that identifies communities of interest. Attempts at quantitative definition have mimicked analogous concepts in the social sciences, especially social clique identification. Although it is accepted that the functional region is based on human interaction, little attention has been paid to interaction modeling in the context of functional regions. This is partly because the spatial interaction model is not equipped to handle behavioral biases.

This research develops a model of spatial interaction that recognizes perceptual affinities or barriers between spatial units. Derived from the Luce and Tversky choice axioms, the Black/White Model relaxes the assumption of isotropism in landscapes, and offers an alternative conceptualization of the functional region. Four classes of the model are distinguished, to address the various conceptual structures traditionally associated with regions in general, and the functional region in particular. It is argued that the calibration of the Black/White model serves two purposes: first, it allows the model to be applied in a predictive capacity; secondly, it identifies regional memberships, and is thus in effect a mechanism for regionalization. A method of calibrating the Black/White model is developed. The technique is initially illustrated on simulated data, and is shown to retrieve embedded regional structures over a range of scenarios. A heuristic calibration method is developed for large problems. The methods are compared under conditions of controlled random error; their performance is documented with various simulated spatial and regional configurations.

The Black/White regionalization method is demonstrated on a data matrix of student migrations in the United States. A number of refinements in the heuristic are effected in the light of the initial results, and there are indications that the algorithm may have identified the optimal Black/White partitioning, given the data.

In conclusion, while further work remains to be done on the detection of complex regional structures, the Black/White model offers a method of binary partitioning, and a fresh outlook on regionalization and spatial interaction, with promising implications for both planning and academic enquiry.

Acknowledgements

Creditors fall into two categories. The first consists of those who contributed academically to this research. Thanks are due to my supervisory committee: Profs Michael Goodchild, William Warntz and Donald Janelle. In addition, I would like to acknowledge the helpful comments, posers and criticisms from Profs Kevin Cox, Milford Green, Edward Taaffe and Charles Whebell. The research staff of Bell Laboratories, AT&T and Bell Canada, in particular, Drs R J Cooper and E N Gilbert, provided useful information in the early stages of this research. A word of gratitude to Trish Chalk for technical advice on the illustrations, and to Bob Ardern, for fixing a number of software problems at short notice.

The second group are those who provided logistical support, contributing towards my physical, economic and emotional health over the past few years. They include my brothers, relatives and friends with culinary talents, numerous Faculty and colleagues at Western (in particular, Bubs, Dan, Gord, Greg and Bev, Kathy, Laura, the Luckmans, Old Bill, the Walls and the Weave), Allan McFee and Victor Borge.

Finally, I would like to dedicate my efforts to my Parents.

Table of Contents

List of Figures

List of Tables

Introduction

The conceptual revolution associated with the advent of quantitative methods in geography has been described by its protagonists and practitioners as a break with the "descriptive morass" and the "nonscientific dark ages of thirty years ago" (Golledge *et al*, 1982, p 558).

The passion for quantitative methods is by no means unanimous. Burton (1963) lists five classes of criticism leveled against the quantitative movement and quantifiers. First, there is the feeling that the development and perfection of methodological tools has taken attention away from building the discipline; second, there is the fear that the movement may lead the discipline into fruitless directions; third, certain phenomena are not amenable to measurement; fourth, techniques are being misused by application to inappropriate problems, data and contexts; the fifth criticism, regardless of its veracity, is directed towards the social behavior of quantifiers: "They are perky, suffer from overenthusiasm, vaulting ambition, or just plain arrogance" (Burton, 1963, p 156). It must be pointed out that Burton was only reporting on views held by others, and he proceeded to refute the criticisms. Porteous (1985), on the other hand, presumably reflected his own views when he wrote:

> The cold, hard-nosed spatial theorists and quantifiers in their white lab coats have created a dull and lifeless discipline.
>
> [Porteous, 1985, p 49]

The polemics between supposedly opposing philosophies may have hindered the advancement of the discipline, but the *debate* has probably been beneficial, forcing adherents of the two schools to examine the issues carefully. There have lately been signs that the dualism may be fading. Quantifiers have begun to examine relatively subjective topics, such as perception and behavior (Golledge and Rushton, 1976); Gould (1980) has investigated "qualitative mathematics" and Q-analysis and Smith (1984) reviewed the development and application of artificial intelligence in geography. For their part, regional geographers now routinely quote summary statistics to emphasize points.

1. The Region

One item of interest of the traditional school that has recently attracted the attention of quantifiers is the notion of the region. Bunge was prompted to remark:

> To see region construction, one of the last preserves of the non or anti-mathematical geographers, crumble away before the ever growing appetite of the computing machines is a little unnerving even for a hard core quantifier.
>
> [Bunge, 1966, p xiv]

The original concept of the region effectively encapsulates and summarizes the general thrust of the ideographic tradition. The professional and recreational delight derived from the exhaustive description of a chosen territory — in a word, exoticism — is bred of regional differentiation. The phenomenon of systematic spatial differentiation, and therefore of spatial classification, suggests itself at the most fundamental level of enquiry: early explorers, from Ulysses to Al-Masudi, were aware of regional differences (see Dickinson and Howarth, 1933, p 52; Odum and Moore, 1938, p 279). The concept of the boundary, enclosing spatial units that share selected characteristics, is similarly one that has an elementary appeal.

At a higher level, one begins to visualize regions not as isolated individual islands of uniformity, but as part of an extended cellular system of boundaries, enveloping all forms of space. The magnitude of the system is intimidating — one is prompted to think about regions as a *phenomenon*, to visualize systems of regions in the abstract, and to attempt to understand their behavior as dynamic, organic entities (that last remark requires a qualification and explanation: to be sure, regions are mental constructs — the matter is dealt with in some detail in Chapter 1 — but any objective attempt to define or discover them, conceptually or spatially, confers upon them a palpability that formalizes their existence).

1.1. History of the Region

The history of the region as a formal geographic concept may have begun with Ptolemy's notions of "chorology" (see Stevenson, 1932). During the 18th and 19th Centuries, the notion was developed by German geographers, notably Ritter and Passarge (Dickinson, 1969).

In North America, the regional concept was slow to develop. Until the middle 19th Century, academic programs in geography emphasized encyclopedic descriptions (Warntz, 1981, p 261), presumably responding to the continuing exploration of the New World. The following conceptual enrichment coincided with the period when governments in the United States and Europe were revising their internal organization as part of the new planning process. Geographers were being asked for their advice in the process. Gilbert (1951) details the experience of France, Germany, Spain, Portugal, Britain, the United States and Australia with redistricting for political administration.

Probably the most spectacular success story in this early geographic consultancy was that of Isaiah Bowman. In 1917, Bowman, as Director of the American Geographical Society, was invited by the U S government to assist in an "Inquiry," prior to its participation in the Peace Conference of Paris. Bowman was named Chief Territorial Specialist of the U S delegation to the conference, and Executive Director of the Section of Economic, Political and Territorial Intelligence (AGS, 1919; Greenough, 1919). His job was succinctly described by President Woodrow Wilson: "Tell me what's right, and I'll fight for it" (quoted in Martin, 1980). The Peace Conference ended in 1919 with the declaration of the Treaty of Versailles.

The White House was evidently pleased with its geographical experience. Bowman was consulted by the administration during its mediation of a border dispute between Guatemala and Honduras in 1933, and helped shape President Franklin Roosevelt's policy on Palestinian and Jewish resettlement, in 1943-1945.

The significance of Bowman's experiences were two-fold. First, his success brought the geographer into the role of the consultant, in the modern sense of the word: his services to the government brought the discipline publicity and respectability outside academia. Secondly, Bowman was essentially constructing appropriate political boundaries around economic regions; if the discipline were to gain from his personal success, here was an opportunity to unify and consolidate the concepts of the region, and formalize the methodologies for delimiting it.

On this note, attention is now devoted to Bowman's mission, and his approach to it. The following remarks are taken from descriptions of Bowman's work. In the first, Martin (1980) describes the Paris negotiations; in the second, he refers to the Guatemala-Honduras conflict; the third quote is from Hartshorne's description of Bowman's book "New World" (1928):

> Bowman was called upon to exercise a detailed knowledge of Europe in his work, especially with regard to Poland and Romania. He studied the history, the rise of urbanization and industrialization, the ethnic composition and the physical environment of these and adjacent countries.
>
> [Martin, p 91]

> Bowman ... proposed an economic survey of the territory in dispute ... and by careful use of the resultant maps and reports the matter was brought to a peaceful conclusion.
>
> [Martin, p 95]

> While this [book] deals with problems which many political scientists have studied, its extraordinarily effective use of maps has made it a well-nigh unique contribution.
>
> [Hartshorne, 1939, p 248]

The recurring references to the *map* are significant. Indeed, it was apparently the elaborate cartographic displays and constant reference to them that gave the American delegation to Paris a measure of objectivity and originality; the problem of creating boundaries was seen primarily as a *geographic* issue, and quite secondarily as political.

1.2. The Region in the Middle 20th Century

Following the S[illegible]rld War, the calculating machine, and then the electronic computer, [illegible]ng a scientific revolution. Multivariate statistical methods were now accessible to social scientists; factor analyses, cluster analyses and seriation techniques were used in sociology and psychology, to analyze problems objectively.

Geographers with quantitative leanings embraced the new technology. Bunge (1962), Berry (1964; 1968) and Grigg (1965; 1967) developed the initial bridges between classification and regionalization (these, and the terminology on regions, are surveyed in Chapter 1); by the late 1960s and early 1970s, acceptable quantitative methods of formal regionalization had been developed (Davies, 1980), by viewing *places* as cases in the cluster model.

By contrast, the *functional* region has never been satisfactorily defined or delimited. There have been numerous attempts to equate the problem of functional regionalization with such diverse imports as clique identification, Markov chain analysis and biproportionalism. None of the associated methodologies contains a specifically geographic component; rather than investigate the history of the regional concept and construct techniques based on it, researchers have tried to reconceptualize the region in the context of the foreign methodologies.

2. Spatial Interaction

Proceeding on quite independent lines has been the study of spatial interaction. Like the theories of Von Thunen and Christaller, the derivation of the interaction model required a number of simplifying assumptions that stripped landscapes of all but the basic variations addressed by the model. Thus, while the spatial interaction model compensated for size effects of the origin and destination, it did not recognize special affinities or conflicts between particular pairs of nodes.

Much effort in quantitative geography is now directed towards relaxing the restrictive assumptions in geographic models. In this context, there is a strong case for examining the adaptability of the spatial interaction model to less controlled environments. In particular, alternative specifications of the model need to be developed for application on heterogeneous landscapes. Such research serves two important purposes. First, the modeling may provide for a fresh understanding of functional regions, since the latter are based on interaction; secondly, there is the obvious contribution to the flexibility of the interaction model itself.

3. Plan of Thesis

This research is intended to explore the possibilities for conceptual and methodological unification of the notions of functional regions and spatial interaction. The first three chapters are devoted to reviews of the literature. Chapter 1 surveys the concepts of the region, and attempts to clarify the associated terminology. The controversies surrounding regions are addressed in detail, in an attempt to identify desirable features of methodologies. The methods for delimiting formal and nodal regions are covered briefly. The application of the theories to practical problems is stressed. Chapter 2 reviews existing methods for delimiting functional regions, with attention focused on the appropriateness of the methods to the traditional concepts of functional regions. The third chapter traces the history of the spatial interaction model, from its early gravitational analogies, through the empirical verification and predictive applications, to the attempts at theoretical derivation.

The Black/White model is introduced in Chapter 4. The model is derived from axioms of choice behavior, and is initially used to illustrate a number of phenomena related to spatial interaction in heterogeneous environments. The basic model is then adapted to address regional structures, and a conceptualization of the functional region is offered. It is argued that the calibration of the Black/White model is in effect a method of functional regionalization. In Chapter 5, methods for calibrating the Black/White model are developed. The techniques are illustrated with simulated data (10 × 10) for the case of two regions, and are shown to retrieve embedded regional structures accurately, over a variety of situations. A method for evaluating the efficacy of regionalization algorithms is developed. The issues surrounding multiple regions are addressed.

The calibration procedure is tested on an actual data set (49 × 49) in Chapter 6. Of particular concern is the size of a realistic interaction matrix, and the practical implications this has on the logistics of computing. The performance of the algorithms developed in Chapter 5 is observed, and enhancements effected. Finally, Chapter 7 reviews the contributions and results of this research.

Chapter 1

THE REGIONAL CONCEPT — REVIEW

1.1. Introduction

Regionalization is now recognized as the geographical manifestation of the general problem of classification (Grigg, 1965; 1967). The problem has been one of the classic concerns of geographical enquiry in the descriptive tradition; only relatively recently, following developments in the related fields of mathematics and statistics, has there been an interest in objective methods of regionalization.

The burgeoning of interest in classification and partitioning may be traced to the advent of computing machinery in the middle of this century. The ambitions of researchers outpaced improvements in technology, so that, ironically, there was a heightened necessity to break down large computational tasks and data sets to manageable sizes (Ando and Fisher, 1963; Rogers, 1976). Developments occurred simultaneously in a variety of disciplines: Rogers (*op cit*) cites Rose and Willoughby (1972), Theil (1972) and Himmelblau (1973), working "in such various fields as process control, structural engineering, systems organization, electrical network theory, and a wide variety of seemingly unrelated problems in economics, mathematics, design and operations research" (Rogers, 1976, p 516). In ecology and the biological and geological sciences, there was increased concern with taxonomy and the identification of core species (Cole, 1949; Krumbein, 1960), and clique identification methodologies were introduced in sociology (Cook, 1945; Forsyth and Katz, 1946; Katz, 1947).

In geography, interest in the regional concept has centered on three principal areas. The earliest work was oriented towards consolidating the concept of the region, as expressed in the writings of travelers such as Marco Polo and Al Masudi: Dickinson and Howarth (1933) and Hartshorne (1939) detail the contributions of Ritter, Hettner and other early scholars of geography, toward the conceptualization of regions.

The second area of concern developed in relatively recent years, addressing the questions of validity of statistical inference under aggregation of geo-statistical units (Warntz, 1957; Haggett, 1964; 1965; Beardwood and Kirby, 1975; Clark and Avery, 1976; 1978). Later research by Openshaw and Taylor (1979) was directed towards modeling across aggregations.

The third area of interest involved the discovery and delimitation of regions, whether out of academic curiosity or for specific consultative tasks. The latter applications were often problems of administrative organization or political arbitration.

This research directly addresses the delimitation of regions, but will necessarily make frequent references to the first named area of interest, namely, the nature of regions.

The geographer's experience with delimitation of regions in the role of consultant is best exemplified by that of Isaiah Bowman, as detailed in the Introduction. It was clear, however, that Bowman's success reflected his personal expertise, rather than a formally developed methodology intrinsic to the discipline. Further, Bowman's role as a mediator of international political disputes was supported by his diplomatic status, and the perception of his country as a neutral third party. The fragility of those circumstances, and the growing drift towards logical positivist thinking in geography, contributed towards the search for objective methods for the delineation of regions. It was necessary to organize and formalize the prevailing notions of regions into a framework that would support objective research and the development of quantitative methodologies. The works of Bunge (1962), Berry (1964; 1968) and Grigg (1965; 1967) in the 1960s were landmarks in this respect. Bunge drew attention to *place* as the counterpart of the individual in classification; Berry offered specific quantitative methodologies for the organization of geographical data; Grigg drew detailed comparisons between classification and regionalization, and laid down ten guidelines for regionalization.

1.2. Taxonomies of Regions

The original notion of the region probably grew out of the need to organize descriptive matter in geographical accounts and travelogs. Odum and Moore (1938) detail the viewpoints of anthropologists, ecologists, economists, political scientists and sociologists, on regionalism and partitions. Thus specified, regions were vaguely defined areal envelopes, characterized by a distinctive *genre de vie* (Grigg, 1967, p 465). They constituted the PAYS that were the basis of the regional monographs of the French school (Demangeon, 1939) and other regional studies (for example, Unstead, 1956). Minshull (1967) notes that the RIVER VALLEY REGION was often a sufficient expression of regional identity in early studies — suggesting that physical environment was the dominant variable in regionalization, correlating strongly with other variables. The implication of environmental determinism was clear even in the absence of an explicit cause-effect relationship. The controversial "natural regions," which will be covered later in this chapter, were a product of this excessive generalization.

On the other hand, the emergence of scientific geographical methods led to a redefinition of regions in terms of more clearly stated sets of variables. Thus there are two distinct taxonomies of regions; they will be compared and contrasted in the following sections.

1.3. Descriptive Taxonomy of Regional Types

1.3.1. Physical Environment

Regional geographies often use PHYSIOGRAPHIC REGIONS to lay the foundations of understanding of areas. Examples of these may be found in Fenneman (1917) and Murphy (1968). Budel (1944[1]; 1948) and Peltier (1950) wrote of MORPHOGENETIC REGIONS, specifically related to geomorphic and climatic *process*, at a macro level; studies of slope genetics (Sharpe, 1938) led to characterizations of micro-relief regions such as the PEDIMENT (King, 1953). Schemes of CLIMATIC REGIONALIZATION of the world have been proposed by Koppen (1884) and Thornthwaite (1948). In addition, one may find, in elementary text books on geography, references to NATURAL REGIONS in the context of global climatic regionalization. This reflects the scheme of Herbertson (1905); the term has since evolved into one which represents inviolate, deterministic regional structures — it will be covered in detail later in this chapter.

---

[1] In this, and other foreign-language bibliographic references, the original vernacular editions, rather than their English translations, have been cited in this research, to maintain an accurate historical perspective.

1.3.2. Human Environment

Odum and Moore (1938) propose a four-way classification of regions relating to human culture. The first are regions of literary or esthetic culture, typified by studies of particular tribes and cultures (for example, Forde, 1934); a second class deals with the "rural nation" and AGRO-ECONOMIC REGIONS (see Kendall, 1939). The third category refers to "metropolitan regionalism," and covers Dickinson's (1947; 1964) concept of the city as a social unit, and the notions of HINTERLAND and UMLAND (Allix, 1922; Kant, 1951). In the context of urban planning, there are concepts of regions relating to the city, such as the CBD (see Jones, 1969; Goddard, 1970; Murphy, 1972), the RURAL-URBAN FRINGE (Kurtz and Eicher, 1958; Pryor, 1968) and "Megalopolis" (Gottmann, 1961).

A fourth regional type concerns the ECONOMIC REGION (see Pokshishevskiy, 1975) and the INDUSTRIAL REGION, both commonly used as devices for organizing regional studies (for example, Leong and Morgan, 1973).

Finally, an interesting variant is Warntz's (1976) musings on the relationship between the University and the region, almost suggesting the existence of INTELLECTUAL REGIONS.

1.3.3. Administrative Districts

The two classes treated above are of primarily academic interest; regions exist in more familiar, everyday contexts. ADMINISTRATIVE REGIONS are established by public bodies (government and civic agencies and defense establishments) and private corporations, as a means of organizing operations and jurisdictions, and deputizing responsibilities. These are most important from the point of view of geography as a profession, in that there are often opportunities for geographic consulting, which emphasize the applied and prescriptive dimensions of the discipline. In a related vein, POLITICAL units are organized for administrative organization, as well as to preserve communal harmony — a variety of objectives of political boundaries are identified in Brigham (1919); also see Boggs (1940).

1.3.4. Statistical Districts

Geographers in the field of marketing and location-allocation have been concerned with the definition of TRADE AREAS. Early studies were oriented solely towards predicting patronage (Reilly, 1929). More recently, market research has been oriented towards providing clients with customer profiles (Compusearch, 1980); organizations have adopted computerized data bases, and current concerns include the design of appropriate geocoding systems to optimize machine-related storage and access to geographic data. The methodologies clearly have the potential for application to the establishment of statistical areas by census organizations, although in that case there is a need for compatibility with the structure of existing political units.

1.3.5. Multiple Criteria

Finally, one may consider composites of regions, defined by a particular combination of the variables treated individually above. The combination may be subjective, or determined by arbitrary or systematic weighting schemes, all of which are necessarily open to debate. As mentioned earlier, there is often a tendency to base these on physical divisions — that is, physical variables are implicitly assigned the greatest weight — suggesting a covariance between human activity and the natural environment (Grigg, 1965).

Unstead (1916) argued that human variables should be considered on a par with physical factors; the incorporation of human considerations led to the concept of the COMPAGE (Whittlesey, 1954, p 36), which recognized homogeneity between regions over a number of variables. Whittlesey's concept was more than a simple expansion of the number of variables; the distinction between multi-criteria regions and the compage may be clarified by analogy to mixtures and compounds in chemistry: although compounds may be formed from the same elements as are mixtures, the constituents are bound together in a unique manner.' Similarly, in Whittlesey's view, the region could be compared with a living organism; it possessed a character or personality of its own. The human variables were accorded greater emphasis, with Man being regarded as the architect of the CULTURAL LANDSCAPE.

1.4. Structural Taxonomy of Regions

The previous section differentiated between regions in terms of description of the regional content. The following taxonomy is based on comparisons and relationships between regions.

Probably the most accepted and best understood dichotomy is between FORMAL and FUNCTIONAL regions. Some authors (for example, Davies, 1980) recognize the NODAL REGION as a third element in this class, while the EQUITABLE REGION has been widely referred to, but rarely recognized (Haggett *et al*, 1977, refer to PLANNING OR PROGRAMMING REGIONS).

1.4.1. Formal Regions

1.4.1.1. Concepts

The concept of the formal region (also known as the UNIFORM or HOMOGENEOUS REGION) is elementary and appealing, since it embodies the most basic notions of spatial uniformity. The earliest mention of the concept has been traced to Claudius Ptolemy (translated by Stevenson, 1932), who introduced the term CHOROGRAPHY, in a broad reference to what is now understood as the study of formal regions.

Minshull, reflecting the popular notions of internal homogeneity, defined the formal region as "the largest area over which a generalization remains valid" (Minshull, 1967, p 38). Uniformity or variability in a landscape, whether physical or cultural, is inferred from the properties of discrete individual observable entities. These form the building blocks that constitute the region. In the taxonomy literature, the building block is called an OPERATIONAL TAXONOMIC UNIT (OTU). Bunge (1962) pointed out that the "individual" (the reference to sociology is significant) in spatial taxonomy is PLACE: by definition, a dimensionless abstraction that possesses only one characteristic — location. As a general observation, Bunge's point is beyond debate. However, for a place to be classified in any respect other than its location, it must possess observable attributes, and in practise, an attribute should remain constant over a certain minimum surrounding area in order to be observed or measured. The extent of this area depends on the scale of the study. Dickinson and Howarth (1933, p 243) thus referred to the "smallest *distinctive* geographical unit" (emphasis added). The ELEMENTARY SPATIAL UNIT (ESU) is now introduced as the geographic counterpart of the OTU. A similar concept was labeled the CHORE by Solch and LANDSCHAFTSTEIL by Passarge (see Dickinson and Howarth, *op cit*, p 243). The geographical area over which chosen variables, observed at the resolution of the ESU, remained constant, was termed the LANDSCHAFT, or the formal region as it is known today. The choice of ESU is a problem of scale and aggregation of data units; it may have a significant effect on the outcome of studies (Openshaw, 1977b; 1978; Openshaw and Taylor, 1979).

For the purposes of this thesis, it will be necessary to define additional terms. The UNIVERSE is the segment of space which contains all the ESUs in question — for a particular study, it may be a country or state. The ESUs may be referred to as ZONES or NODES, depending on the context.

The distinction between nodes and zones introduces the issue of dimensionality. The foregoing discussions suggest that the formal region always has an areal extent, composed as it is of ESUs. One may, however, consider the North and South Poles, two points with similar mathematical and geodesic properties, as constituting a non-contiguous, *punctiform* region; similarly, natural or man-made communications routes, with their unique physical and functional properties, may be reduced to *linear* formal regions at suitable scales; but by far the most popular notion of the formal region involves the characterization of an *area*.

Physiographic, climatic, agricultural and economic regions are popular subjects in atlases and regional textbooks. These representations usually involve an abstraction of continuous variables into binary statements of existence or non-existence (see Cliff, Haggett *et al*, 1975, Chapter 4), as shown in Figure 1-1. This has two important implications; first, homogeneity or uniformity within formal regions is often only relative. A region often has a continuou[illegible] variable *intensity* over space; the representation of the formal region as a discrete spatial set suppresses information on internal variations. Secondly, the boundary of a formal region, although represented as a discrete linear feature, is in fact often a continuum (see Krumbein, 1959). The boundary is merely a line along which the value of a variable, representing the intensity of a regional characteristic, reaches a pre-defined threshold.

1.4.1.2. Delimiting the Formal Region

The demarcation of a formal region involves locating a boundary within which a given phenomenon (on which basis the region is characterized) prevails to an extent that is considered critical. At the most elementary level, one may thus derive a single-variable binary regionalization. This type of spatial bounding appeals to the layman, and the concept has widespread application, for instance, in maps indicating the area of cultivation of a particular agricultural commodity, or the area affected by disease or natural hazard. Representation of the *variation* in the phenomenon is usually not considered. The traditional

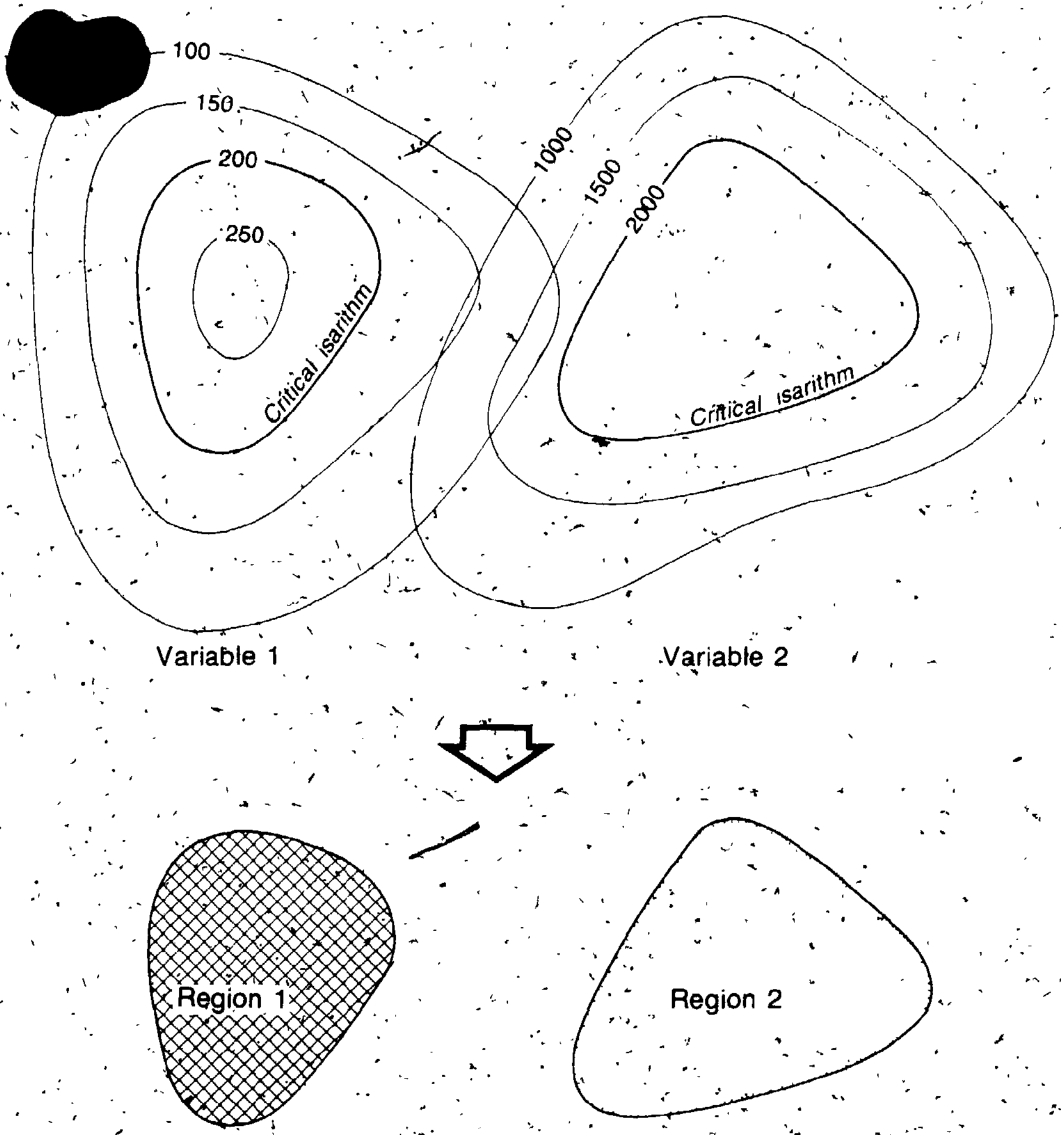


Figure 1-1: Translation of isarithms into formal region boundaries

geographer has relatively little difficulty in drawing or justifying the boundaries to such a region; to the quantifier, identification of ESUs that qualify for inclusion in the region is merely a matter of imposing a threshold value on the variable and examining the value in each ESU for compliance.

The SINGLE-FEATURE REGION (Grigg, 1967, p 467) represents a further step, whereby several grades of the same variable are differentiated — Koppen's climatic regionalization is the classic example. The theoretical aspects of determining the class bounds are of concern in choropleth mapping (see Jenks, 1961). The technique of DENSITY SLICING is now used in the interpretation of remotely sensed data to accomplish similar objectives. Tobler (1973), however, argues against the loss of information involved in such classification.

The treatment of more than one phenomenon introduces the MULTI-FEATURE region, corresponding to the composite region introduced in Section 1.3.5. The delimitation of this would require the identification of the salient variables, and exhaustive descriptions or numerical measurements on each relevant dimension.

Traditional attempts at multi-feature regionalization relied on personal observation and verbal descriptions. The decisions regarding the relative importance of variables, and their intensities in a given area, relied on individual expertise. In the middle 20th Century, inferences were substantiated with extensive numerical evidence (for example, Brookfield, 1962). These were the first tentative attempts at objective regionalization, although they were still essentially descriptive.

The methodologies for distinguishing between regions on the basis of scores on more than one dimension were developed independently in remote sensing and statistics. The technique of PARALLELEPIPED CLASSIFICATION emerged in connection with image processing, as a means of visual separation of dimensions of measurement (see Kelley, 1983). But the single greatest contribution to formal regionalization methodology came from the development of multivariate statistics, specifically cluster analysis. The technique permitted the objective examination of a range of variables, with varying weights applicable if desired. The digital computer, commercial software packages and machine readable information files made the technique readily accessible. The biological and social sciences applied cluster analysis to classification problems (Edwards and Cavalli-Sforza, 1965; Johnson, 1967). In geography, the application of cluster analysis has been discussed by Johnston (1968), Spence and Taylor (1970), Rees (1972), Cliff, Haggett *et al* (1975) and Beavon (1977). In other studies, Semple *et al* (1969) searched for the optimal number of regions. Information theoretic approaches to regionalization, based on variants of Shannon's (1948) information statistic, were proposed by Semple *et al* (1972) and Batty and Sammons (1978). The methodological consensus on the delineation of formal regions, indicated both by theoretical discussions (Johnston, 1968; Spence and Taylor, 1970; Cliff, Martin and Ord, 1975) and by the empirical studies mentioned above, prompted Davies (1980) to declare that the problems surrounding formal regionalization had been "solved."

1.4.2. Functional Regions

The original conceptualization of the functional region was put forward by the traditional German geographers, Penck and Hettner: the LANDESTEIL was a grouping of LANDSCHAFTEN (translated as low-order formal regions), "which, on account of their juxtaposition . . . have mutual relations and a physiognomic unity, and taken together form a geographical unit" (Penck, quoted in Dickinson and Howarth, 1933, p 244). Edward Taaffe, in personal conversations with the author (1985 April 8, May 24), has credited Platt's (1928) study of the Ellison Bay community with being the earliest concrete reference to the functional region (see also Platt, 1935). Notions of functionality were presented independently in the urban literature (Dickinson 1947).

While these concepts have been circulating for several decades, a satisfactory definition of the functional region has been elusive. Verbal definitons have been extended and nebulous:

> ... the functional region is essentially diverse; it is a place where adjacent contrasting environments permit a variety of activities which are complementary in supporting the life of the whole. It is a classic example of unity in diversity; the uniform features of a homogeneous region may just *be* together by chance, but the parts of a functional region *work* together and are to a degree dependent on each other ...
>
> The functional region puts an emphasis on Man's economic activities and in particular this will involve a study of how different areas work together.
>
> [Minshull, 1967, p 40]

Minshull (*op cit*) quotes a variety of other authorities on the issue:

> A cultural area: an assemblage of such forms as have interdependence, and is functionally differentiated from other areas.
>
> [Carl Sauer]

> ... areas within which a higher degree of mutual dependency exists than in relationships outside that area.
>
> [V B Stanberry]

> ... an area whose people are bound together by mutual dependencies arising from common interests.
>
> [American Society of Planning Officials]

To these may be added one of the criteria considered by the American Telephone and Telegraph Co when establishing NUMBERING PLAN AREAS (NPAs) for the continental telephone network:

> Boundaries should be drawn so as to minimize the splitting of *communities of interest* ...
>
> [AT&T, 1980, p 3; emphasis added]

The recurring themes are those of internal diversity, mutual complementarity and independence. There appears to be a consensus that functional regions are a phenomenon of *human* activity, reflecting Whittlesey's (1954) concept of the compage, and that the clue to functional boundaries lies in interactions between the constituent ESUs. Davies (1980) distinguishes between functional regions based on "descriptive allocation" and those based on "empirical derivation." The former are encountered, for example, where the economic functional dominance of a city over its hinterland is observed to diminish with distance — the boundary between the functional regions is judged on the basis of some cutoff values in a set of descriptive variables. While urban function is contained literally within this context, the nature of the regional distinction and the method of delimitation are essentially those of formal regionalization.

In the quantitative literature, the idea of functional coherence of the region led to a tentative definition of the ideal configuration as one which maximized the ratio of within-region to between-region interaction (Brown and Holmes, 1971; Masser and Brown, 1980). The definition has failed to take hold, though. One difficulty is that as the number of regions is decreased, the ratio of within- to between-region interactions can be increased to infinity. Second, there is the problem of non-uniformity in the sizes of the interacting ESUs. Flows into and out of large centers tend to be high on account of zone size, as opposed to complementarity; there is some disagreement on whether and how the flow matrix should be standardized to compensate for the few disproportionately large entries. Despite the technical difficulties, the maximization of the ratio of within- to between-region interaction has emerged as a retrospective definition, and may be treated as one model of the functional region. The issue of delimiting functional regions is treated in greater detail in the next chapter.

A reference was made, in Section 1.4.1.1, to the dimensionality of regions. Whereas formal regions, defined with respect to an environmental variable or a representative ESU-wide mean of a socio-economic index, generally tend to extend over space continuously, functional regions are based on interacting humans with discrete locations in space and time. The essence of the functional region is not the uniformity of characteristics, but a nexus of social and economic tendons that bind the ESUs together. The strength of the bonds may derive from spatial characteristics of the type that define formal regions, but these characteristics are not directly relevant to the working of the functional region. The functional region behaves as if governed by a *force field*, the force being the need or propensity for human interaction. The space pertaining to the region is the communications *network* along which interactions take place. Cartographically, it lends itself better to definition by flow lines and origin and terminal nodes than by enclosing polygons. When the medium of interaction is not cartographically definable, such as in wireless communication, one may think of regions as defined solely by *affiliation* of interacting nodes — in practise, though, these will often be administrative or statistical units such as counties or census tracts, and will therefore lend themselves to continuous cartographic representation. When interactions occur along a spatially definable route, such as a railroad or freeway, the network is an essential component of the representation.

The problem of delimiting functional regions is one of the principal issues of this thesis. The literature on the subject is reviewed in the next chapter.

1.4.3. Nodal Regions

The third structural class of regions was recognized in the 1950s, when Whittlesey (1954) introduced the concept of cores and regional dominance in networks. Philbrick (1957), in an elaborate essay on *functional* regions, wrote of focality and nodal organization; he discussed trade areas and hierarchical inter-connections in the context of regionalization, from the scale of a single household to that of the world. Nystuen and Dacey (1961) translated the concepts into a quantitative methodology for delimiting nodal regions.

Topologically, the nodal and functional concepts of regions are similar, based as they are on interactions. The distinction arises in the concept of relative dominance of the nodes. Taaffe and Gauthier (1973) compared the network to a cobweb: the removal of a junction in the web disturbed the nodal hierarchy of the system since the pattern of dominance was affected — the *functional* ties between the nodes, however, remained unchanged.

Relative dominance is evident in asymmetry of flows. Nystuen and Dacey (1961) defined an INDEPENDENT node as one whose largest outflow was to a smaller city (where size was determined with reference to the marginal total in the interaction matrix); a NODAL FLOW was the largest outflow from a subordinate node. They were able to partition the flow graph into locally dominant components, which defined the structural linkages of the regions. A power series summation of the matrix was used to estimate indirect flows.

While this formulation of the nodal regional structure was quantitative, the concept was more traditional, borrowed from Central Place theory — a dominant center performing functions at a higher level in the hierarchy, and lower order centers being served by local parent nodes. The concepts of central place dominance were also evident in Dickinson's (1947) ideas of cities as regions, and Allix's (1922) and Kant's (1951) concept of umland.

1.4.4. Equitable Regions

Notwithstanding the contentions advanced for or against the natural region, the three regional structures dealt with upto this point — formal, functional and nodal — are academic constructs that may *exist* — albeit in the mind of the researcher — and await discovery by passive regional analysts. In the case of equitable regions, on the other hand, the regionalization is *performed* with a specific objective in mind; the purpose of partitioning is to devise zones between which a set of variables maintain approximately equal values, perhaps subject to some practical constraints, one of which is usually contiguity. Several applications can be cited:

24

- In the case of electoral districts, political and statistical considerations dictate that zones should contain approximately equal voting populations, that minority groups are adequately represented, and that the zones are so compact in shape as to dispel suspicions of gerrymandering.

- The Post Office needs to assign walks to mail delivery personnel in an equitable manner.

- Census organizations attempt to establish statistical areas at various levels. Statistics Canada, in designing its census tracts, includes specifications on population count, cultural and socio-economic homogeneity, and correspondence of tract borders with physical boundaries (Statistics Canada, 1982, p 126).

The problem of deriving equitable regions has generally focused on the design of electoral districts. Algorithms for solution of the problem have been proposed by Nagel (1965), Weaver and Hess (1963) and Garfinckel and Nemhauser (1970). Reviews of the literature appear in Taylor (1979).

The algorithms used in the studies mentioned above are representative of the operations research field. Nagel uses a swap algorithm — a type later employed in the Teitz and Bart (1968) heuristic; Weaver and Hess employ an alternating algorithm that varies very slightly from the Maranzana (1964) location allocation technique, and Garfinckel and Nemhauser (1970) adopt a core-accretion procedure, an approach similar to the random regionalization algorithm of Openshaw (1977a).

A radically different method was proposed by Goodchild and Hosage (1983), who employed a branch and bound algorithm on the arcs constituting the ESU polygons.

1.5. Issues surrounding Regional Structure

The previous sections have introduced the two principal taxonomies of regions. Figure 1-2 summarizes the classifications. Some of the adjectives associated with regions derive from outside the above scheme of classification. These, and the controversies surrounding regions, are addressed in the following sections.

1.5.1. Regions: Objects or Devices?

The earliest concepts of the region were strongly biased in favor of physiographic regions. It had been observed that physiographic features exerted a strong influence on climate, and that these factors in turn affected natural vegetation. Further, early civilizations and culture hearths (for example, Indus, Nile and Euphrates) were associated with river valleys, prompting scholars of the region to attribute the observed covariance between physical environment and human culture to a deterministic cause-effect relationship. Forde (1934) exemplifies this environmental determinism in his accounts of the Khirgiz of Central Asia and the Masai of Kenya. The predominant basis of Forde's regionalization was the natural environment, although he was in fact observing *cultural* regions. The NATURAL REGION was embraced by few Anglo-American geographers (Herbertson, 1905, was notable). Hartshorne (1939) refuted Hettner's arguments in favor of natural classification, and Minshull (1967) and Hart (1982) have been emphatic in their rejection of the concept:

> The region is an intellectual concept created by the selection of features relevant to the interest of the geographer or the problem in hand. . . . The word selection applies not only to the area or areas under consideration, but also to the number and kind of phenomena which will be included. Thus to apply this device one chooses certain topics or phenomena from the whole range of possibilities in order to give the required groupings and regions; and one disregards all topics which one considers irrelevant.
>
> [Minshull, 1967, p 122]

Descriptive Taxonomy	Structural Taxonomy
Physical ●	● Formal
Human ●	● Functional
Administrative ●	● Nodal
Trade Areas ●	● Equitable
Total Region (Compage) ●	

Figure 1−2: Taxonomies of regions

> Regions are subjective artistic devices, and they must be shaped to fit the hand of the individual user. There can be no standard definition of a region, and there are no universal rules of recognizing, delimiting and describing regions.
>
> [Hart, 1982, p 21-22]

On the other hand, the natural region has been a centerpiece of the Soviet Bloc view of geography. Kalesnik (1961) speaks for the Russians when he writes:

> . . . the earth's surface has been divided in a natural manner . . . into a multitude of sections differing qualitatively from one another in an aggregate of many interrelated external features and internal characteristics. . . .
>
> A landscape is an actually existing and genetically homogeneous section of the earth's surface; it is framed by natural boundaries; . . . it is characterized by territorial integrity, that is, it cannot consist of parts isolated by the territory of another landscape . . . the general style of combination of the heterogeneous components and structural peculiarities is preserved unchanged within the landscape.
>
> [Kalesnik, 1961, p/24-25]

His compatriot Solntsev (1962) indicates the passion with which the Soviets regard their academic territory:

> . Abroad, and especially in bourgeois geography, the study of natural areal units has still not attracted the attention of geographers. . . .
>
> It may seem to some comrades that [the controversy over natural regions] is simply a terminological argument that does not touch the substance of the matter . . . But if you look carefully at what the argument is all about, it soon becomes clear that it is not so much a question of the best use of the term 'landscape' as of the very meaning of landscape science.
>
> [Solntsev, 1962, p 4-6]

A conclusive argument against the concept of a system of natural regions, waiting for geographers to discover it, is the limit to spatial covariance between variables. One would be justified in searching for such a system only if there was a complete, [illegible] at least significant, correlation between all spatial variables — this implies an extreme form of determinism, which does not sit well with the Communist view of voluntarism.

The controversy over the existence of a natural classification is not peculiar to geography — Whittaker (1962, p 14) describes a similar debate in botany.

1.5.2. Aggregation and Decomposition

Grigg (1965, p 468) notes that classification is inherently different from division, although the net effect of either operation is to create subsets of the universe. While classification proceeds by *aggregation* of ESUs on the basis of common characteristics, "logical division" *divides* the universe into sub-categories at each step. There is no suggestion that either method is superior; Grigg merely points out that the two procedures are distinguishable and therefore are appropriate under differing circumstances. The regions derived in each case are termed SYNTHETIC and ANALYTICAL respectively.

1.5.3. Location and Contiguity

Classification on the basis of variables with a strong spatial autocorrelation would *usually* lead to regional systems that exhibited a strong degree of contiguity; but is contiguity a necessary characteristic of a region? Grigg implies that it is:

> . . . a geographical classification which neglects location is of limited value. . . .
>
> . . . it is assumed that all the individuals which are similar are also contiguous. If they are not, then a region is not formed.
>
> [Grigg, 1965, p 476]

Unstead *et al* (1937), distinguish between GENERIC and SPECIFIC regional systems. In the former, ESUs are aggregated purely on the basis of their non-spatial properties, and contiguity is not necessary.

29

1.5.4. Regional Hierarchies

Grigg (1965; 1967), in outlining the logical basis of hierarchical regionalization, identified a variety of dendrograms (Figure 1-3). Several authors of regionalization algorithms (for example, Slater 1976; Masser and Brown 1975; Clark 1973) have adopted the "binary aggregation" structure. Regionalization procedures should, however, recognize the variety of arrangements available, and exploit the characteristics of each option.

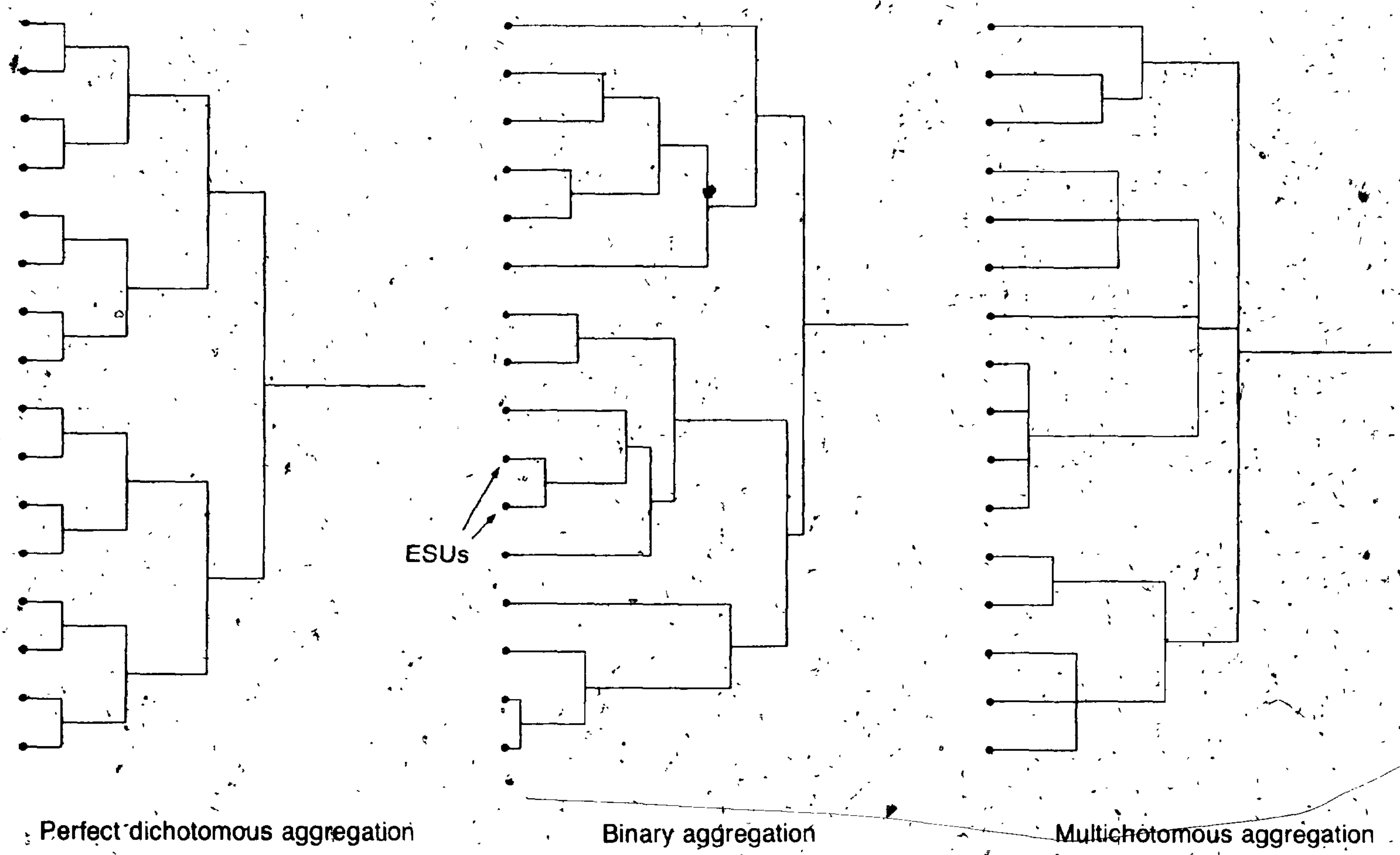
1.6. Applied Regionalization

It should be evident from the foregoing discussions that regions exist at so many scales, and for purposes that are so different, that a reconciliation into a single global system of regions will not be an easy task, and will involve compromises between criteria of one system or the other. This section examines the role of regionalization in various aspects of everyday organization.

Administrative regionalization seeks a rational basis for the division of power and responsibility. The objectives of private and public sector clients differ in that the latter may be concerned with equity while the former operate to maximize cost efficiency. Public sector organization of space may exert some control over private sector schemes — for example, the numbering plan areas of the continental telephone network respect state (but not county) boundaries.

