ON SOME ASPECTS OF MODELS FOR THE ANALYSIS OF SPATIAL PROCESSES

Thesis for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY

ANTHONY V. WILLIAMS

1968

This is to certify that the

thesis entitled

ON SOME ASPECTS OF MODELS FOR THE ANALYSIS

OF SPATIAL PROCESSES

presented by

Anthony V. Williams

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Geography

Date Oc

O-169

ABSTRACT

ON SOME ASPECTS OF MODELS FOR THE ANALYSIS OF SPATIAL PROCESSES

by Anthony V. Williams

Many spatial patterns, the study of which has been a major objective of geographical research, are increasingly characterized by rapid change. The study of spatial processes which focuses on the measurement, description, and analysis of changes in patterns over time should become, therefore, more and more central to the field. Unfortunately, the development of the theoretical bases for this aspect of the discipline has not received attention commensurate with its present, and potentially greater future importance.

We do have, as a most important foundation for such development, the work begun by Hägerstrand and continued by those influenced by him. The focus of this work is particularly on the process of diffusion -- of people, ideas, and things. But the effectiveness of this research as a paradigm for spatial processes is weakened by some conceptual shortcomings and by the lack of a general theoretical substructure which would serve to unify disparate research on many spatial processes and provide the basis for the search for fruitful analogies which are so often central to further work.

We have attempted here to provide such a theoretical substructure which can, hopefully, serve adequately as the basis for

future research on spatial processes. It is argued that there are two fundamental components of all spatial systems: an attribute space which defines places by their respective properties and associated intensities, and a position space which defines the relative location of places. These two components, though, provide only the static elements of any system and the study of spatial processes must consider as an additional component the set of rules that define the nature and intensity of the dynamic linkages between the attribute and position spaces. The properties of such rule sets are considered and we take the position that their specification is most useful when they are defined stochastically. For pragmatic reasons we also advance a basic method for combining these components into a working model of any system under study.

As a demonstration of the basic steps that have to be taken to utilize the proposed structure, we present a simplified example of a spatial process -- the spread of a disease in an isolated region. Successive steps of defining attribute and position space, determining the rules under which the system operates and then integrating these components into a working model are illustrated for this simple system.

The interpretation and validation of the output of models of spatial processes present formidable difficulties. These arise partly from the difficulties in measurement and imprecision of definitions that are common to much research in the social sciences. But they are also due to the fact that the fundamental assumptions of most statistical tests regarding independence of observations are usually violated in spatial process models. Tentative suggestions for dealing with this problem, including a general approach and then an approach for dealing specifically with the validation problems peculiar to spatial systems, are presented.

Technical discussions are appended regarding the selection of computer languages must suitable for particular applications and on the program used in carrying out the calculations for the example mentioned above.

Approved: Gerard Rushton

Date: 24 October 1968

ON SOME ASPECTS OF MODELS FOR THE ANALYSIS OF SPATIAL PROCESSES

by

Anthony V. Williams

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Geography

1968

ACKNOWLEDGMENTS

Any merit this study may have is due, in large part, to the help, encouragement and inspiration afforded by Dr. Gerard Rushton of the Department of Geography at Michigan State University and by Dr. Julian Wolpert, now of the Department of Regional Science at The University of Pennsylvania. Dr. Wolpert, in a series of courses and seminars and in many informal discussions, stimulated my interest in the use of dynamic models in geography which led directly to the decision to explore the problems of their application. Dr. Rushton, who served as my dissertation advisor during the later stages of its preparation, offered advice and constructive criticism at just those times when it was most sorely needed.

Thanks are due also to Dr. Lawrence Sommers for making available to me the resources of the Department of Geography at Michigan State University and to Dr. Donald Blome who was my original advisor at the department. The university's Computer Institute for Social Science Research provided financial support during my doctoral program and more importantly provided a stimulating and congenial intellectual environment. My participation in the Michigan Inter-university Community of Mathematical Geographers made it possible to clarify my ideas of what geography is and what it should be.

Finally, I wish to express my appreciation for the fortitude of my wife, Donna, without whose support and encouragement this research would have been impossible.

This investigation was supported, in part, by a Public Health Service Fellowship (number 1-F1-GM-28, 714-01) from the Division of General Medical Services.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The field of geography can be fruitfully thought of as the study of spatial patterns and processes. Many workers in geography have accepted this definition fairly explicitly[1] and other well-accepted definitions of the field[2] differ from it, it seems to me, only in terms of phrasing and greater emphasis on either the study of patterns or of process. This broad view of the field of course completely begs some of the controversies in geographic methodology such as the question of whether areas are "unique" or should be considered as individual cases.[3]

A consequence of this definition is the implicit division of geographical models into two types -- static or dynamic -- with a common subset formed by combining these types. This is the dichotomy used in this paper although there are others of equal

---

[1] For example Gunnar Olsson, Distance and Human Interaction: A Review and Bibliography (No. 2, Bibliography Series; Philadelphia: Regional Science Research Institute, 1965), p. 1; John Leighly, "What Has Happened to Physical Geography?" Annals of the Association of American Geographers, XXXXV (1955), p. 318; Edward Ackerman, Geography as a Fundamental Research Discipline (Department of Geography Research Paper No. 53; Chicago: University of Chicago, 1958), p. 28.

[2] For an exposition of these, a good source is Richard Hartshorne, Perspective on the Nature of Geography (Association of American Geographers, Monograph Series; Chicago: Rand McNally, 1959).

[3] William Bunge, Theoretical Geography (Lund Studies in Geography; Series C: General and Mathematical Geography, No. 1; Lund, Sweden: C. W. K. Gleerup, 1962), pp. 6-13.

validity, for as usual our purpose governs our classification system.

The most-used static model in geography is the map which serves the unique function of providing both a beginning and ending point for most research in the field. Other examples of static models are the various methods used for regionalizing -- factor and pattern analysis,[4] linear graphs -- and much of the work based on central place theory.[5] Most applications of these models involve analysis of why phenomena occur where they do, the implications of the distribution(s