1.6.1. Private Sector

Large retail concerns establish distribution centers to maximize the efficiency of distribution of goods to outlets, as well as to maintain a certain minimum accessibility to customers. The definition of trade areas as a basis for market profiles and publicity efforts is of concern to managements. A related problem is that of school busing and the location of community centers and places of worship — these issues are subject to ethical and political considerations,



Perfect dichotomous aggregation

Binary aggregation

Multichotomous aggregation

Figure 1—3: Selected hierarchical structures centering on whether the function of such facilities is to perpetuate existing boundaries or to help break them down.

1.6.2. Public Sector

In the public sector, the most evident partitioning problem is that of devising political units such as provinces and counties. The regular outlines of states and provinces in North America — and, for that matter, the international boundary between the United States and Canada — may suggest that political boundaries can be drawn quite arbitrarily. The shapes of these particular boundaries may be a result of the youth of the civilization and the state of political harmony on this continent. This is not the case with most other countries; in India, for instance, state boundaries effectively divide the country into religious and linguistic regions, and inter-state antagonisms are often quite pronounced.

The scope for application of geographic theory and method in establishing political boundaries is necessarily limited by the number of opportunities that arise: both international and provincial frontiers tend to be stable through time because of the sentiments associated with them and the costs of reorganization. Attempts to redefine them must consider both the historical evolution of existing boundaries and the arguments for change.

The second application of regionalization by the public sector is closely related to the first. Governments worldwide are concerned with organizing censuses, and this necessitates setting up statistical regions. There are advantages to be gained from organizing these so as to minimize conflict with political districts, although the criteria for the two systems may be different. The expense of censuses has led to recent attempts to estimate urban populations from remotely sensed imagery (Holz, 1973). While such methods are likely to de-emphasize established boundaries of statistical areas, their accuracy is necessarily limited; the need for high-resolution data, especially for market analysis studies, will probably sustain the continued need for traditional boundaries.

Governments are concerned with the establishment of jurisdictional areas for courts and other public services such as utilities (see Goodchild and Massam, 1969), or for particular types of radio and television programming and broadcasting; civic and emergency services and ration stores. A districting problem subject to close public scrutiny is the design of electoral districts. Despite the availability of objective procedures for districting, recent events in California (see Kelly, 1982) testify that the problem is still a pressing one. A variety of applications of regional concepts in planning may be found in Masser and Brown (1978).

1.6.3. Semi-Private Sector Geocoding

The two regional organization systems with perhaps the greatest impact on everyday life are the geocoding systems of the Post Office and the telephone network. The criteria used to derive the systems have little in common. While the Post Office is concerned with efficiency in routing and distribution, the Bell system has to operate within the confines of the specifications of its switching equipment; it attempts to maintain certain patterns in the volume of telephone traffic (AT&T, 1980), and establishes its independent geocoding and tariff structure towards this end.

1.6.4. Other Applications

One could identify more ambitious and exotic applications for regionalization:

- Equitable zoning could be applied to the worldwide pool of athletic talent, to institute a depoliticized Olympics.

- It might be argued that the International Date Line should be located along the meridian across which communication is at a minimum, with its implications for the prime meridian and the time standard (the popularity of the Americentric universal map projections in North America suggests a discontent with Greenwich).

Some of the applications outlined above may be better expressed as location allocation or transportation problems. Figure 1-4 lists a number of regionalization problems, and suggested strategies for their solution.

1.7. Regionalization: Practical Considerations

The development of methods for applied regionalization must keep in mind the objectives of the user. The following are some of the features that might be considered:

- What is the primary purpose of the districting: organization? Minimization of distribution or administrative costs? Preservation of communities?
- Can the number of regions be specified by the user or is it discovered by the procedure? If the latter, can it be overridden?
- Can the system be altered dynamically to conform to exogenous requirements? Can two regional configurations be compared and evaluated with respect to costs?
- Can the system handle several variables at once?
- How does the system accommodate the future?

1.8. Concluding Remarks

This chapter was intended to review the conceptual and theoretical background of regions, and to highlight the practical issues involved in developing methodologies for application to real world problems.

In regard to functional regions, which are the focus of this research, the concepts are vaguely expressed notions of complementarity. A plausible concept of the region is presumably one that could be translated into a technique for *delimiting* regions. The next chapter will review existing methods of functional regionalization.

Problem	Solution
Retail distribution	Multi – step routing Transportation Problem
Trade area definition	Empirical Survey
Churches School busing	Functional Region Location – Allocation
Courts Police/Fire Services: Hydro, Garbage	Equitable Region Transportation Problem Location – Allocation
Radio broadcasting	Formal/Equitable Region
Telephone	Functional/Equitable Region
Postal Codes	Routing
Date Line	Functional Region
Political frontiers	Formal/Functional Region
Depoliticized Olympics	Equitable Region

Figure 1–4: Applied Regionalization

Chapter 2

DELIMITING FUNCTIONAL REGIONS — REVIEW

2.1. Introduction

The concept of the functional region has been introduced and discussed in broad terms in Chapter 1, with attention particularly focused upon the structural distinctions between formal, functional and nodal regions. It was pointed out that no rigorous definition of the functional region had yet been developed, and that the conceptualization was largely postdicted by quantitative methodologies for regionalization. The methods for formal and nodal regionalization were covered briefly, noting the general agreement on purposes and methods.

The aim of this chapter is to review the literature on quantitative methods of functional regionalization and to evaluate the conformity of the techniques with the verbal concepts of spatial functionality as they have been expressed in the past.

36

2.2. Data Pre-Processing

A number of verbal conceptualizations of the functional region were surveyed in Chapter 1. In summary, they agreed on the notion of functional complementarity *within* the region, and independence *between* regions. The implicit consensus is that the data requirement be a square matrix of connections, or human interactions or flows, between the n ESUs. Analysis need not be restricted to a single measure of flow: Berry (1968) argued in favor of simultaneous consideration of several measures of interaction between ESUs; this could be accomplished by treating each of the n^2 elements of the interaction matrix as a single case in a DYADIC data structure, so that the various flow indices could be manipulated as variables.

The concept of the functional region could now seemingly be translated into a search for that allocation of ESUs to regions, that maximized the ratio of within-region to between-region flows. However, practical problems associated with the data matrix may limit the variety of mathematical options available to the analyst. These are detailed below.

2.2.1. Interaction with Self

It is not always possible to find a conceptual justification for flows from an ESU to itself. One might, in some cases, expand the scale of investigation within each ESU, and consider flows within the ESU as the required measure of interaction with oneself — one clearly encounters a limit to such expansion at the level of the individual interacting person. Since the required change of scale often cannot be accommodated by agencies responsible for gathering and maintaining data, one may be faced with missing values in the diagonal cells of the matrix. Some investigators have arbitrarily set these at zero (for example, Illeris and Pederson, 1968; Clayton, 1974; 1977a). Alternately, the missing data may be treated as the result of sampling error: Savage and Deutsch (1960) and Goodman (1963) have offered algorithms to estimate the entries.

2.2.2. The Number of Regions

Given that the square data matrix **I** is acceptable, or has been suitably processed to compensate for sampling error, a simple minimization of the ratio of between-region to within-region flows will yield a single region, containing the entire universe. Clearly, there must be a constraint on the number of regions to be distinguished. Pocock and Wishart (1969) and Semple *et al* (1969) have suggested means of discovering the "natural" number of groupings in spatial clusters, in the context of formal regionalization. On purely pedagogic grounds, there is undoubtedly a case for discovering the natural groupings, although one may argue that, given the fragility of regions with respect to variables and time, it is inappropriate to think in terms of a finite number of regions; in planning applications, on the other hand, one may be required to partition a universe into a given number of regions, for a specific political or administrative purpose.

2.2.3. Corrections for Node Size

In practise, a large proportion of the total interaction in a universe centers around a few locally dominant population concentrations. In any given environment, there will be a rank ordering of central places (Zipf, 1949; Johnston and Scott, 1965; Davies, 1966), and therefore magnitudinal differences and asymmetries in the flows (Nystuen and Dacey, 1961). This makes it difficult to justify the minimization of the above ratio. While there is general agreement on the need for standardization of the flows to control for the size effects, the methods of achieving this have been subject to controversy and confusion. The following discussion is organized chronologically, but at some cost of coherence; the reader's patience is solicited.

Ng (1969), using data on migration between the "chawads" of Thailand, calculated a WORKING MATRIX **J** as the change in net migration per 100 residents, over a given time period. Such a compensation assumes that migration varies linearly with zone population — this runs counter to the specifications of recent interaction models (for example, Ewing, 1984).

Hollingsworth (1971) assumed that under the null hypothesis of independence between the row and column marginal totals of **I**, the expected interaction matrix **I′** could be derived by treating **I** as a contingency table, so that

$$I'_{ij} = \frac{\sum_i I_{ij} \sum_j I_{ij}}{\sum_i \sum_j I_{ij}} \qquad \ldots (2.1)$$

He proceeded to calculate "mobility indices" as the ratio of observed interactions **I** to the corresponding entries in the expected table, **I′**, and constructed the working matrix **J** as follows:

$$J_{ij} = \frac{I_{ij}}{I'_{ij}} + \frac{I_{ji}}{I'_{ji}} \qquad \ldots (2.2)$$

Blau and Duncan (1967) noted that it was nearly impossible to find two identical matrices of mobility ratios, as defined by Hollingsworth; therefore, since such matrices could not be compared, the contingency-based null hypothesis was invalid. Tyree (1973) demonstrated that mobility indices failed to correct the inequalities in the marginal distributions. She proposed a procedure of successive standardization of the rows and columns of **I**, converging on a matrix in which the marginal totals were simultaneously equal.

Masser and Brown (1975), presumably extending the ideas of Hollingsworth, calculated **J** directly from the expected interactions, measuring the difference, as opposed to the ratio, between observed and expected interactions:

$$J_{ij} = [I_{ij} - I'_{ij}] + [I_{ji} - I'_{ji}] \qquad \ldots (2.3)$$

Slater (1976), referring to Hollingsworth's working matrix, raised the criticism that J was forced into symmetry, thereby suppressing structural information contained in **I** — the criticism may clearly be extended to the Masser and Brown formulation. Borrowing techniques from the literatures of mathematics (Fienberg, 1970) and economics (Bacharach, 1970), he employed an "iterative proportional fitting procedure" (IPFP), similar to that advocated by Tyree (*op cit*), to standardize **I**. While the convergence and other properties of the doubly standardized matrix are accepted (Macgill, 1977), Holmes (~~1977~~) felt that the procedure may not preserve the asymmetries in **I**, and thereby discard information.

Hirst (1977) criticized Masser and Brown for considering *arithmetic* differences between observed and expected flows; since the larger entries in **I** would usually generate greater differences, the working matrix would effectively fail to even out the distribution of values in the marginal totals. Hirst suggested compensating for the unequal differences by dividing by the appropriate entry in **I'**, so that

$$J_{ij} = \frac{I_{ij} - I'_{ij}}{I'_{ij}} \qquad (2.4)$$

with no specific reference to symmetry of **J**. This was similar to the "relative acceptance" (RA) score in the "transaction-flow model" of Savage and Deutsch (1960) and Goodman (1963), which had been applied in a spatial context by Brams (1966). Goodchild (1981) pointed out ~~that~~ Hirst's proposal amounted to subtracting 1.0 from Hollingsworth's mobility index, with no apparent benefit other than conformity with other social sciences. He suggested that a method of compensating for zone size could be considered satisfactory only if it delivered identical results under an arbitary halving of the origin or destination size, or both — this would in effect make the procedure independent of the geocoding system in use. In that regard, both Hollingsworth's and Hirst's specifications were acceptable, but Masser and Brown's index was not.

It would be fair to infer that the problem of correcting for size remains unsolved. Following an analysis of Tasmanian telephone data, Holmes (1978) concluded that both IPFP and RA scores displayed a "systematic bias, which is most extreme for matrices with high variability in scores and in row and column sums" — which precisely described the condition that the procedures were intended to eliminate.

2.2.4. Indirect Flows

The matter of whether or not indirect flows not represented in the data matrix should be included in the analysis is again the subject of disagreement. One may argue that communications routed along particular links reflect the strength of those links; on the other hand, one may be more concerned with the ties between the origin and destination *nodes* rather than with the links (see Stephenson, 1974).

2.2.4.1. Matrix Powering

The standard graph theoretic method of estimating indirect flows has been by means of a power series expansion of a matrix. The method has been discussed in a sociological context by Luce and Perry (1949). Nystuen and Dacey (1961) applied the technique in the context of nodal regions, freely admitting its limitations:

> It is extremely doubtful that the matrix . . . is the most appropriate measurement of the total direct and indirect influences. It is essentially a measure of chance indirect contact . . . . It does, however, have a greater appeal than the matrix [I], which incorporates only the direct influences.
>
> [Nystuen and Dacey, 1961, p 37]

Stephenson (1974) criticized the Nystuen and Dacey application of the technique, suggesting that concern should be directed towards flows rather than towards the properties of the network, and that direct flows should receive greater weight than indirect contacts, in accordance with theoretical or empirical rules; Nystuen and Dacey had raised the latter issue themselves (p 36), but had not arrived at a rational weighting scheme. Tinkler (1976) showed that the Nystuen and Dacey method did assign progressively lower weights to less direct flows, but that the weighting scheme was linked to the power expansion through a mathematical procedure, rather than an empirically established scheme. The lack of a justifiable operational method of including indirect flows strengthens Stephenson's case for their exclusion.

2.2.4.2. Markov Chains

A radically different interpretation of the issue was provided by Brown and Holmes (1970) and Brown and Horton (1970) (see also Brown 1970; Brown, Odland and Golledge, 1970).[2] Since **I** was a sample of interactions at a given point in time, they viewed it as a probabilistic measure of connection between the nodes. Against that background, indirect flows were reflected in transitional states in a Markov process (Tinkler, *op cit*, attempted to establish a conceptual link with Nystuen and Dacey's power series method): the matrix of mean first passage times (MFPT) "has no meaning in terms of real world time or real world phenomena, but provides an abstract measure or index of functional distance from node *i* to node *j*" (Brown and Holmes, 1971, p 61).

Predictably, the MFPT method invited a barrage of criticism. Masser and Scheurwater (1980) noted that the procedure did not correct for differences in node size. Davies (1980) criticized the lack of practical relevance of the measures of functional distance. Goodchild (1981) pointed out that diagonal entries were necessarily involved in the calculation of the MFPT matrix, and that the method would therefore not be applicable in situations where those entries were missing. Further, under Goodchild's criterion of effective standardization (Section 2.2.3), the MFPT method was found unsatisfactory, and the matrix was therefore not considered amenable to further processing.

---

[2] In all these studies, the stated objective was the discovery of "nodal" regions; that appears to be a terminological problem.

2.3. Regionalization Strategies

The above sections covered the issues and methods relating to the prior processing of the interaction matrix **I**. Given that a satisfactory method has been adopted to overcome the inherent problems with **I**, it is now assumed that a working matrix **J** is available for examination.

2.3.1. Preliminary Issues

The process of regionalization could adopt either of two strategies. First, one could search for structural weaknesses in the working matrix, and *partition* the universe in a "top-down" sequence — the method of Brams (1966) was an example of this "divisive" or "decompositional" approach. Alternately, one could identify strengths between pairs of ESUs in the matrix, and *cluster* them on that basis. This "agglomerative" strategy has generally been more popular (for applications, see Hollingsworth, 1971; Masser and Brown, 1975), probably due in part to the similarity to Ward's (1963) algorithm for cluster analysis. In the context of regionalization, the distinction between the two approaches is technical and academic (Rogers, 1976, reviews the finer differences between the concepts). The terms "partitioning" and "clustering" are used synonymously, and with general reference to the problem of allocating ESUs to regions, rather than the particular strategy for doing so. The issue is raised here to provide a conceptual link with the "synthetic" and "analytical" regional types discussed in Chapter 1.

A second concern in regionalization methods has centered on the contiguity and compactness of derived regions. Some of the methods reviewed in the following sections explicitly provide for constraints on contiguity; on the issue of compactness, Spence and Taylor (1970, p.23), in a broad review of regionalization methods, repeatedly refer to the tendency of some methods to "chain," that is, to agglomerate ESUs in extended elongated structures, as a "problem condition," reflecting the difficulty many researchers have with accepting the plausibility of regions that are not compact.

A third issue has arisen in the wake of the methodologies for regionalization that have been proposed in recent years. Judging by the objective functions in use, two fundamentally distinct models of the functional region seem to have emerged. The following sections treat each of them in turn.

2.3.2. Similarity of Interaction Patterns

The early attempts at functional regionalization (Garrison and Marble, 1963; McConnell, 1967; Illeris and Pederson, 1968), applied a conventional factor analysis to the interaction matrix (due to the dimensional symmetry of I, the distinction between R-mode and Q-mode analysis is irrelevant at this point) in an attempt to cluster the nodes in the matrix with others that exhibited similar interaction patterns, or connective structures. Russett (1967, pp 101, 130) criticized the mathematical objective as being inconsistent with the purpose of functional regionalization, which was to aggregate ESUs on the basis of *mutual* relationships, rather than relationships with all other ESUs. Russett allowed, however, that DIRECT factor analysis (in which the clustering was based not on a matrix of correlation coefficients, but on the interaction matrix itself, normalized for convenience so that values ranged from 0.0 to 1.0) would not be subject to the criticism. Goodchild (1981) notes a difference in the term "direct factor analysis" between applications in sociology and those in geographic regionalization.

The factor analytic tradition was strengthened by Berry's (1968) attempt at conceptual unification of formal and functional regional structures, on the basis of the organization of geographic data matrices; the technique was subsequently applied by Berry and Rees (1969), Goddard (1970), Clark (1973) and Clayton (1974; 1977a; 1977b), and the extension to DYADIC factor analysis (in which several interaction variables could be considered simultaneously) was operationalized by Davies and Thompson (1980). Recently, Davies (1980) and Nader (1981) investigated HIGHER ORDER factor analyses.

Goodchild (1981), in evaluating the relevance of the factor model to functional regionalization, notes that the outcome of the analysis would be dependent on the size of either the origins (in the case of Q-mode analysis) or the destinations (R-mode), and would therefore violate the fundamental assumptions regarding zone size. A second criticism of the factor model concerns the criteria for assigning columns to factors. This is a crucial step in the regionalization, and its reliance on subjective assessment of the factor loadings seems to contradict the purpose of objective analysis.

Finally, it is worth noting that the conceptual and empirical arguments against a distinction between formal and functional structures (Berry, 1968, and Clark, 1973, respectively), have come from proponents of the factor model.

2.3.3. Maximizing Within-Region Flows

The above heading is often quoted as a shorthand for the problem of maximizing the ratio of within-region to between-region flows, or the ratio of within-region flows to all movements in the system. Operationally, a procedure may be a hierarchically organized search for groups of ESUs that are most similar, or closest together, depending on the context of the working matrix **J**. Three approaches to the procedure are discussed below.

2.3.3.1. Decomposition

The decompositional approaches of Brams (1966) and Slater (1976) began by isolating the larger cells in the matrix. Entries higher than a certain threshold value were set to unity, and the remaining cells emptied. Successive increases in the threshold were used to partition the matrix: Brams described a program that searched for the most "salient" links, while Slater considered the binary matrix as a digraph, and employed matrix multiplication to extract the "strong components" (Harary, 1965) at each level of the hierarchy.

2.3.3.2. Seriation

Another decompositional technique is SERIATION, also known as the "block diagonal" method, "block modeling" or "matrix decomposition." The method was originally used in sociology (Cook, 1945; Forsyth and Katz, 1946; Katz, 1947) as a means of identifying cliques, but with little mathematical rigor. The theoretical aspects of the operation, and the relevant algorithms, were developed by Luce and Perry (1949) and Beum and Brundage (1950); Borgatta and Stolz (1963) and Ling (1973) produced popular computer programs for the operation. For an updated discussion of the technique see Arabie (1978). Briefly, the procedure is to rearrange the rows and columns of a matrix, so that the higher entries are concentrated in the immediate vicinity of the diagonal. The technique has been applied to the analysis of input-output tables in economics (see Ando and Fisher, 1963; Simpson and Tsukui, 1965; Yan and Ames, 1965); Gale *et al* (1982) used the method to evaluate a number of spatial interaction models and Muller (1983) explored its use in geographic information systems. Roistacher (1974) reviewed the method in the context of sociometric modeling.

Ng (1969) was one of the first to apply seriation to the problem of regionalization. He stratified the values in the working matrix, using a variety of pictorial symbols, which were then rearranged in block-diagonal form, by a combination of subjective inspection and automated processes, to reveal the regional structure. O'Sullivan (1977; 1978) seriated matrices to investigate partitions of intra-urban transport networks.

2.3.3.3. Agglomeration

The technique of agglomerative regionalization proceeds by a series of searches for the nearest, or most similar, pair of ESUs. There are essentially two forms of the algorithm; they differ in respect of the structure of the resulting hierarchy. Hollingsworth (1971) created pairs from all ESUs in one pass through the proximity matrix, before recomputing the matrix based on the centroids of the pairs; Masser and Brown's (1975) "Intramax" procedure, on the other hand, recomputed successively smaller matrices as each pair was found. Thus, while Hollingsworth's dendrogram corresponded to the "perfect dichotomous" structure (recall Figure 1-3), Masser and Brown derived the "binary aggregation" dendrogram. The results of the operation are clearly dependent on the algorithm. In choosing between the options, one must presumably either subscribe to the view that one's choice is inherently superior for universal application, or establish a justification in the particular geographic context. In neither of the above cases did the authors distinguish between the possible structural forms.

2.4. Conclusions

2.4.1. Evaluation of Techniques

2.4.1.1. Validity of Methods

A researcher in need of a "black box" method of discovering functional regions is unlikely to have found the above review reassuring. There is clearly a considerable divergence of views on a number of crucial issues. First, while there is agreement on the need to filter the data matrix to correct problems due to sampling and measurement, the methods of accomplishing this have not met with universal approval. The second problem involves the algorithms for clustering. There are clearly major difficulties with the most fundamental notions of functional regional structure — namely, the applicability of the factor model — that have led the field into two widely differing schools of thought.

Masser and Scheurwater (1980) attempted to compare three of the above procedures (Intramax, IPFP and Markov chains) in an empirical study. That raises the question of evaluation of techniques. It would be possible to pass judgements on the relative merits of the various approaches and algorithms only if there was a known result against which the outcomes could be compared.

However, in the absence of a stochastic null hypothesis governing the interaction matrix in a completely undifferentiated universe, and a model of the effect of regional structure on interaction, there can be no rigorously defined TARGET configuration of ESUs. The field has therefore to rely on subjective assessments of results, made with reference to particular application contexts, therefore involving biases of specific geographic circumstances and of the authors' perceptions of local environments. One must question the scientific validity of verdicts expressed in terms such as:

> These results agree *fairly well* with the regions used by the government for planning and statistics . . . '.
>
> [Hollingsworth, 1971, p 2758; emphasis added]

or

> The functional regions produced by the approach outlined above . . . conformed extremely closely to an *intuitive knowledge* of the study area.
>
> [Brown and Holmes, 1971, p 65; emphasis added]

Without a target, and a measure of deviation from it, the above statements could accurately describe a variety of results.

2.4.1.2. General Comments

Two peripheral issues seem to pervade the general discussions of regionalization. The first is the matter of hierarchies. Masser and Scheurwater (1980) worried that seemingly similar hierarchical procedures being applied to matrices were derived from a variety of clustering processes, and that the function of aggregation in each case was therefore different. In addition, the structural forms of the hierarchies varied. Hollingsworth (1971) and Masser and Brown (1975) both searched for hierarchical systems, but their dendrogram-forms were entirely different, as demonstrated earlier. One is led to suspect that structure is being imposed upon regions artificially, at the time of *analysis*, without due concern for the *genesis* of hierarchies in the flows.

Secondly, methodologists have clung to the notion of functional regions necessarily being contiguous. This may be a legacy of the long history of concern with *formal* regions. As explained in Chapter 1, formal regions are continuous space phenomena. They may therefore be scrutinized for single or multiple bounds. Functional regions are a topologically distinct phenomenon. They are not tied together by similarities observed across polygonal boundaries; the forces of adhesion act longitudinally along links that terminate in discrete space. One must therefore be prepared to accept functional regions that are spatially fragmented (admittedly, it is more difficult to justify the absence of contiguity, particularly in consultative situations, but that is poor justification for blurring the integrity of a concept).

2.4.2. The Need for an Alternative Model

A disturbing feature of the literature on functional regions is the lack of a consistent geographic theory relating concepts and methods. Techniques applied to matrices in other disciplines have been imported for application to interaction matrices. One necessarily achieves similar results, and therefore might infer a lack of disciplinary identity. Holmes observes:

> The rationale for matrix manipulation has been poorly expressed, usually being subsumed under loosely stated considerations such as plausibility, simplicity, dominance, salience, regional ties, nodal structures, hierarchical systems, and so on. Methodology remains fragmented and derived from a disparate array of research problems and disciplinary contexts.
>
> [Holmes, 1978, p 326]

It is clear that insufficient attention has been paid to the underlying geographic concepts of the functional region, and in particular, to the distinction between *form* and *process* in the context of human behavior. Without discarding current notions regarding functional structures, it will be helpful to reconsider the fundamental issues.

The methodologies described above treat the interaction matrix as a contingency table, and calculate the "expected" interaction I′ from (2.1). Given the present understanding of spatial interaction, flows are known to be sensitive to *distance*. Since measures of expected interaction do not compensate for the friction of distance, dyads are likely to register high mobility ratios purely on the basis of their proximities; as a result, functional regions constructed in this way are likely to be contiguous, reflecting the uncorrected influence of distance. The spatial interaction model offers a means of compensating for distance.

A major difficulty with many of the methodologies reviewed in this chapter was the issue of uneven zone sizes in the interaction matrix. Again, the spatial interaction model considers origin and destination size factors, and recognizes asymmetries in the matrix; it may offer clues as to appropriate strategies for modeling.

Finally, investigations of phenomena related to human behavior should be concerned with both the *revealed* spatial behavior of individuals, and the abstracted *preferences* that for some reason cannot be realized. The methodologies discussed in this chapter consider only revealed behavior, in the terms of the interaction matrix; there is a clear need for an alternative conceptualization of the functional region, from the vantage point of individual perceptions of spatial structures and decision making processes. In particular, one needs to recognize the role of distance as a deterrent to interaction; this will be a key discriminant between revealed and preferred behavior.

The modeling of spatial interaction will be reviewed in the next chapter.

50

Chapter 3

SPATIAL INTERACTION MODELS — REVIEW

3.1. Introduction

Spatial interaction has never been a dominant focus of geographic thought. It is not an immediately evident aspect of spatial differentiation, since it involves the dyadic dimension of activity and *relationship*, beyond the traditional unary dimensions. Early geographical thought focused on description; the discipline was slow to embrace the notion of relationships, whether conceptual, between descriptive parameters, or communicative, between geographical entities.

Significantly, the theoretical contributions of Von Thünen (1826), Weber (1909) and Christaller (1933), which were being accepted early in the twentieth century as cornerstones of deductive reasoning in geography, had all emphasized transportation. Crowe (1938) responded with a plea for increased attention to movement. Meanwhile, the field of urban planning was developing, and transportation was emerging as a sub-specialty of engineering; the potential for objective analysis was recognized at a time when the quantitative revolution in geography was in its infancy. Spatial interaction came to be associated with modeling, and quickly acquired a respectability.

3.2. The Gravity Model

The events leading to Isaac Newton's formulation in 1687 of the Universal Law of Gravitation are well known. The fall of an apple was orchestrated into a relationship detailing the forces between bodies in space:

$$F = G \frac{M_1 M_2}{d_{12}^{\ 2}} \qquad \ldots (3.1)$$

where F is the force of attraction between the bodies, G represents the universal gravitational constant (now known to be 6.67×10^{-11} N m^2 kg^{-2}), M_1 and M_2 the masses of the bodies and d_{12} the distance separating them.

The first documented social analogy to Newton's law was made by Carey (1858):

> . . . we have here the great law of molecular gravitation as the *indispensable* condition of the existence of the being known as man . . . the greater the number collected in a given space, the greater is the attractive force that is there exerted . . . Gravitation is here, as everywhere in the material world, in the direct ratio of the mass, and in the inverse one of distance.
>
> [Carey, 1858, p 42-43]

Ravenstein (1885; 1889) formalized the relationship in connection with urban migration in Britain:

$$I_{ij} = \frac{f(P_j)}{d_{ij}} \qquad \ldots (3.2)$$

wh[illegible]

I_{ij} is the flow of people from origin i to destination j;

$f(P_j)$ is a function of the population of j; and

d_{ij} is the geographical distance between i and j.

Significantly, Ravenstein chose to express distance in its raw form, whereas the Newtonian formulation considered the *square* of distance.

3.3. Developments Pre-1960

The 1940s were marked by the rise of the "social physics" concepts of Stewart (1942; 1947), and subsequent attempts to identify physical analogies that related to geographic concepts. Zipf (1946b) likened the interaction between population centers to a *force* between them, represented by **F** in

$$F_{ij} = K \frac{P_i P_j}{d_{ij}^2} \qquad \ldots (3.3)$$

where K was a constant balancing factor, analogous to G in (3.1). Since energy was the product of force and distance,

$$Q_{ij} = K^* \frac{P_i P_j}{d_{ij}} \qquad \ldots (3.4)$$

where Q_{ij} was the energy between the two centers, and K^* a balancing factor. Subsequent researchers were not able to agree on the distinction between force and energy in regard to human interactions; this resulted in some disagreement over whether or not the distance term in the denominator of the gravity model should be squared (see Hagerstrand, 1957, p 118-19; Stewart and Warntz, 1958).

The above relationships formed the basis for numerous studies in following years. One body of research was concerned with empirical examination of the model in various contexts, and attempted to derive universal constants that would play the same role in (3.3) as the gravitational constant G did in (3.1); these studies eventually led to the refinement and generalization of the formulation of the model. A second body of research was devoted primarily to the theoretical development and extension of the gravity concepts. Each of these areas of research will now be discussed in turn.

3.3.1. Empirical Studies

The studies of the pre-1960 period were notable for both the variables that they attempted to fit to the model, and their numerical findings: Bossard (1932) noted the effect of intervening distance (stratified into counts of city blocks) in the selection of spouses in Philadelphia; Zipf (1946a; 1946b) studied railroad and airline traffic; Cavanaugh (1950) examined several variables for fit with the model, including tourist traffic into national parks and student migration to colleges. Isard and Peck (1954) applied the modeling to the movement of industrial goods and Ikle (1954) considered traffic flows.

The experience with the gravity model failed to produce a set of universal constants such as those associated with the Newtonian equation. Price (1948) found that the power of distance varied with direction around a node (Price studied migrations between nine regions of the United States, and failed to correct the observed interactions for the varying populations of the zones; it was only to be expected that a variation of the effect of "distance" would be observed, but to attribute this exclusively to variation in direction was naive). Dodd (1950), in his "Interactance Hypothesis," attempted to explain the observed variations in the effect of distance by suggesting that distance ordinarily carried an exponent of unity, but that the non-uniformity of opportunities over space caused the observed value to vary; further, he argued that the laws of human interaction, like the laws of molecular and celestial gravitation, were governed by probabilities, and that the coefficients in the latter were found to be more stable principally on account of the large numbers of molecules involved in the movements, as compared with the number of participants in human interactions. He applied weights γ_i and λ_j to the origin and destination populations; to represent the "per capita activity" of the population centers, introduced a

54

variable exponent β for the distance term and proposed the generalized relationship:

$$I_{ij} = K \frac{\gamma_i P_i \lambda_j P_j}{d_{ij}^{\beta}} \quad \ldots (3.5)$$

Cavanaugh (1950) applied this form of the gravity model to a number of flow variables, but assumed, presumably for computational simplicity, that the origin and destination specific terms γ and λ were constants. These studies predated other attempts at generalizing the interaction equation, notably by Anderson (1955, 1956).

The generic equation proposed by Dodd implicitly acknowledged that the constants K and β lacked universality, in contrast to those in Newton's gravitational equation. In other studies, Cavanaugh had obtained values of β ranging between 0.23 and 1.48, and correlation coefficients between 0.50 and 0.96, with various data sets. Carroll (1955) found that β took values between 2.0 and 3.5 in his studies of telephone calls and highway traffic in the Michigan area. Hagerstrand (1957) recovered a value of 3.0 for migration in Sweden.

3.3.2. Offshoots of the Gravity Modeling

The gravity model offered a fresh means of thinking of human interactions, and provoked the development of related concepts. Reilly (1929) applied the model to the delimitation of areas of urban influence, and thereby of trade areas, predicting the point of indifference between alternative facilities, given their attractions and distances. Stewart (1941) sought explanations for interaction in the concept of POTENTIALS of population; in this particular application he was proposing no more than a special case of the gravity model, with a single destination and a distance exponent of 1.0. Harris (1954) extended Stewart's ideas on potential, to market, manufacturing and farm potentials, and Warntz

(1957; 1959) examined the role of income potentials in explaining economic activity in the United States.

3.3.3. Pre-1960 in Review

The verdict of the experimentation prior to 1960 established quite firmly that the gravity model was not destined to become an overnight panacea for the growing pains of an adolescent discipline. The parameters clearly did not possess the universality required of a scientific formulation — they would therefore have to be determined in advance for a given type of interaction in a given environment, before they could be applied in a prescriptive capacity. The gravity model was therefore perceived, rightly or not, as more an empirical regularity than a physical reality. Its credibility as a scientific *theory* could be maintained only if it could be derived deductively — this became the focus of research in the 1960s.

Even as an empirical regularity (and particularly due to the lack of a theoretical justification), a universal structure for the model could not be satisfactorily determined. The following sections detail some of the difficulties.

3.3.3.1. A Function for Distance

The most elusive secret was a correct functional form of distance. Most researchers had agreed to a powering of distance to an exponent β. This form posed a difficulty when the diagonal cells of an interaction matrix were considered; since the corresponding distance was zero, the formulation produced an undefined value of I_{ij} for this case. In theory one could argue that *inter*-action must necessarily occur over some minimum distance. Court (1966) later attempted to derive a universal correction factor, and Dutton *et al* (1971) got around the problem, in this and other contexts, by assuming distance to self to be one half the distance to one's nearest neighbor.

A second objection to the general form of the gravity model was raised much later by Haynes (1975), who pointed out that the dimensionality of the balancing factor K varied with the empirically derived value of β. Unless it could be accepted that there was in fact a universal, (albeit elusive) β, K would not have a fixed dimensionality, and would therefore be an unacceptably defined component in a quantitative model.

The answer to both these difficulties seemed to lie in considering an exponential, as opposed to power, functional form of distance. If the model were expressed as

$$I_{ij} = K \frac{\gamma_i P_i \lambda_j P_j}{e^{\beta d_{ij}}} \qquad \ldots (3.6)$$

rather than (3.5), interactions over zero distances could readily be defined, and the dimensionality of K was constant. On the other hand, β would depend on the unit of measurement of distance; if it were a universal constant, it would have to be stated with respect to a particular unit.

The exponential form in (3.6) may be derived from the simple condition that the *rate of decrease* of interaction at a given distance is directly proportional to the interaction at that distance; there is no clear reason why this should be so.

In practise, there is little difference between the predictions of the exponential and power forms of distance decay. The critical distinction arises when diagonal entries are present in the data; in many cases they are not available, or for reasons peculiar to the scale of the data, may not exist (for example, phone calls to oneself). Given the difficulty in resolving the issue satisfactorily, it appears reasonable to expect a researcher to examine both functional forms for fit with a given data set.

3.3.3.2. Interpretation of Other Model Parameters

The exponential form of distance had merely *stabilized* the dimensionality of K, but on account of the other empirically defined and derived parameters in the model (that is, γ and λ), none of which had a clear unit of measurement, the dimensions of K were as yet undefined. It was not clear whether γ and λ represented definitive measures of activity, and therefore contributed meaningfully to the dimensional coherence of the equation, or were purely numerical compensating elements. If one agreed that they fulfilled the latter role, then it was clear that the population terms in the equation were in fact being superseded in their predictive power by an unknown composite of causal factors. It would then be appropriate to aggregate the origin-specific and destination-specific parameters separately, leading to the formulation

$$I_{ij} = E_i A_j f(d_{ij}) \qquad \ldots (3.7)$$

where **E**, the emissivity of the origin, represented all the functional elements particular to the origin, and **A** represented the attraction parameters of the destination. Since, given only **I**, **d** and the function f, the equation in the form (3.7) could be calibrated, **E** and **A** could be determined empirically.

3.4. The 1960s: The Search for Theory

Despite the considerable theoretical shortcomings of the gravity model, there was little doubt expressed as to its empirical predictive power. The model found application in urban transportation forecasting in a number of cities (see, for example, Carroll and Bevis, 1957). However, if geography as a discipline were to capitalize on the growing popularity of the technique, it would need to reinforce the empirical performance of the model with a credible theoretical derivation. Three studies in the 1960s attempted to fill this void.

3.4.1. The Huff Model

In the context of individual shopping behavior, Huff (1963) postulated a trade-off between attraction of a store and its distance from the consumer. A shopper would be prepared to travel a greater distance (or spend more time traveling) to a store that was more likely to offer the selection of goods he or she required. The shopper would therefore weigh the alternatives with regard to their attractiveness (measured by floor area, variety of goods or a similar surrogate) and distance. The probability of choosing a given store j, given a choice set S, would be

$$p(j|S) = \frac{A_j d_{ij}^{-\beta}}{\sum_{m \in S} A_m d_{im}^{-\beta}} \qquad \ldots (3.8)$$

where

A_j is the attraction of store j;

d_{ij} is the distance of store j from a patron located at i; and

β is a friction of distance parameter.

Then the number of persons at i who visited store j would be

$$I_{ij} = w_i \frac{A_j d_{ij}^{-\beta}}{\sum_m A_m d_{im}^{-\beta}} \qquad \ldots (3.9)$$

where w_i is the number of consumers at i. This equation can be made to resemble (3.7) by substituting

$$\xi_i = \frac{w_i}{\sum_m A_m d_{im}^{-\beta}} \qquad \ldots (3.10)$$

Then (3.9) may be written as

$$I_{ij} = \xi_i A_j d_{ij}^{-\beta} \qquad (3.11)$$

which corresponds perfectly with (3.7), with ξ interpreted as emissivity.

Huff's work served two important purposes. First, it conferred a respectability on the gravity model, by reinforcing empiricism based on a purely physical analogy with indigenous behavioral theory. Secondly, it provided a fresh outlook on the problem of delimiting trade areas: a significant advantage of this method over Reilly's (1929) "breakpoint" concept was that by casting shopping behavior in a probabilistic framework, it recognized that individuals might travel to facilities other than the nearest or most attractive.

The derivation invited criticism, principally from Bucklin (1971), that while it was oriented towards decision-making processes at the level of the individual, the data that were usually fit to the model pertained to *aggregate* distributions (see McFadden and Reid, 1974).

3.4.2. Entropy Maximizing Derivation

Wilson (1967) approached the gravity model with another analogy, this time from the methods of analyzing particulate arrangements in statistical mechanics. He defined a MACROSTATE of interaction as the sum of the number of movements in a universe, that is, $\sum_i \sum_j I_{ij}$. A given configuration of **I**, that satisfied the constraints of the macrostate, may be termed the MESOSTATE.[8] A MICROSTATE referred to any one of the ways in which individuals could allocate themselves among the cells of **I**. Now, given the marginal totals of **I** and the total travel in the system, $\sum\sum I_{ij}$, as constraints, the most probable mesostate would be that which was produced by the greatest number of microstates. Wilson demonstrated that the most probable outcome of **I** would be

---

[3] The term "mesostate" was not used in Wilson's original study; it has since crept into the literature.

$$I_{ij} = U_i O_i V_j D_j e^{-\beta d_{ij}} \qquad \ldots (3.12)$$

where O_i and D_j were the marginal totals of **I**:

$$O_i = \sum_j I_{ij} \qquad \ldots (3.13)$$

$$D_j = \sum_i I_{ij} \qquad \ldots (3.14)$$

and **U** and **V** were balancing vectors:

$$U_i = \left[\sum_j V_j D_j e^{-\beta d_{ij}}\right]^{-1} \qquad \ldots (3.15)$$

$$V_j = \left[\sum_i U_i O_i e^{-\beta d_{ij}}\right]^{-1} \qquad \ldots (3.16)$$

Clearly, (3.12) could be reduced to the generic interaction equation. Wilson emphasized that the mechanical analogy would be complete only if the interacting population were completely homogeneous, in the same way that molecules of a gas were identical, implying that the equation should be applied only to samples that were appropriately stratified.

According to the entropy maximizing derivation, (3.12) holds only when both the marginal vectors **O** and **D** are constrained, or known in advance. Although Wilson provided a separate derivation for the singly-constrained model, and in the particular context of journey to work trips to which he was referring, it may have been appropriate to apply constraints on the marginal totals, his derivation did not adequately address the unconstrained gravity model.

3.4.3. Utility Theoretic Derivation

A third attempt at theoretical derivation of an equation for movement was made by Niedercorn and Bechdolt (1969). They began with the assumption that the utility of a trip between a given origin-destination pair, for a given individual, was a function of the number of such trips made by the individual. They then proceeded to find the number of trips made by all individuals between the origin-destination pair, so as to maximize utility, subject to the constraints of available funds and time for travel. This was expressed as

$$I_{ij} = \frac{A_i}{t} \cdot \frac{P_j / \sum_m P_m}{d_{ij}} \qquad \ldots (3.17)$$

where A_i was the total amount of money that all individuals at i were willing to spend on travel to all destinations, per unit time, and t was the cost of travel per unit distance. The derivation assumed a logarithmic relationship between trip totals and utility; a slightly different result was offered for a power relationship.

The weakest point in the derivation was the circular assumption that utility was a function of the number of trips made. Mathur (1970) pointed out that trips were generated by opportunities at the destination, and favorable characteristics of the route or mode. The authors responded (Niedercorn and Bechdolt, 1970) by revising their definition of utility as the "amount of the attribute or characteristic consumed." This was in effect no more than a cosmetic redefinition of the centerpiece of the derivation, and in a later paper (Niedercorn and Bechdolt, 1972), the authors failed to respond meaningfully to the criticisms.

3.4.4. The 1960s in Review

The various approaches taken to the theoretical derivation of the spatial interaction model helped establish the integrity of the model from a variety of viewpoints. Researchers tended to embrace the derivation that was most relevant to their field of endeavor. Huff's derivation was applied in market research, while Wilson's work was applauded in planning circles — according to Gould (1972a, p 696), the entropy maximizing derivation "raises the gravity model phoenix-like from the ashes, and places it upon a secure theoretical foundation for the first time."

The derivations differed slightly in the conceptualization of the determinants of interaction, but the common thread running through the specifications was the identification of three classes of determinants: those specific to the origin, those peculiar to the destination, and the dyadic element; each of these was controlled by parameters such as β, that were supposedly universal. It would be a fair generalization, then, to infer that the overall algebraic structure of the generic equation (3.7) was now accepted, but that the individual elements were subject to varying interpretations and fine tuning, depending upon the context of study. Of particular importance was the appropriateness of constraints on the total flows out of and into origins and destinations. Marketing applications, for instance, were unlikely to require rigid constraints on patronage of stores, whereas planners for centrally administered facilities such as school busing would face very tangible limits, such as classroom capacities.

3.5. Recent Developments

The 1970s were years of consolidation of the concepts of spatial interaction. Efforts focused on five principal areas. First, there was a scrutinization of the concepts of interaction, with particular reference to the interpretation of the model parameters under various distributions of population over space (Ingram, 1971; Curry, 1972; Cliff, Martin and Ord, 1974; Kirby, 1974; Vickerman, 1974; Johnston, 1975; Kirby and Leese, 1978; Sheppard, 1978; Kau and Sirmons, 1979). In particular, this group of studies addressed the theoretical underpinnings of the model with respect to autocorrelation in zone size. A sobering conclusion was the recognition of "the large volume of work which remains to be undertaken" (Cliff, Martin and Ord, 1976, p 342).

A second area of concentration was the development of calibration procedures for various classes of interaction models (for example, Batty, 1971; Evans, 1971; Batty and Mackie, 1972; Stetzer, 1976; Williams, 1977; Pirie, 1979; Tobler, 1979). McFadden (1973), Baxter (1979) and Southworth (1981) explored the use of logit transformations in interaction modeling.

A third focus of research was the study of the parameters of the interaction model under aggregation of the origin and destination nodes. These included the works of Cesario (1974), Beardwood and Kirby (1975), Openshaw (1976;1977b), Masser (1977) and Webber (1980). Fourth, the pattern of decay of interaction with distance was examined by Taylor (1971) and Openshaw and Connolly (1977).

Finally, Fotheringham (1983) proposed a broad class of competing-destinations models, extending Stouffer's (1940) concepts of intervening opportunities. Passing mention must also be made of Ellis and Van Doren (1966) and Ellis (1967), who drew analogies between human interaction and electrical flow, and Griesinger (1979), who reexamined the gravitational concepts.

3.6. Conclusions

This chapter has traced the history of the spatial interaction model, from its beginnings as a gravitational analogy, through a series of empirical experiences and theoretical generalizations, and finally the attempts at deductive derivation. There is a consensus on the generic form of (3.7); within that framework, a number of refinements of the specification have been made in particular application contexts. Gale *et al* (1982) evaluated a number of contemporary models of movement, but indicated that the traditional spatial interaction equation was most satisfactory.

Like the theories of Von Thunen and Christaller before it, the spatial interaction model operates in an abstracted space. The model takes account of characteristics of the origin and destination nodes, but does not recognize other variations in the landscape. Distance (or standard observable surrogates, such as time or cost) is the only dyad-specific factor in the equation. Perceptual and attitudinal affinities or prejudices peculiar to origin-destination pairs are not considered.

In the context of functional regions, there is a need to explore the applicability of interaction modeling to non-uniform environments. Any additional terms in the equation will necessarily have to be specific to the dyad (since constants, or terms specific to either the origin or the destination, may be absorbed within the emissivity and attraction), and thus complement the distance term.

Baxter and Ewing (1984) suggested two reasons why distance alone may be an inadequate representation of the dyad-specific factors in interaction; one was the problem of spatial structure (this will be covered in Chapter 5).

The second reason was complementarity between origin and destination characteristics. Given that the classical concepts of the functional region revolve

65

around notions of communities and affinities, it would appear that the inclusion of dyad-specific factors in the interaction equation may produce an appropriate model for interaction in heterogeneous environments, and thus shed light on the origins of functional regions. These ideas will be developed in the next chapter.

Chapter 4

THE BLACK/WHITE MODEL

4.1. Introduction

One of the conclusions of Chapter 2 was that the spatial interaction model, in recognizing scale differences and the effect of distance on interactions, had a role to play in regionalization theory. It is clear from the review in Chapter 3, however, that except for the variables subsumed under the emissivity and attraction of the origin and destination, the interaction model in its traditional form is bound by assumptions of isotropism, and is therefore not suitably equipped for application across functional regions. The purpose of this chapter is to examine how the model can be adapted to take account of qualitative incompatibilities between nodes, that is, to address cultural, political and economic *barriers* to interaction, that may result from a variety of independent dissimilarities between nodes, and which are the bases of functional segregation. The spatial organization of these variations will be incorporated into a model for the behavior of functional regions, establishing their theoretical and conceptual basis.

Several types of regions were identified in Chapter 1, under the descriptive taxonomy. By way of recapitulation, these included, for example, physiographic, climatic, cultural and economic regions. A structural taxonomy was described too, under which scheme the functional region was introduced. For purposes of illustration, it will be convenient to link the structurally defined functional region with a corresponding item from the descriptive taxonomy. Since the functional region is a phenomenon of *human* activity, with a strong behavioral undercurrent in its conceptualization, its most obvious descriptive counterpart is the cultural region. While culture is by no means the sole basis of the functional region, it will be used to illustrate the development of the reasoning in the following sections.

4.2. Derivation of the Black/White Model

4.2.1. The Compatibility Factor

The derivation of the spatial interaction model has been covered in some detail in Chapter 3. Now consider a similar derivation in a situation of cultural conflict — differing cultural groups are distinguished by the arbitrary labels BLACK and WHITE, hence the title of the model. Let a COMPATIBILITY FACTOR R_{ij} between any origin-destination pair, i and j, represent the emotional affinity[4] of i for j. In a situation of cultural homogeneity, $R_{ij} = 1.0$; beyond the confines of cultural isotropism, compatibility may fall below 1.0, to a minimum of 0.0. Except inasmuch as the compatibility may in certain cases theoretically exceed a value of 1.0, it may be compared with a probability that an individual at i will be willing to overcome the cultural barriers to interact with j, given a set of alternatives perfectly compatible with i.

---

[4] The term "*emotional* affinity" is specific to the context of this example; in practise it may be substituted with other dyad-specific factors not included in the distance term, such as political or economic affinities.

4.2.2. Luce Choice Structure

Under the Luce (1959) "irrelevant alternatives" choice axiom, an individual makes a decision based on simultaneous consideration of all the properties of the available alternatives, weighing advantages on one variable against disadvantages on another. Huff's (1963) derivation of the spatial interaction model (Chapter 3) assumed a Luce choice structure: a customer evaluated options by weighing physical attraction against distance. Introducing compatibility as a third factor in the choice process, (3.8) becomes

$$p(j|S) = \frac{A_j\, f(d_{ij})\, R_{ij}}{\sum_m A_m\, f(d_{im})\, R_{im}} \qquad \ldots (4.1)$$

using the same notation as in (3.8). Then

$$I_{ij} = w_i \frac{A_j\, f(d_{ij})\, R_{ij}}{\sum_m A_m\, f(d_{im})\, R_{im}} \qquad \ldots (4.2)$$

$$= \xi^*_i\, A_j\, f(d_{ij})\, R_{ij} \qquad \ldots (4.3)$$

where

$$\xi^*_i = \frac{w_i}{\sum_m A_m\, f(d_{im})\, R_{im}} \qquad \ldots (4.4)$$

ξ^*_i is again origin-specific, and may therefore be considered as an alternative expression of emissivity.

4.2.3. Tversky Choice Axiom

An alternative approach to individual choice theory is that of elimination by aspects (Tversky, 1972). Under this scheme, an individual makes a series of decisions about the alternatives, at *several* levels. In choosing a shopping destination, for instance, one might first isolate the set of facilities that offered the good one wished to purchase. The second decision level would perhaps involve the choice between competing store chains; the final level of choice may be between stores of a given chain, the selection being made on the basis of distance, appearance or management.

The principal distinction between the Luce and Tversky approaches to choice theory lies in the way the probabilities are distributed among the alternatives. In Tversky's view the choices are arranged *exclusively* and therefore hierarchically, so that the probability of visiting a given facility is the compound probability of selection of each of its characteristics (assuming that those characteristics are measured on dimensions that are mutually independent).

To apply the Tversky choice axiom to the derivation of the Black/White model, the universe S must first be stratified into p_i internally homogeneous cultural subsets, $a_{i1}, a_{i2}, \ldots a_{ip_i}$, for each origin i, such that

$$a_{i1} \cup a_{i2} \cup \ldots \cup a_{ip_i} = S \quad \forall i \qquad \ldots (4.5)$$

The probability of choosing a given destination j is now the compound probability of choosing the subset to which j belongs, given i — that is, $p(a_{ij})$ — and the probability of choosing j, given a_{ij}. The first of those probabilities is simply R_{ij}. The second level of choice operates independently of the shared characteristics of the origin and destination, and is in that respect common to the spatial interaction model:

$$p(j|a_{ij}) = \frac{A_j\, f(d_{ij})}{\sum_{m \in a_{ij}} A_m\, f(d_{im})} \qquad \ldots (4.6)$$

Multiplying the probabilities,

$$p(j|S) = p(a_{ij})\, p(j|a_{ij}) \qquad \ldots (4.7)$$

$$= R_{ij} \frac{A_j\, f(d_{ij})}{\sum_{m \in a_{ij}} A_m\, f(d_{im})} \qquad \ldots (4.8)$$

The interaction from i to j is

$$I_{ij} = w_i R_{ij} \frac{A_j\, f(d_{ij})}{\sum_{m \in a_{ij}} A_m\, f(d_{im})} \qquad \ldots (4.9)$$

The denominator in this case is specific to the *dyad*, rather than to the origin alone. Substituting

$$R^*_{ij} = \frac{R_{ij}}{\sum_{m \in a_{ij}} A_m\, f(d_{im})} \qquad \ldots (4.10)$$

(4.9) becomes

$$I_{ij} = w_i A_j\, f(d_{ij})\, R^*_{ij} \qquad \ldots (4.11)$$

Since the emissivity and attraction terms are in practise never recognized in tangible terms, but only approximated by population, income or other surrogates, their algebraic components (which are one distinction between the specifications 4.3 and 4.11) have no bearing on their interpretability. Both (4.3) and (4.11) represent interaction in terms of three categories of variables: those specific to the origin, those specific to the destination and those peculiar to the dyad. In the traditional spatial interaction model, the only variable in the third category was distance; the derivations above introduce **R** and **R*** as additional elements.

Since **R*** is composed of accumulated unknowns, and is dyad-specific, its role in predictive modeling (and thence the utility of the Tversky-derived Black/White model) at this level is questionable — this criticism will be nullified in the spatially aggregated versions of the model, to be introduced later in this chapter.

Nevertheless, it would be useful to generalize the Black/White model specification as

$$I_{ij} = E_i \ A_j \ f(d_{ij}) \ R_{ij} \qquad \ldots (4.12)$$

where R_{ij} represents the compatibility factor (under the Luce derivation) or R^*_{ij} (under the Tversky derivation). It must be noted that (4.12) represents the most generic form of the model, and that the unknowns are not interpretable *across* derivations: that is, parameters recovered in (4.3) are not valid in (4.11).

4.3. Properties of the Model

4.3.1. Theoretical Properties

4.3.1.1. Symmetry of R

The compatibility factor was introduced as a variable specific to an origin-destination pair. Assuming that all destinations also act as origins, and vice versa, **R** will be a square matrix. Reciprocity of emotional affinity is never to be assumed in any social situation, and accordingly in the Black/White formulation, R_{ij} does not necessarily equal R_{ji}. In practise, however, the emissivity, attraction and compatibility variables in the model are empirically derived quantities; asymmetry in observed interactions cannot conclusively be ascribed to asymmetry in **R**, since it may be the result of idiosyncratic variation on other unidentified dimensions, and therefore treated as noise. One may therefore consider **R** to be symmetric in calibration procedures.

4.3.1.2. Transitivity

It is customary to examine the topological consistency of an interaction model by testing for transitivity in predicted flows. The test may be stated as follows: in any closed triangular circuit in the network,

$$I_{ij} > I_{ji}\,,\ I_{jk} > I_{kj} \Rightarrow I_{ik} > I_{ki} \qquad \ldots (4.13)$$

Since $I_{ij} > I_{ji}$ and $I_{jk} > I_{kj}$,

$$\frac{I_{ij}}{I_{ji}} \cdot \frac{I_{jk}}{I_{kj}} > 1.0 \qquad \ldots (4.14)$$

In the case of the spatial interaction model,

$$\frac{I_{ij}}{I_{ji}} \cdot \frac{I_{jk}}{I_{kj}} = \frac{E_iA_jf(d_{ij})}{E_jA_if(d_{ji})} \cdot \frac{E_jA_kf(d_{jk})}{E_kA_jf(d_{kj})} \qquad \ldots (4.15)$$

$$= \frac{E_iA_k}{E_kA_i} = \frac{I_{ik}}{I_{ki}} > 1.0 \qquad \ldots (4.16)$$

and therefore $I_{ik} > I_{ki}$. A similar proof can be constructed for the Black/White model:

$$\frac{I_{ij}}{I_{ji}} \cdot \frac{I_{jk}}{I_{kj}} = \frac{E_iA_jf(d_{ij})R_{ij}}{E_jA_if(d_{ji})R_{ji}} \cdot \frac{E_jA_kf(d_{jk})R_{jk}}{E_kA_jf(d_{kj})R_{kj}} \qquad \ldots (4.17)$$

$$= \frac{E_iA_k}{E_kA_i} = \frac{I_{ik}}{I_{ki}} > 1.0 \qquad \ldots (4.18)$$

The transition from (4.17) to (4.18) is contingent on **R** being symmetric. When **R** is asymmetric, transitivity may or may not exist, depending on the degree of asymmetry of **R**. As pointed out earlier, **R** can usually be assumed to be symmetric, so that in practise, predictions of the Black/White model are transitive.

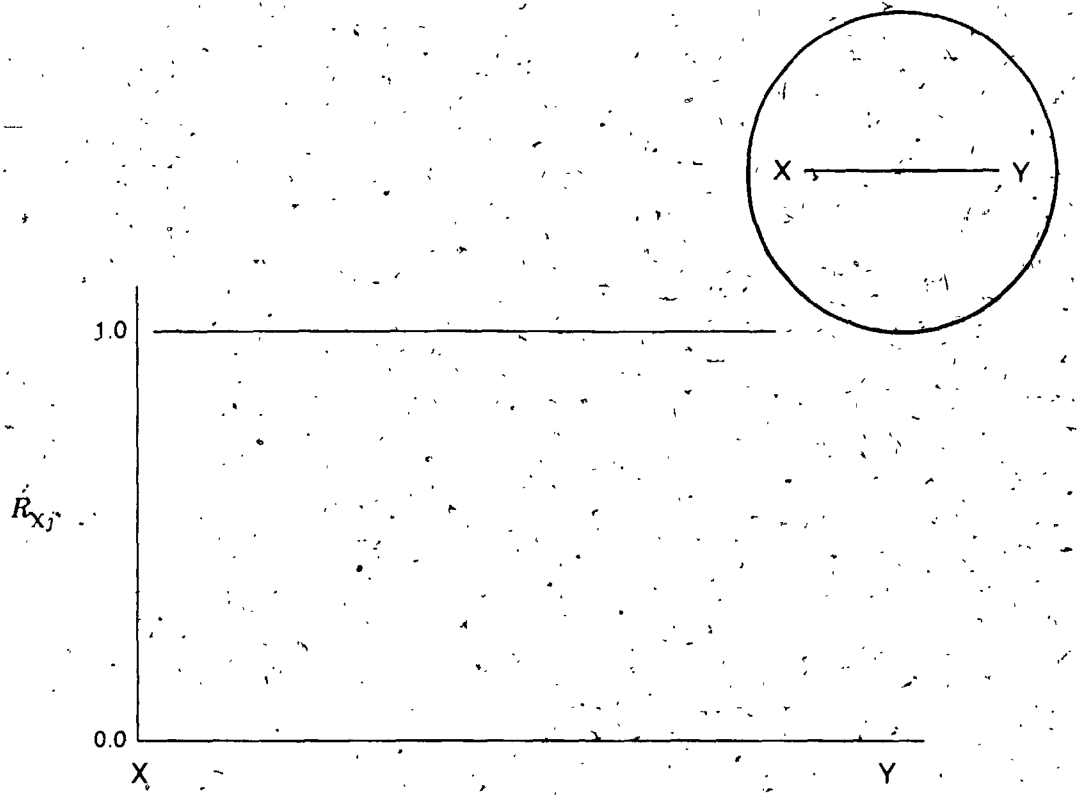
4.3.2. Operational Properties

The compatibility factor R_{ij} can be thought of as constituting (or derived from) a surface of destination compatibilities, for any given origin, and vice versa (since **R** addresses a *human* landscape, which is essentially discrete, the surface may not be continuous). When superimposed upon the abstracted isotropic landscape of the traditional interaction model, the surface determines the pattern of interactions of all destinations with the origin under consideration. The interactions of all origins with all destinations in a universe are, by extension, determined by the combined influences of the compatibility surfaces associated with all origins. The following sections illustrate the workings of the Black/White model in various compatibility scenarios, with one and more origins. Rather than engage in an exhaustive cartographic documentation of the Black/White effect, attention will be focused on the conceptual revelations and practical significance of each case.

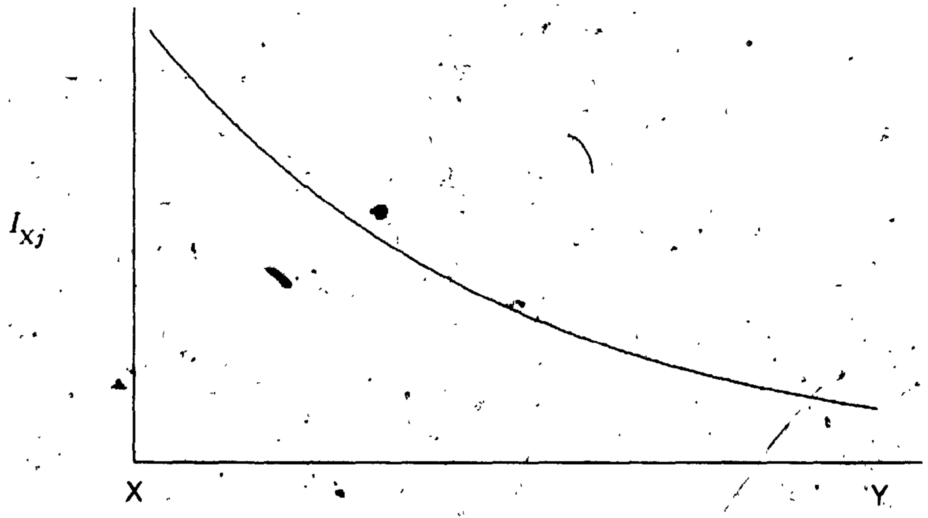
4.3.2.1. Case 1 : Single Origin, One Dimension

Under the terms of the standard interaction model, compatibility is invariant across space — that is, its profile is flat (Figure 4-1a). A plot of interactions (corrected for emissivity and attraction) against distance reveals an exponential decay (Figure 4-1b), which may be linearized by appropriate transformation of the axes.

Now consider a situation where the more distant reaches of the universe exhibit a lower compatibility with the origin than do the immediate surroundings

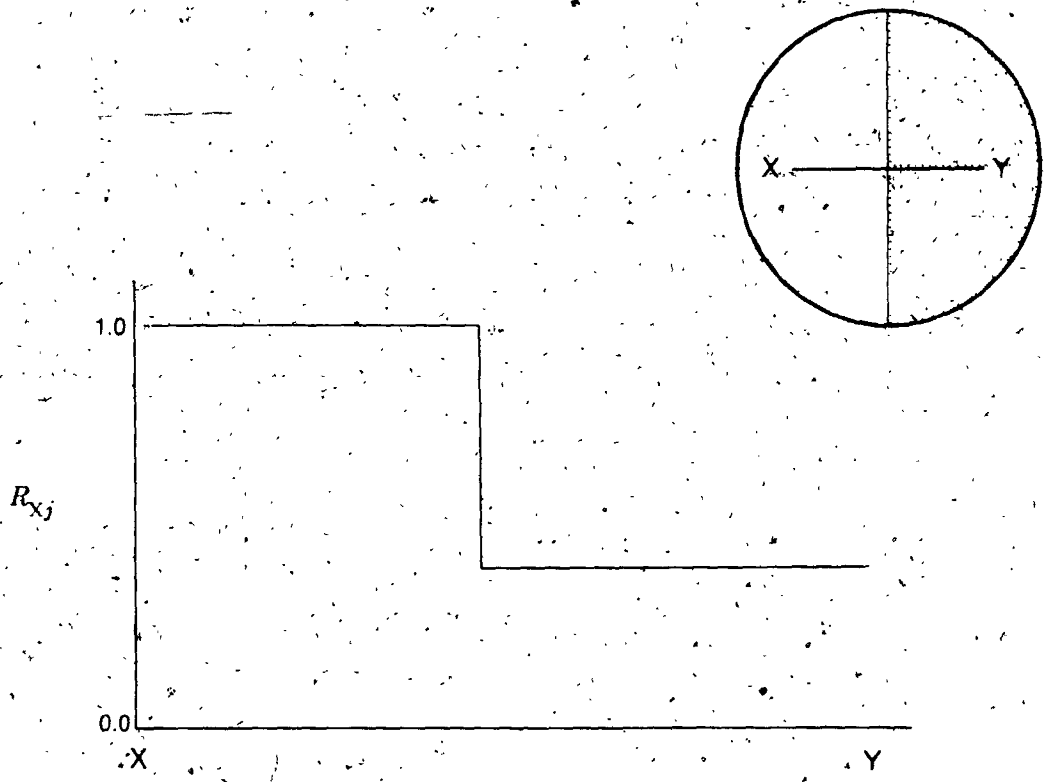


a: Flat compatibility profile

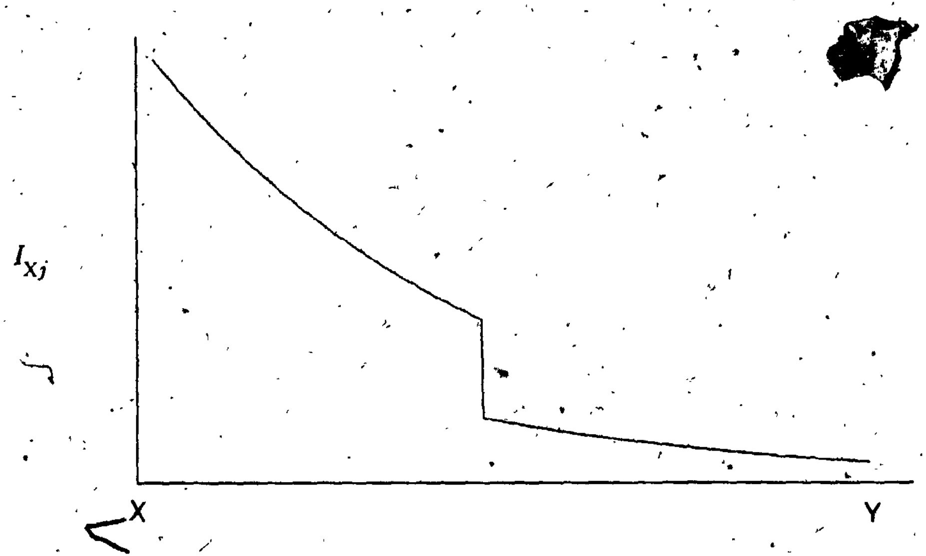


b: Exponential decay of interaction

Figure 4–1: Conventional interaction curve



a: Compatibility profile across cliff-type boundary



b: Cliff-type interaction profile

Figure 4-2: Cliff-type boundary

of that origin (Figure 4-2a). The corresponding interaction profile is shown in Figure 4-2b. In practise the regional boundary may be fuzzy (Figure 4-3), or the landscape may be more complex: Figure 4-4 illustrates the case of spatially disjoint cultural concentrations, as manifested in the settlement patterns of Native American and other cultural minorities in North America, or in the political situation of West Berlin; Figure 4-5 illustrates the compatibility and interaction profiles across several regional types.

4.3.2.2. Case 2 : Multiple Origins and Destinations, One Dimension

The spatial interaction model is undoubtedly a useful concept in purely academic terms; it fulfills an important function too in predicting flows in economic situations, and is thereby a planning tool. It is in this context that the following scenario should be viewed.

The Black/White modeling may be applied in a context similar to that described by Hotelling (1929) in regard to competitive pricing. A continuum of points is assumed to lie along a bounded line. Each point serves as both a demand location and a facility; emissivities and attractions remain constant along the line. Interactions (which may be thought of as shopping trips) between the points may be simulated using the spatial interaction model. The problem draws attention to the importance of position along the line. Points located in the center of the line are able to draw their clientele from both directions, and therefore register the greatest patronage, whereas those at the extremities of the line segment draw their customers from a relatively restricted area. This illustrates the economic advantage of a central location (Figure 4-6).

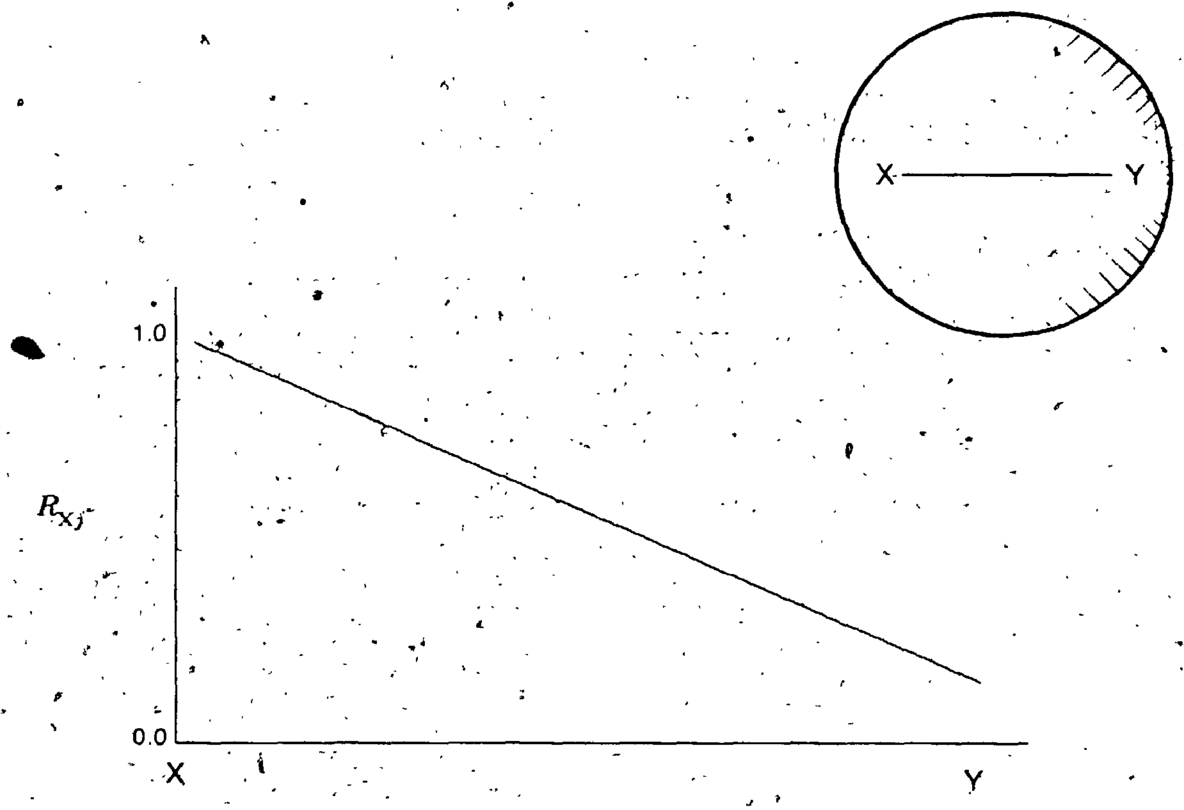
When the interactions along this line are simulated with the Black/White model (assuming a regional boundary at the center) the profile takes on a BUTTERFLY form: it splits into two sub-curves, each a miniature replica of the original curve (Figure 4-7). Depending on the value of the compatibility factor, of that origin (Figure 4-2a). The corresponding interaction profile is shown in Figure 4-2b. In practise the regional boundary may be fuzzy (Figure 4-3), or the landscape may be more complex: Figure 4-4 illustrates the case of spatially disjoint cultural concentrations, as manifested in the settlement patterns of Native American and other cultural minorities in North America, or in the political situation of West Berlin; Figure 4-5 illustrates the compatibility and interaction profiles across several regional types.

4.3.2.2. Case 2 : Multiple Origins and Destinations, One Dimension

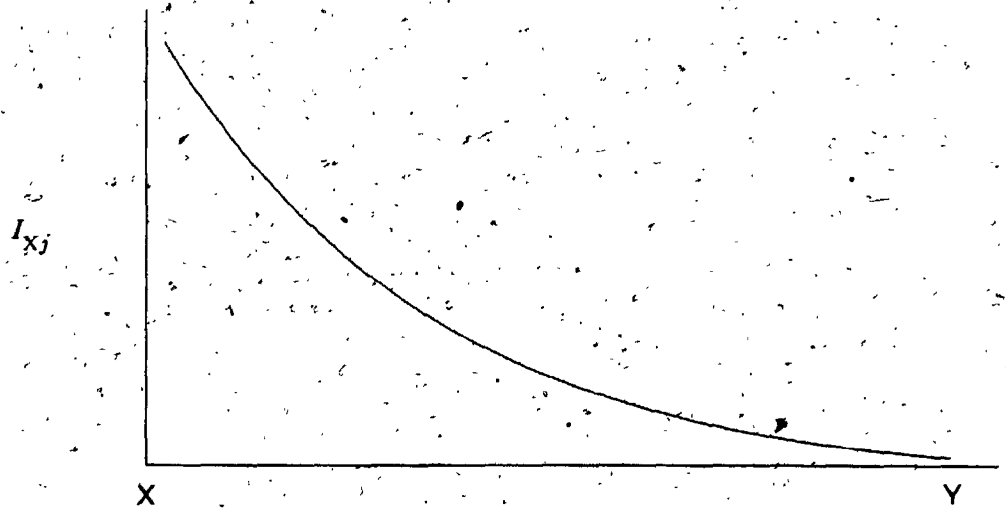
The spatial interaction model is undoubtedly a useful concept in purely academic terms; it fulfills an important function too in predicting flows in economic situations, and is thereby a planning tool. It is in this context that the following scenario should be viewed.

The Black/White modeling may be applied in a context similar to that described by Hotelling (1929) in regard to competitive pricing. A continuum of points is assumed to lie along a bounded line. Each point serves as both a demand location and a facility; emissivities and attractions remain constant along the line. Interactions (which may be thought of as shopping trips) between the points may be simulated using the spatial interaction model. The problem draws attention to the importance of position along the line. Points located in the center of the line are able to draw their clientele from both directions, and therefore register the greatest patronage, whereas those at the extremities of the line segment draw their customers from a relatively restricted area. This illustrates ~~the~~ economic advantage of a central location (Figure 4-6).

When the interactions along this line are simulated with the Black/White model (assuming a regional boundary at the center) the profile takes on a BUTTERFLY form: it splits into two sub-curves, each a miniature replica of the original curve (Figure 4-7). Depending on the value of the compatibility factor,

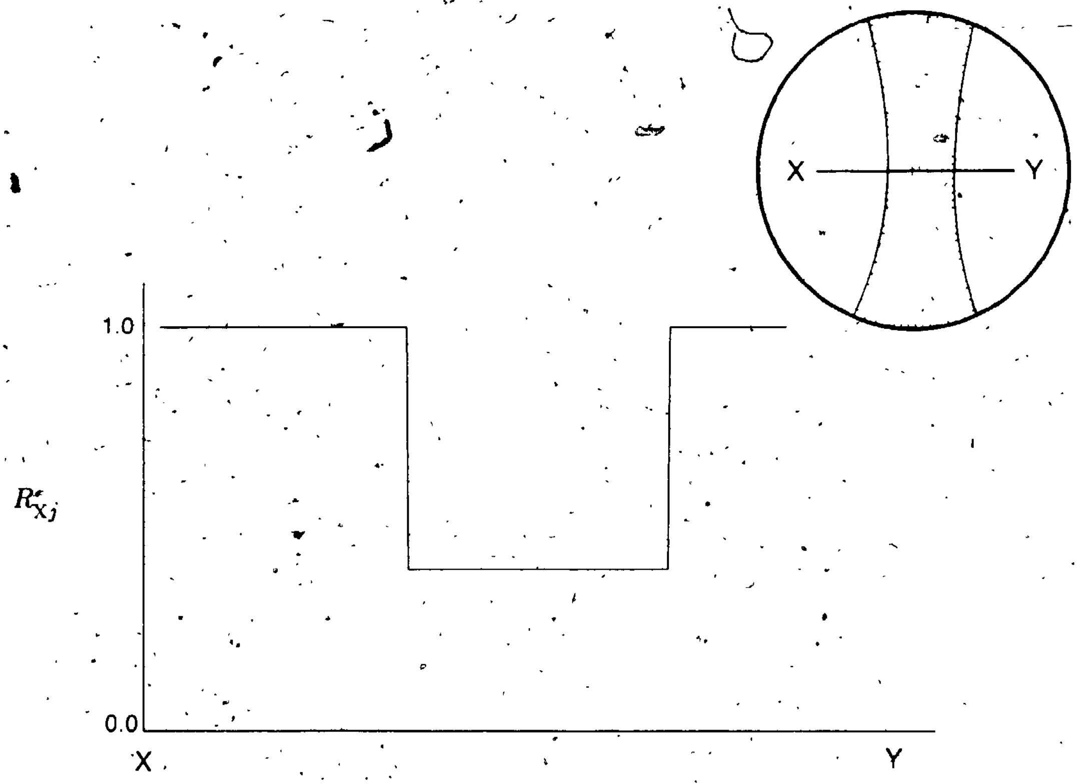


a: Compatibility profile across fuzzy boundary

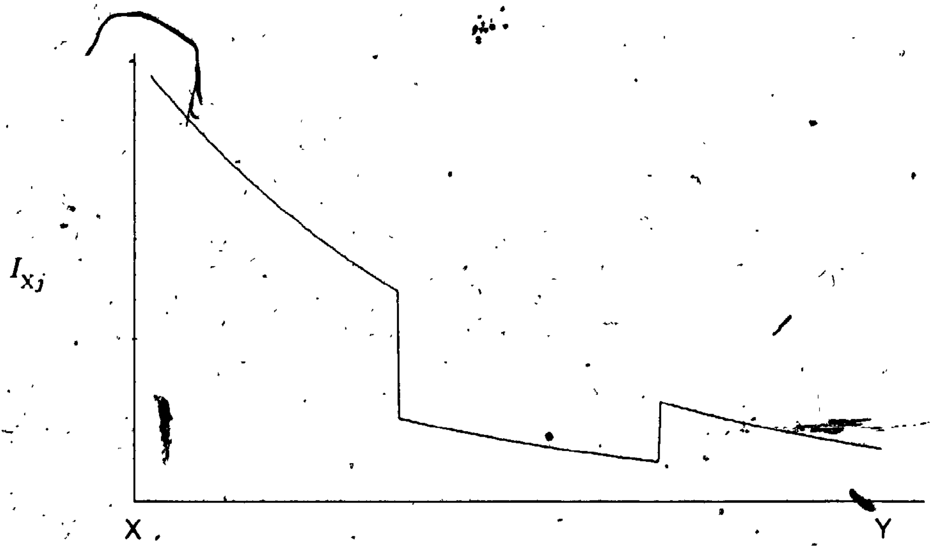


b: Fuzzy interaction profile

Figure 4–3: Fuzzy boundary

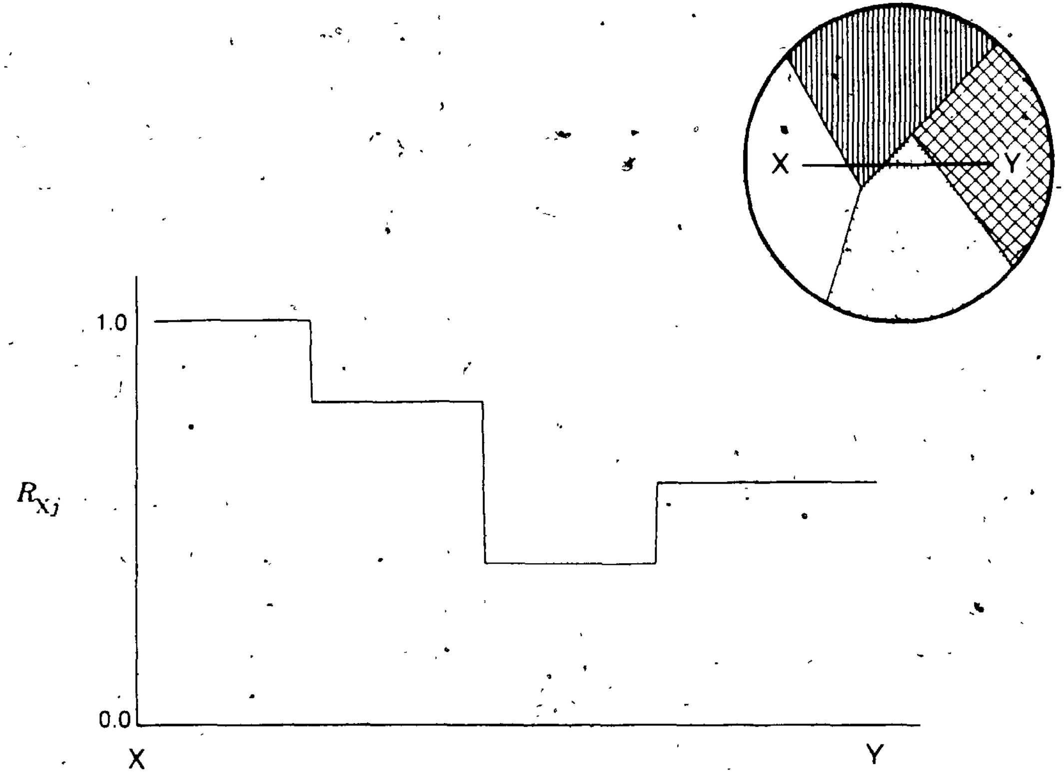


a: Compatibility profile across multiple bounds (Berlin)

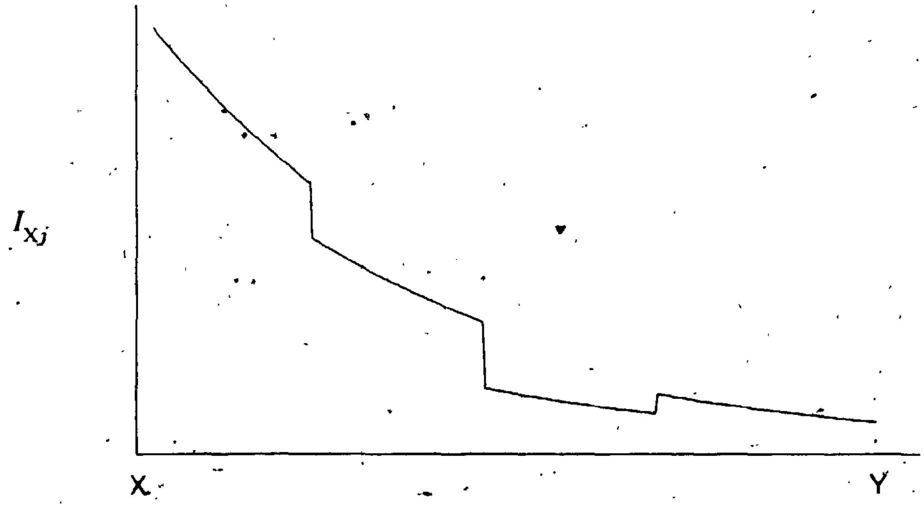


b: Berlin interaction profile

Figure 4–4: Berlin boundary



a: Compatibility profile across multiple regions



b: Multiple region interaction profile

Figure 4–5: Multiple boundaries

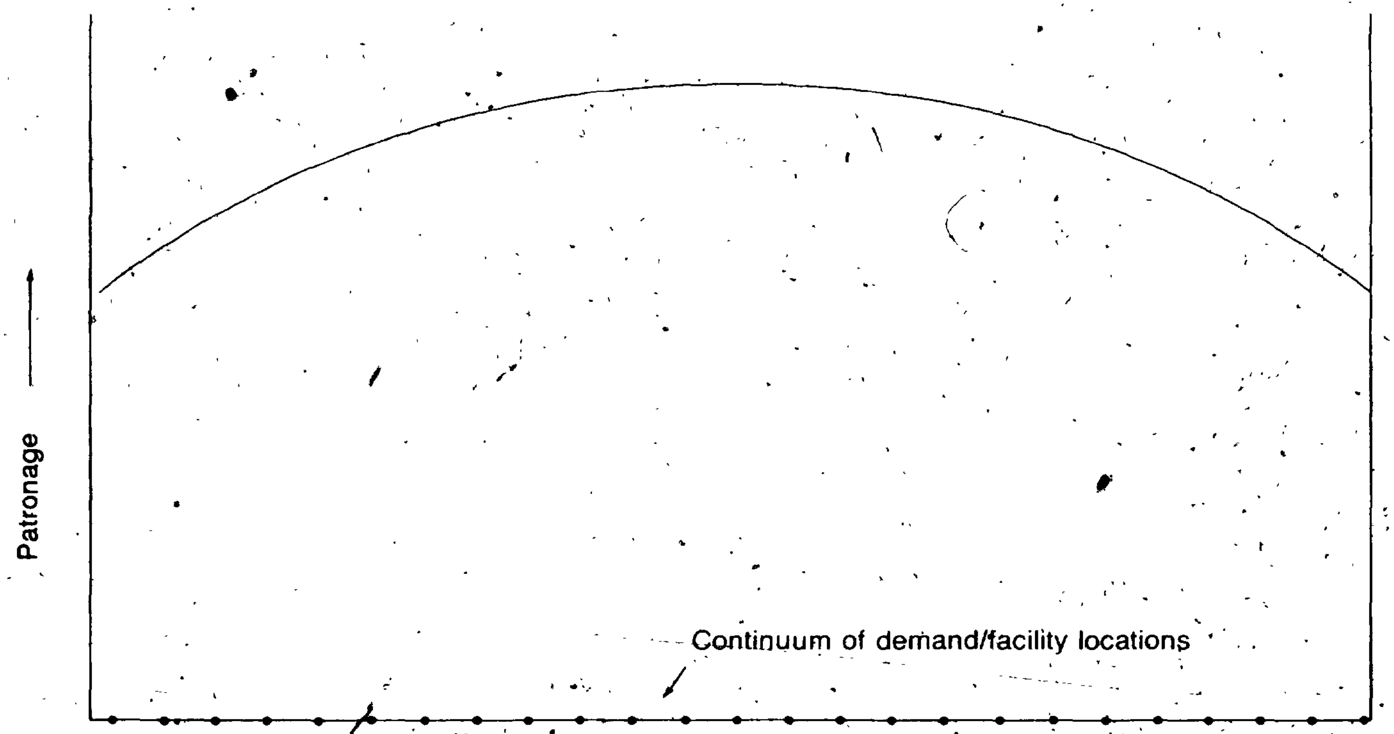


Figure 4–6: Utility of a central location

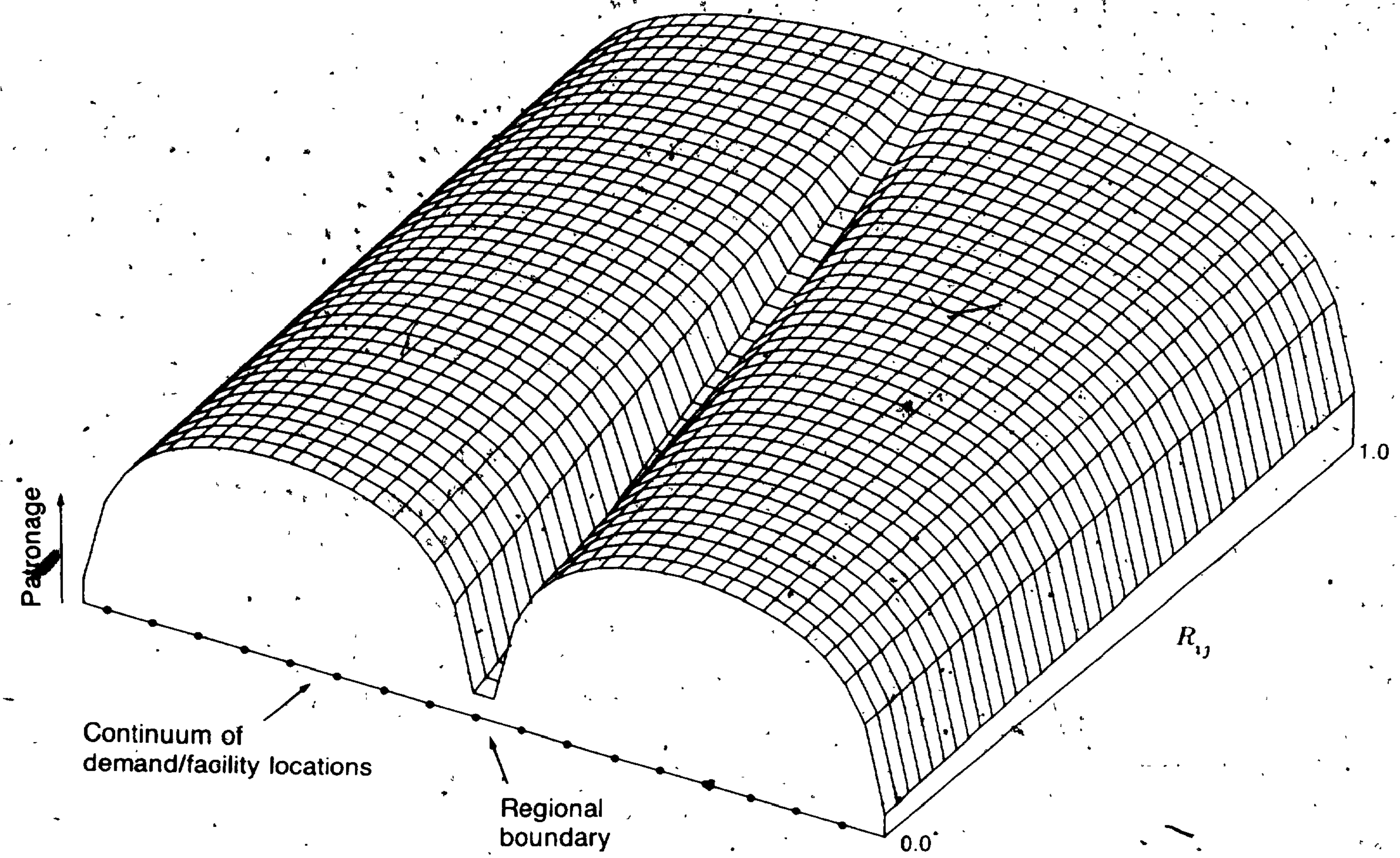
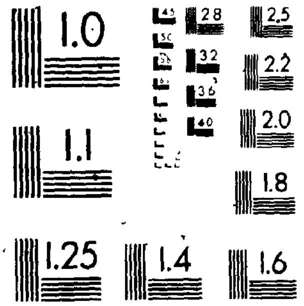


Figure 4 – 7: Butterfly effect

2

MICROCOPY RESOLUTION TEST CHART
NBS 1010a
ANSI and ISO TEST CHART No 2



patronage (defined as the sum of interactions with all other places on the line segment) at the center may decline to the level of that of the peripheral points. The points receiving the greatest number of patrons are those located towards the centers of the undivided segments of the line. The Butterfly effect is one reason why areas in the vicinity of politically and militarily sensitive borders are economically depressed.

4.4. Geographical Roles for the Black/White Model

4.4.1. Functional Distance and Black/White Topology

In the two cases considered above, it was relatively simple to construct cartographic illustrations of the Black/White effect, since the origins and destinations were located along a line. The study of Black/White effects in two dimensions is considerably more complex, as it involves the superimposition of 3-dimensional compatibility surfaces for each origin in the universe. A complete appreciation of the behavior of Black/White systems entails a study of the transformations imposed by them on distance.

The geographical literature on spatial interaction contains frequent references to permeable boundaries, perceived distance and functional distance. Gould (1972b) examines mathematical transformations of distance that correspond with human perceptions; and Brown and Holmes (1971) suggest that the matrix of mean first-passage times in their Markov chain analysis of interactions represents the matrix of functional distance (the technique was reviewed in Chapter 2). Mackay (1958) very clearly makes the connection between functional distance and regional bias in interaction:

> . . . an Ontario city interacts with a Quebec city as if it were five times as far away as it really is by comparison with a Quebec city of the same population and separation.

> On the average, an Ontario city receives roughly a fifth to a tenth the telephone traffic of a Quebec city . . .
>
> [Mackay, 1958, p 3-5]

In the first statement, Mackay alludes to a concept of functional distance:

$$I_{ij} = E_i A_j f(d_{ij} \times 5.0) \qquad \ldots (4.19)$$

whereas in the second he suggests the presence of a dyad-specific multiplier in the equation:

$$I_{ij} = E_i A_j f(d_{ij}) \times 0.2 \qquad \ldots (4.20)$$

Except under the assumption of an inverse linear function of distance, the two explanations of the same phenomenon are inconsistent with each other.

The concept of functional distance in the case of the Black/White model is illustrated in Figure 4-8. The functional distance between X and Y2 is derived from a comparable relationship in a single-region landscape: the combined effect of distance and compatibility on the interaction between X and Y2 is equivalent to the effect of distance alone on interactions between X and Y1. The functional distance $d^*_{X,Y2}$ between X and Y2 is therefore equal to the geographic distance $d_{X,Y1}$ between X and Y1. Mathematically, the relationship may be defined as

$$d^*_{ij} = f^{-1}[f(d_{ij})\, R_{ij}] \qquad \ldots (4.21)$$

In the literatures of transportation and economics, the interaction model is usually quoted with a "separation" or "cost" term (c_{ij}) in place of distance (d_{ij}). The Black/White equation may be accommodated within this specification by means of a simple maneuver:

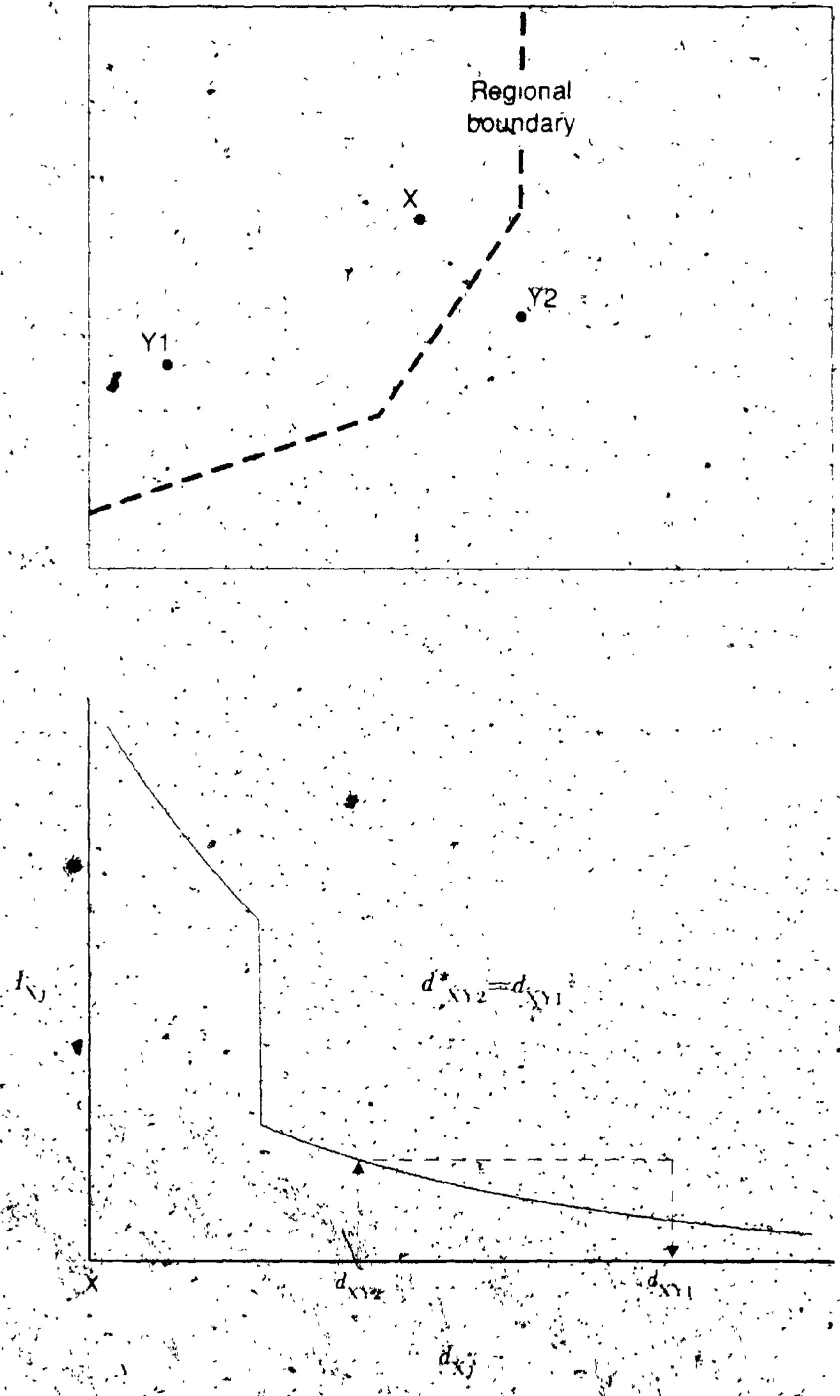


Figure 4 – 8: Black/White concept of functional distance

$$f(c_{ij}) = g(d_{ij}, R_{ij}) \qquad \ldots (4.22)$$

so that

$$I_{ij} = E_i A_j f(c_{ij}) \qquad \ldots (4.23)$$

4.4.2 Implications of Black/White Topology

In the theoretical study of trade areas, one encounters the concept of Thiessen polygons, or Dirichlet regions. They are the geometric duals of the lines joining pairs of facilities, and are derived by finding the perpendicular bisectors of those lines; under the assumption of normative, nearest-place behavior, the bisectors define the spatial limits of patronage of a given facility.

Thiessen polygons are traditionally constructed in Euclidean space; they may, however, be derived under a functional distance topology (see Getis, 1963), where the effective distance between any two points in space is equated with functional distance, as defined above.

4.4.3 Other Implications

The field of location-allocation is concerned with prescribing locations for facilities, and allocating demand to those facilities in order to satisfy some objective, which is usually the minimization of total distance traveled. Many applications of the modeling assume inelastic demand, whereas in reality, individuals situated close to a facility are more likely to make use of it. There have been attempts to make the allocation of demand to facilities more realistic: Cooper (1972) employed a transportation algorithm when service capacities of facilities were limited; Goodchild and Booth (1980) used observed interactions to calibrate an interaction model, and respecified the objective function in terms of probability of utilization of facilities. Hodgson (1978) described an independent solution procedure.

Having recognized that the likelihood of patronage of facilities should be taken into account when locating them, the issue of compatibility needs to be considered. Several prominent researchers routinely conduct facility location programs in less developed countries, where social and cultural friction is often a matter of some importance. A recognition of social affinities (by means of calibrating the Black/White model) would give prescriptions, particularly in these environments, more authenticity and credibility.

4.5. The Black/White Model and Functional Regions

The Black/White model was developed specifically to address the shortcomings of the traditional spatial interaction model. In the previous section, the model was introduced in continuous space; attention is now devoted to its role in the context of functional regions. Compatibility as defined in the foregoing sections has been assumed to vary continuously over space (except in the Tversky derivation of the Black/White model). In contrast, regionalization, by definition, involves a generalization over *discrete* compartments of space; in that light, space may be viewed as continuous when contained within a single ESU or region.

In the context of regions, compatibility is an attribute of a pair of *regions*, whereas the interaction model addresses pairs of *points* (or ESUs). Therefore the compatibility term in the Black/White model must be adapted to address the *regions* to which the points (or ESUs) belong.

4.5.1. Classes of Black/White Structures

The manner in which regions are addressed should be able to accommodate the various regional structures described in the literature on functional regions (Chapter 1) and in the foregoing discussions. Briefly, these are:

- there may be two or more regions constituting the universe
- functional regions defined on several bases may coexist upon the same landscape

Four alternative CLASSES of the Black/White model are presented in Figure 4-9. Twelve ESUs (that is, $n=12$), are shown in various aggregated configurations (the boundaries of the ESUs are immaterial in the present discussion, and are not shown in Figure 4-9; however, since whole ESUs are being aggregated, all the regional boundaries shown must lie along ESU borders). Regions may exist on different BASES, such as language, race, religion, etc, designated $k=1, 2, \ldots r$. On each basis there may exist p_k regions within the universe (for example, on a linguistic basis, a differentiation between English, French, Gaelic and Latin would yield $p_k=4$).

4.5.1.1. Class 1

In the simplest case of a biregional split on a single basis (Figure 4-9a), $r=1$ and $p=2$ ($p=1$ would represent a completely undifferentiated universe). For purposes of explanation, consider a hypothetical linguistic split between English and French. It may be assumed that within the context of a single study, the compatibility between English and French areas is constant, say 0.4 (assume here that **R** is symmetrical). Let this UNIVERSAL compatibility factor for a given pair of regional types be ρ. Whether or not ρ plays an active role in a given dyad is determined by a binary dissimilarity exponent, σ_{ij}, which takes a value of zero if i and j are in the same region, or a value of one otherwise. In this particular case, R_{ij} can take one of only two values, 1.0 and 0.4; the application of an appropriate exponent to ρ generates the required value:

$$R_{ij} = \rho^{\sigma_{ij}} \qquad (4.24)$$

The utility of this algebraic arrangement will be apparent later.

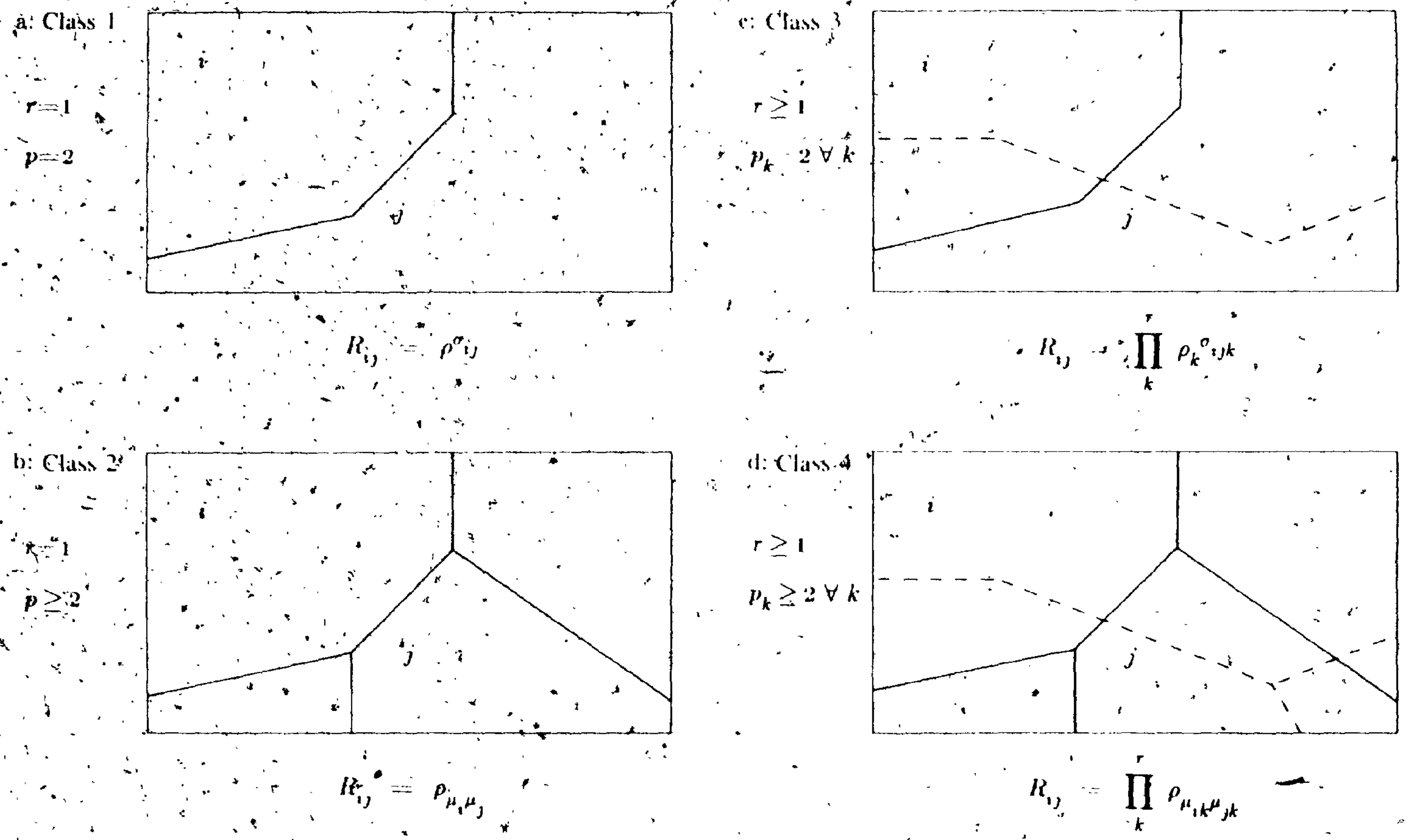


Figure 4-9: General classes of the Black/White Model

4.5.1.2. Class 2

Figure 4-9b illustrates the case of more than two regions differentiated on a single basis, for example, a linguistic distinction between English, French, Gaelic and Latin, which yields $r=1$, and $p=4$. Each of the $p(p-1)/2$ pairwise language combinations will be associated with its own value of ρ, which may therefore be subscripted with the identities of the *regions* to which the origin and destination belong. Let u_i represent the region to which i belongs in its role as an origin, and let v_j represent the membership of j in its role as a destination -- in general the membership of an ESU will be independent of its role, so that the vectors **u** and **v** will be equal; one may thus represent the spatial membership of an ESU i, regardless of its role as origin or destination, as μ_i: this gives the modeling a geographic rather than an economic or sociological slant. Now, if ESU 7 lies in a Gaelic region, and ESU 9 is French, and since $\mathbf{u}=\mathbf{v}=\mu$,

$$R_{79} = \rho_{\mu_7\mu_9} \qquad (4.25)$$

$$= \rho_{\text{Gaelic, French}} \qquad (4.26)$$

4.5.1.3. Class 3

When considering multiple bases of differentiation, the effective value of R_{ij} for a given dyad i-j is the product of the compatibility factors associated with each basis [illegible] which a differentiation exists between i and j. Figure 4-9c shows a possible situation where $r=2$, and $p_1=p_2=2$. One may think of $k=1$ (continuous line) as representing a linguistic dichotomy as before, and $k=2$ (broken line) as representing religious differentiation, say between Catholics and Protestants. The individual two-way configurations for the two bases of differentiation may be quite different. Let the compatibility on the basis of religion be 0.2. Then $\rho_1=0.4$, as before, and $\rho_2=0.2$. The exponent matrix σ may now be given the additional subscript k, which specifies the basis to which the exponent value refers: In this case ρ_1 or ρ_2. When a dyad i-j straddles both linguistic and religious boundaries, $R_{ij} = \prod_k \rho_k^{\sigma_{ijk}} = 0.4 \times 0.2 = 0.08$; if the ESU's are similar on both bases, R_{ij} is 1.0.

4.5.1.4. Class 4

Similar reasoning can be applied to the final possibility (Figure 4-9d), where each basis of differentiation is associated with multiple categories. The subscripting is inevitably more complex. Each zone i has k memberships, one for each basis of differentiation. The effective value of R_{ij} is $\prod_k \rho_{\mu_{ik}\mu_{jk}}$.

In the scheme described above, the cases involving only two regions per basis are really special circumstances of the respective multiple-region conditions. However, there is an operational advantage to be gained from employing the binary exponent matrix σ to control a single universal value of ρ – this will be evident when calibrating the model in the next chapter. Further, the representation gives the model a resemblance to the Multiplicative Competitive Interactive (MCI) models used in market research (see Nakanishi and Cooper, 1974; Achabal, Gorr and Mahajan, 1982).

4.5.2. Regional Boundary Profiles

In the treatment of the Black/White modeling of discrete regional groups, it has been assumed that boundaries between regions are clear-cut — this appears to be a popular characteristic (and shortcoming) of the treatment of regions in the literature, both in the conceptualization and the objective demarcation studies — a notable exception is the recent work of Leung (1984) on fuzzy sets. The profile of compatibility may not always take the form of a cliff. Regional boundaries may be fuzzily defined, without affecting the *extent* of decline of compatibility. Ideally, the boundary could assume a continuous range of forms, from a gently sloping profile (Figure 4-10, Curve a), through a cliff (Curve d) and

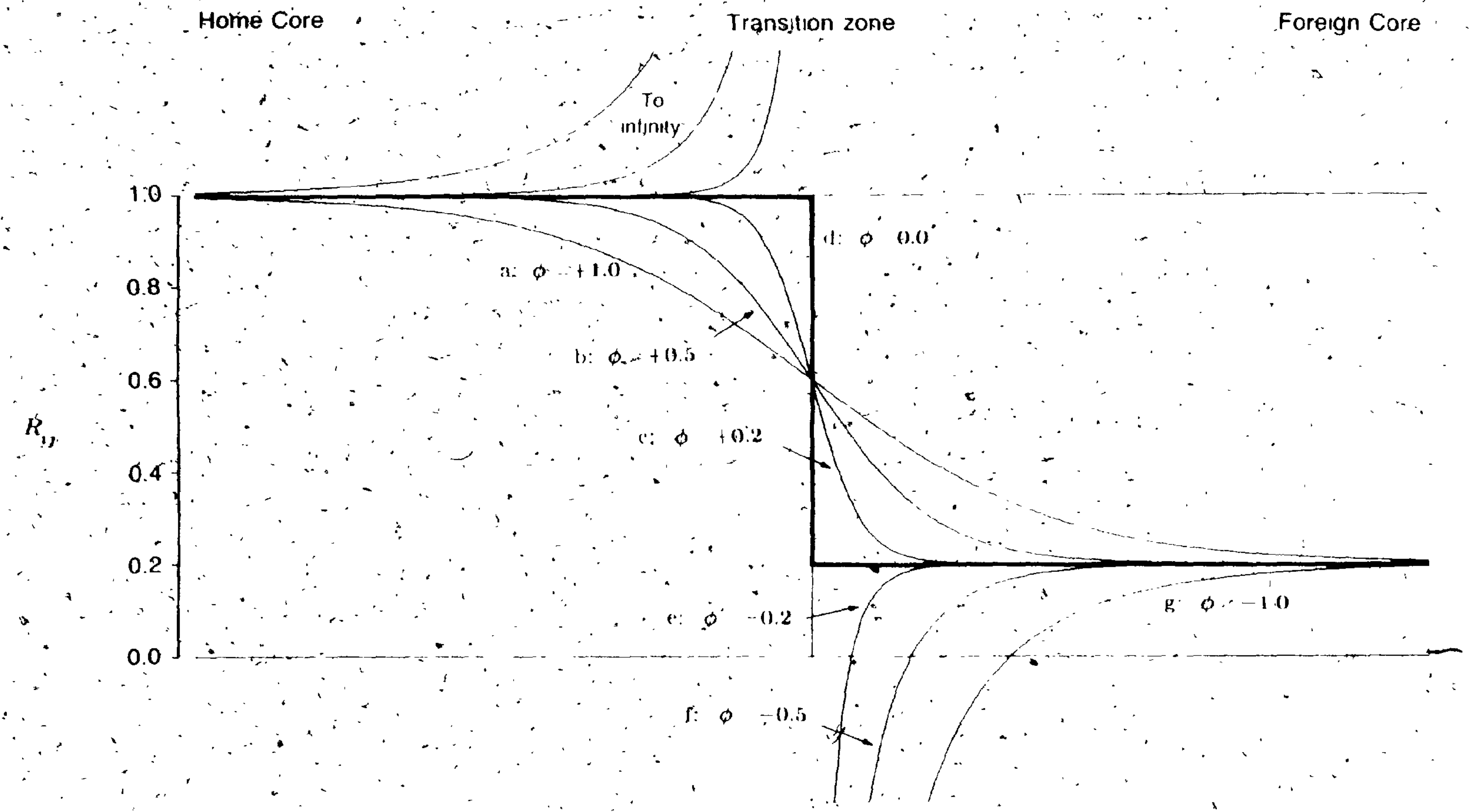


Figure 4-10: Regional boundary profiles

into extreme extrapolations (Curves e, f, g). The compatibility factor in the Black/White formulation may assume values greater than 1.0, as pointed out earlier (Section 4.2.1); this allows the model to cope with the effects in Curves e, f and g.

The extreme forms of the profile (Curves a and g) are, respectively, represented by the generic logistic functions

$$y = \frac{1}{1 + e^{x}} \qquad \ldots (4.27)$$

and

$$y = \frac{1}{1 - e^{x}} \qquad \ldots (4.28)$$

The transition from Curve a to Curve c is accomplished by introducing a FUZZINESS FACTOR ϕ ($\phi > 0.0$) into (4.27) as follows:

$$y = \frac{1}{1 + e^{x \; \phi}} \qquad \ldots (4.29)$$

and the transition from Curve e to Curve g may similarly be generated by writing (4.28) as

$$y = \frac{1}{1 - e^{x \; \phi}} \qquad \ldots (4.30)$$

The entire range of profiles may then be generated by the generic function

$$y = \frac{1}{1 + (\phi/|\phi|)e^{x \; |\phi|}} \qquad \ldots (4.31)$$

Positive values of ϕ generate Curves a thru c; negative values are associated with Curves e thru g. The case of ϕ=0.0, which corresponds to a cliff, produces an undefined value; it is better handled by a separate step function.

When considered in this form, the discrete representation of the affinity factor as $R_{\mu_i\mu_j}$ is clearly invalid. Compatibility has now to be specified as a continuous function of the distance s from the boundary. In a Class 1 situation,

$$R_s = \rho + \frac{1-\rho}{1+(\rho/\phi)e^{s\phi}} \qquad \ldots (4.32)$$

where R_s is the compatibility at a distance s from the theoretical regional boundary, measured *away* from the HOME region. Since the function is doubly asymptotic, it constitutes a slight approximation of the affinity factor in the cores of the home and foreign regions.

The value of ϕ, though theoretically independent of ρ, may in practise be found to be directly related to the compatibility factor that prevails between the cores. A highly fuzzy boundary will be more difficult to detect, and will negate the purpose of Black/White differentiation.

4.5.3. Contiguity of Black/White Regions

It must be emphasized that the Black/White model recognizes regions that are non-contiguous. (while the legitimacy of non-contiguous regions is generally accepted by scholars in the field, existing methods of regionalization tend to produce contiguous aggregations since, as explained in Chapter 2, they do not compensate for the effect of distance). On a single basis of regional differentiation ($r=1$), the model can generate four different scenarios, as illustrated in Figure 4-11. Two regions ($p=2$) may be contiguous (Figure 4-11a) or non-contiguous (Figure 4-11b; a chessboard is another example). Multiple regions could be arranged as in Figure 4-11c (contiguous) or Figure 4-11d (non-contiguous).

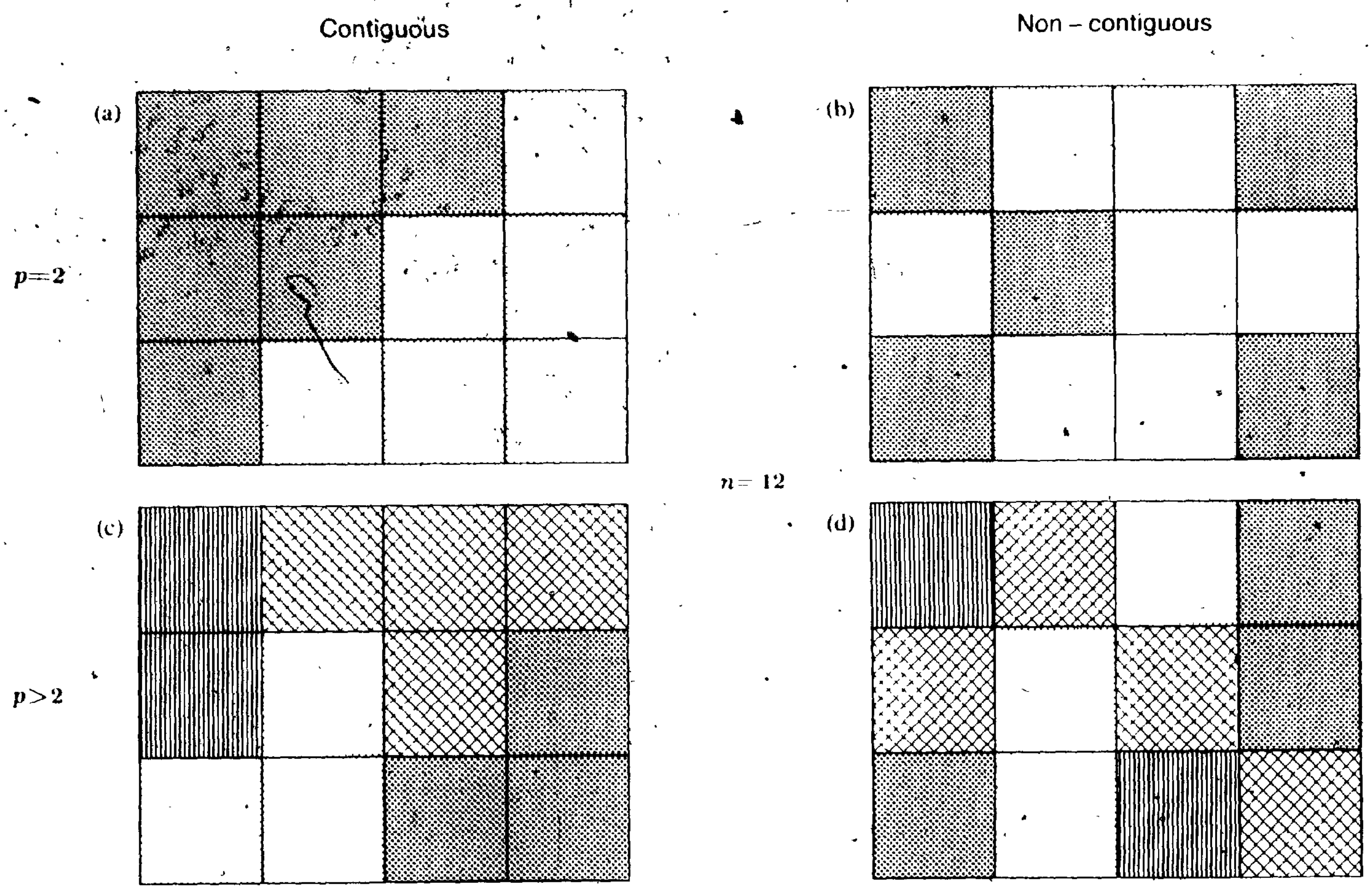


Figure 4 – 11: Selected Black/White regional schemes

4.5.4. A Mechanism for Functional Regionalization

Consider a set of n nodes, with emissivities **E** and attractions **A**, located in a universe consisting of two regions (Figure 4-12). Let a matrix of interactions **I**, satisfying the Black/White Interaction model, connect the nodes. In the absence of a regional bias ($\rho=1.0$), the interactions (corrected for differing emissivities and attractions of their origins and destinations) would plot as a decreasing function of distance (Figure 4-13).

If the interactions were sensitive to the regional boundary ($\rho < 1.0$), all interactions that crossed the boundary would suffer a multiplicative decline determined by the value of ρ. The interactions would now plot against distance as two distinct curves: the upper curve would represent dyads contained within the same region (regardless of which region that was), while the lower curve would contain dyads that straddled the regional boundary (Figure 4-14). A simple logarithmic transformation of the axes would yield two parallel straight lines in place of the curves (Figure 4-15), the vertical distance between the lines reflecting the magnitude of the regional dissimilarity. In general, there will be one line for each active value of ρ (including $\rho=1.0$), so that the number of lines, κ, generated by p regions (assuming that each paired combination of regions is associated with a unique value of ρ) is $\{p(p-1)/2\}+1$.

Startling empirical support for this purely deductive modeling has already been provided by Mackay (1958), who studied the effect of boundaries on various kinds of interactions in Canada. He found that interactions *within* the confines of individual Canadian provinces plotted on lines parallel to and above those composed of cross-boundary dyads (his axes were not defined in exactly the same way as those described in the preceding paragraphs, but it can easily be proved that he was observing an identical effect). Mackay analyzed phone calls originating in the province of Quebec, and marriages of brides from Vancouver.

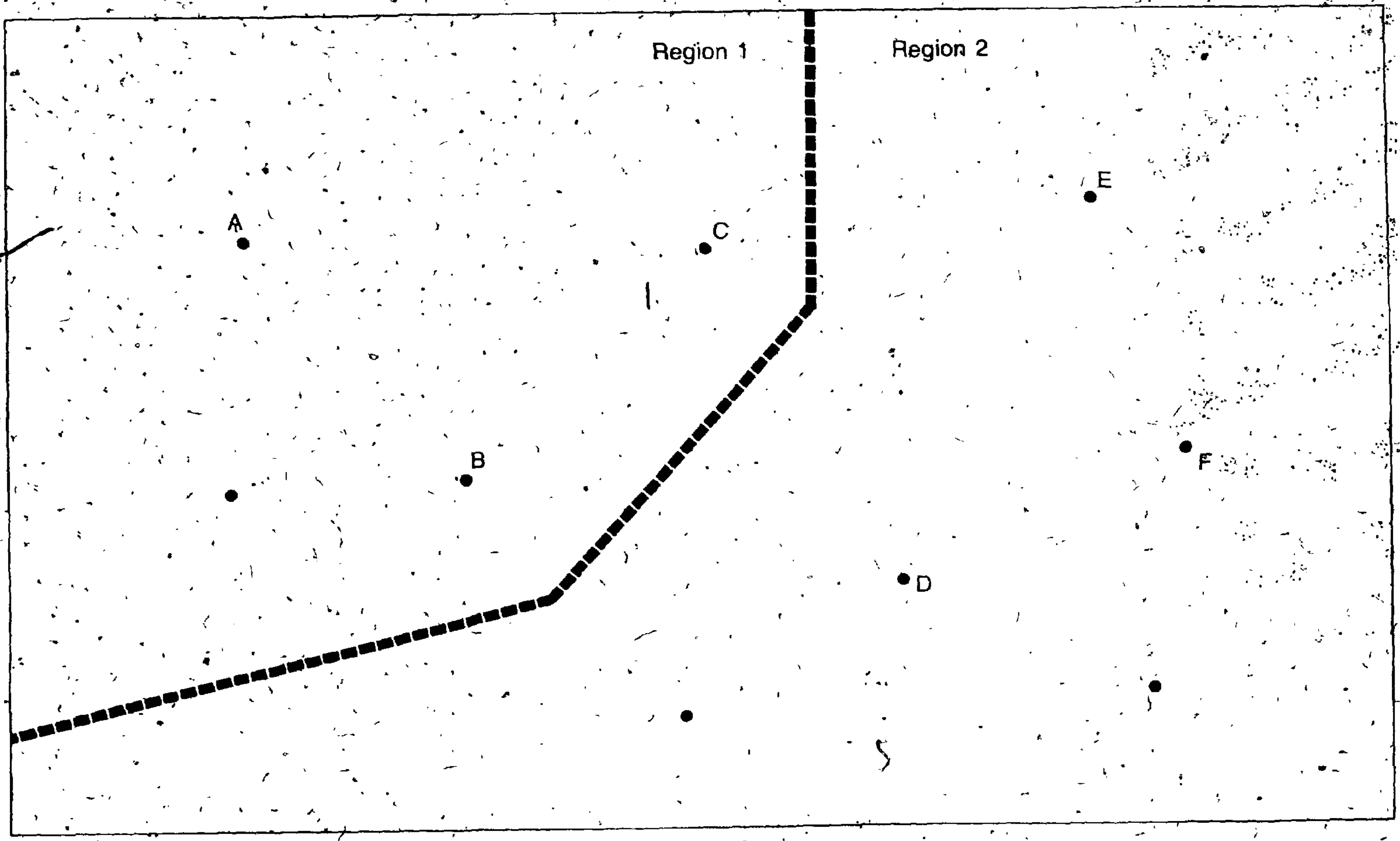


Figure 4–12: Hypothetical two–region universe

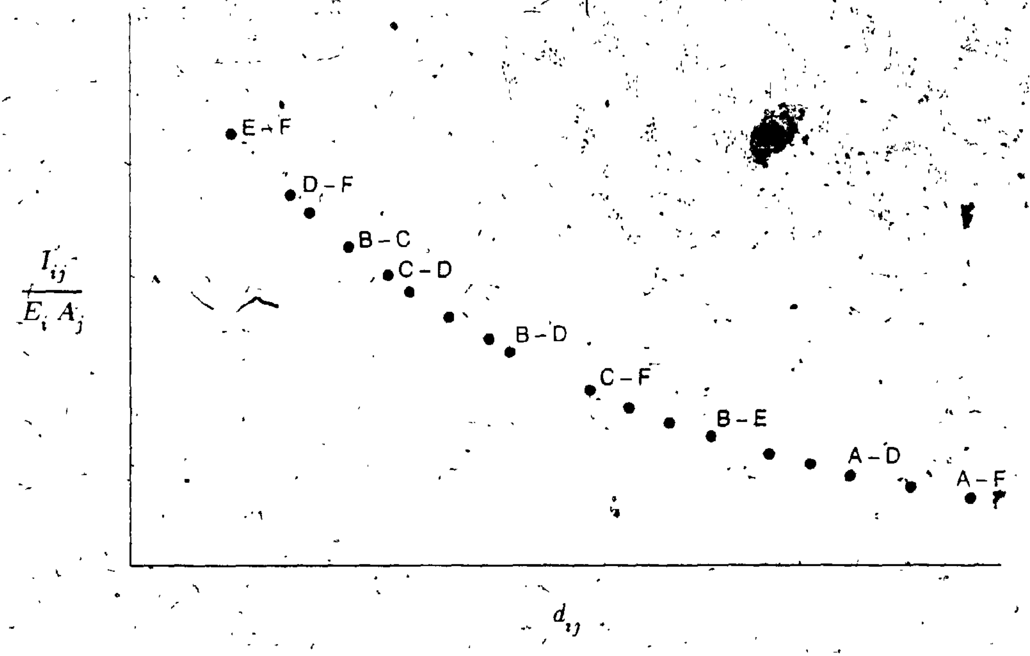


Figure 4 – 13: Single – region interaction profile

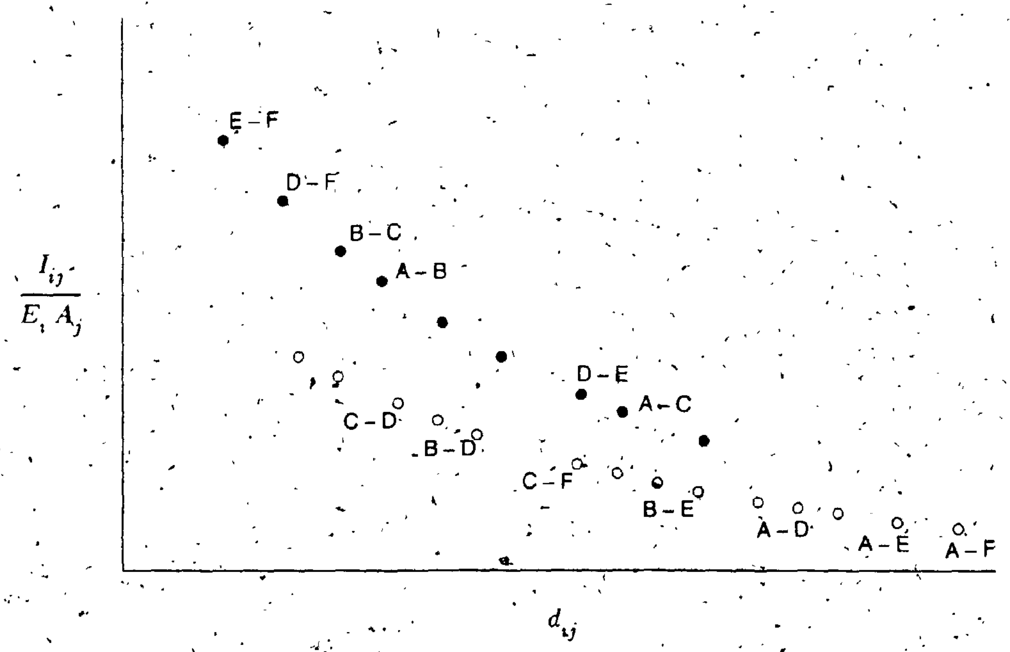


Figure 4 – 14: Two – region interaction profiles

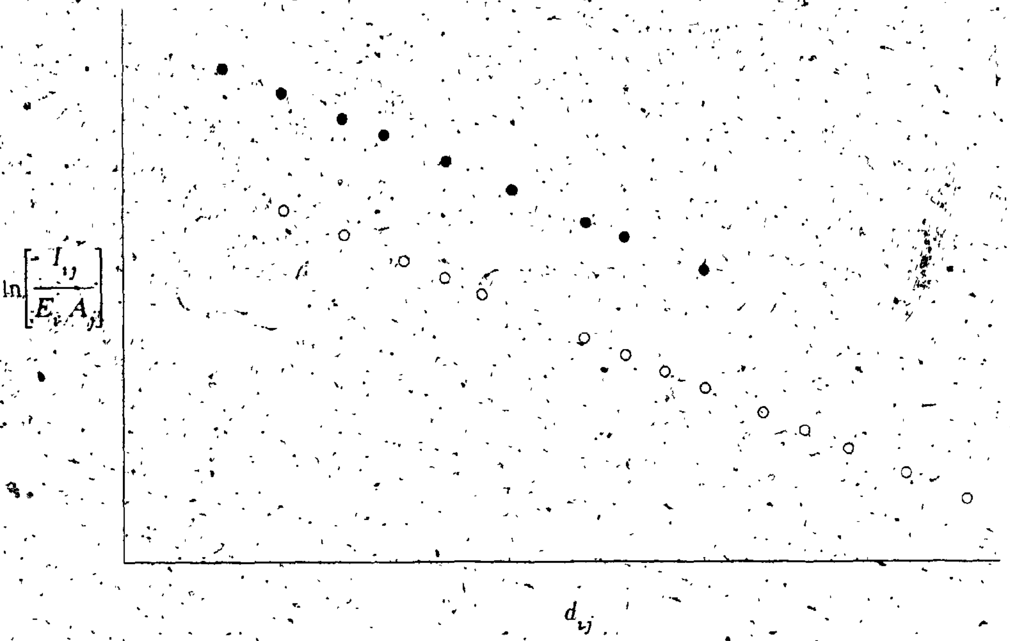


Figure 4–15: Linearized two–region interaction profiles

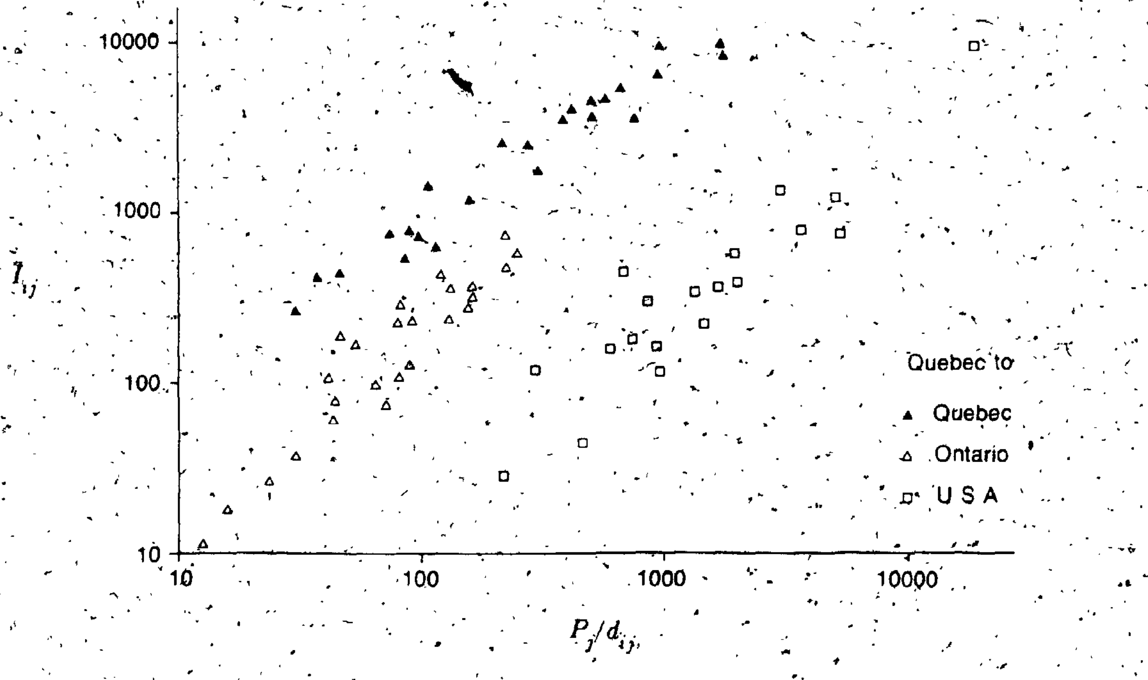


Figure 4–16: Phone traffic between Canada and USA (after Mackay)

The most dramatic effects were observed in the case of long-distance phone calls from Montreal to other North American cities. The plot revealed three distinct lines: the uppermost represented calls to points within the province of Quebec; a central line was composed of destinations in Ontario and other Canadian provinces, while the lowest contained destinations in the United States (Figure 4-16). Personal communication with Mackay (letter dated 1984 May 20) revealed that his data had since been destroyed, but a rough analysis of his plot indicated values of p to be 0.2 between Quebec and Ontario, and 0.025 between Quebec and the United States (it was in this connection that Mackay had made the remarks quoted in Section 4.4.1).

The significance of the form of the interactions plotted in Figure 4-15, and empirically confirmed in Figure 4-16, is clear: since the lines are differentiated on the basis of the regional memberships of the contributing points, it follows that, in the absence of prior knowledge of the regional affiliations of the nodes, **an analysis of the dyads contributing to each line would reveal the regional memberships.** This is the basis of the method of functional regionalization, derived from the Black/White model, that is proposed in this research. It constitutes an alternative to the traditional model of maximizing within-region flows, and by compensating for the effect of distance on interaction, reflects *preferred* as opposed to *revealed* behavior.

4.6. Concluding Remarks

The Black/White model offers a fresh conceptualization of human interactions in environments where relationships between dyads are not uniform. Although the derivation and the examples in this chapter have referred to *cultural* conflict, it is worth emphasizing that the model could just as easily be applied to situations where perceptual barriers (for example, physical features, political boundaries, opinions, etc) affect the propensity to interact, in dyad-specific patterns.

In the field of marketing and retailing, consumer preferences and loyalties (for instance, in the use of credit cards) are amenable to the modeling. From the point of view of marketability of the theory and the model, this seems to be a promising area of application. Figure 4-17 is a scheme within which the model may be applied in areas beyond purely academic pursuits.

Two key questions are of immediate concern. First, what are the means by which the existence of regions, as opposed to a null hypothesis of non-differentiation, can be established? Secondly, assuming that a regional structure does exist, the classical interaction model is incorrectly specified within that environment; is an alternate calibration process available to estimate the parameters of the Black/White model?

The calibration of the model, including the mechanism for identifying the memberships indicated in Figure 4-15 and evaluating the significance of the results, is the subject of the next chapter.

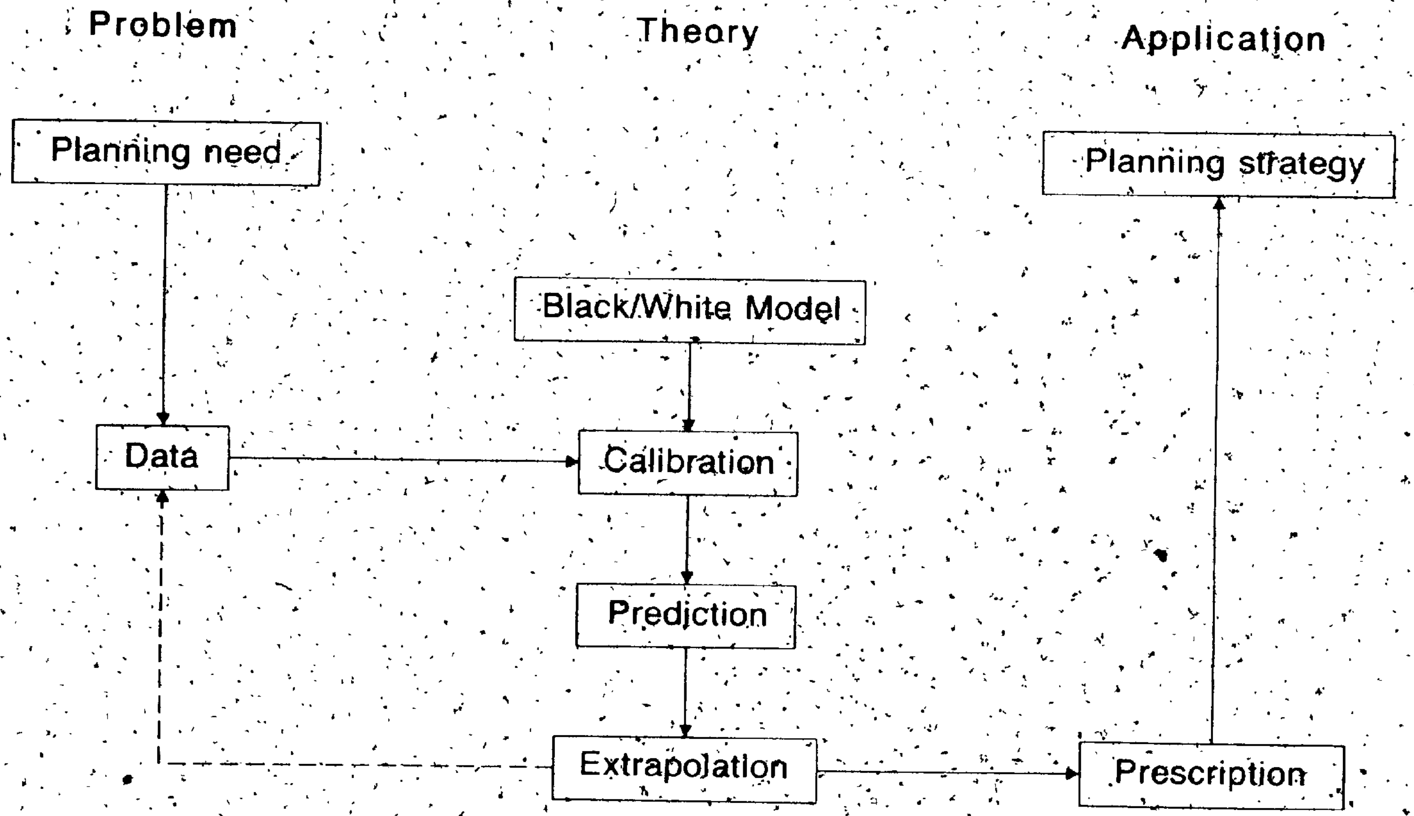


Figure 4–17: Application scheme for the Black/White Model

Chapter 5

CALIBRATING THE BLACK/WHITE MODEL

5.1. Introduction

This chapter is concerned with the dual problems of parameter estimation for the Black/White model, for purposes of application of the model in a predictive capacity, and of discovering the regional structure present in the interaction matrix under study. If a complete calibration of the model,

$$I_{ij} = E_i A_j f(d_{ij}) R_{ij} \qquad (5.1)$$

were possible, the recovered values in $\mathbf{R}$, (which are the compatibilities between individual ESUs) could be processed to reveal the regional memberships. Since the right hand side of (5.1) contains an unknown (R_{ij}) specific to the same dyad that appears in the left hand side, it will not be possible to calibrate the equation by conventional methods.

This chapter will be restricted to the first two classes of the model (Figure 4-9), so that R_{ij} may be expressed as $\rho_{\mu_i \mu_j}$. Now, with prior knowledge of the regional membership vector μ, the calibration of the Black/White model is accomplished with only a slight modification of the method used to calibrate the spatial interaction model. When μ is not known, the procedure is more complex. The calibration of the spatial interaction model is treated in the next section, as a prelude to the calibration of the Black/White model.

5.2. Calibration of the Spatial Interaction Model

As pointed out in Chapter 3, the earliest conceptualizations of the spatial interaction model assumed emissivity and attraction to be equated with the populations of the origin and destination nodes respectively. As such, the only unknown in the model was the function of distance. If distance was assumed to be powered to an exponent, $-\beta$ ($\beta > 0.0$), the model could be written in the form

$$\frac{I_{ij}}{E_i A_j} = d_{ij}^{-\beta} \qquad (5.2)$$

and it was possible to identify the required value of β by means of a simple logarithmic regression procedure.

In later years, when population was thought to be an unacceptable surrogate for emissivity and attraction, the vectors **E** and **A** had to be determined from the observed matrix. The technique of multiple regression made it possible to achieve complete calibration of the model, given only the interaction and distance matrices, and an *a priori* notion of the form of the function of distance. Given an $n \times n$ interaction matrix, the value in each cell (logged to linearize the equation) is represented as the dependent variable for a single case, and the corresponding value of distance (with an appropriate linearizing transformation, if necessary) is set up as one of the independent variables. Two dummy arrays of n binary variables each, designated $\mathbf{E}^*$ and $\mathbf{A}^*$, are set up to represent the emissivity and attraction vectors; for each case, the element in each dummy vector, with an index corresponding to the current cell, is set at 1, while all others are set at 0. By way of illustration, when a 10×10 array is listed row-wise, the sixteenth case represents interactions between origin 2 and destination 6;

accordingly, for this case, E^*_2 and A^*_6 each takes a value of 1, while the remaining elements in the dummy arrays are set at zero. A multiple regression now produces a coefficient for each element in the two dummy arrays. Since the regression model is linear and additive, the corresponding multiplicative elements in $\mathbf{E}'$ and $\mathbf{A}'$, the estimates of the TARGET vectors $\mathbf{E}$ and $\mathbf{A}$ respectively, are the antilogarithms of the recovered coefficients.

In practise, there will appear to be a self-compensating error in the recovered vectors, due to which the recovered emissivities may be higher than their known values, and the attractions proportionately lower, or vice versa. This is because $\mathbf{E}$ and $\mathbf{A}$ in the model are defined to an arbitrary constant, and therefore do not have fixed values. While this condition may be disconcerting, the predictions of the model are not affected.

A discussion of the theoretical requirements for calibrating spatial interaction models may be found in Kirby (1974). Ewing (1982) demonstrates the use of the linear modeling package GLIM (Baker and Nelder, 1978).

5.2.1. Simulated Illustration

The procedures outlined above are now illustrated with a simulated data set. Ten nodes are located at random within a rectangular area measuring 1.0 × 1.0 units (Figure 5-1). Emissivities (arbitrarily scaled between 0.0 and 50.0) and attractions (between 0.0 and 45.0) are assigned to the nodes by a random number generator (Table 5-1). There is no programmed correlation between the emissivity and attraction values, and no attempt is made to rationalize the locations or characteristics of the nodes in terms of geographical logic (for example, see Diggle and Matern, 1980), so that the distribution of emissive or attractive forces over the area may be erratic and incongruous. The simulations are thus minimally structured, and procedures to deal with them will have to be

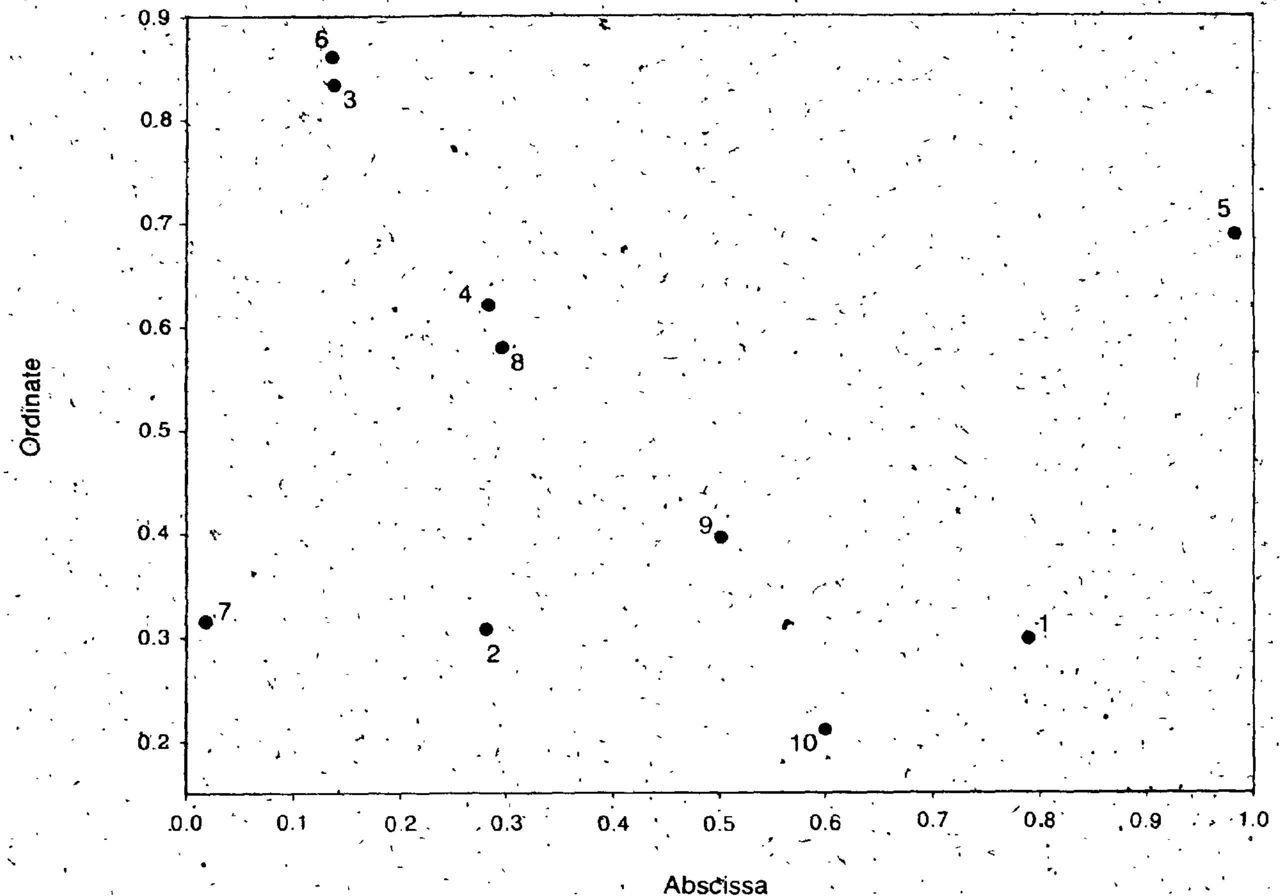


Figure 5 – 1: Locations of nodes in simulated universe

Table 5-1: Parameters of simulated ESUs

i	Abscissa	Ordinate	E_i	A_i
1	0.786	0.298	29.006	42.773
2	0.278	0.306	22.685	0.282
3	0.133	0.832	34.455	17.220
4	0.277	0.620	29.149	4.438
5	0.979	0.694	4.175	44.567
6	0.131	0.863	46.724	9.544
7	0.019	0.314	40.945	24.332
8	0.289	0.585	38.259	42.369
9	0.497	0.396	45.151	39.901
10	0.595	0.210	43.591	41.900

exceptionally robust.[5]

Pythagorean distances are measured between the nodes (Table 5-2), and a 10 × 10 interaction matrix is generated (Table 5-3), in accordance with the spatial interaction model, in the form

$$I_{ij} = E_i A_j e^{-\beta d_{ij}} \qquad \ldots (5.3)$$

where β is arbitrarily assigned a value of 2.0.

5.2.1.1. Case 1: Emissivity and Attraction Vectors Known

Although there is no exogenous surrogate variable that can completely replace emissivity and attraction as defined in the spatial interaction model, several authors (for example, Mackay, 1958) have used population in both capacities. The value of β is found from the equation for the gradient of a regression line forced through the origin:

$$b = \frac{\sum xy}{\sum x^2} \qquad \ldots (5.4)$$

(using the conventional regression notation), and in this simulated example evaluates within 10^{-6} of the target value of 2.0.

---

[5] In one sense, this statement runs contrary to the remarks of Curry *et al* (1975) and Sheppard *et al* (1976); they maintain that the gravity model is misspecified in the real world due to spatial autocorrelations in city sizes, and that the parameters are more likely to be recovered accurately in environments *lacking* a coherent spatial structure.

Table 5-2: Inter-nodal distances

	j									
i	1	2	3	4	5	6	7	8	9	10
1	0.00000	.51070	.84406	.60344	.44075	.86554	.76775	.57416	.30611	.21020
2	.51070	0.00000	.54525	.31480	.80361	.57553	.25708	.27968	.23887	.33386
3	.84406	.54525	0.00000	.25560	.85762	.03082	.53017	.29224	.56734	.77535
4	.60344	.31480	.25560	0.00000	.70663	.28271	.40034	.03767	.31402	.52010
5	.44075	.80361	.85762	.70663	0.00000	.86530	1.03290	.69858	.56695	.61796
6	.86554	.57553	.03082	.28271	.86530	0.00000	.55977	.31974	.59264	.80153
7	.76775	.25708	.53017	.40034	1.03290	.55977	0.00000	.38282	.48491	.58608
8	.57416	.27968	.29224	.03767	.69858	.31974	.38282	0.00000	.28024	.48442
9	.30611	.23887	.56734	.31402	.56695	.59264	.48491	.28024	0.00000	.21139
10	.21020	.33386	.77535	.52010	.61796	.80153	.58608	.48442	.21139	0.00000

Table 5-3: Simulated Interactions, ρ=1.0

	j									
i	1	2	3	4	5	6	7	8	9	10
1	0.00	2.94	92.34	38.51	535.38	49.03	151.99	389.78	627.45	798.21
2	349.40	0.00	131.27	53.64	202.65	68.48	330.09	549.36	561.35	487.49
3	272.45	3.26	0.00	91.72	276.28	309.19	290.36	813.70	442.02	306.20
4	372.95	4.38	301.05	0.00	316.13	158.05	318.48	1145.37	620.65	431.60
5	73.96	.24	12.94	4.51	0.00	7.06	12.87	43.75	53.61	50.83
6	353.92	4.16	756.48	117.81	368.94	0.00	371.12	1044.39	569.84	394.05
7	377.15	6.90	244.20	81.60	231.24	127.57	0.00	806.75	619.43	531.32
8	519.04	6.16	367.22	157.48	421.67	192.64	432.92	0.00	871.56	608.40
9	1047.03	7.89	249.98	106.93	647.50	131.72	416.55	1092.20	0.00	1239.57
10	1224.59	6.30	159.21	68.37	564.50	83.74	328.49	700.95	1139.64	0.00

5.2.1.2. Case 2: Emissivity and Attraction Vectors Unknown

This case is more typical of an everyday encounter with the problem of calibrating the model. The interaction matrix is analyzed using the multiple regression facility of the statistical package SPSS (Nie *et al*, 1975), with arrays **E*** and **A*** established as dummy variables, as described in Section 5.2. The results of the calibration are shown in Table 5-4. Note that the emissivities have been over-estimated by a constant factor of 44.57, while the attractions have been under-estimated by the same multiple. Nevertheless, the recovered parameters are as effective as the target parameters in predicting the interactions in any given cell of the matrix, and therefore the constraints of the marginal totals and grand total are fully met. For practical applications, then, this method of calibrating the model may be considered satisfactory.

There are, however, theoretical objections to the calibration of spatial interaction models using the method of ordinary least squares demonstrated above. An extended dialog between Curry (1972), Curry *et al* (1975) and Sheppard *et al* (1976) on the one hand, and Cliff, Martin and Ord (1974; 1975; 1976) on the other, dwelt on the problem of spatial autocovariance leading to distortions in the recovered value of β (see also Johnston, 1975). Cliff, Martin and Ord (1974) found this to be a problem with interactions over shorter distances, and therefore a factor to be considered when studying intra-urban, rather than inter-urban flows. They raised the possibility that the function of distance was better represented by the half-normal distribution curve than the exponential or power function. Fotheringham (1983) and Ewing (1984) lent some strength to that last suggestion when they raised the issue of *modal split* (that is, the fragmentation of interaction among alternative modes of transportation and communication) in interactions over short distances. Meanwhile, Goodchild and Smith (1980) isolated the problem of heteroscedasticity in the least squares approach, which they attributed to the powering of distance; they circumvented the problem by transforming the conventional least squares objective function

Table 5-4: Parameter estimates, ρ=1.0

i	E_i	E'_i	E'_i/E_i	A_i	A'_i	A_i/A'_i
1	29.006	1292.702	44.567	42.773	0.960	44.567
2	22.685	1011.005	44.567	0.282	0.006	44.601
3	34.455	1535.575	44.568	17.220	0.386	44.567
4	29.149	1299.091	44.567	4.438	0.100	44.566
5	4.175	186.075	44.569	44.567	1.000	44.567
6	46.724	2082.345	44.567	9.544	0.214	44.567
7	40.945	1824.826	44.568	24.332	0.546	44.566
8	38.259	1705.102	44.567	42.369	0.951	44.567
9	45.151	2012.252	44.567	39.901	0.895	44.568
10	43.591	1942.731	44.567	41.900	0.940	44.568

Correlations: **E** with **E′** 1.000

A with **A′** 1.000

β: Target: 2.000

Recovered: 2.000

$$Z = \sum_{i,j} (L'_{ij} - L_{ij})^2 \qquad \ldots (5.5)$$

into

$$Z = \sum_{i,j} \frac{(L'_{ij} - L_{ij})^2}{L'_{ij}} \qquad \ldots (5.6)$$

where

L_{ij} is the log of the observed interaction between i and j and

L'_{ij} is the log of the expected or predicted interaction

effectively weighting each observation inversely by its assumed Poisson error variance (Goodchild and Kwan, 1978).

5.3. Calibration of the Black/White Model

5.3.1. Defining the Problem

Four classes of the Black/White model were introduced in Chapter 4 (recall Figure 4-9). The calibration strategies are clearly unique to each class, although Class 4 is downwardly compatible with any of the other three. The procedures developed in this research are directed primarily towards the first two classes. The Class 1 specification will be dealt with in detail in this chapter; the techniques will be illustrated with a real problem in Chapter 6. The Class 2 specification may in general be calibrated in the same manner, but there are logistical difficulties barring the testing of the procedures at this time. The Class 1 technique will therefore be introduced with frequent references to generalization to Class 2 problems.

The scope of this calibration effort may appear limited, in that it is difficult to justify the *existence* of only two regions, isolated from multi-dimensional effects of other geographic characteristics. But there have been situations, political and socio-economic, past and present, where the purpose of regionalization has been to identify a single two-way division, regardless of the dimensionality of existing regional affinities. This is most often true in the case of political disputes (where the consultative role of the geographer is most obvious). One can cite, as boundaries that have been established recently, the Green Line in Beirut, the cease-fire line in Korea and the Berlin Wall. The division of the Indian subcontinent into India and Pakistan in 1947 was a classic example of binary non-contiguous partitioning.

5.3.2. Configuration Addressing

To facilitate the description and testing of functional regional structures, it will be necessary to develop a means of concise encoding of a given spatial configuration, as well as the ability to generate a regional pattern at random.

It is assumed that an inviolate case-wise structure is associated with the interaction data matrix. That is, the columns and rows of the matrix are ordered identically, perhaps with reference to a popular or favored list. For example, counties constituting a state may be ordered alphabetically, and it would be operationally desirable to maintain that ordering.

A particular assignment of n ESUs to p regions may be viewed as a STRING of regional membership codes. Let the abscissa 0.5 divide the simulated universe in Figure 5-1 into two regions. The nodes 1, 5 and 10, comprising one distinct region, may be differentiated from the rest of the universe by the ordered alphabetic codings *abbbabbbba* or *baaabaaaab*, or either of the ordered binary digital strings 0111011110 or 1000100001. The decimal equivalents of the digital strings are 478 and 545 respectively. It follows that every decimal integer between 0 and 2^n-1 (inclusive) has a binary representation that could be interpreted as a two-way regional configuration (by extension, p-way spatial configurations may be represented by base-p strings, and their associated integers;

the issues surrounding multiple regions are developed later in this chapter). Since the strings can be toggled without meaningfully affecting the configurations that they represent, and since the extreme integers 0 and 2^n-1 represent undifferentiated universes, the set of possible configurations is restricted to the binary equivalents of the integers between and including 1 and $2^{n-1}-1$. Since there is a one-to-one correspondence between regional configurations and decimal integers in the given range, the former may be generated at random by referencing the corresponding integers at random; similarly, a sequential enumeration of the integers effectively addresses all possible combinatorial arrangements of the ESUs in two regions.

5.3.3. Calibration Strategy

The essential difference between the Black/White model and the spatial interaction model is the presence of the additional variable R_{ij} in $n(n-1)$ of the n^2 interaction equations (it is assumed that $R_{ii}=1.0 \; \forall \; i$). Using the representation of (4.24), however, the single-basis, two-region form of the Black/White model is better expressed as

$$I_{ij} = E_i A_j f(d_{ij}) \rho^{\sigma_{ij}} \qquad \ldots (5.7)$$

where

$$\sigma_{ij} = \begin{cases} 0 & \text{if } \mu_i = \mu_j \\ 1 & \text{otherwise} \end{cases}$$

or

$$\sigma_{ij} = | \mu_i - \mu_j | \qquad \ldots (5.8)$$

The problem now reduces to finding a single universal value of ρ, and the binary array σ. Given σ, the calibration is a simple extension of the procedure for parameter estimation in the spatial interaction model: the cells in the matrix are represented as a dummy scalar variable for each case, and the regression coefficient for that variable may be processed to reveal ρ.

5.3.3.1. One-Piece Calibration

When neither σ nor ρ is known, the calibration may conceivably be achieved by differentiating the least squares objective function

$$W = \sum_{i}^{n}\sum_{j}^{n}[A_{ij} - E'_{i}A'_{j}\, f'(d_{ij})\, \rho'^{\sigma}{}_{ij}]^2 \qquad \ldots (5.9)$$

where the primes (′) indicate estimated values. The Newton-Raphson (steepest descent) method of optimization might then be used to estimate the unknowns. Since σ is discrete, W is not continuous, and therefore not differentiable with respect to σ.

The alternative is to use unconstrained optimization techniques (that is, where the objective function is not necessarily continuous or differentiable). Cooper and Steinberg (1970) present several such methods; prominent among them are the "patterned search" methods of Hooke and Jeeves (1961) and the quasi-Newtonian method of Fletcher and Powell (1963). These techniques are likely to arrive at a minimum value of the objective function W as defined in (5.9), but they do not readily provide for constraints on the solution: there are important characteristics of σ that have to be preserved in a solution; these will be detailed later, in Section 5.3.3.4.

Given that the Black/White model can be calibrated when σ is known, the approach that will now be developed is to pursue the problem of finding σ, independently of the estimation process for the other parameters in the equation. The operation is in effect a partitioning, and thus takes on a specifically geographical complexion. Since σ is fundamentally linked with its companion parameters, the operation cannot proceed in ignorance of the latter. However, in the absence of σ (and ρ), the Black/White equation is identical to the spatial interaction model; the latter may therefore be loosely interpreted as the null hypothesis[6] of an undifferentiated universe; the search for σ may now be directed towards finding the spatial structure contained in the departure from that null hypothesis.

5.3.3.2. Preliminary Estimates

A calibration process such as that described in Section 5.2 assumes that the interaction matrix obeys the assumptions of the traditional interaction model. In a Black/White interaction environment, the traditional model is misspecified; the recovered parameters assume an undifferentiated universe, and are therefore incorrect. However, they can be used as preliminary estimates, and be refined by further processing.

A fresh matrix of interactions was simulated for the universe in Section 5.2.1, with ρ=0.4, and a target regional configuration of 478_{10}. The results of a calibration run are shown in Table 5-5. The correlations between the target and recovered values of the emissivity and attraction vectors are only 0.973 and 0.994 respectively, and the recovered value of β, at 3.265, is substantially at odds with the target value of 2.0. These estimates correspond to the null hypothesis of an undifferentiated universe, so that their deviation from the target values constitutes a measure of the rejection of that hypothesis (in practise there are other dimensions of variation in the data, which are the source of noise that prevents this conclusion from being drawn so readily). One may now calculate the interactions $\mathbf{I}'$ expected on the basis of the traditional model:

$$I'_{ij} = E'_i \, A'_j \, e^{-\beta' d_{ij}} \qquad \ldots (5.10)$$

and the multiplicative residuals $\Delta\mathbf{I}'$:

---

[6] Since the spatial interaction model does not usually contain a random error term, it cannot be considered a stochastic model; therefore it is not a *strictly* defined null hypothesis.

Table 5-5: Parameter estimates, ρ=0.4

i	E_i	E'_i	E'_i/E_i	A_i	A'_i	A_i/A'_i
1	29.006	1282.120	44.202	42.773	0.761	56.211
2	22.685	1298.312	57.232	0.282	0.006	43.447
3	34.455	2251.788	65.354	17.220	0.453	38.019
4	29.149	1560.030	53.519	4.438	0.096	46.424
5	4.175	232.770	55.753	44.567	1.000	44.567
6	46.724	3149.314	67.402	9.544	0.259	36.862
7	40.945	2807.501	68.568	24.332	0.671	36.237
8	38.259	2020.015	52.798	42.369	0.900	47.060
9	45.151	2465.550	54.607	39.901	0.877	45.502
10	43.591	1757.434	40.316	41.900	0.680	61.630

Correlations:	**E** with **E′**	0.973
	A with **A′**	0.994
β:	Target:	2.000
	Recovered:	3.265

$$\Delta I'_{ij} = \frac{I_{ij}}{I'_{ij}} \qquad \ldots (5.11)$$

For the purpose of this discussion, the matrix $\Delta\mathbf{I}$ is also defined; it contains the residuals of observed interactions on those generated by the interaction model given the target vectors:

$$\Delta I_{ij} = \frac{I_{ij}}{E_i\, A_j\, e^{-\beta d_{ij}}} \qquad \ldots (5.12)$$

It is clear that $\Delta\mathbf{I}=\mathbf{R}$.

5.3.3.3. Seriation of the Residuals Matrix

The technique of seriation or block modeling was introduced in Chapter 2. Several popular algorithms are available for the purpose, notably the iterative intercolumnar correlation procedure (McQuitty, 1968; McQuitty and Clark, 1968), or CONCOR (Breiger, Boorman, and Arabie, 1975), and COBLOC (Carrington and Heil, 1979). In the present context, the purpose of applying a seriation algorithm is to *polarize* a matrix by assigning the cells to either of two classes; the CONCOR algorithm achieves this with minimal effort. The technique is simply to construct a matrix of Pearson correlation coefficients between the rows (or columns) of the object matrix. The matrix of coefficients is then subjected to the same procedure iteratively; on each iteration the coefficients tend further towards the extreme values of $+1.0$ or -1.0. At convergence, the sum of the squared coefficients in any row (or column) of the matrix will be equal to n. The progress towards convergence on a given iteration is therefore indicated by the grand total of the row (or column) sums of the matrix of squared correlation coefficients, expressed as a fraction of n^2 — this quantity will be referred to in the following paragraphs as the convergence ratio, or C. The method is applicable to square matrices, and is more justifiable in the case of symmetric matrices, since the procedure, whether applied to the rows or to the columns of a symmetric matrix, achieves identical results.

The CONCOR algorithm was applied to the problem of polarizing the matrix $\Delta\mathbf{I}$ (generated from $\mathbf{E}$ and $\mathbf{A}$). It took 7 iterations to converge (convergence was defined as $1.0-C < 10^{-12}$), correctly recovering the target solution, namely, *abbbabbbba*.

When faced with the residuals matrix $\Delta\mathbf{I}'$, (generated from $\mathbf{E}'$ and $\mathbf{A}'$), convergence was again achieved in 7 iterations, but with an erroneous result, *aabbbbabaa*. Evidently, the error in $\mathbf{E}$, $\mathbf{A}$ and β (due to model misspecification) overwhelmed the algorithm. This particular simulated example (target$=478_{10}$, ρ=0.4) was judged to be representative of a typical application; the failure of CONCOR to identify the regions correctly ruled out the technique of seriation as an appropriate treatment of $\Delta\mathbf{I}'$.

5.3.3.4. Separating the Regression Lines

It was argued in the previous chapter that interactions across a boundary, corrected for emissivity and attraction, would plot against distance as two straight lines, when subjected to an appropriate linearizing transformation (Figure 4-15), and that regional memberships could be recovered from an analysis of the plot. The following discussion is based on interactions across a single boundary (Class 1 in Figure 4-9), but may clearly be extended to multi-region situations (Class 2). / It must be emphasized that although the method is *applicable* to multiple regions, it is being *developed* and *tested* in this chapter for the simpler case of two regions.

Traditional techniques for fitting regression lines to multi-linear scatters (Sprent, 1969) are not equipped to handle the Black/White scatter, due to the peculiar properties of the latter. The first problem is the allocation of points to lines. In the example in Figure 4-15, the ordinates of the points are calculated on the basis of the true values of **E** and **A**, and therefore plot neatly on two lines. When the ordinates **y** are computed from the *estimated* vectors, **E′** and **A′**:

$$y_{ij} = \ln\left[\frac{I_{ij}}{E'_i A'_j}\right] \qquad \ldots (5.13)$$

the result is a PARALLEL SCATTER. Points that belong on the upper line (by virtue of representing within-region interactions) may fall closer to the lower line, and vice versa (Figure 5-2). A low value of ρ (associated with greater regional conflict) produces a wider vertical separation between the scatters, which may help reduce the error in allocation, but this is offset by more severe errors in the estimates **E′** and **A′**. Secondly, due to the assumed symmetry of **R**, points representing complementary dyads (i-j and j-i) plot on the same line. Thirdly, there is the property of triangular relationships: in a two-region system, a knowledge of any two compatibility values for a given triple (ijk) of ESUs unambiguously establishes the third value, as follows:

$$R_{ij} = \begin{cases} 1.0 \text{ if } R_{ik}=R_{jk} \\ \text{Min}\{R_{ik},R_{jk}\} \text{ if } R_{ik}\neq R_{jk} \end{cases} \qquad \ldots (5.14)$$

One could conceivably constrain a parallel-line regression procedure to obey all the properties outlined above; however, the number of points in the scatter is very large, varying almost as the square of the number of ESUs (specifically, $n(n-1)/2$), and the isolation and examination of allocations of all possible triads of these points to the two lines would be cumbersome and extremely expensive. Because of these properties of the Black/White model, standard regression methods, the techniques introduced in Section 5.3.3.1 as well as discriminant analyses, are not favored.

A preferable method is the systematic minimization of the sum of squared

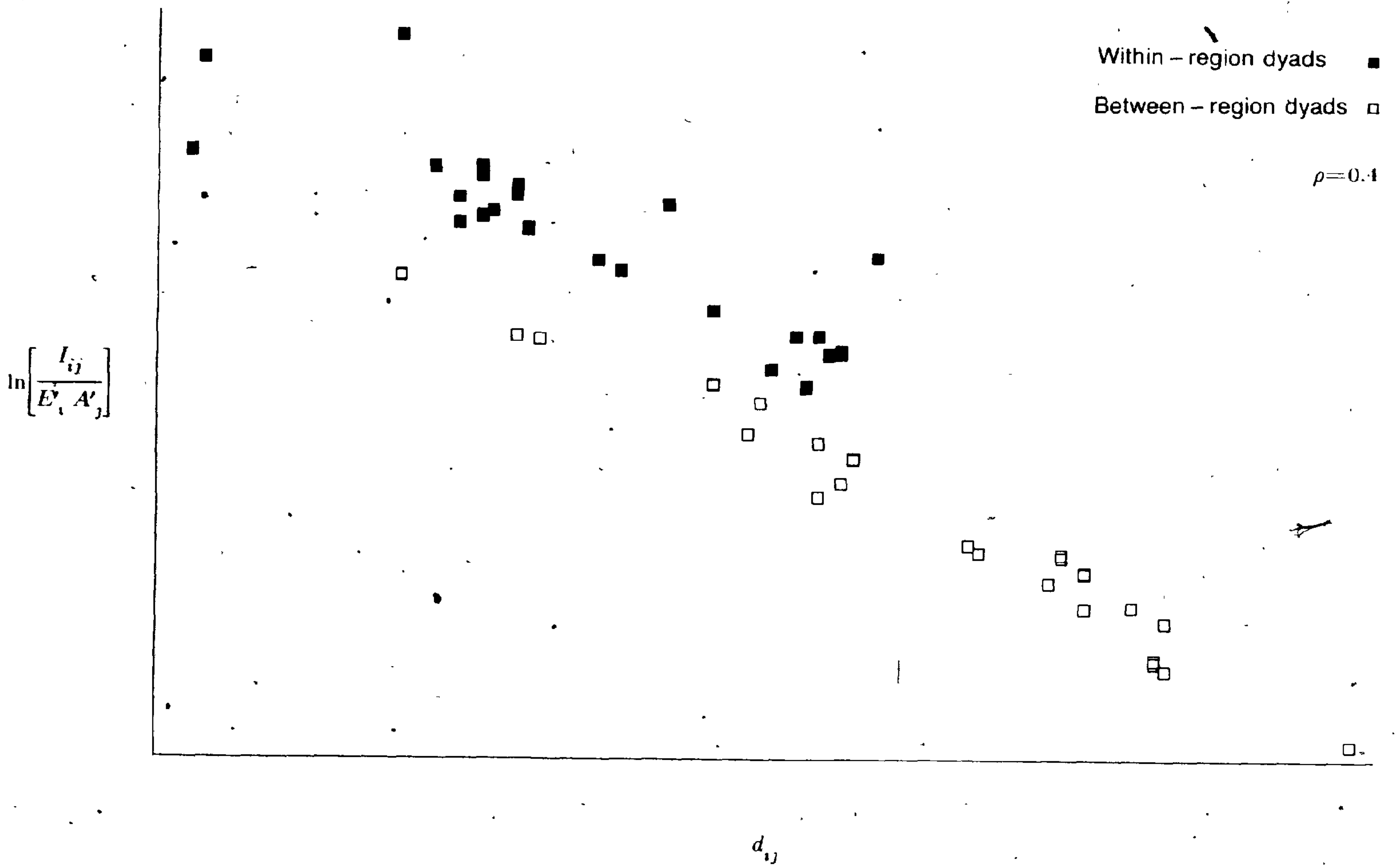


Figure 5–2: Parallel scatter

differences between the ordinate (y_{ij}) of each point in the scatter and the ordinate of its projection on the line to which it is assigned. The objective function may be simplified by the substitution

$$z_{ij} = \ln\left[\frac{I_{ij}}{E'_i A'_j}\right] - (-b)d_{ij} \qquad \ldots (5.15)$$

In graphical terms, this "slides" the points up a line of gradient $-b$ (that slope corresponds to the value of $-\beta$; since β is positive, the net effect of the transformation is positive) to the vertical axis. In mathematical terms, it can easily be shown that **z** represents the log of the ratio of observed interactions to those expected under the null hypothesis of an undifferentiated universe. While this provides a conceptual link to the notions of "expected" interactions reviewed in Chapter 2, the null hypothesis here is based on the spatial interaction model, which, due to Huff (1963) and Wilson (1967), is adequately founded upon geographic theory.

Let $l=1,2,\ldots\kappa$ enumerate the individual scatters (in the case of two regions, $\kappa=2$). For a given configuration, points representing dyads within the same region are known to plot on the upper scatter, while the other dyads will plot on the lower scatter. Let $\delta_{ijl}=1$ if the point i-j is allocated to scatter l, or zero otherwise. A unidimensional separation of the points on the vertical axis is now achieved by minimizing the objective function

$$Z = \sum_{l}^{\kappa}\sum_{i}^{n}\sum_{j}^{n}\left[\delta_{ijl}\, z_{ij} - \frac{\sum_{g}^{n}\sum_{h}^{n}\delta_{ghl}\, z_{gh}}{\sum_{g}^{n}\sum_{h}^{n}\delta_{ghl}}\right]^2 \qquad \ldots (5.16)$$

which, for computational efficiency, may be expressed as

$$Z = \sum_{l} \left[\sum_{i,j} \delta_{ijl}\, z_{ij}^{2} - \frac{\left(\sum_{g,h} \delta_{ghl}\, z_{gh}\right)^{2}}{\sum_{g,h} \delta_{ghl}} \right] \qquad \ldots (5.17)$$

If it can be assumed that the individual scatters are normally distributed in y and d, then the minimization of Z corresponds to the maximum likelihood partition.

Unfortunately, the true value of β is only *approximated* by β', so that **z**, as calculated in (5.15), is at best an estimate. The practical task of minimizing the objective function would have to be achieved by evaluating Z over a range of b.

If, as will often be the case, the regionalization tends strongly towards contiguity, the preliminary calibration procedure will produce an excessively high estimate of β'. A glance at Figures 4-15 and 5-2 will reveal why this is so: within-region interactions plot nearer to the vertical axis, while between-region interactions tend to be associated with greater distances, and therefore have higher abscissas in the scatters; in an attempt to fit a single line to a parallel scatter, where the points are *not* evenly distributed horizontally between the two scatters, the estimated gradient will be steepened to accommodate the vertical and horizontal lag between the scatters — where contiguity is not as strong, the horizontal lag will be less pronounced. Thus if the regions are suspected to tend toward contiguity, the estimate β' should act as a ceiling for experimental values of b.

The following sections are devoted to establishing the validity of the proposed procedure, using the set of simulated data introduced earlier in this chapter.

5.4. Tests with Simulated Regional Structures

5.4.1. Development of Algorithms

Data were simulated in Section 5.2.1 primarily to illustrate the technique of calibrating the spatial interaction model. The same universe will now be treated by systematically assigning the nodes between the regions, and monitoring the outcome of the calibration procedures described in Section 5.3.3.4. The purpose of these sections of the research is to demonstrate that the target regional configuration corresponds to the minimum objective function as defined in (5.17). To establish this empirically, it would suffice to show that there exists no alternative configuration that yields a lower objective function than does the target. In order to explore the configurations exhaustively, one must adopt an explicit enumeration procedure.

5.4.1.1. Algorithm 'A'

The initial batch of investigations operated on the simulated interaction matrix mentioned in Section 5.3.3.2, with ρ set at 0.4, and the regional configuration at 478_{10}. Values of Z were computed for each of the 511 possible two-region configurations, with b assuming values from 1.0 to 3.3, in intervals of 0.1: this pattern of search for the optimum is termed Algorithm 'A'.

The procedure correctly identified the target, with an objective function that exhibited an overwhelming "primacy" over its competitors. The five best arrangements are shown in Table 5-6. Similar results were achieved with ρ taking values from 0.01 to 0.99, clearly establishing the efficacy of the method over a complete range of possibilities.

One might have suspected that the cost (in mis-allocation of ESUs to regions) of *not* identifying the optimum would increase monotonically with the rank of the corresponding objective function. Figure 5-3 is a plot of the percentage of ESUs

Table 5-6: Five best results, Algorithms A and B

Rank	ID (decimal)	Binary string	Number misidentified	Z	b at critical point
1	478	0111011110	0	2.68573	2.5
2	462	0111001110	1	4.88595	3.0
3	350	0101011110	1	4.89879	3.1
4	470	0111010110	1	5.37045	3.0
5	414	0110011110	1	5.42923	3.1

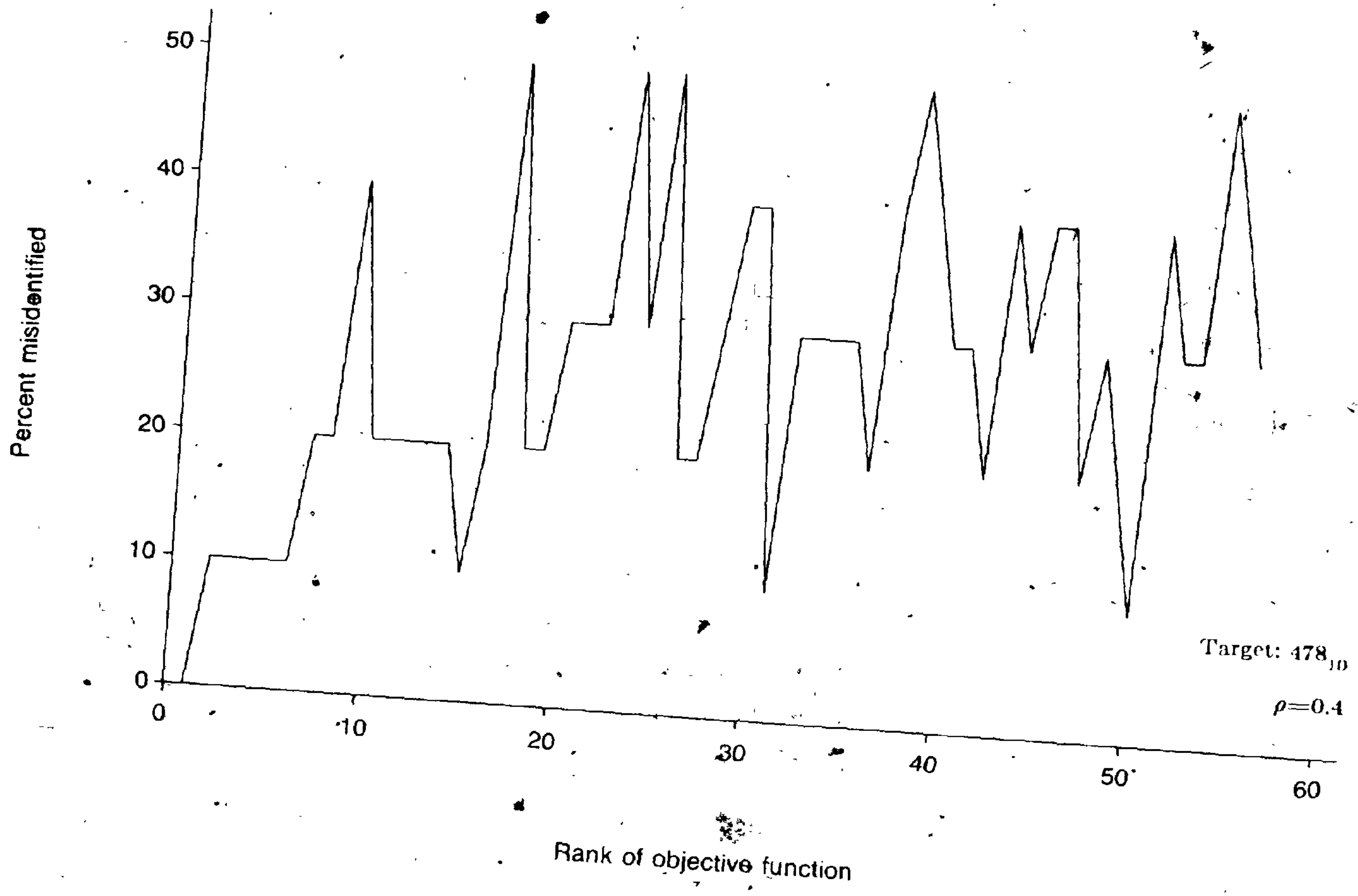


Figure 5–3: Variation of recovery error with rank of objective function

misallocated (since the regionalization is binary, this cannot exceed 50%) versus the rank of the objective function for that configuration. It is clear that there is no pattern.

5.4.1.2. Algorithms 'B' and 'Betamax'

Over repeated experiments, it was observed that the target registered the lowest Z only within a certain favored window of b. A plot of the objective function against b (Figure 5-4) uncovered the reason: each configuration was associated with its own Z-curve, which reached a minimum at a value of b peculiar to the configuration. Figure 5-4 reveals two distinctive characteristics of the curve for 478_{10}. Apart from registering the lowest optimal value of Z, the CRITICAL POINT of the curve occurs at a lower value of b. The explanation for the first observation is that the multiple regression procedure which determines the initial estimates of **E** and **A** seeks to optimize those values in ignorance of a regional structure; that structure finally reveals itself in the target curve, in the form of an exceptional value of Z.

The second observation, namely, the value of b at which the target curve optimizes, can be best explained with reference to Figure 4-15, where the parallel lines were introduced, and Section 5.3.3.4 above, where the issue of contiguity was raised. A target configuration such as 478_{10}, with a Moran autocorrelation statistic (Cliff and Ord, 1973) of 0.196, causes a significant horizontal lag between the parallel scatters (Figure 5-2 was generated from this configuration).

The observations on Figure 5-4 suggest two further lines of development: first, since each configuration is associated with a unique convex curve, one could seek optimization more effectively by unidimensional searches directed towards each curve in sequence; under the assumption of strict convexity, one may assume a single minimum or critical point, and employ standard methods of optimization for unconstrained unimodal functions (Cooper and Steinberg, 1970). One is thus

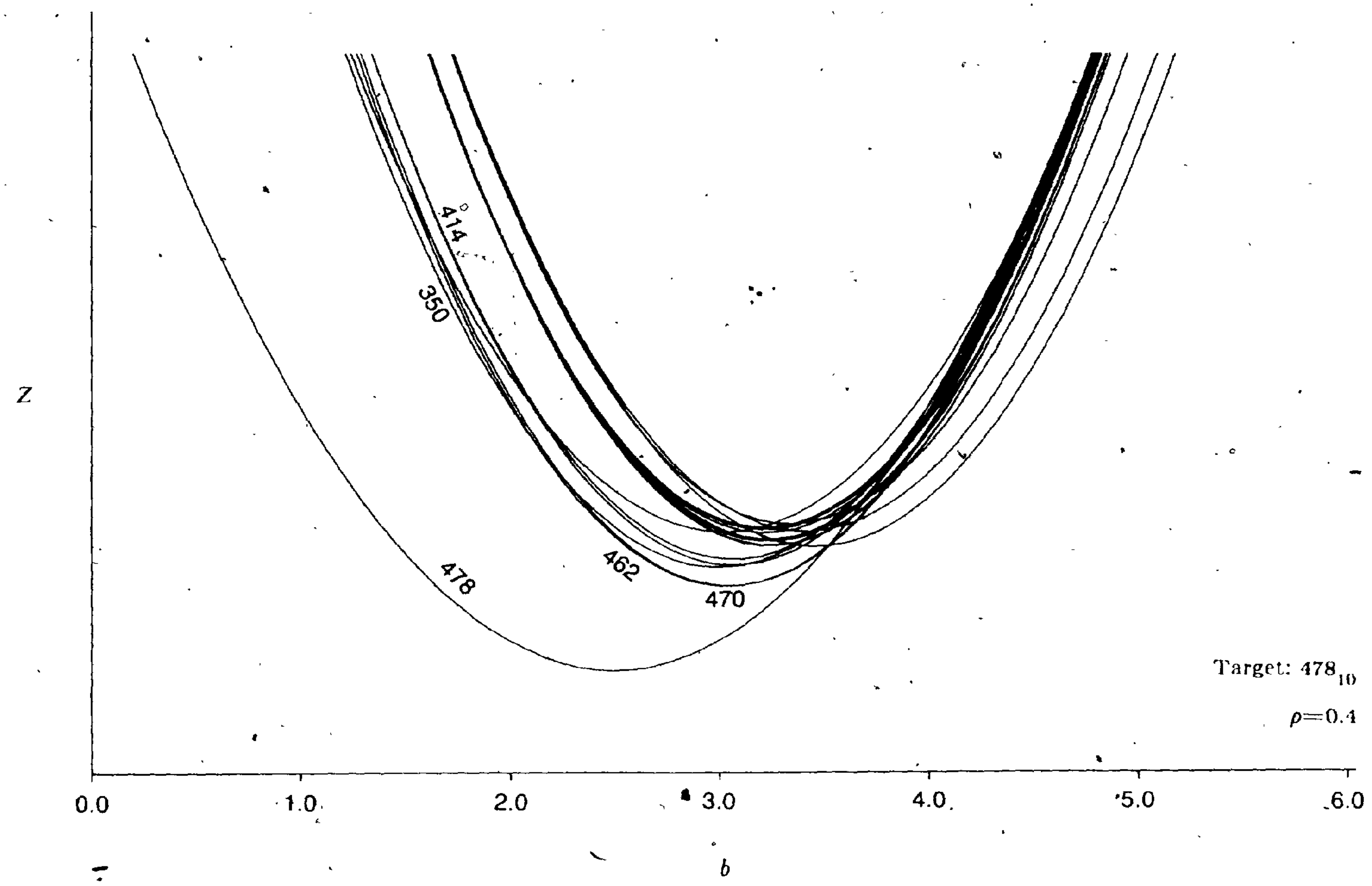


Figure 5–4: Z–curves

freed from the possibility of identifying local (as opposed to global) optima, and may benefit from the reduced domain of the search (once a critical point is identified for a given configuration, it is known to be the only one). This approach was operationalized as algorithm 'B'. The search method employed is similar to that of Tornqvist (1971). The objective function is evaluated at regular intervals within a chosen window of b; when a change in the sign of the gradient is detected, the search direction is reversed, and the step size reduced; the operation is terminated when the step size reaches a critical floor. The user provides the lower and upper limits of the initial search window, as well as the initial step size, the step reduction ratio and the critical step size. These five SEARCH PARAMETERS will be labeled b_1 thru b_5. By careful selection of these values, algorithm B can be made to converge on the solution in fewer iterations than does the EXHAUSTIVE SEARCH method of algorithm A at a comparable resolution, but the former is more expensive since **z** has to be computed repeatedly as b is varied (since b oscillates in irregular patterns, there is little benefit to be gained from storing the **z**'s in core).

The second possibility was that strongly autocorrelated targets could be retrieved by searching for the curve that "turned" at the lowest value of b, rather than searching for the lowest Z. This approach was the basis of the BETAMAX algorithm.

5.4.1.3. Heuristic Algorithms

Given the documented performance of the algorithms in the above tests, one may now assume that the task of empirically demonstrating the efficacy of the Black/White method of regionalization has been accomplished, and that further efforts should be directed towards streamlining the method for practical applications. The most obvious shortcoming of the test data in this respect has been the restricted size of the interaction matrix (10 × 10). The use of explicit enumeration on problems of this size is quite reasonable (see Edwards and

Cavalli-Sforza, 1965). When looking at practical problems such as the regionalization of the United States at the state level, one should be prepared to encounter matrices of the order of 50 × 50; based on the observations of computing time recorded in the simulated data (3.5 milliseconds per iteration) on a CDC Cyber 170-835 computer, it was estimated that a solution to a 50 × 50 problem by explicit enumeration (64 ms per iteration) would require, for each sampled value of b, 3.6×10^{13} seconds, or 1.14 million years, in CPU time. Apart from the practical problems this presents, the cost of this processing, at current billing rates and currency values, would be over 3 trillion dollars, or about 30 times the current (1985) Canadian federal budgetary outlay. Such expenditure of resources would probably be viewed as excessive, and establishes fairly conclusively the need for a more economical algorithm.

Heuristic algorithms are meant to provide efficient methods of solution to intractable computational problems. While not guaranteed to provide the optimal solution, they are meant to operate according to *rules* (intuitive strategies appropriate to the given problem), and thus provide educated guesses. In the partitioning of the Black/White scatter, a heuristic could take either of two approaches. First, one could use rules of intelligence to allocate the points in the scatter to their respective lines (that is, in the configuration of δ), and proceed by means of swap operations (for example, Teitz and Bart, 1968) to improve the initial objective function. The rules in such a strategy (including a reasonable stopping rule) are unclear, and further, this approach would require, prior to each evaluation of the objective function, a test for compliance with the properties of the scatter outlined in Section 5.3.3.4.

A second possibility is to apply heuristic selection procedures to the vector of regional memberships, μ. Being a unidimensional array, there are fewer permutations possible, and they do not have to be tested for logical consistency.

An algorithm was developed as follows:

- The entire array of ESUs is assigned to one region (corresponding to the configuration 0_{10}), dubbed the SUBDOMINANT CLUSTER.
- Each cell in turn is assigned to the second region, and the objective function evaluated.
- The cell registering the lowest value of Z is fixed in the second region, thus being admitted as a permanent member of the DOMINANT CLUSTER.
- The search continues in scan CYCLES for further admissions to the dominant cluster, on the same basis. At the end of each cycle, the lowest objective function reached for that cycle (that is, a PARTIAL SOLUTION) is recorded; when that fails to improve upon the INCUMBENT SOLUTION, the search is terminated.

The algorithm is similar to the "greedy-add" heuristic, employed in operations research. It emphasizes the notion of *companionship* in its approach to selection of entities into the dominant cluster, and may therefore be considered particularly appropriate to the task of regionalization. In contrast, binary linear programming techniques are inappropriate to this problem since the objective function is non-linear in μ. Further, and particularly in light of the expense of calculating the objective function in the Black/White scatter, the greedy heuristic is relatively economical when compared with other search methods such as the vertex substitution algorithm of Teitz and Bart (1968).

The heuristic described above would require at most $n^2/2$ evaluations of Z for each b sampled. At 64 ms per evaluation, a 50×50 problem would require 80 seconds of CPU time; a 10×10 problem would be solved in 0.175 seconds by the heuristic, compared with 1.8 seconds by Algorithm A.

In a series of tests, the heuristic recovered the target 478_{10} correctly for values of ρ ranging from 0.01 to 0.93.

132

5.4.2. Evaluation of Algorithms

A number of strategies for partitioning the Black/White scatter have now been outlined. There are valid reasons for adopting or rejecting each approach under particular circumstances, and one may argue in favor of preserving such a variety of options. On the other hand, one needs to evaluate the performance of the methods, under the various circumstances with which they will be faced in practical applications. The following sections examine the performance of the algorithms.

5.4.2.1. Simulated Error

The interactions simulated thus far have been in perfect accordance with the Black/White model. In reality one cannot expect human behavior to conform to the specifications of a mathematical equation. Real data sets are likely to contain variances on other dimensions, collectively referred to as "noise." If a method of regionalization is to be proposed for application in the real world, it will have to be robust enough to cope with seemingly random components. This suggests a strategy for the evaluation of the algorithms outlined above: if random error could be introduced into the simulations in measured PACKETS, the most effective algorithm under a given set of circumstances would be the one which could correctly identify the target under the greatest "stress," or noise.

5.4.2.2. Error Tolerance Tests

The performance of Algorithm A on configuration 478_{10} has already been detailed. Since algorithm B is merely an operational variant on its predecessor, it produces the same results, and Figure 5-4 demonstrates that the Betamax algorithm too identifies the target correctly. The experiments detailed in this section are meant to provide a general overview of the suitability (measured by tolerance of simulated error) of the algorithms in various situations. It is worth emphasizing that these are *not* meant to be benchmark runs, since the production of a commercial package is not the purpose of this research; rather, these tests numerically corroborate the statements made while introducing the algorithms in Section 5.4.1.

Error was introduced into the simulations by generating normally distributed random deviates with a mean of 0.0 and a variance of 1.0, using the IMSL (1975) subroutine GGNML. The deviates were multiplied by a variable ERROR MEASURE, ϵ; the natural exponentiation of this quantity was introduced as a multiplier in the equation generating the interactions:

$$I_{ij} = E_i A_j f(d_{ij}) R_{ij} e^{\epsilon \theta_{ij}} \quad \ldots (5.18)$$

where θ_{ij} is the random deviate returned by GGNML. In effect, $\epsilon = 0.0$ nullifies the effect of the error (since $e^{0.0} = 1.0$), while $\epsilon = 1.0$ gives the deviates an effective variance of about 3.0. The arrangement precludes the possibility of negative interactions.

Figure 5-5 shows the amount of error that could be tolerated by algorithms A and B, with ρ varying from 0.01 to 0.99. Intuitively, regions that are hostile to each other (low ρ) would be more readily distinguished. The observations bear this out: at low compatibility values, the algorithm delivers the correct solution despite high error components in the simulations; when the regions are not clearly differentiated (high ρ), the procedure cannot tolerate as much error.

In Figure 5-6a, the tolerance curves of algorithms B and Betamax are compared with that of the heuristic. Note that the performance of the heuristic is erratic, but is generally comparable with the results of algorithm B. Betamax easily outperforms the other two algorithms.

The error tests were repeated with alternative target configurations that had extremely high and extremely low spatial autocorrelations. Given the locations of

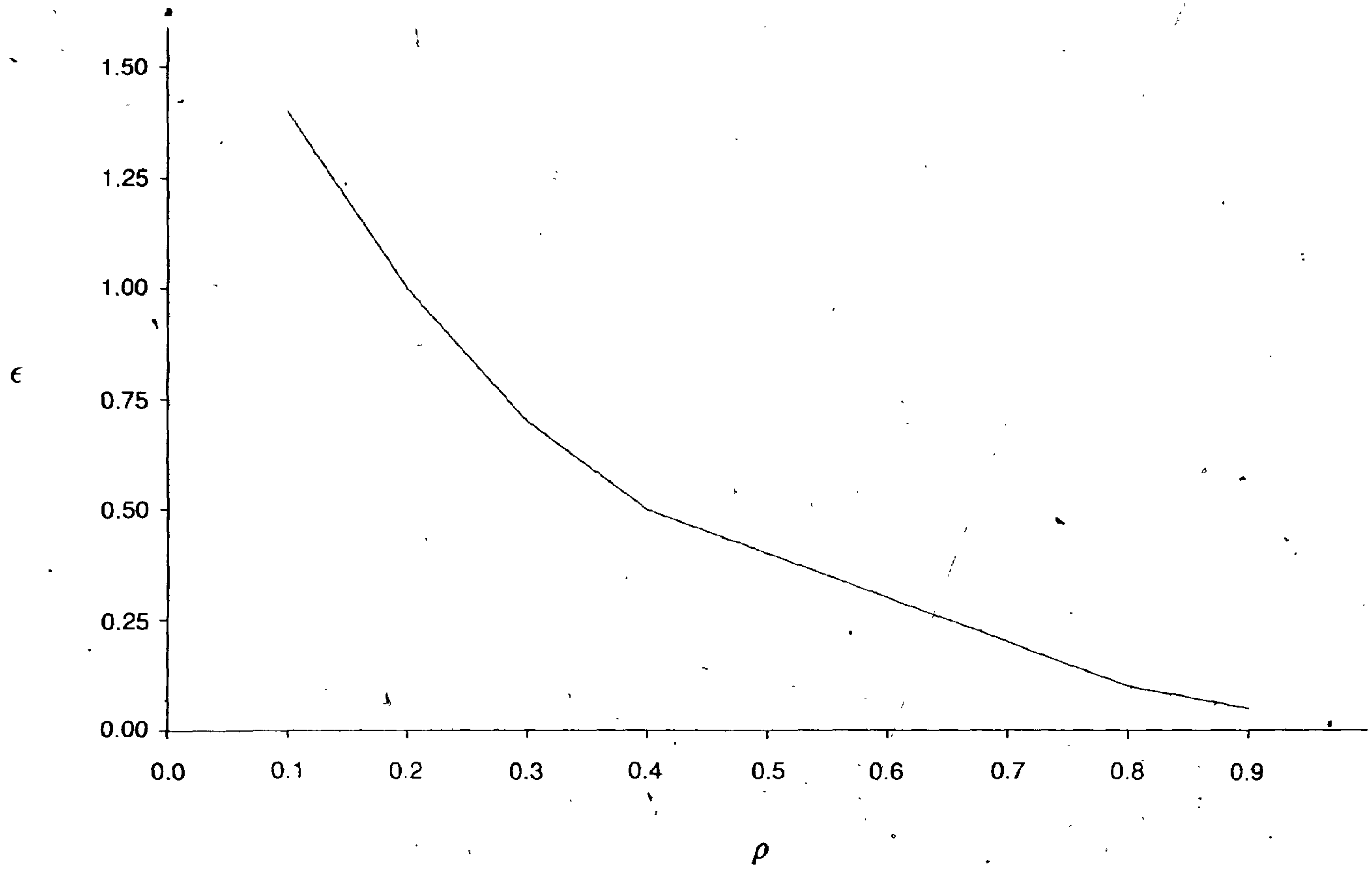


Figure 5–5: Error tolerance: Algorithms A and B

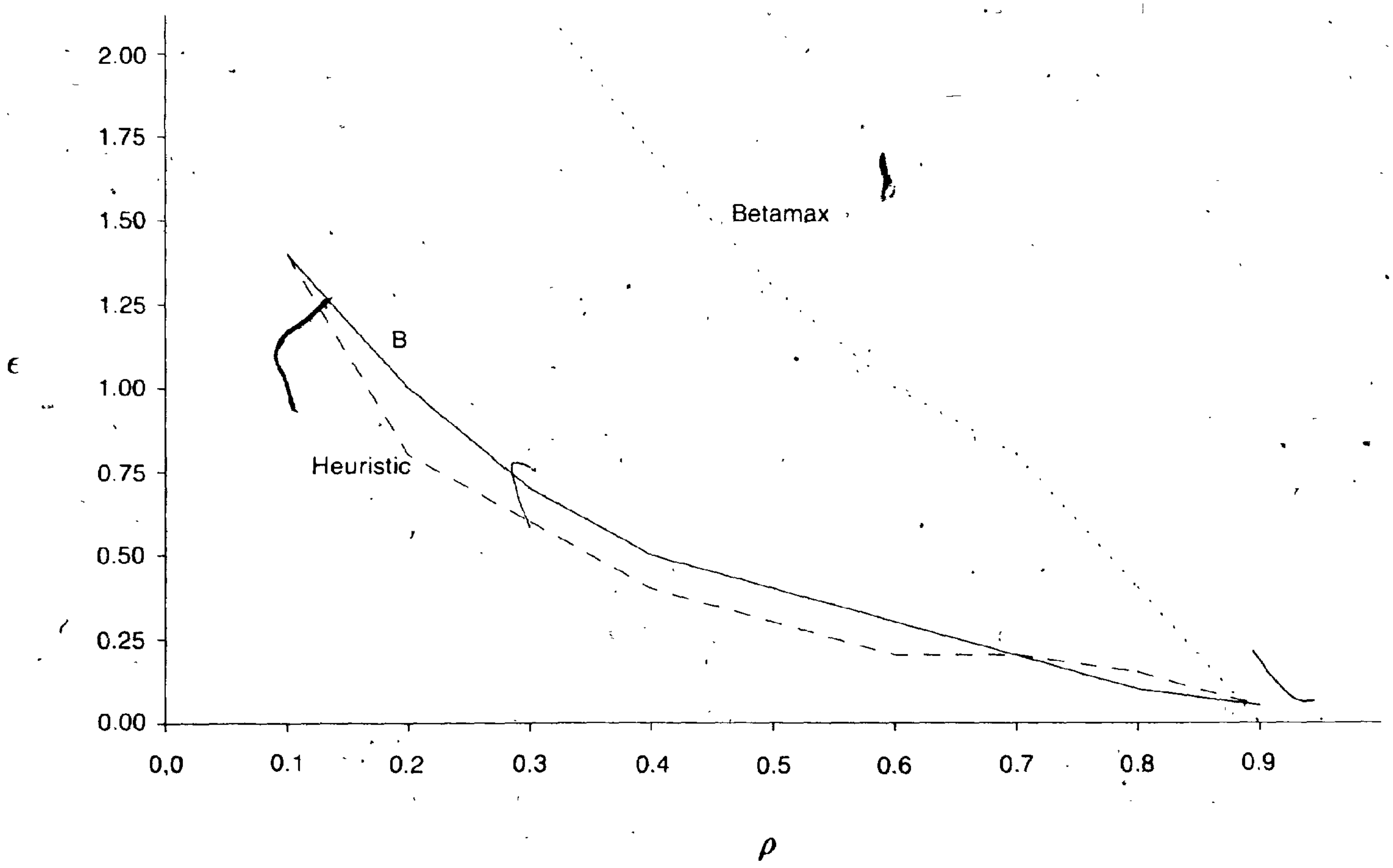


Figure 5−6a: Error tolerance: B, Beta and Heuristic. Target$=478_{10}$

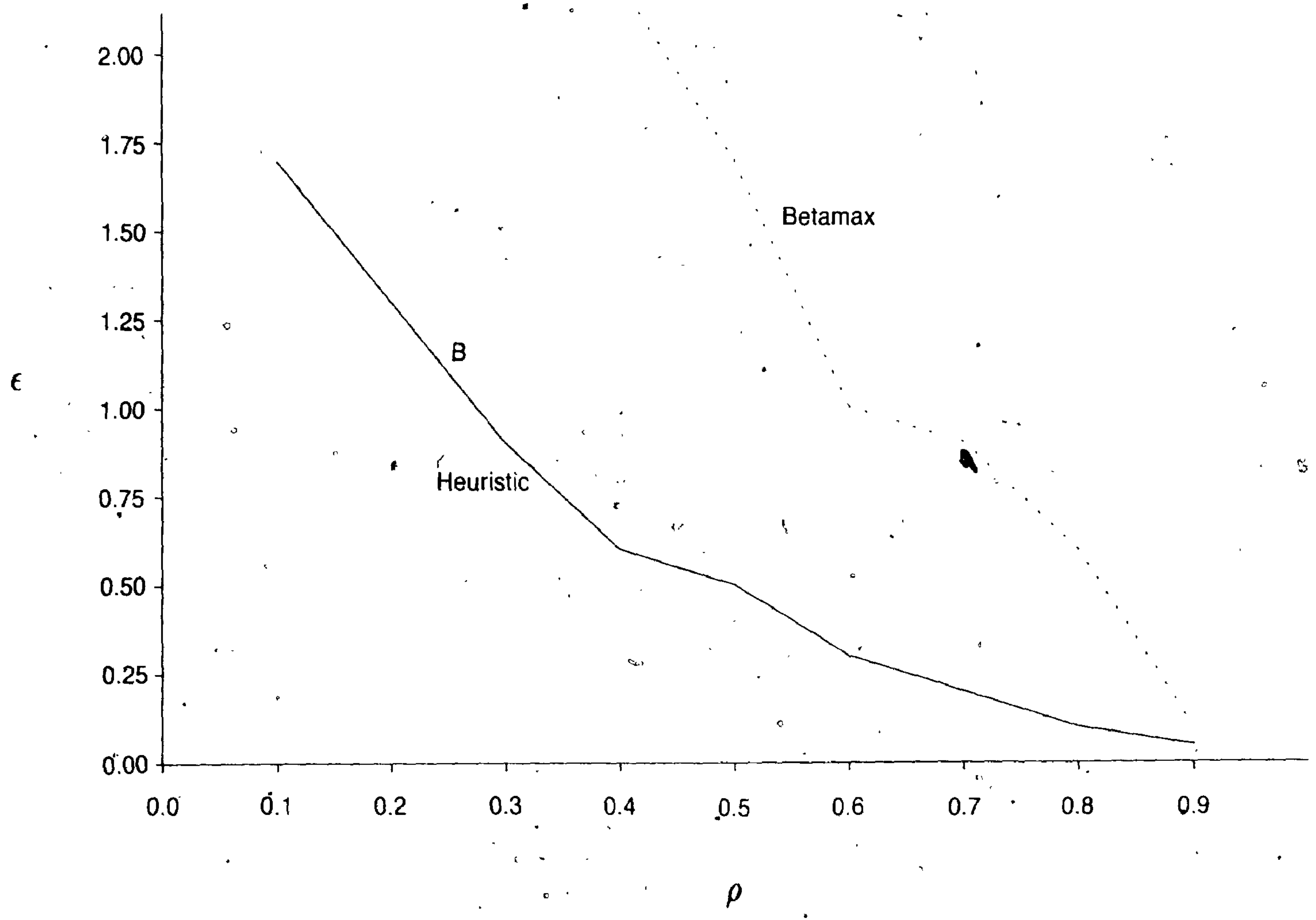


Figure 5-6b: Error tolerance: B, Beta and Heuristic. Target$=220_{10}$

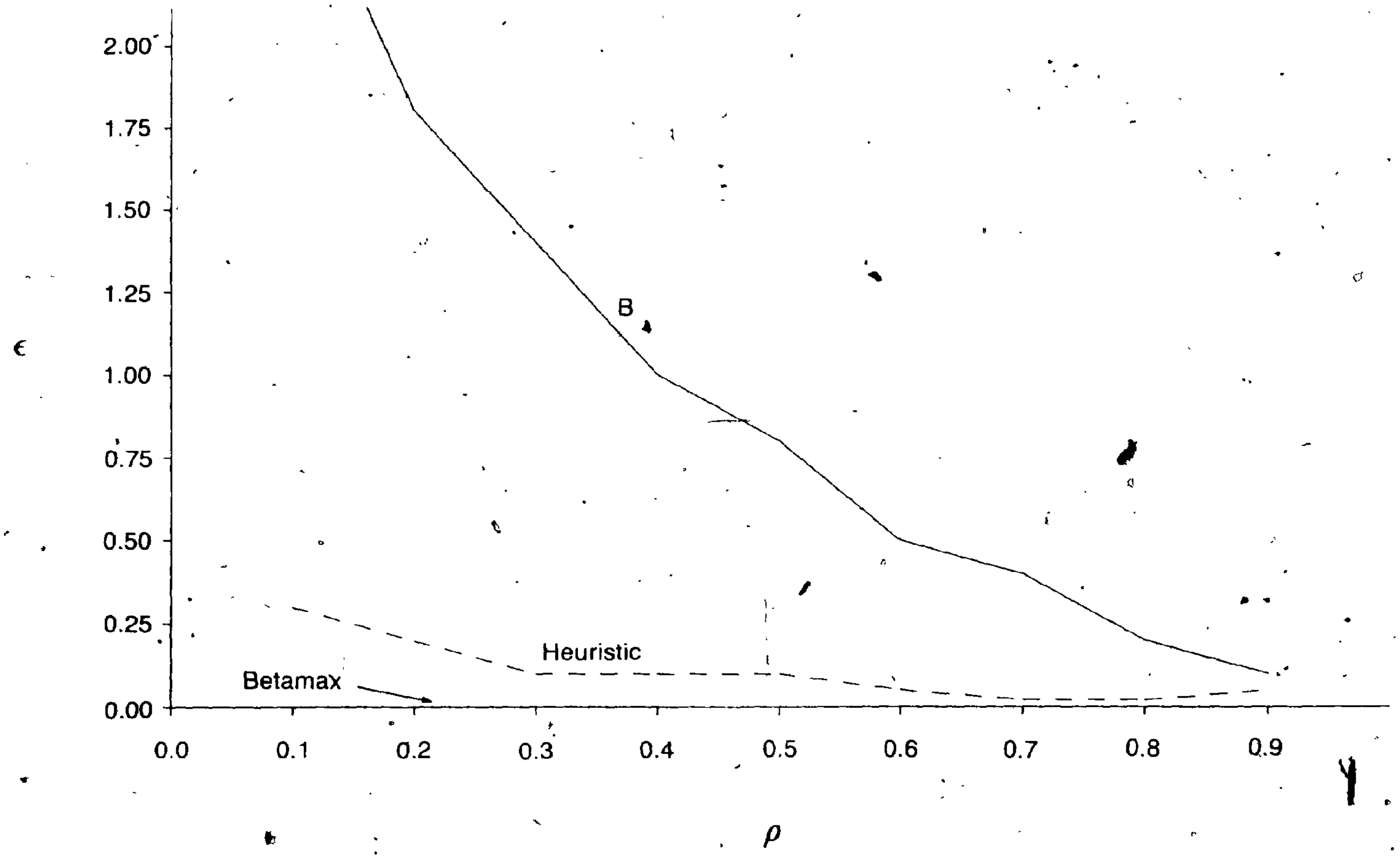


Figure 5−6c: Error tolerance: B, Beta and Heuristic. Target=421_{10}

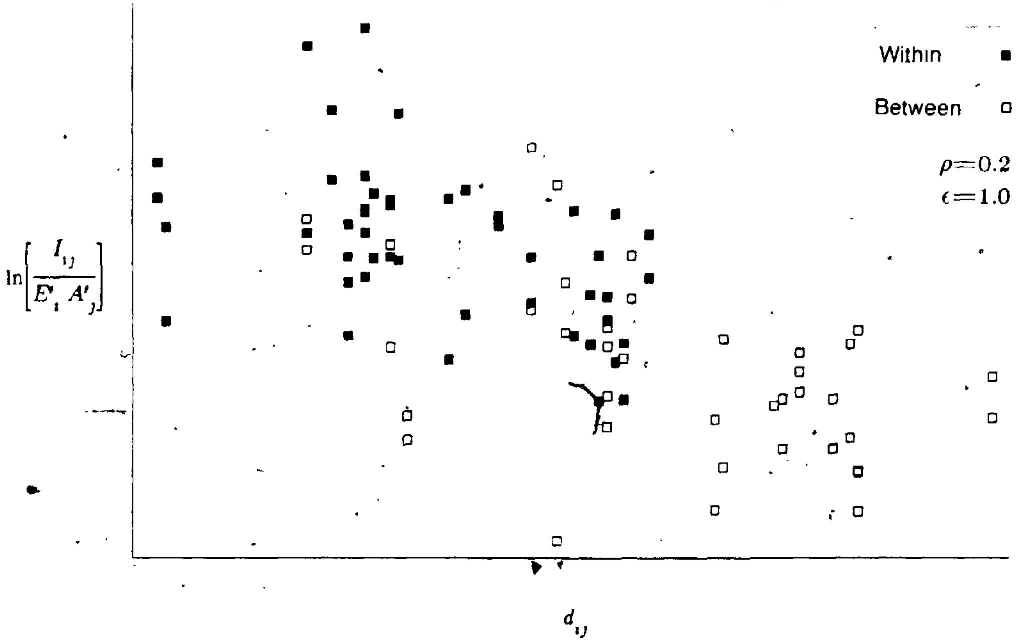
the simulated nodes, Moran statistics were calculated for all possible Black/White configurations. With a statistic of 0.396, configuration 220_{10} (0011011100) had the strongest autocorrelation among arrangements with a 5-5 Black-White split; the lowest statistic (−0.460) was registered by 421_{10} (0110100101). The tolerance curves for these configurations are shown in Figures 5-6b and 5-6c respectively. Betamax was outstanding on the highly autocorrelated 220_{10}, but failed to recover target 421_{10}, even with no simulated error. The heuristic performed as well as did algorithm B on 220_{10}, but it was not as effective in recovering 421_{10}. The tolerance of B was comparable across configurations. The tests indicate that strongly autocorrelated targets have the best chance of being correctly recovered by all the algorithms.

5.4.2.3. ϵ in Perspective

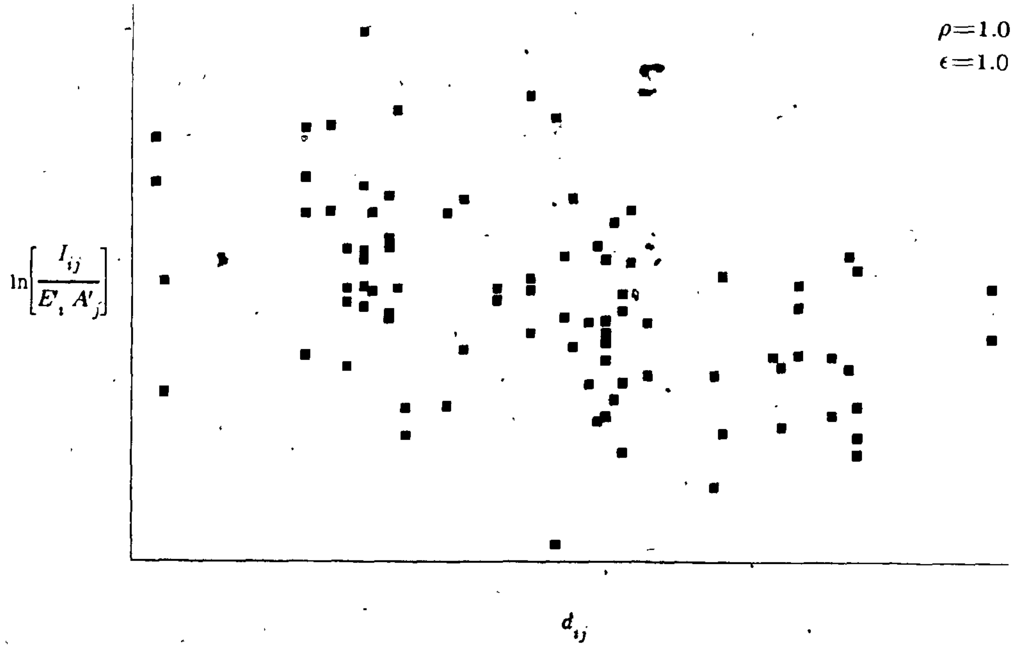
The foregoing comments have referred to ϵ as a measure of error, on a scale that is not clearly defined. To give the reader a feel for the values of ϵ, a visual illustration of Black/White scatters with threshold error levels, and ρ assuming values of 0.2, 0.4, 0.6 and 0.8, appears in Figures 5-7a thru 5-10a. Single straight lines, treated with equivalent packets of noise, produced the scatters in Figures 5-7b thru 5-10b, with Pearson correlation coefficients of −0.442, −0.687, −0.840 and −0.972 respectively.

The methods described are clearly able to tolerate noisier data than are usually encountered in quantitative modeling. The reader may recollect that Mackay's (1958) study of telephone calls pointed to compatibility factors of about 0.2 between Quebec and Ontario, and 0.025 between Quebec and the United States. Mackay's scatter (Figure 4-16), on a purely visual basis, is more cleanly defined than is Figure 5-7a.

To the extent that a regional structure does exist, then, these methods are likely to identify it correctly; there is no certainty, however, that a given solution

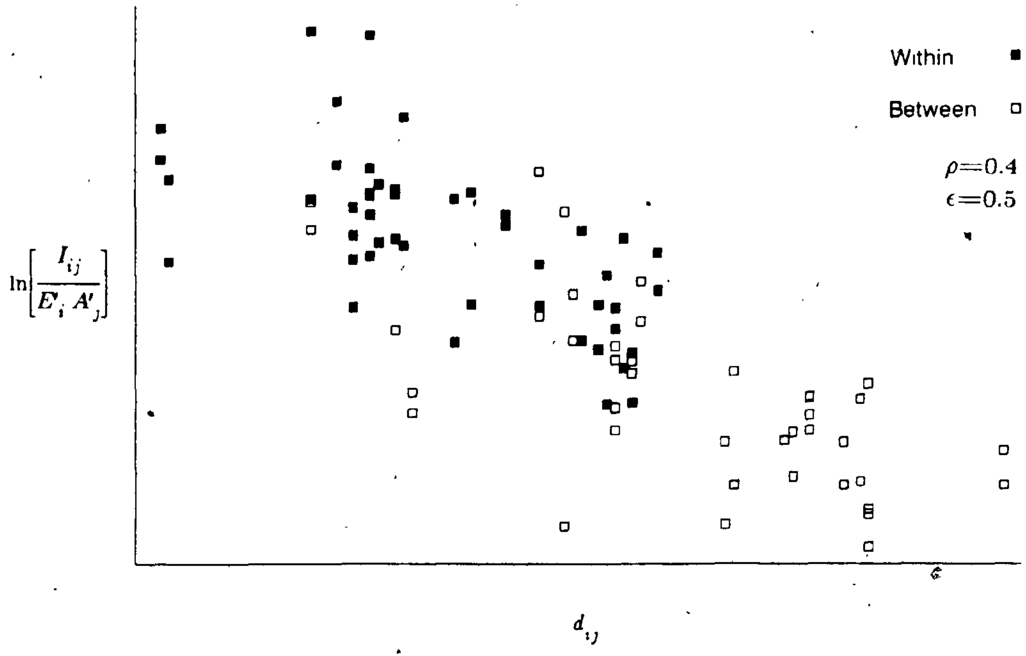


a: Parallel scatter at threshold error level, $\rho=0.2$

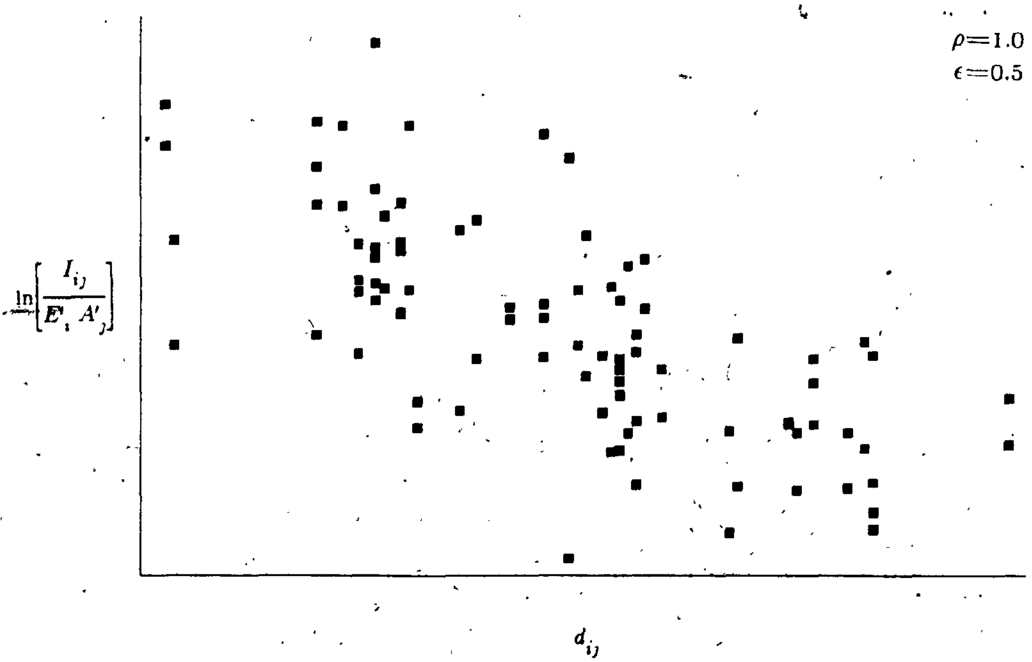


b: Threshold error applied to straight line

Figure 5-7: Threshold error: $\rho=0.2$

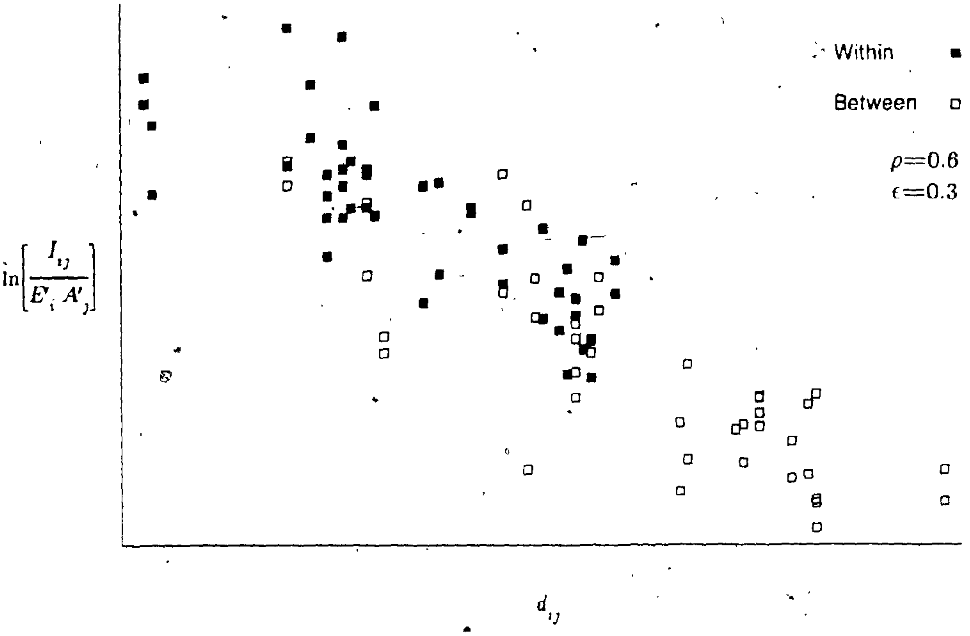


a: Parallel scatter at threshold error level, ρ=0.4

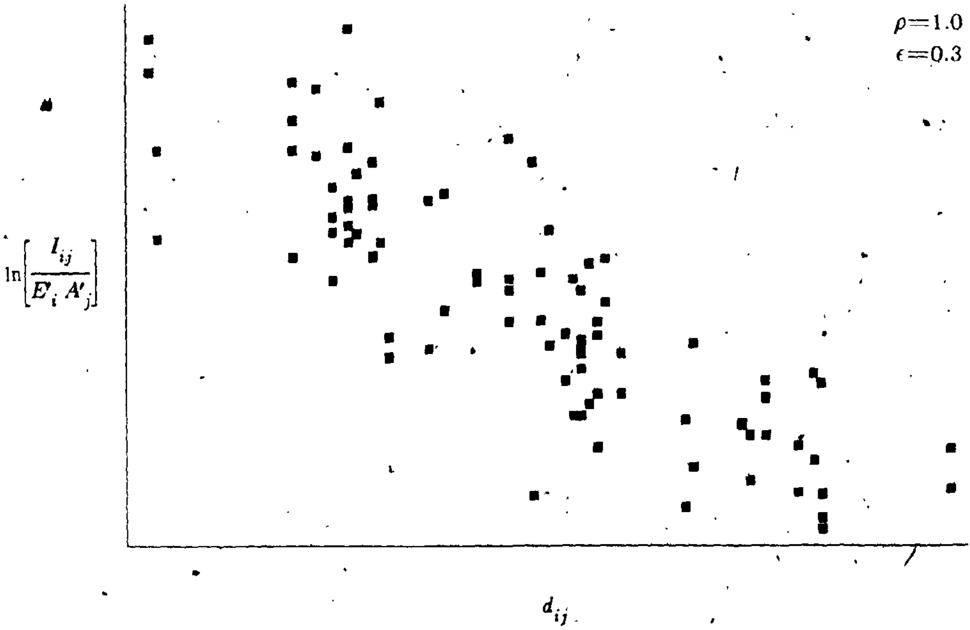


b: Threshold error applied to straight line

Figure 5–8: Threshold error: ρ=0.4

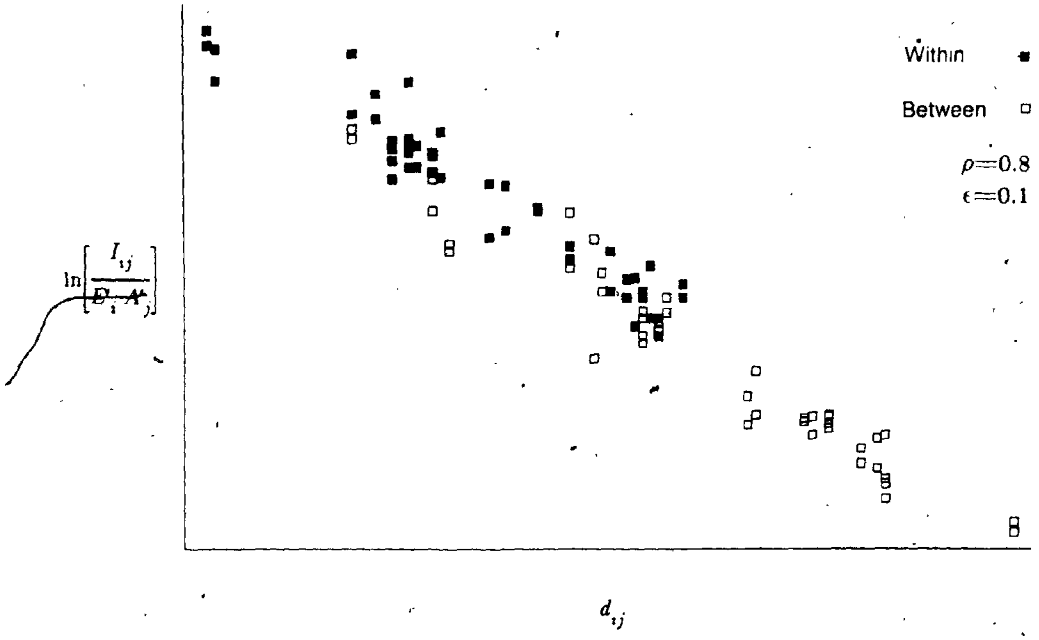


a: Parallel scatter at threshold error level, $\rho=0.6$

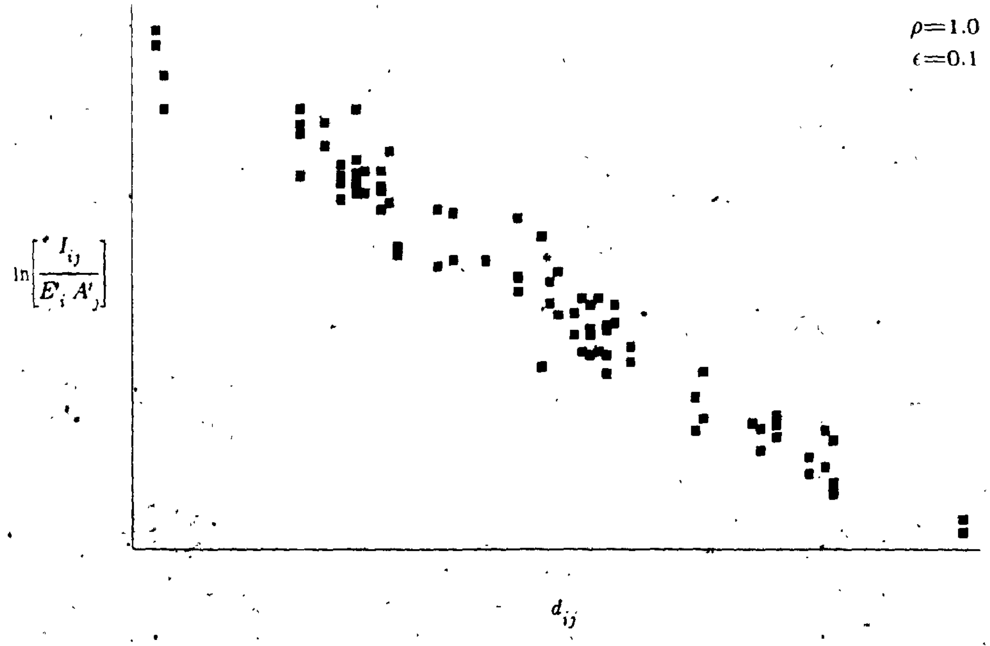


b: Threshold error applied to straight line

Figure 5–9: Threshold error: $\rho=0.6$



a: Parallel scatter at threshold error level, $\rho=0.8$



b: Threshold error applied to straight line

Figure 5 - 10: Threshold error: $\rho=0.8$

represents a regional structure actually contained within the data, and not a chance optimal partitioning of the Black/White scatter with no claim to the underlying existence of regions.

One would suspect a role for parametric statistical tests. Spence and Taylor (1970, p 42-44), referring to Zobler's (1958) use of the *F*-statistic in regionalization, discuss the unfulfilled assumptions required by the technique. As noted earlier, there can be no null hypothesis regarding the spatial interaction model, since it does not provide for stochastic variation. It is clear that the results of the operations introduced above should be interpreted with caution, and that for the present, inferences of regional structure should be conditional upon a clean visual split in the Black/White scatter.

5.5. Multiple Regions

The focus thus far has been on the problem of two regions, or Class 1 in Figure 4-9. While most of the methodological issues are applicable to multi-region universes, there are significant operational and computational differences.

5.5.1. Structures of Multiple Regions

As strictly defined, Class 2 multiple regions exist at one level (that is, they are non-hierarchical), as in the linguistic illustration in Chapter 4. An attempt to view the classification of ESUs as hierarchical must provide for a distinct *basis* of classification for each junction in the hierarchical tree, and is therefore covered by Class 4 of the Black/White regional scheme. In detecting a Class 2 split, one must guard against misinterpreting what might actually be a Class 4 structure.

The literature on regionalization contains several examples of multiple regions being treated as a series of hierarchical binary splits (for example, Clark, 1973;

Hirst, 1975; Slater, 1976). Given the Black/White interpretation of regions, such a construct is not necessarily valid. Consider a two-tiered regional system, where the first-order differentiation is between U and V, and a second-order split exists between V1 and V2. Let the compatibilities between the regions be distributed as shown in Figure 5-11. It is visually clear that an attempt to partition the scatter along the lines of the primary (U-V) split, using the method of least squares described in Section 5.3.3.4, will produce an erroneous result. When the compatibilities are as indicated in Figure 5-12, however, the method would probably be successful. The success of the method thus relies on the distribution of compatibilities between the regions. This raises questions about the nature of hierarchies. It appears that the distribution of compatibilities has to conform to certain patterns in order to be interpreted correctly. Berry states:

> ... regions may be arranged into a hierarchy by the successive splitting of larger regions into smaller regions; and the smaller regions should display less internal differences than the larger regions into which they are grouped.
>
> [Berry, 1968, p 421]

(also see Davies, 1980). Interestingly, Mackay's Quebec-Ontario-US scatter (Figure 4-16) exhibits the structure of Figure 5-12 rather than that of Figure 5-11. In general, suspected hierarchical splits may have to be partitioned as Class 2 scatters initially, and the structure of the compatibilities analyzed for evidence of hierarchy.

5.5.2. Permutations of Class 2 Structures

The general "balls and boxes" problem of permuting articles between containers is well understood (for example, Cohen, 1966). The corresponding problem of regionalization is that of assigning n ESUs to p regions, such that every region is allocated at least one ESU. TOGGLED IMAGES (such as *aabbccd* and *ccaaddb*) are geographically undistinguishable, and should be considered as redundant in the enumeration of permutations. In addition, geographical

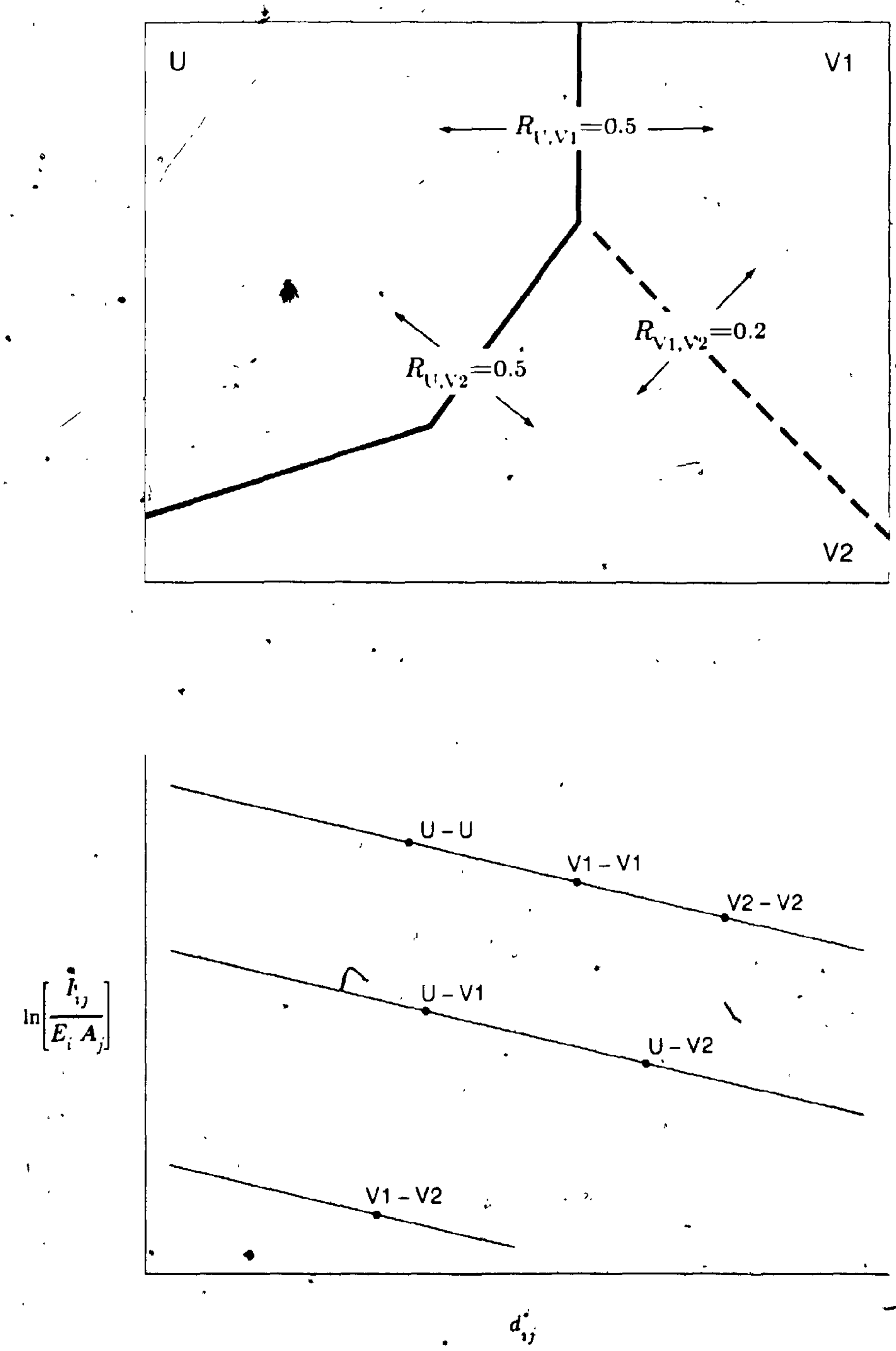
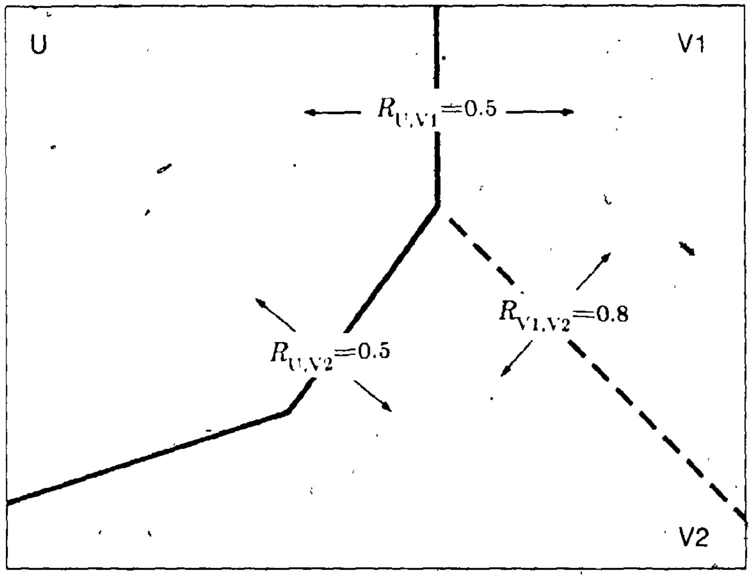


Figure 5-11: Poorly defined hierarchy and scatter



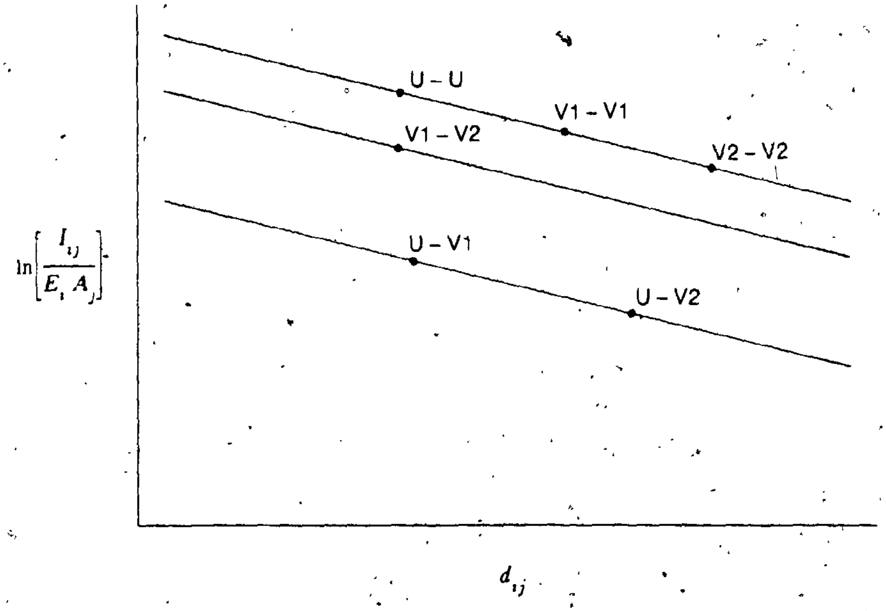


Figure 5 – 12: Well defined hierarchy and scatter

regionalization sometimes requires that regions be contiguous. The exact number of permutations of contiguous regions that may be generated from n ESU's will depend on the contiguity matrix of the ESUs, and cannot be determined independently. In Chapter 5, the expression $2^{n-1}-1$ was derived for $p=2$, assuming no contiguity constraints. Cliff and Haggett (1970) have developed means of calculating the number of possible permutations for two special cases: in the totally unconstrained case there is no contiguity constraint, so that any ESU may be tied with any other; in the second, maximally constrained case, the ESUs are positioned to form a closed loop, so that each ESU is contiguous with exactly two neighbors, those preceding and following it in the ordered list (with accommodations for the extremities).

The number of permutations possible in the absence of contiguity constraints is in general known to be very large. It can be seen from Table 5-7 that the number of permutations of ESUs in three regions is about 20 times the number possible in two regions. The computational resources required to perform tests such as those in Section 5.4, for three and more regions under various conditions, were decided to be unreasonably expensive at the present time.

5.6. Concluding Remarks

This chapter has treated the calibration of the Black/White model from a theoretical standpoint, and illustrated the methodologies with simulated data sets of a limited size. Only experience with real data will illustrate the problems that will confront a researcher in the application of these methods; the feedback may prompt refinements of the techniques. This will be the focus of Chapter 6.

Table 5-7: Permutations generated by n and p

	p							
n	1	2	3	4	5	6	7	8
1	1							
2	1	1						
3	1	3	1					
4	1	7	6	1				
5	1	15	25	10	1			
6	1	31	90	65	15	1		
7	1	63	301	350	140	21	1	
8	1	127	966	1701	1050	266	28	1
9	1	255	3025	7770	6951	2646	462	36
10	1	511	9330	34105	42525	22827	5880	750

Chapter 6

PARTITIONING THE UNITED STATES

6.1. Introduction

The foregoing chapters have been devoted to the development of the Black/White model as a mechanism for regionalization. This chapter has two purposes: first, to illustrate the operation of the model and regionalization procedure on a real data set; second, to discover, and attempt to rectify, operational difficulties and inadequacies of the procedure. The principal concern will be the size of the typical data matrix: since the tests in Chapter 5 operated on 10 × 10 matrices, the opportunities to observe the performance of the heuristic algorithm were limited; a realistic data set might be of the order of 50 × 50.

A second concern will be the choice of data. The reviews in Chapters 2 and 3 drew attention to the problems encountered with interaction matrices, for example, missing diagonals, sparse matrices and modal splits (see Davies, 1980, p 685). These issues will have to be addressed in the choice of data.

6.2. Background

The movements of students from their places of residence to their institutions of study was the focus of early studies on population potentials (Stewart, 1942). In a more recent study of universities in the United States, Warntz (1976) detected a north-south parochialism in the pattern of student migrations. The observation suggested a possibly *literal* Black/White split, along racial lines.

From the point of view of this research, the problem has several attractive features. First, given the politics of university funding and administration (in particular, the standardization of systems within states arising from state control of most major universities), data at the resolution of the state are ideal. With 48 contiguous states and the District of Columbia, the U S migration matrix is of the preferred size for testing the heuristics. Second, student migration is influenced by characteristics of individuals and institutions, and is minimally sensitive to the travel route — the Black/White model is most appropriate in this context. Third, since the movements are monitored at the terminal point, the data are not biased by modal splits. Fourth, since the migrations are across a large country, the gravity modeling escapes the criticisms advanced by Cliff *et al* (1974) regarding the appropriateness of the exponential or power distance decay functions in intra-urban environments. Fifth, large numbers of students are involved in the migrations, producing an extremely dense matrix. Sixth, interaction with self can readily be interpreted as students attending institutions within their home state. Seventh, the problem of indirect flows does not arise. Finally, reliable data are readily available.

The National Center for Education Statistics (NCES), an office of the U S Department of Education, maintains detailed figures on flows of students between states. The information is collected by means of the Higher Education General Information Survey (HEGIS) questionnaire, which has been distributed to the more than 3000 post-secondary institutions in the United States in the Fall of every year since 1966. The Residence and Migration (R&M) information was sought in 1968, 1972, 1975, 1979 and 1981. The response rate to the questionnaire was 91.2% in 1981. Data for non-responding institutions are imputed, based on the known characteristics of the institutions. NCES publishes the data in the form of printed volumes of processed tables, or as machine-readable files (about 10 megabytes each) containing the coded responses to the questionnaires. In the latter form, each of the 3000-odd institutions is identified individually, and selected characteristics listed, for example:

- institution type:
    - university/2-year/4-year institution
    - main/branch campus
    - highest level of offering
- predominant race of student body
- sex of student body
- type of control:
    - private or public
    - specific controlling organizations
- calendar system

Counts of students enrolling in a program for the first time are listed by state of residence, and further broken down by program level (undergraduate, graduate and first professional) and a distinction is made between part-time and full-time enrollment in each category.

The data for 1979 and 1981 were obtained for this study. Student flows satisfying the conditions of origin and destination shown in Table 6-1 were aggregated into a 49 × 49 asymmetrical origin-destination matrix.

Table 6-1: Selected characteristics of institutions and students

	Included	Excluded
Institution	✓ University ✓ Other 4-year institution ✓ 4-year branch of institution	× 2-year institution × 2-year branch of 4-year institution
	✓ Located in 48 contiguous states and District of Columbia	× Located in Alaska, Hawaii or outlying areas
Student	✓ Full time	× Part-time
	✓ Undergraduate ✓ Graduate	× First-professional × Unclassified
	✓ Resident of 48 contiguous states and District of Columbia	× Resident of Alaska, Hawaii or outlying areas, or foreign student

6.3. Selection of Variables and Model Form

As a preliminary step in this analysis, the various flow variables were examined for conformity with four alternative model specifications, with separations between states expressed as the great circle distances between their population centroids. The results are shown in Table 6-2.

Notwithstanding the remarks made earlier regarding the availability and interpretability of diagonal data, arguments could be advanced against the use of those entries. The first has to do with the nature of student migration. Universities routinely apply differential fees and admission standards to out-of-state students, leading to a large proportion of students remaining within their home state. Of the 1,618,640 students entering university in 1981 (and who satisfied the characteristics listed in Table 6-1), 1,240,918, or 77%, stayed within their home state. The corresponding figure for graduate students was 62%.

Second, there is the problem of an appropriate distance measure for in-state attendance: zero-distances may be used in the exponential, but not in the power model. For the experimental record, the power model was calibrated with half the distance to the nearest neighbor substituted in the diagonal cells.

For all the variables investigated, the power transformation, leading to the relationship

$$I_{ij} = E_i A_j d_{ij}^{-\beta} \qquad \ldots (6.1)$$

yielded the highest multiple correlations, outperforming the exponential models

$$I_{ij} = E_i A_j e^{-\beta d_{ij}} \qquad \ldots (6.2)$$

The inclusion of diagonal flows led to poorer model fits, in the case of all variables (Table 6-2).

Table 6-2: Model fits for various data sets

Student group	Year	Model*	Code	Multiple Correlation	β'
Undergraduates	1979	H	UG79H	0.97830	1.818
		P	UG79P	0.98127	1.601
		D	UG79D	0.96085	0.135E-2
		E	UG79E	0.97262	0.113E-2
	1981	H	UG81H	0.97656	1.831
		P	UG81P	0.97932	1.619
		D	UG81D	0.95813	0.135E-2
		E	UG81E	0.96985	0.114E-2
Graduates	1979	H	G79H	0.97120	1.167
		P	G79P	0.97547	0.948
		D	G79D	0.95383	0.832E-3
		E	G79E	0.96823	0.658E-3
	1981	H	G81H	0.96896	1.168
		P	G81P	0.97369	0.938
		D	G81D	0.95053	0.818E-3
		E	G81E	0.96601	0.640E-3
Total: all students	1979	H	T79H	0.98646	1.678
		P	T79P	0.98920	1.459
		D	T79D	0.97460	0.122E-2
		E	T79E	0.98359	0.102E-2
	1981	H	T81H	0.98575	1.684
		P	T81P	0.98849	1.465
		D	T81D	0.97314	0.121E-2
		E	T81E	0.98230	0.101E-2

* Models:

H	Power transformation: distance to self=0.5 × distance to nearest
P	Power transformation: diagonal entries excluded
D	Exponential transformation: diagonals included with $d_{ii}=0\ \forall\ i$
E	Exponential transformation: diagonals excluded

By itself, this observation is not a sufficient basis on which to infer the superior predictive power of a particular transformation, since the provisional calibration assumes no regional distinctions — a low correlation could be the result of a wide regional split, rather than poor model fit. More significantly, when the estimates of interaction, based on the parameters recovered in these provisional calibration runs, were plotted against distance as a parallel scatter, the inappropriateness of the exponential transformation became clearly apparent, and the power model emerged as decidedly superior (Figure 6-1). Excluding from the analysis, for reasons outlined above, those students who remained in their home state, the principal objection to the power model was eliminated; by contrast, the estimated parameters of exponential models would be biased by omission of these cases.

6.4. Preliminary Partitioning

The parallel scatter for the variable UG81P (Figure 6-1b) did not exhibit the clear split that Mackay (1958) had observed in his data on telephone calls. This was partly due to the large number of points (2352), and therefore a proportionately greater amount of visual noise in the data. But it raised the issues of whether the internal divisions in the United States were significant, whether or not student migration was a typical instance of regional bias in interaction and whether the amount of noise was overwhelming.

The logistical arguments against explicit methods of analyzing the scatter for a problem of this size have been dealt with in the previous chapter. Of the heuristic methods, the choice was between the strategies of algorithms A and B. It was noted in the last chapter that configurations with widely differing spatial representations may yield only slightly different objective functions. In practical terms, this means that the cost of imprecision may be high, and that an expensive algorithm could pay off in terms of accuracy. The search strategy of Algorithm B was therefore preferred over Algorithm A.

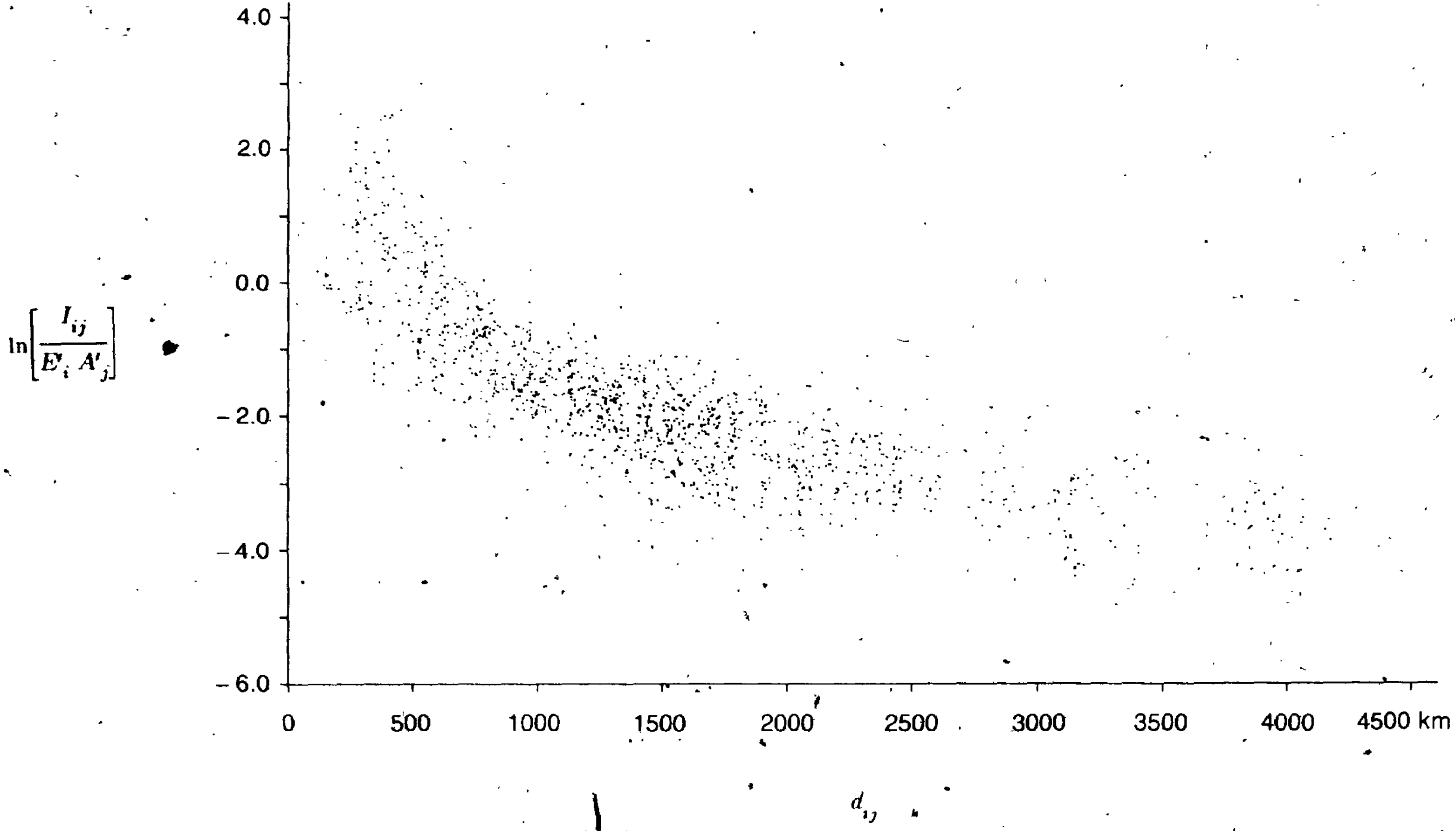


Figure 6–1a: Parallel scatter: UG81E

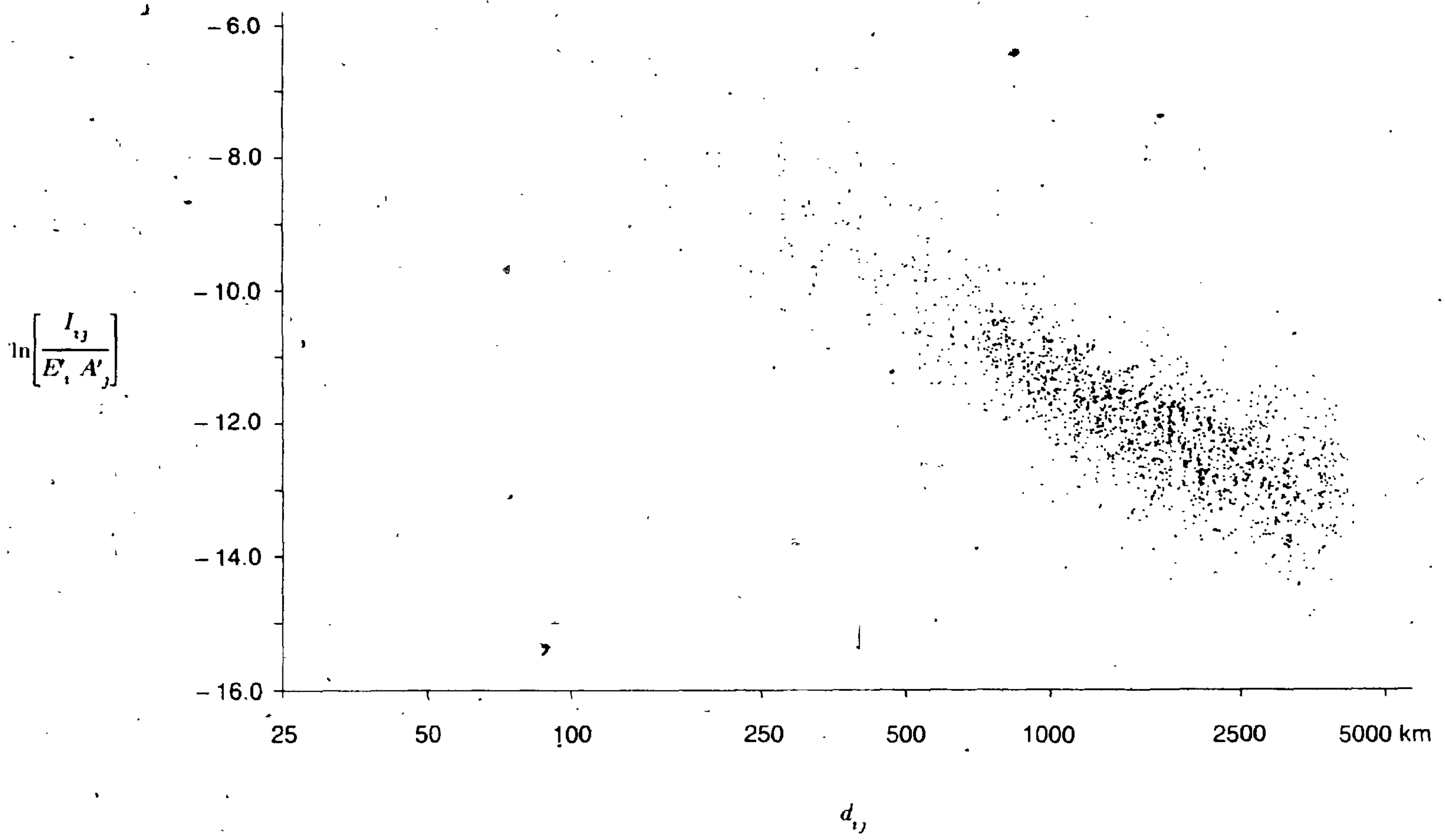


Figure 6 – 1b: Parallel scatter: UG81P

In analysis with the heuristic, the state of Delaware was the first admission to the dominant cluster. It was joined, in order, by the District of Columbia, Maryland, New York, Connecticut, New Jersey, New Hampshire, Massachusetts and Rhode Island, forming a strong Northeastern core, before aggregating Michigan, Virginia, and some of the Midwest and South. The value of the objective function for the final configuration (85997873634098-0)[7] was 1056.83 (Figure 6-2).

The geographical patterns in the solution did not lend themselves to ready explanation. It was difficult to associate Minnesota, Iowa and Missouri with West Virginia, the South and the West. Again, it was not easy to explain why Arkansas, Tennessee, Kentucky and Virginia were associated with the Northeast and part of the industrial Midwest.

6.5. Improvements to Heuristic

The troublesome question in this situation was how far the heuristic solution was from the actual optimum. The options available to improve the value of the objective function were

- search the beta axis at a finer resolution, by decreasing the critical step size parameter (b_5)
- modify the algorithm so that the operation could build upon an intelligent, intuitive starting solution provided by the user
- introduce a facility in the algorithm to REVERSE out of local minima

The first option, while operationally straightforward, would have led to greatly increased processing costs, and was least likely to improve the solution.

---

[7] Configurations in this chapter are identified by a decimal integer, followed by the binary code for Alabama. The configurations 85997873634098-0 and 85997873634098-1 are geographically indistinguishable; in the first, the Alabama cluster is subdominant, while in the second it is dominant. The distinction is relevant only with respect to the methodology.

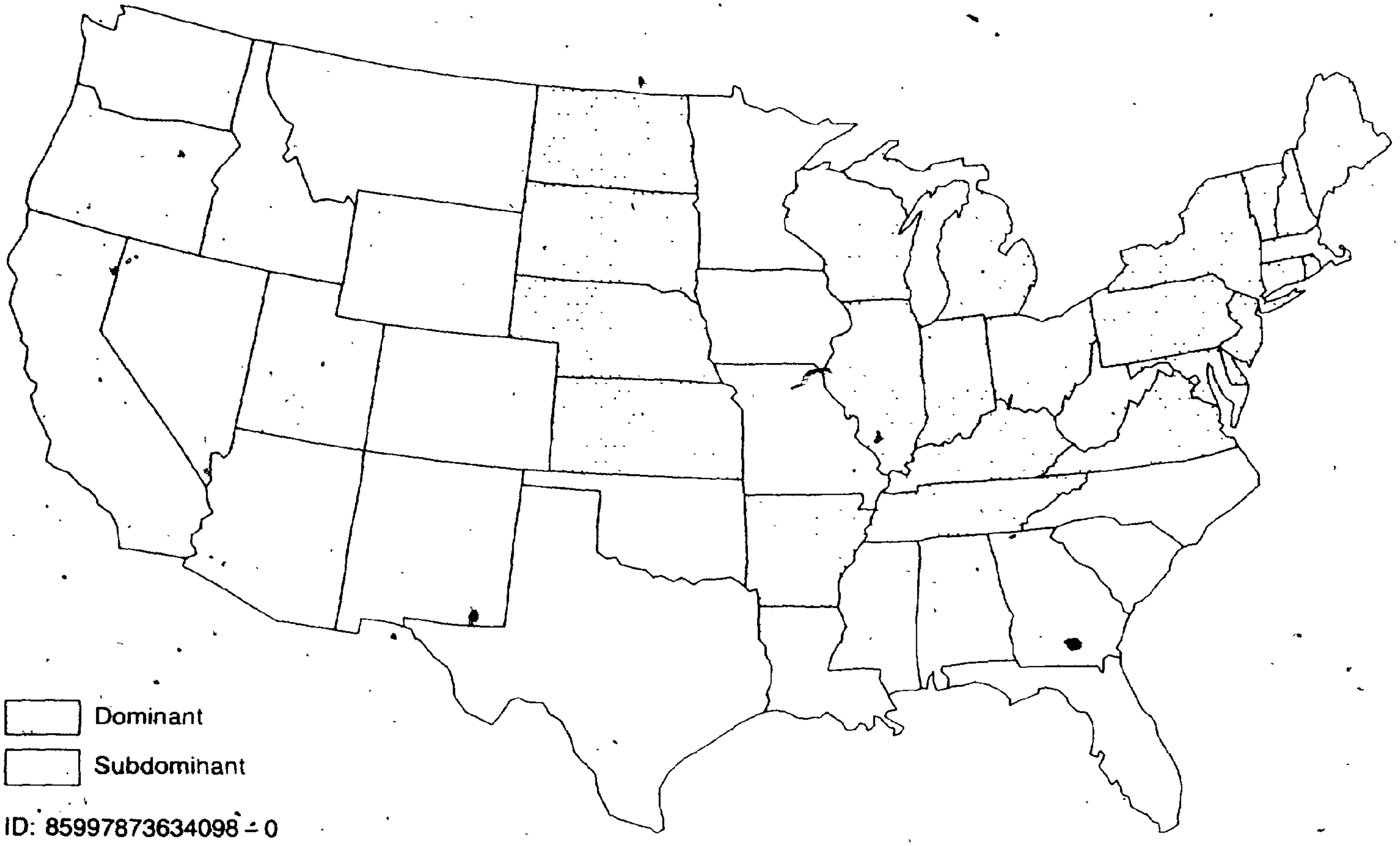


Figure 6−2: Preliminary result: Algorithm B, UG81P

Attention was therefore devoted to the latter two options, both of which required modifications of the algorithm.

6.5.1. Polar Aggregation

The development of this option was prompted by the tendency of the heuristic to get locked in a partial solution, refusing to admit further ESUs into the dominant cluster. As it now stood, the criterion for stopping the process of aggregation of ESUs to the dominant cluster was a failure to improve the objective function in one scan cycle. However, the check for improvement was effectively bypassed in the initial cycle, since the incumbent objective function at the start of the operation was arbitrarily set at a very large constant. Thus, if the search for the optimum commenced with a pre-set dominant cluster, the algorithm would search the ESUs in the first scan without a prior evaluation of the objective function; it would therefore necessarily register a decrease in the objective function, *forcing* the admission of the most eligible ESU.

The heuristic was modified to offer the option of specifying a number of nuclear ESUs, or POLES, and hence dubbed Algorithm 'P'. The poles would constitute the dominant cluster at the commencement of the search process; and at least one compatible ESU would be admitted to the cluster before the algorithm terminated.

6.5.1.1. Single Pole

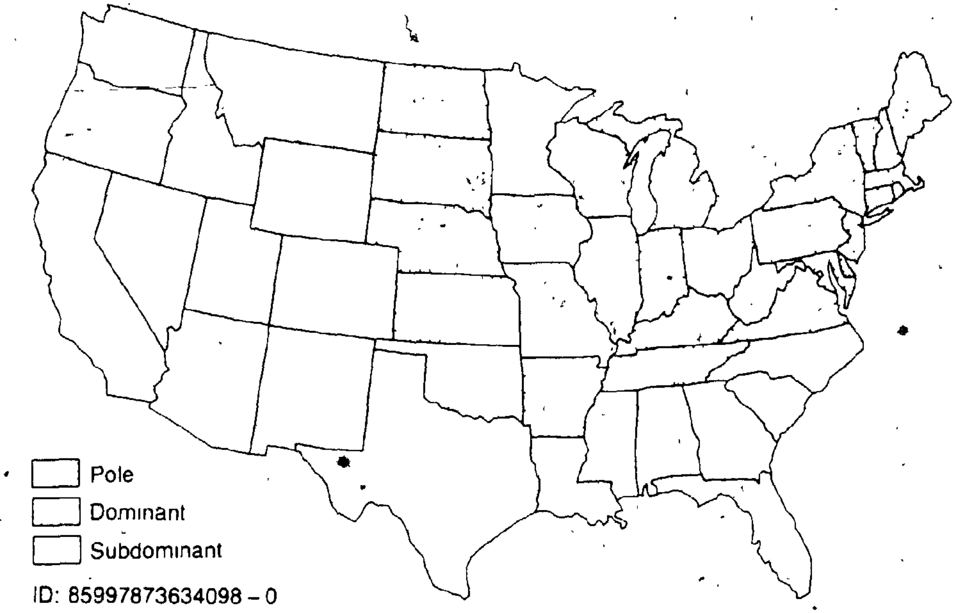
It is tempting to argue that when establishing a pole consisting of a single ESU, the choice of the ESU is immaterial, since the aggregation process will identify only its compatible companions: that is, if the pole were naturally "black," as opposed to "white," the search would identify the most compatible black ESU for aggregation, and vice versa. However, the particular sequence of combination of ESUs may be crucial to the outcome of the heuristic, as illustrated by the following example. In a trial run with no poles, Delaware was the first ESU to enter the dominant cluster, and at the end of the operation, California was one of the states in the opposing (sub-dominant) cluster. In the next run, California was established as an ANTI-NATURAL pole. The results of the two runs differed substantially, as documented in Figure 6-3 and Table 6-3. This clearly indicates a role for *some* geographic expertise on the part of the user.

6.5.1.2. Multiple Poles

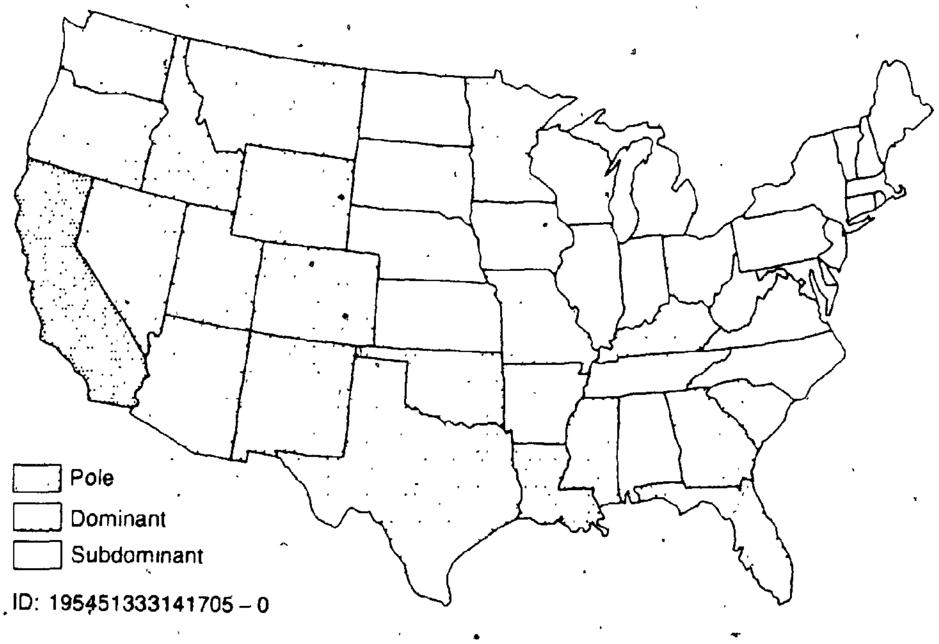
In general, the introduction of poles helps establish the direction in which the solution will evolve, and saves some computing time by obviating the earlier iterations. Table 6-4 records the computing time of the algorithm with the number of poles varying from zero to four. The poles chosen were NATURAL, that is, they were the first entries into the solution for the previous problem. Thus Delaware (ESU 7), used in the one-pole run, was chosen because it was the first to be admitted in the test with no poles; ESU 8 was the first admission in the one-pole run, and was therefore used as a pole in the two-pole run; etc. It is noteworthy that although the execution time in the multi-pole problem was reduced by nearly 20% when 4 poles were used, the solution itself did not change — that is, it did not matter whether a particular ESU was naturally *discovered* or artificially *established* as a pole.

Table 6-5 illustrates how a variety of solutions may be arrived at by varying the pole sets. In contrast to the previous illustrations (Table 6-4), where the selection of poles was natural, the poles in these runs were INTUITIVE. For example, a selection of Southern states was established in some runs, while the New England states were set up in another.

Table 6-5 demonstrates not only that the choice of poles is important, but that in the case of this example, an intuitive selection has actually led to an improvement in the objective function over the figure arrived at without the pole facility.



a: Pole: Delaware



b: Pole: California

Figure 6–3: Effects of pole selection (single pole)

Table 6-3: Effects of pole selection (single pole)

Data: UG81P

$b_1 - b_5$: 0.5, 2.0, 0.25, 0.2, 0.01

Computer: Cyber 170-835

Run		Poles	Basis	Z	Solution ID	CPU (secs)
1	1	Delaware	Natural	1056.831	85997873634098	2048
2	1	California	Anti-natural	1058.720	195451333141705	1837

Table 6-4: Performance of Algorithm B with natural poles

Data: UG81P

$b_1 - b_5$: 0.5, 2.0, 0.25, 0.2, 0.01

Computer: Cyber 170-825

Run	Poles		Basis	Z	Solution ID	CPU (secs)
1	0			1056.831	85997873634098	3558
2	1	Delaware	Natural	1056.831	85997873634098	3350
3	1	Delaware	Natural	1056.831	85997873634098	3180
	2	DC				
4	1	Delaware	Natural	1056.831	85997873634098	2994
	2	DC				
	3	Maryland				
5	1	Delaware	Natural	1056.831	85997873634098	2809
	2	DC				
	3	Maryland				
	4	New York				

Table 6-5: Effects of intuitive pole selection

Data:	UG81P
$b_1 - b_5$:	0.5, 2.0, 0.25, 0.2, 0.01
Computer:	Cyber 170-835

Run		Poles	Basis	Z	Solution ID	CPU (secs)
1	1	Alabama	South	1073.197	86023702289170	2055
2	1	Alabama	South	1052.819	204770009246251	1836
	2	Georgia				
3	1	Alabama	South	1052.819	204770009246251	1733
	2	Georgia				
	3	Mississippi				
4	1	Alabama	South	1052.819	204770009246251	1629
	2	Florida				
	3	Georgia				
	4	Mississippi				
5	1	Alabama	South	1052.819	204770009246251	1422
	2	Florida				
	3	Georgia				
	4	Louisiana				
	5	Mississippi				
	6	S Carolina				
6	1	Connecticut	N England	1059.184	195442751759565	1539
	2	Maine				
	3	Massachusetts				
	4	N Hampshire				
	5	Rhode Island				
	6	Vermont				

In practise, a regionalization proceeds from some *a priori* notions of the anticipated result. The provision of the pole facility permits a researcher to steer a solution towards its expected outcome by providing a meaningful set of growth nuclei. While this may be criticized for introducing subjectivity into the outcome, it may be pointed out that location-allocation algorithms routinely use "starting solutions," and that the ultimate purpose is, after all, to improve the value of the objective function.

The map corresponding to the improved solution (Figure 6-4) lends itself to interpretation much more easily (it may be pointed out that this was the lowest value of the objective function found in the entire range of tests, with this particular variable). The South emerges as a cohesive entity, stretching far enough north to include the states of Ohio, Pennsylvania, New York, New Jersey, Delaware and Maryland — it is significant that the District of Columbia is not included within this cluster. On the other hand, it is surprising that Oklahoma is excluded. The anomaly in this configuration, however, is the sub-cluster of Western states — Idaho, Nevada and Utah — that emerge as compatible with the South rather than the North. Under the terms of the Black/White model, these do *not* represent a third independent group, as one is tempted to infer; the correct interpretation is that the Western states have a greater affinity with the South than with the North. This could arise on account of a common perception of abandonment by the rest of the North. An exhaustive contextual explanation of the pattern is beyond the scope of this exercise.

6.5.2. Reversing Algorithm

The algorithm as described so far proceeds by aggregating ESUs to the dominant cluster until a cycle fails to improve the objective function. There is the possibility, though, that with the successive aggregation of compatible ESUs to the dominant cluster, the character of the cluster changes to the point where

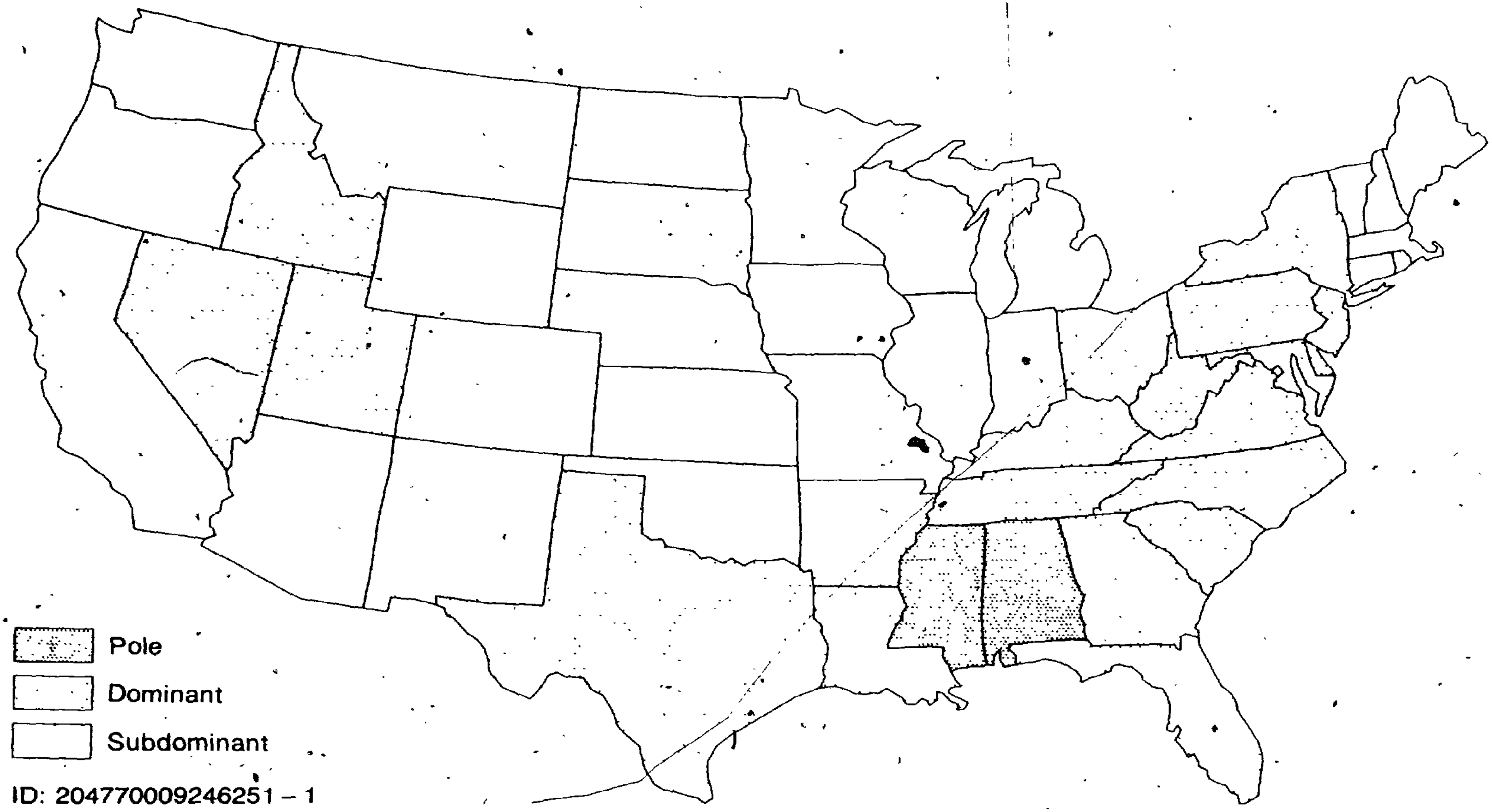
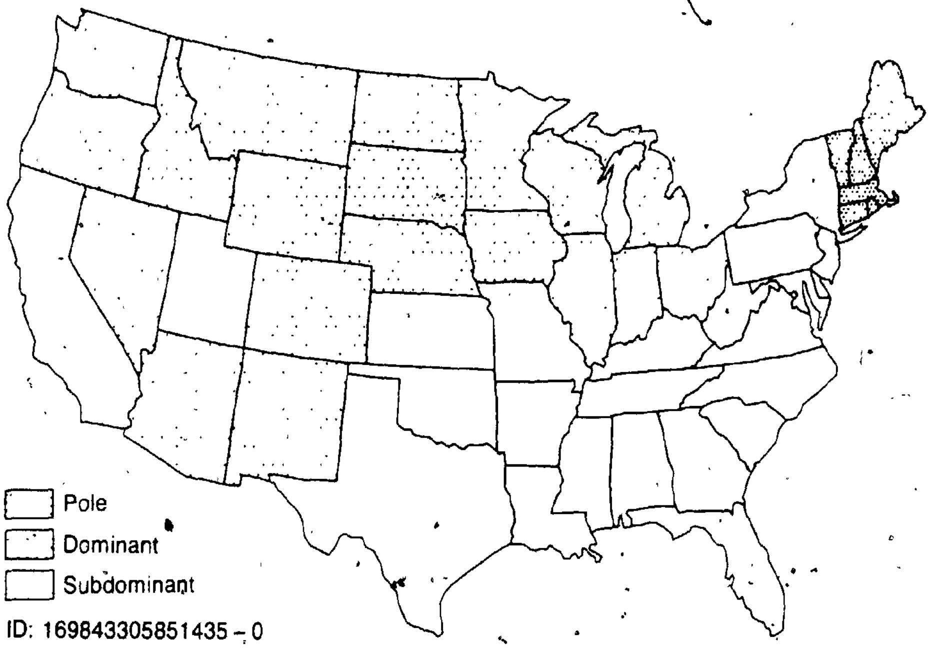


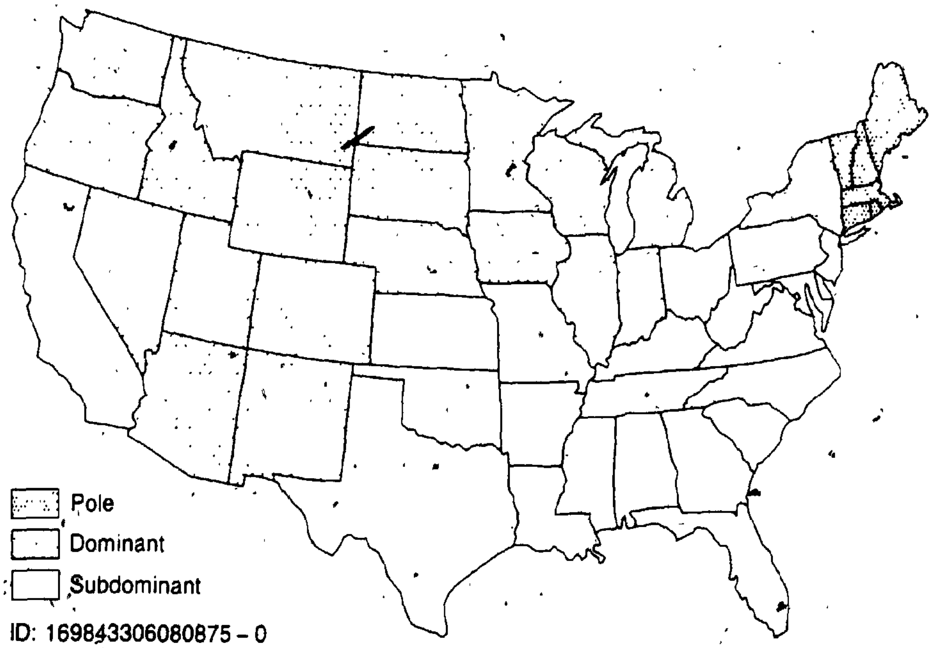
Figure 6–4: Optimal configuration, UG81P

one or more of the ESUs admitted earlier no longer belong to the pack. It would therefore be appropriate, when the algorithm has terminated, to identify the most expendable ESU in the dominant cluster (the one whose removal causes the least increase in the objective function), and attempt to replace it with an ESU currently in the sub-dominant cluster. If the replacement qualifies for admission (by lowering the objective function established before the reversal), the algorithm may resume, until either an expended ESU is immediately replaced by itself, or the most expendable ESU is the one last admitted into the dominant cluster. When this reversing feature was incorporated into the regionalization algorithm (now dubbed Algorithm 'R'), it failed to produce an improvement in the objective function for this particular variable (Undergraduates, 1981). Maryland was identified as the most expendable ESU, but was then found to be the most eligible entry into the dominant cluster. In runs with another variable (All students, 1981), though, the objective function reached a low of 879.15 before reversal (Figure 6-5a). Ohio was dropped from the dominant cluster, and replaced by New York and Utah, resulting in a decrease in the objective function, to 875.79 (Figure 6-5b).

Clearly, there is scope for more exhaustive examination of the dominant cluster for the possibility of substitution, perhaps along the lines of the Teitz and Bart (1968) vertex substitution algorithm used in location allocation modeling (also see Goodchild and Noronha, 1983). Unfortunately, the examination of each fresh configuration requires the identification of the most eligible or expendable ESU, and therefore up to *n* evaluations of the objective function; while this increases computing time substantially, experience with the data has indicated that the solution is not always improved. Future generations of computing equipment may make such enhancements of the heuristic practically feasible.



a: Before reversal



b: After reversal

Figure 6–5: Effects of reversal

6.6. Final Results

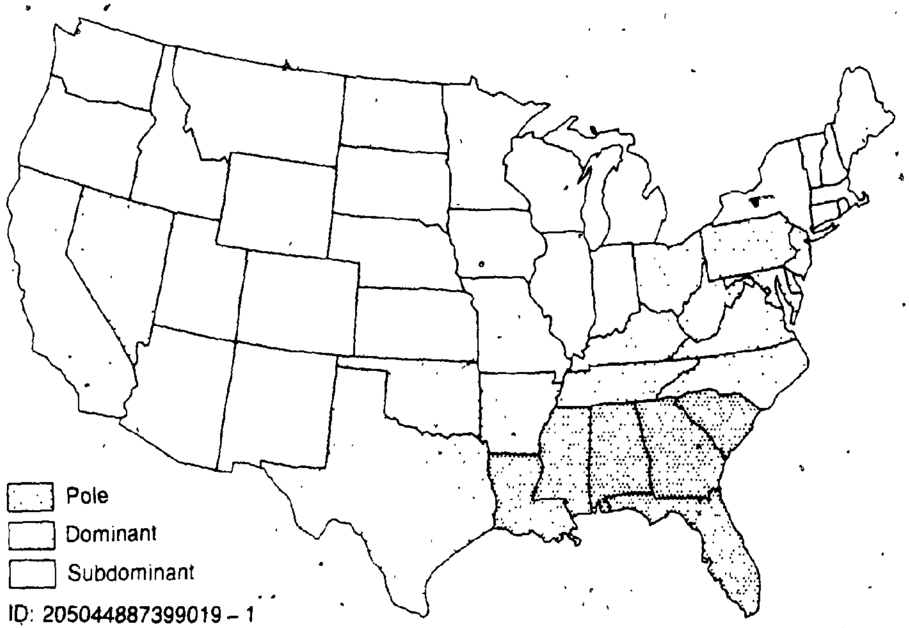
Data sets T79P and T81P produced the highest correlations in the original regression runs (Table 6-2); however almost all the illustrations of the algorithms have been with reference to UG81P. For the record, Figures 6-6 and 6-7 show the splits resulting from the analysis of these two sets (the variable T81P was cited earlier while illustrating the reversing feature, but that analysis proceeded from a New England pole, and did not produce the best outcome; a Southern pole produced a better objective function).

The analysis of T79P produced an unexpected bonus: both the New England and the Southern poles produced the same configuration. It has been pointed out earlier that this is an unlikely result; its occurrence in this case suggests that the heuristic may indeed have hit upon the global optimum — that suspicion is reinforced by the inability of the solution to reverse.

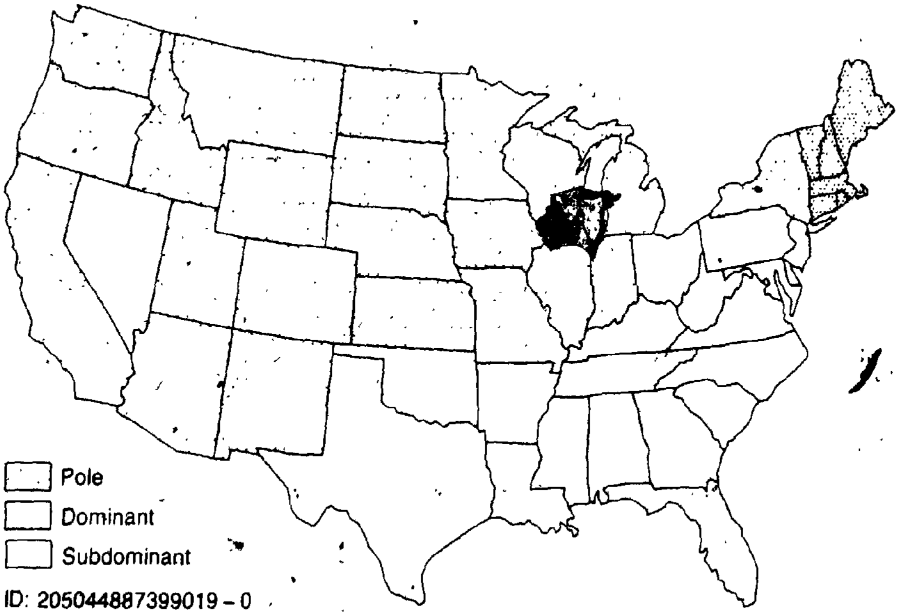
All the analyses indicate an outlier of the South, in the Nevada-Utah area. This suggests that it is not an anomaly of the data, but a structural secret that may demand an explanation. Significantly, it is a non-contiguous solution, that would not likely have been arrived at, or perhaps would have been constrained from occurring, in classical regionalization techniques. However, in the absence of a tested means of interpreting the parametric statistics, and particularly since the scatters offered no visual corroboration, one should be cautious in one's interpretation of these results.

6.7. Concluding Remarks

In retrospect, it was *not* discouraging to be unable to extract a clear verdict regarding the regional structure of the United States, from the data. The basis for suspecting a clear regional dichotomy at this point in the history of that country was tenuous, and it was not certain that the flows of students reflected



a: Six Southern poles



b: Six New England poles

Figure 6-6: Optimal configuration, T79P

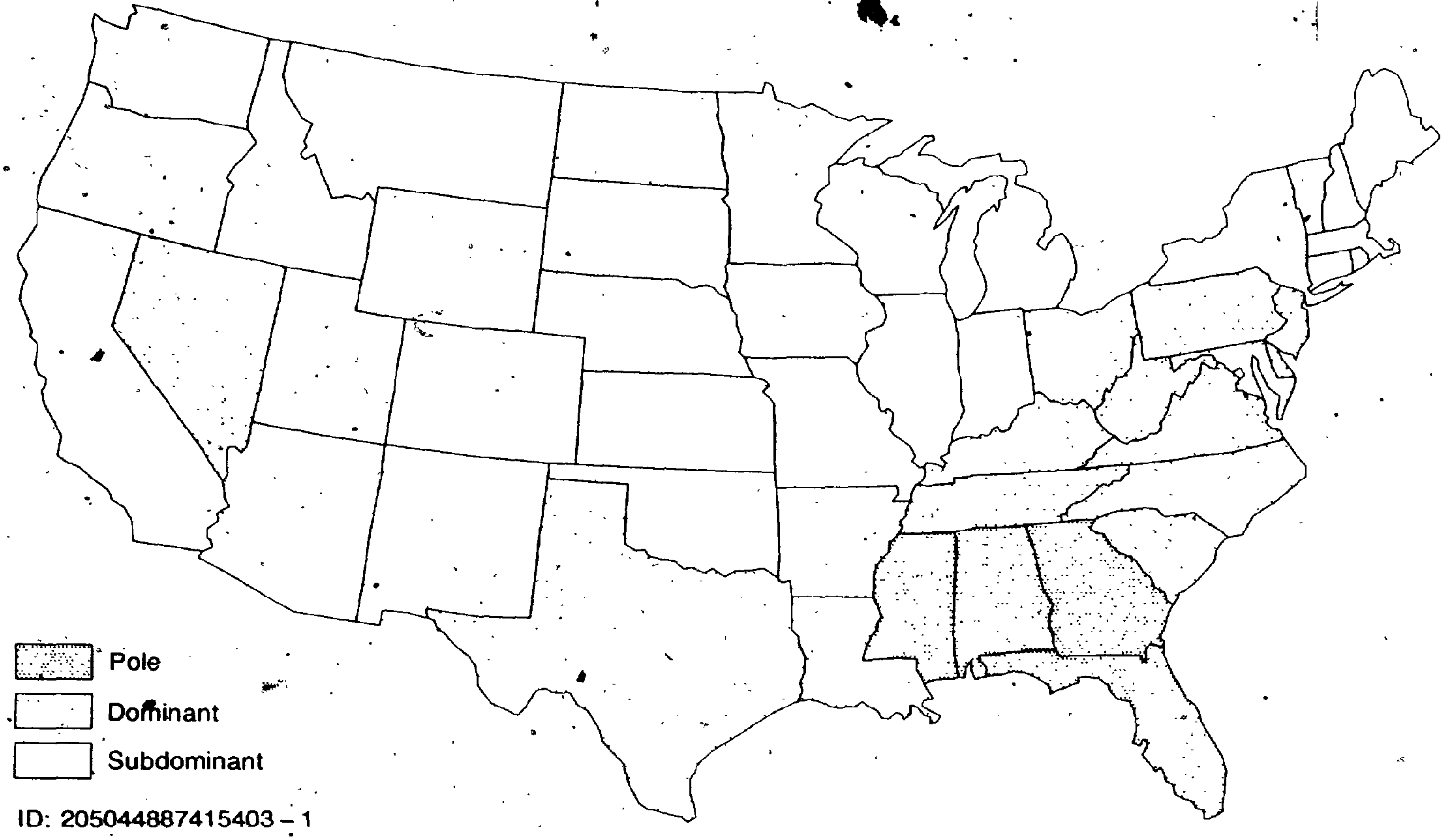


Figure 6–7: Optimal configuration, T81P

such a partitioning, if it did exist. The compatibility factors for the various data sets ranged from approximately 0.5 to 0.9 — considerably higher than the 0.2 observed by Mackay between Ontario and Quebec.

The purpose of this chapter in the research was not necessarily to contribute to the present understanding of American regionalism. Rather, it was meant to highlight the practical issues involved in discovering regional structures, and to suggest strategies for improving the algorithms. In both respects, the experience has proved extremely rewarding.

In regard to the practical issues, these included selection of data. A researcher has to cope with matters such as modal split, interactions with self, density of the matrix, and, not least, the reliability of the data. In all these respects, the NCES data were the ideal choice.

In terms of examining the efficacy of the algorithms, the experience with the data fulfilled its purpose. The algorithms produced fairly consistent results across variables; in one case (T79P), the heuristic appeared to have hit upon the optimum solution; in all cases, the algorithms produced non-contiguous solutions, reflecting the emphasis on preferred, as opposed to revealed, behavior, and thence the principal advantage of the Black/White concept as a model of the functional region, and as a mechanism for regionalization.

Chapter 7

EPILOGUE

7.1. General Review

This research was intended to break new ground in the reconciliation of divergent approaches in geography. On the one hand, it was meant to explore the concept of the functional region in some detail, and to identify the structural regularities and behavioral rules associated with the phenomenon. It was an attempt to bring that improved understanding of the region to bear in a fresh set of objective methodologies for delimiting functional regions. On another level, the research was expected to develop means of adapting quantitative models to real environments, in ways that have not been adequately explored, especially in the recognition of the operational subtleties that variations in human culture impose on landscapes. In effect, the intention was to weld together two divergent approaches to geography, as well as two independent quantitative literatures, namely the technique of functional regionalization and the modeling of spatial interaction. Finally, the tone of the research was set by a commitment to *application* of geographical theory in the real world.

This concluding chapter is devoted to a critical appraisal of the research, and will identify directions for further development of the concepts.

7.1.1. Summary of Content

The classical notions of regionalization were examined in the first chapter of this thesis. Two independent taxonomies of regions were identified, namely the descriptive and the structural. The terminology and controversies associated with the various concepts of regions were surveyed within the framework of the two taxonomies. Methodologies for delimiting regions were reviewed briefly. It was noted that while formal, nodal and equitable regions were adequately conceptualized, and the methods for their delimitation were accepted, there were considerable operational problems with the methods of delineating functional regions. Chapter 1 also drew attention to the application potential of the regional concept, and outlined a number of desirable features of objective methods of regionalization.

Chapter 2 was devoted to a survey of methods of functional regionalization. While there was agreement that the data requirement should be a matrix of interactions, the methods of pre-processing the data to correct inherent problems were subject to debate. Two fundamentally different approaches to the treatment of the processed matrix were identified. On the one hand, the factor analytic methods isolated groups of ESUs which had similar patterns of interaction with all other ESUs in the system; the other class of techniques aggregated ESUs by maximizing the ratio of within-region to between-region interaction. The conclusion of the review was that techniques had been borrowed from other disciplines, and applied to the problem of regionalization, with little concern for the causal origins of functional regions. It was argued that the clues to the structure of interaction matrices should be investigated from the point of view of spatial interaction modeling.

The purpose of Chapter 3 was to review the concepts and models of spatial interaction. The development of the gravity model was traced from its origins as a physical analogy, where the mass terms were interpreted as populations, to the modern generic form (equation 3.7) where origin and destination attributes are aggregated into the emissivity and attraction terms respectively. The theoretical derivations of the model were examined briefly, emphasizing the credibility that they lent to what was otherwise merely an empirical regularity. In conclusion, it was noted that the spatial interaction model, while recognizing size and other unary characteristics of the origin and destination nodes, otherwise assumed culturally isotropic applications environments.

The Black/White model was introduced in Chapter 4. While algebraically similar to the spatial interaction model, it was derived from accepted axioms of choice behavior. The principal element that it added to the spatial interaction model was the compatibility factor R_{ij}, representing a dyad-specific variable. The factor was shown to be amenable to a variety of algebraic forms, to address various regional structures. It was argued that the Black/White model could be applied in predictive situations; in regard to the purpose of this thesis, it offered an alternative model of the functional region, and, for the first time, a method of functional regionalization based on geographic theory.

Chapter 5 was devoted to methods of calibrating the Black/White model. The calibration of the spatial interaction model was treated first, illustrating the techniques on simulated data. Various options were examined for calibrating the Black/White model; the method finally selected was based on obtaining preliminary estimates of the parameters (using the procedures appropriate to the spatial interaction model), and examining the multiplicative residuals between the observed matrix of interactions and that predicted by the initial parameter estimates. The method was shown to deliver accurate results over a range of scenarios, with simulated data. A number of variants of the basic algorithm were developed, including a heuristic method for large problems. They were tested on the simulated data, and were shown to produce correct results. Random error was introduced into the simulations, to test the tolerance of the algorithms. The performance of the algorithms, with varied error inputs, was documented with three representative regional configurations. While the above methods were able to identify binary regional structures correctly, the extension of the techniques to the problem of delimiting multiple regions was only briefly examined; it was decided that extensive testing with current computing facilities would be unreasonably expensive. This problem would be a promising area of further research.

Finally, it remained to test the techniques on real data. Chapter 6 was intended to examine the problems encountered with selection of data, and to demonstrate the heuristic algorithms developed in Chapter 5. Tables of the flows of students in the United States were treated with the Black/White methods. A number of improvements in the heuristic algorithms were effected. The final results broadly reflected suspicions of the functional divisions in the United States, but there were some interpretational problems, suggesting that the data did not reflect a clear regional structure.

7.2. Appraisal of Contributions

7.2.1. To Interaction and Modeling

Quantitative treatments of geographic phenomena and attempts at modeling geographic processes have, since Von Thunen, begun by simplifying and abstracting landscapes. While this facilitates conceptual development of the skeletal issues, models constructed in these rarefied environments constitute misspecifications of the real world. Other dimensions of variation manifest themselves as noise. Developmental research in quantitative modeling is largely directed towards identifying subsidiary dimensions, and systematically incorporating variances based on them into the formulation of the models.

Despite the professed concern with human activity, geographic modeling has tended to ignore crucial, and behaviorally significant, characteristics of the human landscape. The Black/White model offers a fresh look at interaction, and has implications in several related fields of geography. These include retailing (loyalty to store chains or brands and the use of credit cards), migration and population forecasting, and transport and communications network planning. The model raises fresh questions about the validity of potentials of population, income and markets. More fundamentally, it explodes the myth of cultural isotropism, while offering a means for incorporating a new dimension of variation into traditional models of human spatial behavior.

7.2.2. To Regions and Regionalization

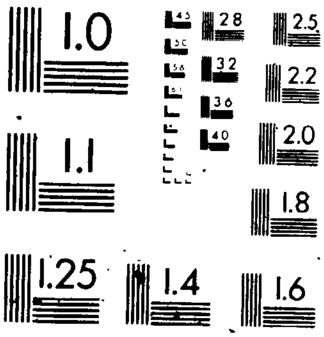
The Black/White model uncovers new horizons in the *understanding* of functional regions. First, it offers an alternative conceptual and numeric model, in place of what has until now been a nebulous subjective notion of communities and functional complementarity. Although the model is deductive, and relies on geographical and behavioral theory, it has a well established empirical precedent in the work of Mackay.

As a planning tool, the Black/White model would have to fulfill a predictive role, and to that end, it should be demonstrated that the model can be calibrated for a given set of observed data, and that the parameters are portable to situations where data are not available. The first of these requirements has been fulfilled by the development of a viable calibration method; in regard to the second requirement, parameters measuring regional affiliation may be generalized to neighboring parcels of space, but such extrapolations are necessarily to be employed with caution.

The calibration method developed in this research makes more significant contributions. First, given the fresh conceptual framework afforded by the

3 OF/DE 3

MICROCOPY RESOLUTION TEST CHART
NBS 1010a
ANSI and ISO TEST CHART No. 2



model, it is now possible to *simulate* interactions across functional regions, and hence to *verify* the outcome of a proposed calibration procedure, or in general any regionalization process, against the simulated structure. Secondly, as a result of this ability to simulate data, this research has introduced a means of *evaluating* alternative methods for calibration under various scenarios, based on the ability of the method to deliver the correct solution under the stress of measured error. Such a system of "accreditation" of algorithms is particularly appropriate when evaluating methods for regionalization, where the permutations are so numerous that reliance must often be placed on heuristics.

It may be argued that the Black/White model is based on the gravity model, and does not reflect more recent thinking on modeling spatial interaction (for example, Fotheringham, 1983). Further, Wilson's theoretical derivation of the spatial interaction model, which has conferred some respectability on the gravity model (on which the Black/White model appears to be based), assumes binding constraints on the marginal totals of the interaction matrix, whereas in reality, it is rare that both marginal vectors would constitute constraints on interactions. The theoretical foundations of the interaction model, and thence of the Black/White model, are thus called into question. In rejecting those arguments, it may be pointed out that the Black/White model has been derived primarily on the grounds of the Luce and Tversky choice axioms, with incidental reference to the most generic statement of the spatial interaction model. The calibration procedure for the Black/White model assumes that the basic form of an interaction model is accepted as a valid null hypothesis for interaction in a given environment, and can be calibrated satisfactorily. The regional partitioning process proceeds from the ratio of observed interactions to those expected under that prior model.

There are inadequacies in the capabilities of the methodology as it now exists, and operational and interpretational hazards of which one must be aware. The most obvious shortcoming of the calibration procedures is the lack of documented ability to handle multiple regions. Although the applications of two-way splits are numerous, there remains a compelling need to identify more complex structures. It seems reasonable to expect that advances in computer technology will make further investigations along these lines practically feasible in the foreseeable future.

The algorithms developed for two-way regionalization could probably be improved. Although the heuristic developed in Chapter 5, and enhanced in Chapter 6, has performed well on 10 × 10 simulations, it needs to be tested more extensively, and on larger problems. With faster processors, it should be possible to incorporate more complex operations in the heuristic.

In practical applications, the conclusions drawn by a researcher are only as good as the data. The temptation to draw conclusions regarding the regional structure of the United States in Chapter 6, and to search for explanations for the unexpected outcome, was resisted. It is discouraging, after processing large volumes of data, to arrive at a result that defies intuition, or that is not significant enough to bear positive interpretation; on the other hand, it is ethically unacceptable to draw artificial inferences under such circumstances. There is a clear role for further research on the significance of solutions, although the theoretical problems with null hypotheses have been noted.

Interaction matrices themselves may be a poor representation of human contacts. Apart from the problems posed by diagonal entries, it is not clear how the matrices should represent interactions of the conference type, that is, meetings set up between two parties at a third place. One has to be aware of the resultant measurement errors.

It appears that, with more exhaustive behchmark testing, and experimentation with alternative objective functions, it may be possible to develop an operation manual for the proper and effective application of the technique in realistic situations, such as political and planning applications. While such a chore is beyond the scope of this thesis, it is satisfying that the research has produced a methodological innovation with immediate tangible application potential.

7.2.3. To Geography!

The final contribution of this research has been to break down the barriers between opposing arms of geography. The diversity of the discipline is a source and sign of strength and resilience; the human behavioral dimensions of this diversity unfortunately act against the better interests of the discipline.

It is hoped that this research, in attempting to explore and build upon common interests, will contribute to the process of reconciliation and constructive dialog between regional and quantitative geography.

Appendix A

Notation

TERM	SUBSCRIPTS	DESCRIPTION	INTRODUCED IN CHAPTER
A	1	Attraction of destination	3
A*	1	Dummy attraction vector (binary)	5
C		Convergence ratio in seriation	5
D	1	Marginal (column) total of **I**	3
E	1	Emissivity of origin	3
E*	1	Dummy emissivity vector (binary)	5
F	2	Force of attraction	3
G		Gravitational constant	3
I	2	Interaction matrix	2
J	2	Working matrix	2
K		Balancing factor	3
*K**		Balancing factor	3
L	2	ln **I**	5
M	1	Mass of body (Newton)	3
O	1	Marginal (row) total of **I**	3
P	1	Population	3
Q	2	Energy	3
R	2	Compatibility factor	4
R*	2	Effective compatibility (Tversky)	5
S		Choice set	3
U	1	Balancing vector	3

$\mathbf{V}$	1	Balancing vector	3
W		Objective function	5
Z		Objective function	5
$\mathbf{a}$	2	Choice subsets (Tversky)	4
b		Gradient (generic)	5
b^*		Working gradient	5
$b_1 - b_5$		Search parameters (Algorithm B)	5
$\mathbf{c}$	2	Cost factor	4
$\mathbf{d}$	2	Distance matrix	3
$\mathbf{d}^*$	2	Functional distance matrix	4
e		Naperian logarithm base	3
f		Function of distance	3
g		Function of distance and compatibility	4
g		Temporary dummy subscript	5
h		Temporary dummy subscript	5
i		Origin identity	2
j		Destination identity	2
k		Basis identity	4
l		Line/scatter identity	5
m		Temporary dummy subscript	3
n		Number of ESU's	2
p		Probability	3
r		Number of bases of differentiation	4
s		Distance from home core	4
t		Time	3
$\mathbf{u}$	1	Origin region	4
$\mathbf{v}$	1	Destination region	4
$\mathbf{w}$	1	Weight of origin	3
x		Abscissa (generic)	4
$\mathbf{x}$	1	Data vector (generic)	5
y		Ordinate (generic)	4

y	2	Ordinate on parallel scatter	5
z	2	Intercept of parallel scatter	5
Δ		Residual (prefix)	5
Λ	1	Available resources	3
$\prod$		Product	4
$\sum$		Sum	2
β		Friction of distance	3
γ	1	Balancing factor	3
δ	3	Line/scatter membership (binary)	5
ϵ		Error multiple	5
θ	2	Random deviate	5
κ		Number of scatters	4
λ	1	Balancing factor	3
μ	1	Membership vector	4
ξ	1	Emissivity (Huff)	3
ξ^*	1	Emissivity (Luce)	4
ρ		Universal compatibility factor	4
σ	2	Dissimilarity matrix (binary)	4
ϕ		Fuzziness factor	4

Primes (′) as postscripts indicate estimated values

References

ACHABAL D D, GORR W L, MAHAJAN V, 1982: "Multiloc: a multiple store location decision model" *Journal of Retailing* **58** (2) 5-25

AGS, 1919: "The American Geographical Society's contribution to the Peace Conference" *Geographical Review* **7** 1-10

ALLIX A, 1922: "The geography of fairs" *Geographical Review* **12** 532-69

ANDERSON T R, 1955: "Intermetropolitan migration: a comparison of the hypotheses of Zipf and Stouffer" *American Sociological Review* **20** (3) 287-91

ANDERSON T R, 1956: "Potential models and spatial distribution of population" *Papers and Proceedings, Regional Science Association* **2** 175-94

ANDO A, FISHER F M, 1963: "Near decomposability, partition and aggregation and the relevance of stability discussions" *International Economic Review* **4** 53-67

ARABIE P, 1978: "Constructing block models -- how and why" *Journal of Mathematical Psychology* **17** 21-63

AT&T, 1980: *Notes on the Network. Section 2: Numbering Plan and Dialing Procedures* (American Telephone and Telegraph Company)

BACHARACH M, 1970: *Biproportional Matrices and Input-Output Change* (Cambridge University Press)

BAKER R J, NELDER J A, 1978: *The GLIM System: Generalized Linear Interactive Modeling. Release 3* (Numerical Algorithms Group, Royal Statistical Society)

BATTY M, 1971: "Exploratory calibration of a retail location model using search by golden section" *Environment and Planning A* **3** 411-32

BATTY M, MACKIE S, 1972: "The calibration of gravity, entropy and related models of interaction" *Environment and Planning A* **4** 205-233

BATTY M, SAMMONS R, 1978: "On searching for the most informative spatial pattern" *Environment and Planning A* **10** 747-79

BAXTER M J, 1979: "The application of logit regression analysis to production constrained trip distribution models" *Journal of Regional Science* **19** (2) 171-78

BAXTER M J, EWING G O, 1984: "A framework for the exploratory development of recreational trip distribution models" *Mimeopaper, McGill University*

BEARDWOOD J E, KIRBY H R, 1975: "Zone definition and the gravity model: the separability, excludability and compressibility properties" *Transportation Research* **9** 363-69

BEAVON K S O, 1977: *Central Place Theory: a Reinterpretation* (Longmans, London)

BERRY B J L, 1964: "Approaches to regional analysis: a synthesis" *Annals, Association of American Geographers* **54** 2-11

BERRY B J L, 1968: "A synthesis of formal and functional regions using a general field theory of spatial behavior" in Berry B J L, Marble D F (Eds) *Spatial Analysis in Geography* (Prentice Hall, Englewood Cliffs)

BERRY B J L, REES P H, 1969: "The factorial ecology of Calcutta" *American Journal of Sociology* **74** 445-91

BEUM C O, BRUNDAGE E G, 1950: "A method for analyzing the socio-matrix" *Sociometry* **13** 141-145

BLAU P M, DUNCAN O D, 1967: *The American Occupational Structure* (Wiley, New York)

BOGGS S W, 1940: *International Boundaries: a Study of Boundary Functions and Problems* (Columbia University Press, New York)

BORGATTA E F, STOLZ W, 1963: "A note on a computer program for rearrangement of matrices" *Sociometry* **26** 391-92

BOSSARD J H S, 1932: "Residential propinquity as a factor in marriage selection" *American Journal of Sociology* **38** (2) 219-44

BOWMAN I, 1928: *The New World: Problems in Political Geography* (World Book Co, Yonkers on Hudson, New York)

BRAMS S J, 1966: "Transaction flows in the international system" *American Political Science Review* **60** 880-98

BREIGER R L, BOORMAN S A, ARABIE P, 1975: "An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling" *Journal of Mathematical Psychology* **12** 328-83

BRIGHAM A P, 1919: "Principles in the determination of boundaries" *Geographical Review* **7** (4) 201-209

BROOKFIELD H C, 1962: "Local study and comparative method: an example from central New Guinea" *Annals, Association of American Geographers* **52** (3) 242-54

BROWN L A, 1970: "On the use of Markov chains in movement research" *Economic Geography* **46** 393-403

BROWN L A, HOLMES J H, 1971: "The delimitation of functional regions, nodal regions and hierarchies by functional distance approaches" *Journal of Regional Science* **11** 57-72

BROWN L A, HORTON F E, 1970: "Functional distance - an operational approach" *Geographical Analysis* **2** 76-82

BROWN L A, ODLAND J, GOLLEDGE R G, 1970: "Migration, functional distance and the urban hierarchy" *Economic Geography* **46** 472-79

BUCKLIN L P, 1971: "Retail gravity models and consumer choice: a theoretical and empirical critique" *Economic Geography* **47** 489-98

BUDEL J, 1944: "Die morphologischen wirkungen des eiszeitklimas in gletscherfrein gebiet" *Geologische Rundschau* **34** 482-519

BUDEL J, 1948: "Die klima morphologischen zonen der polarlander" *Erdkunde* **2** 25-53

BUNGE W, 1962: "Theoretical geography" *Lund Studies in Geography, Series C, General and Mathematical Geography, No 1* (Gleerup, Lund)

BUNGE W, 1966: "Theoretical geography" *Lund Studies in Geography, Series C, General and Mathematical Geography, No 1.* (Gleerup, Lund). Second Revised and Enlarged Edition

BURTON I, 1963: "The quantitative revolution and theoretical geography" *Canadian Geographer* **7** (4) 151-62

CAREY H C, 1858: *Principles of Social Science, Vol 1* (Lippincott, Philadelphia)

CARRINGTON P J, HEIL G H, 1979: "COBLOC: a hierarchical method for blocking network data" *Structural Analysis Project Report, Department of Sociology, University of Toronto*

CARROLL J D, 1955: "Spatial interaction and the urban metropolitan description" *Traffic Quarterly* **9** (2) 149-61

CARROLL J D, BEVIS H W, 1957: "Predicting local travel in urban regions" *Papers and Proceedings, Regional Science Association* **3** 183-197

CAVANAUGH J A, 1950: "Formulation, analysis and testing of the interactance hypothesis" *American Sociological Review* **15** 763-66

CESARIO F J, 1974: "The interpretation and calculation of gravity model zone to zone adjustment factors" *Environment and Planning A* **6** 247-57

CHRISTALLER W, 1933: *Die Zentralen Orte in Suddeutschland: Eine Okonomisch-geographische Untersuchung Uber Die Gesetzmassigkeit Der Verbreitung Und Entwicklung Der Siedlungen Mit Stadtischen Funktionen* (Jena)

CLARK D, 1973: "The formal and functional structure of Wales" *Annals, Association of American Geographers* **63** 29-42

CLARK W A V, AVERY K L, 1976: "The effects of data aggregation in statistical analysis" *Geographical Analysis* **8** 428-38

CLARK W A V, AVERY K L, 1978: "Aggregation and statistical analysis: a comment" *Geographical Analysis* **10** 194-95

CLAYTON C, 1974: "Communications and spatial structure" *Tijdschrift Voor Economische En Sociale Geografie* **65** 221-27

CLAYTON C, 1977a: "Interstate population migration process and structure in the United States, 1935 to 1970" *Professional Geographer* **29** 177-81

CLAYTON C, 1977b: "The structure of interstate and interregional migration 1965 to 1970" *Annals of Regional Science* **11** 109-22

CLIFF A D, HAGGETT P, 1970: "On the efficiency of alternative aggregations in region building problems" *Environment and Planning A* **2** 285-91

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

CLIFF A D, MARTIN R L, ORD J K, 1974: "Evaluating friction of distance parameters in gravity models" *Regional Studies* **8** 281-86

CLIFF A D, MARTIN R L, ORD J K, 1975: "Map pattern and friction of distance" *Regional Studies* **9** 285-88

CLIFF A D, MARTIN R L, ORD J K, 1976: "A reply to the final comment" *Regional Studies* **10** 341-42

CLIFF A D, ORD J K, 1973: *Spatial Autocorrelation* (Pion, London

COHEN J E, 1966: "A model of simple competition" *Annals, Computation Laboratory of Harvard University* **41** Harvard University Press

COLE L C, 1949: "The measurement of interspecific association" *Ecology* **30** 411-24

COMPUSEARCH, 1980: *Tradarea Update No 8, April 1* (Compusearch Market and Social Research Limited, Toronto)

COOK L A, 1945: "An experimental sociographic study of a stratified 10th grade class" *American Sociological Review* **10** 250-61

COOPER L, 1972: "The transportation-location problem" *Operations Research* **20** 1-94

COOPER L, STEINBERG D, 1970: *Methods of Optimization* (W B Saunders, Philadelphia)

COURT A, 1966: "Population distributions and self potentials" *Seminar Series on Quantitative Geography, University of Michigan, Ann Arbor*

CROWE P R, 1938: "On progress in geography" *Scottish Geographical Magazine* **54** 1-19

CURRY L, 1972: "A spatial analysis of gravity flows" *Regional Studies* **6** 131-47

CURRY L, GRIFFITH D A, SHEPPARD E S, 1975: "Those gravity parameters again" *Regional Studies* **9** 289-96

DAVIES W K D, 1966: "The ranking of service centers: a critical review" *Transactions, Institute of British Geographers* **40** 51-65

DAVIES W K D, 1980: "Higher order factor analysis and functional regionalization: a case study in south Wales, 1971" *Environment and Planning A* **12** 685-701

DAVIES W K D, THOMPSON R R, 1980: "The structure of interurban connectivity: a dyadic factor analysis of Prairie commodity flows" *Regional Studies* **14** 297-311

DEMANGEON A, 1939: *The British Isles* (Heinemann)

DICKINSON R E, 1947: *City, Region and Regionalism: a Geographical Contribution to Human Ecology* (Routledge and Kegan Paul, London)

DICKINSON R E, 1964: *City and Region — a Geographical Interpretation* (Routledge and Kegan Paul, London)

DICKINSON R E, 1969: *The Makers of Modern Geography* (Routledge and Kegan Paul, London)

DICKINSON R E, HOWARTH O J R, 1933: *The Making of Geography* (Clarendon Press, Oxford)

DIGGLE P J, MATERN B, 1980: "On sampling designs for the study of point-event nearest neighbor distributions in *r* squared" *Scandinavian Journal of Statistics* **7** 80-84

DODD S C, 1950: "The interactance hypothesis: a model fitting physical masses and human groups" *American Sociological Review* **15** 245-57

DUTTON G, KIERNAN K, KINGSBURY D, WARNTZ W, 1971: "Tabulations of data on area, population, income and certain derived quantities for the 3070 counties of the 48 coterminous states of the United States, 1967" *Harvard Papers in Theoretical Geography, Geography of Income Series, No 2* (Cambridge, Mass)

EDWARDS A W F, CAVALLI-SFORZA L L, 1965: "A method for cluster analysis" *Biometrics* **21** 362-75

ELLIS J B, 1967: *A Systems Model for Recreational Travel in Ontario, a Progress Report* (Department of Electrical Engineering, University of Waterloo, Project W6)

ELLIS J B, VAN DOREN, 1966: "A comparative evaluation of gravity and system theory models for statewide recreational traffic flows" *Regional Science* **6** (2) 57-70

EVANS A W, 1971: "Calibration of trip distribution models with exponential or similar cost functions" *Transportation Research* **5** 15-38

EWING G O, 1982: "Using GLIM for exploratory modeling of camper trip distributions" *Association of American Geographers, Annual Meetings, San Antonio*

EWING G, 1984: "Gravity models and behavioral inference" *Mimeopaper, McGill University*

FENNEMAN N M, 1917: "Physiographic divisions of the United States" *Annals, Association of American Geographers* **6** 19-98

FIENBERG S E, 1970: "An iterative procedure for estimation in contingency tables" *Annals of Mathematical Statistics* **41** 907-17

FLETCHER R, POWELL M J D, 1963: "A rapidly convergent descent method for minimization" *Computer Journal* **6**

FORDE C D, 1934: *Habitat, Economy and Society* (Methuen, London)

FORSYTH E, KATZ L, 1946: "A matric approach to the analysis of sociometric data - preliminary report" *Sociometry* **9** 340-47

FOTHERINGHAM A S, 1983: "A new set of spatial interaction models: the theory of competing destinations" *Environment and Planning* **A 15** 15-36

GALE N, HUBERT L J, TOBLER W R, GOLLEDGE R G, 1982: "Combinatorial procedures for the analysis of alternate models: an example from interregional migration" *Regional Science Association, 29th North American Meetings, Pittsburgh*

GARFINKEL R S, NEMHAUSER G L, 1970: "Optimal political districting by implicit enumeration techniques" *Management Science* **16** B495-B508

GARRISON W L, MARBLE D F, 1963: "Factor analytic study of the connectivity of a transportation network" *Papers and Proceedings, Regional Science Association* **12** 231-38

GETIS A, 1963: "The determination of the location of retail activities with the use of a map transformation" *Economic Geography* **39** 1-22

GILBERT E W, 1951: "Geography and regionalism" in Taylor T G (Ed) *Geography in the 20th Century* (Methuen, London)

GODDARD J B, 1970: "Functional regions within the city center - a factor analytic study of taxi flows in central London" *Transactions, Institute of British Geographers* **49** 162-82

GOLLEDGE R G *et al*, 1982: "Commentary on 'The highest form of the geographer's art'" *Annals, Association of American Geographers* **72** 557-58

GOLLEDGE R G, RUSHTON G (Eds), 1976: *Spatial Choice and Spatial Behavior* (Ohio State University Press, Columbus)

GOODCHILD M F, 1981: "Regionalization and aggregation from interactions and flows" *Association of American Geographers, Annual Meetings, Los Angeles*

GOODCHILD M F, BOOTH P J, 1980: "Location and allocation of recreation facilities: public swimming pools in London, Ontario" *Ontario Geography* (15) 35-52

GOODCHILD M F, HOSAGE C M, 1983: "On enumerating all feasible solutions to polygon aggregation problems" *Modeling and Simulation* **14** Part 3

GOODCHILD M F, KWAN M Y C, 1978: "Models of hierarchically dominated spatial interaction" *Environment and Planning A* **10** 1307-17

GOODCHILD M F, MASSAM B H, 1969: "Some least cost models of spatial administrative systems in southern Ontario" *Geografiska Annaler* **52** (B) 86-94

GOODCHILD M F, NORONHA V T, 1983: *Location Allocation for Small Computers* (Monograph 8, Department of Geography, University of Iowa)

GOODCHILD M F, SMITH T R, 1980: "Intransitivity, the spatial interaction model, and U S migration streams" *Environment and Planning A* **12** 1131-44

GOODMAN L, 1963: "Statistical methods for the preliminary analysis of transaction flows" *Econometrica* **31** 197-208

GOTTMANN J, 1961: *Megalopolis* (MIT Press, Cambridge, Mass)

GOULD P R, 1972a: "Pedagogic review: entropy in urban and regional modeling" *Annals, Association of American Geographers*, **62** 689-700

GOULD P R, 1972b: "The black boxes of Jonkoping" in Downs R M, Stea D (Eds) *Cognitive Mapping: Images of Spatial Environment* (Aldine, Chicago)

GOULD P R, 1980: "Q-analysis or a language of structure: an introduction for social scientists, geographers and planners" *International Journal of Man-machine Studies* **12** 169-99

GREENOUGH J, 1919: "Geographical record" *Geographical Review* **7** 110-111

GRIESINGER D W, 1979: "Reconsidering the theory of social gravity" *Journal of Regional Science* **19** (3) 291-302

GRIGG D B, 1965: "The logic of regional systems" *Annals, Association of American Geographers* **55** 465-91

GRIGG D B, 1967: "Regions, models and classes" in Chorley R J, Haggett P (Eds) *Integrated Models in Geography* (Methuen, London)

HAGERSTRAND T, 1957: "Migration and area: survey of a sample of Swedish migration fields and hypothetical considerations of their genesis" *Lund Studies in Geography, Series B, Human Geography, No. 13* 27-158

HAGGETT P, 1964: "Regional and local components in the distribution of forests in southeast Brazil — a multivariate analysis" *Geographical Journal* **130** 365-80

HAGGETT P, 1965: "Scale components in geographical problems" in Chorley R J, Haggett P (Eds) *Frontiers in Geographical Teaching* (Methuen, London)

HAGGETT P, CLIFF A D, FREY A, 1977: *Locational Analysis in Human Geography* (Edward Arnold, London)

HARARY F, NORMAN R Z, CARTWRIGHT D, 1965: *Structural models* (Wiley, New York)

HARRIS C D, 1954: "The market as a factor in the localization of industry in the United States" *Annals, Association of American Geographers* **44** (4) 315-48

HART J F, 1982: "The highest form of the geographer's art" *Annals, Association of American Geographers* **72** 1-29

HARTSHORNE R, 1939: *The Nature of Geography* (Association of American Geographers, Lancaster, Pa)

HAYNES R M, 1975: "Dimensional analysis: some applications in human geography" *Geographical Analysis* **7** 51-68

HERBERTSON A J, 1905: "The major natural regions: an essay in systematic geography" *Geographical Journal* **25** 300-12

HIMMELBLAU D M, 1973: *Decomposition of Large Scale Problems* (North Holland, Amsterdam)

HIRST M A, 1975: "Telephone transactions, regional inequality and urban growth in East Africa" *Tijdschrift Voor Economische En Sociale Geografie* **66** 277-93

HIRST M A, 1977: "Hierarchical aggregation procedures for interaction data — a comment" *Environment and Planning, A* **9** 99-103

HODGSON M J, 1978: "Towards more realistic allocation in location-allocation models" *Environment and Planning A* **10** 1273-86

HOLLINGSWORTH T M, 1971: "Gross migration flows as a basis for regional definition — an experiment with Scottish data" *Proceedings, IUSSP Conference, London, 1969* **4** 2755-65

HOLMES J H, 1977: "Hierarchical regionalization by iterative proportional fitting procedures — a comment" *IEEE, Systems Man and Cybernetics* **7** 474-77

HOLMES J H, 1978: "Transformation of flow matrices to eliminate the effects of differing sizes of origin-destination units — a further comment" *IEEE, Systems Man and Cybernetics* **8** 325-32

HOLZ R K, HUFF D L, MAYFIELD R C, 1973: "Urban spatial structure based on remote sensing imagery" in Holz R K (Ed) *The Surveillant Science* (Houghton Mifflin, Boston)

HOOKE R, JEEVES T A, 1961: "Direct search solution of numerical and statistical problems" *Journal of ACM* **8**

HOTELLING H, 1929: "Stability in competition" *Economic Journal* **39** 41-57

HUFF D L, 1963: "A probabilistic analysis of shopping center trade areas" *Land Economics* **39** 81-90

IKLÉ F C, 1954: "Sociological relationship of traffic to population and distance" *Traffic Quarterly* **8** 123-36

ILLERIS S, PEDERSEN P O, 1968: "Central places and functional regions in Denmark — factor analysis of telephone traffic" *Lund Studies in Geography, Series B, No 31* (Gleerup, Lund)

IMSL, 1975: *IMSL Library Reference Manual* (International Mathematical Subroutine Libraries Inc, Houston, Texas)

INGRAM D R, 1971: "Concept of accessibility: search for an operational form" *Regional Studies* **5** 101-107

ISARD W, PECK M J, 1954: "Location theory and international and interregional trade theory" *Quarterly Journal of Economics* **68** 97-114

JENKS G F, KNOS D, 1961: "The use of shading patterns in graded series" *Annals, Association of American Geographers* **51** 316-34

JOHNSON S C, 1967: "Hierarchical clustering schemes" *Psychometrika* **32** 241-54

JOHNSTON R J, 1968: "Choice in classification: the subjectivity of objective methods" *Annals, Association of American Geographers* **58** 575-89

JOHNSTON R J, 1975: "Map patterns and friction of distance parameters: comment" *Regional Studies* **9** 281-83

JOHNSTON R J, SCOTT P, 1965: "The measurement of a hierarchy of central places" *Australian Geographer* **10** 315-17

JONES F L, 1969: *Dimension of Urban Structure* (Australian National University Press, Canberra)

KALESNIK S V, 1961: "The present state of landscape studies" *Soviet Geography* **2** (2) 24-34

KANT E, 1951: "Umland studies and sector analysis" *Studies in Rural-urban Interaction. Lund Studies in Geography, Series B* (Gleerup, Lund)

KATZ L, 1947: "On the matrix analysis of sociometric data" *Sociometry* **10** 233-41

KAU J, SIRMONS C, 1979: "The functional form of the gravity model" *International Regional Science Review* **4** 127-36

KELLEY P S, 1983: "Digital analysis in remote sensing: from data to information" in Richason B F (Ed) *Introduction to Remote Sensing of the Environment* (Kendall and Hunt, Dubuque, Iowa) Second Edition

KELLY J, 1982: "The zigzag art of politics" *Time* **119** (11) 15-16

KENDALL M G, 1939: "The geographical distribution of crop productivity in England" *Journal of the Royal Statistical Society* **102** 21-62

KING L C, 1953: "Canons of landscape evolution" *Bulletin, Geological Society of America* **64** 721-52

KIRBY H R, 1974: "Theoretical requirements for calibrating gravity models" *Transportation Research* **8** 97-107

KIRBY H R, LEESE M N, 1978: "Trip distribution calculations and sampling error — some theoretical aspects" *Environment and Planning A* **10** 837-51

KOPPEN W, 1884: "Die warmezonen der erde, nach der dauer der heissen, und kalten zeit und nach der wirkung der warme auf die organische welt betrachtet" *Metereologische Zeitschrift* **1** 215-226

KRUMBEIN W C, 1959: "Trend surface analysis of contour type maps with irregular control point spacing" *Journal of Geophysical Research* **64** 823-34

KRUMBEIN W C, 1960: "The geological population as a framework for analyzing numerical data in geology" *Living and Man Geological Journal* **2** 341-68

KURTZ R A, EICHER J B, 1958: "Fringe and suburb — a confusion of concepts" *Social Forces* **37** 32-37

LEONG G C, MORGAN G C, 1973: *Human and Economic Geography* (Oxford, Hong Kong)

LEUNG Y, 1984: "Towards a flexible framework for regionalization" *Environment and Planning A* **16** (12) 1613-32

LING R F, 1973: "A computer generated aid for cluster analysis" *Communications of the ACM* **16** 355-61

LUCE R D, 1959: *Individual Choice Behavior: a Theoretical Analysis* (Wiley, New York)

LUCE R D, PERRY D, 1949: "A method of matrix analysis of group structure" *Psychometrika* **14** 95-116

MACGILL S M, 1977: "Theoretical properties of biproportional matrix adjustments" *Environment and Planning A* **9** 1033-42

MACKAY J R, 1958: "The interactance hypothesis and boundaries in Canada : a preliminary study" *Canadian Geographer* **11** 1-8

MARANZANA F E, 1964: "On the location of supply points to minimize transport costs" *Operations Research Quarterly* **15** 261-70

MARTIN G J, 1980: *The Life and Thought of Isaiah Bowman* (Archon Books, Hamden, Ct)

MASSER I, 1977: "A comparative analysis of spatial representation in doubly constrained interaction models" *Environment and Planning A* **9** 759-69

MASSER I, BROWN P J B (Eds), 1978: *Spatial Representation and Spatial Interaction* (Martinus Nijhoff, Leiden, Holland)

MASSER I, BROWN P J B, 1975: "Hierarchical aggregation procedures for interaction data" *Environment and Planning A* **7** 509-23

MASSER I, SCHEURWATER J S, 1980: "Functional regionalization of spatial interaction data — an evaluation of some suggested strategies" *Environment and Planning A* **12** 1357-82

MATHUR V K, 1970: "An economic derivation of the gravity law of spatial interaction: a comment" *Journal of Regional Science* **10** 403-405

MCCONNELL J E, 1967: "The Middle East — competitive or complementary" *Tijdschrift Voor Economische En Sociale Geografie* **58** 82-93

MCFADDEN D, 1973: *Conditional Logit Analysis of Qualitative Choice Behavior* (University of California, Berkeley)

MCFADDEN D, REID F, 1974: *Aggregate Travel Demand Forecasting From Disaggregated Behavioral Models* (Institute of Urban and Regional Development, University of California, Berkeley)

MCQUITTY L L, 1968: "Multiple clusters, types and dimension from iterative intercolumnar correlational analysis" *Multivariate Behavioral Research* **3** 465-78

MCQUITTY L L, CLARK J A, 1968: "Clusters from iterative intercolumnar correlational analysis" *Educational and Psychological Measurement* **28** 211-38

MINSHULL R, 1967: *Regional Geography: Theory and Practise* (Hutchinson, London)

MULLER J C, 1983: "Geographic seriation revisited" *Professional Geographer* **35** (2) 196-202

MURPHY R E, 1968: "World structural regions" *Annals, Association of American Geographers* **58** (1) Map Supplement

MURPHY R E, 1972: *The Central Business District* (Aldine, Chicago)

NADER G A, 1981: "Delineation of a hierarchy of nodal regions by means of a higher order factor analysis" *Regional Studies* **15** (6) 475-92

NAGEL S S, 1965: "Simplified bipartisan computer redistricting" *Stanford Law Review* **17** 863-99

NAKANISHI M, COOPER L G, 1974: "Parameter estimation for a multiplicative competitive interaction model — least squares approach" *Journal of Marketing Research* **11** 303-11

NG R C Y, 1969: "Recent internal migration in Thailand" *Annals, Association of American Geographers* **59** 710-30

NIE N H, HULL C H, JENKINS J G, STEINBRENNER K, BENT D H, 1975: *Statistical Package for the Social Sciences* (Mcgraw Hill, New York)

NIEDERCORN J H, BECHDOLT B V, 1969: "An economic derivation of the gravity law of spatial interaction" *Journal of Regional Science* **9** (2) 273-82

NIEDERCORN J H, BECHDOLT B V, 1970: "An economic derivation of the gravity law of spatial interaction: reply" *Journal of Regional Science* **10** (3) 407-10

NIEDERCORN J H, BECHDOLT B V, 1972: "An economic derivation of the gravity law of spatial interaction: a further reply and a reformulation" *Journal of Regional Science* **12** (1) 127-36

NYSTUEN J D, DACEY M F, 1961: "A graph theory interpretation of nodal regions" *Papers and Proceedings, Regional Science Association* **7** 29-42

O'SULLIVAN P, 1977: "The spread of traffic effects and regions of a network" *Environment and Planning A* **9** 917-26

O'SULLIVAN P, 1978: "Regions of a transport network" *Annals, Association of American Geographers* **68** 396-401

ODUM H W, MOORE H E, 1938: *American Regionalism: a Cultural-historical Approach to National Integration* (Henry Holt, New York)

OPENSHAW S, 1976: "An empirical study of some spatial interaction models" *Environment and Planning A* **8** 23-41

OPENSHAW S, 1977a: "Algorithm 3 — a procedure to generate pseudo random aggregations" *Environment and Planning A* **9** 1423-29

OPENSHAW S, 1977b: "Optimal zoning systems for spatial interaction models" *Environment and Planning A* **9** 169-84

OPENSHAW S, 1978: "An optimal zoning approach to the study of spatially aggregated data" in Masser I, Brown P J B (Eds) *Spatial Representation and Spatial Interaction* (Martinus Nighoff, Leiden, Holland)

OPENSHAW S, CONNOLLY C J, 1977: "Empirically derived deterrence functions for maximum performance spatial interaction models" *Environment and Planning A* **9** 1067-79

OPENSHAW S, TAYLOR P J, 1979: "A million or so correlation coefficients — three experiments on the modifiable areal unit problem" in Wrigley N (Ed) *Statistical Methods in the Spatial Sciences* (Pion, London)

PELTIER L, 1950: "The geographic cycle in periglacial regions as it is related to climatic geomorphology" *Annals, Association of American Geographers* **40** 214-36

PHILBRICK A K, 1957: "Principles of areal functional organization in regional human geography" *Economic Geography* **33** 299-336

PIRIE G H, 1979: "Measuring accessibility — review and proposal" *Environment and Planning A* **11** 299-312

PLATT R S, 1928: "A detail of regional geography: Ellison Bay community as an industrial organism" *Annals, Association of American Geographers* **18** (2) 81-126

PLATT R S, 1935: "Field approach to regions" *Annals, Association of American Geographers* **25** 153-74

POCOCK D C D, WISHART D, 1969: "Methods of deriving multi-factor uniform regions" *Transactions, Institute of British Geographers* **47** 73-98

POKSHISHEVSKIY V V, 1975: "On the Soviet concept of economic regionalization" *Progress in Geography* **7** 1-52

PORTEOUS J D, 1985: "Why is geography so dull and boring?" *The Operational Geographer* (6) 48-50

PRICE D O, 1948: "Distance and direction as vectors of internal migration, 1935-40" *Social Forces* **27** (1) 48-53

PRYOR R J, 1968: "Defining the rural urban fringe" *Social Forces* **47** 202-15

RAVENSTEIN E G, 1885: "The laws of migration" *Journal of the Royal Statistical Society* **48** 167-235

RAVENSTEIN E G, 1889: "The laws of migration" *Journal of the Royal Statistical Society* **52** 241-305

REES P H, 1972: "Problems of classifying sub-areas within cities" in Berry B J L (Ed) *City Classification Handbook: Methods and Applications* (Wiley, New York)

REILLY W J, 1929: "Methods for the study of retail relationships" *University of Texas Bulletin, 2944*

ROGERS A, 1976: "Shrinking large scale population projection models by aggregation and decomposition" *Environment and Planning A* **8** 515-41

ROISTACHER R C, 1974: "Review of mathematical models in sociometry" *Sociological Methods and Research* **3** 123-71

ROSE D J, WILLOUGHBY R A, 1972: *Sparse Matrices and Their Applications* (Plenum Press, New York)

RUSSETT B M, 1967: *International Regions and the International System* (Rand McNally, Chicago)

SAVAGE R I, DEUTSCH K W, 1960: "A statistical model of the gross analysis of transaction flows" *Econometrica* **28** 551-72

SEMPLE R K, CASSETTI E, KING L J, 1969: "Determination of the optimum number of groupings in classification problems" *Discussion Paper #10, Ohio State University*

SEMPLE R K, YOUNGMANN C E, ZELLER R E, 1972: "Economic regionalization and information theory : an Ohio example" *Discussion Paper #28, Ohio State University*

SHANNON L E, 1948: "A mathematical theory of communication" *Bell System Technical Journal* **27** 379-423, 623-656

SHARPE C F S, 1938: *Landslides and Related Phenomena: a Study of Mass Movement of Soil and Rock* (Columbia University Press, New York)

SHEPPARD E S, 1978: "Theoretical underpinnings of the gravity hypothesis" *Geographical Analysis* **10** 386-402

SHEPPARD E S, GRIFFITH D A, CURRY L, 1976: "A final comment on mis-specification and autocorrelation in those gravity parameters" *Regional Studies* **10** 337-39

SIMPSON D, TSUKUI J, 1965: "The fundamental structure of input output tables: an international comparison" *Review of Economics and Statistics* **47** 434-46

SLATER P B, 1976: "A hierarchical regionalization of Japanese prefectures" *Regional Studies* **10** 123-32

SMITH T R, 1984: "Artificial intelligence and its applicability to geographical problem solving" *Professional Geographer* **36** (2) 147-57

SOLNTSEV N A, 1962: "Basic problems in Soviet landscape science" *Soviet Geography* **3** (6) 3-15

SOUTHWORTH F, 1981: "Calibration of multinomial logit models of mode and destination choice" *Transportation Research* **15A** 315-25

SPENCE N A, TAYLOR P J, 1970: "Quantitative methods on regional taxonomy" *Progress in Geography* **2** 1-64

SPRENT P, 1969: *Models in Regression and Related Topics* (Methuen, London)

STATISTICS CANADA, 1982: *Products and Services of the 1981 Census of Canada* (Statistics Canada, Ottawa)

STEPHENSON L K, 1974: "On functional regions and indirect flows" *Geographical Analysis* **6** 383-85

STETZER F, 1976: "Parameter estimation for constrained gravity model: a comparison of six methods" *Environment and Planning* **8** 673-83

STEVENSON E L, 1932: *The Geography of Claudius Ptolemy* (New York Public Library, New York)

STEWART J Q, 1941: "The gravitation, or geographical drawing power, of a college" *Bulletin, American Association of University Professors* **27** 70-75

STEWART J Q, 1942: "A measure of the influence of population at a distance" *Sociometry* **5** 63-71

STEWART J Q, 1947: "Suggested principles of social physics" *Science* **106** (2747) 179-80

STEWART J Q, WARNTZ W, 1958: "Physics of population distribution" *Journal of Regional Science* **1** 99-123

STOUFFER S A, 1940: "Intervening opportunities: a theory relating mobility and distance" *American Sociological Review* **5** (6) 845-57

TAAFFE E J, GAUTHIER H L, 1973: *Geography of Transportation* (Prentice Hall, Englewood Cliffs)

TAYLOR P J, 1971: "Distance transformations and distance decay functions" *Geographical Analysis* **3** 221-38

TAYLOR P J, JOHNSTON R J, 1979: *The geography of elections* (Croom Helm and Penguin, London)

TEITZ M B, BART P, 1968: "Heuristic methods for estimating the generalized vertex median of a weighted graph" *Operations Research* **16** (9) 955-61

THEIL H, 1972: *Statistical Decomposition Analysis* (North Holland, Amsterdam)

THORNTHWAITE C W, 1948: "An approach toward a rational classification of climate" *Geographical Review* **38** 55-94

THUNEN J H VON, 1826: *Der Isolierte Staat in Beziehung Auf Landwirtschaft Und Nationalokonomie* (Hamburg)

TINKLER K J, 1976: "On functional regions and indirect flows" *Geographical Analysis* **8** 205-14

TOBLER W R, 1973: "Choropleth maps without class intervals?" *Geographical Analysis* **3** 262-65

TOBLER W R, 1979: "Estimation of attractivities from interactions" *Environment and Planning A* **11** 121-27

TORNQVIST G, 1971: "Manufacturing goods, servicing people and processing information at multiple locations" *Lund Studies in Geography, Series C General, Mathematical and Regional Geography, No 12* (Gleerup, Lund)

TVERSKY A, 1972: "Elimination by aspects: theory of choice" *Psychological Review* **79** (4) 281-99

TYREE A, 1973: "Mobility ratios and association in mobility tables" *Population Studies* **27** 577-88

UNSTEAD J F, 1916: "A synthetic method of determining geographical regions" *Geographical Journal* **48** 230-49

UNSTEAD J F, 1956: *The British Isles* (University of London Press, London)

UNSTEAD J F, MYRES J L, ROXBY P M, STAMP L D, 1937: "Report on classification of regions of the world" *Geography* **22** 253-82

VICKERMAN R W, 1974: "Accessibility, attraction and potential: a review of some concepts and their use in determining mobility" *Environment and Planning A* **6** 675-91

WARD J H, 1963: "Hierarchical grouping to optimize an objective function" *Journal of the American Statistical Association* **58** 236-44

WARNTZ W, 1957: "Contribution toward a macroeconomic geography" *Geographical Review* **27** 420-24

WARNTZ W, 1959: *Toward a Geography of Price* (University of Pennsylvania Press, Philadelphia)

WARNTZ W, 1976: "The university and the region" *Northeast Regional Science Review* **6** 18-47

WARNTZ W, 1981: "*Geographia Generalis* and the earliest development of American academic geography" in Blouet B (Ed) *The Origins of Academic Geography in the United States* (Archon Books, Hamden, Ct)

WEAVER J B, HESS S W, 1963: "A procedure for non-partisan districting" *Yale Law Journal* **73** 288-308

WEBBER M J, 1980: "A theoretical analysis of aggregation in spatial interaction models" *Geographical Analysis* **12** 129-41

WEBER A, 1909: *Uber Den Standort Der Industrien* (Tubingen)

WHITTAKER R A, 1962: "The classification of natural communities" *The Botanical Review* **28** 1-160

WHITTLESEY D, 1954: "The regional concept and the regional method" in James P E, Jones C F (Eds) *American Geography: Inventory and Prospect* (Syracuse University Press)

WILLIAMS I N, 1977: "Algorithm 1: three-point rational function interpolation for calibrating gravity models" *Environment and Planning A* **9** 215-21

WILSON A G, 1967: "A statistical theory of spatial distribution models" *Transportation Research* **1** 253-69

YAN C, AMES E, 1965: "Economic interrelatedness" *Review of Economic Studies* **32** 299-310

ZIPF G K, 1946a: "The P_1P_2/d hypothesis on intercity movement of persons", *American Sociological Review* **11** 677-86.

ZIPF G K, 1946b: "The P_1P_2/d hypothesis : the case of railway express" *Journal of Psychology* **22** 3-8

ZIPF G K, 1949: *Human Behavior and the Principle of Least Effort* (Addison Wesley Press Inc, Cambridge)

ZOBLER L, 1958: "Decision making in regional construction" *Annals, Association of American Geographers* **48** 140-48

END

08 04 87

FIN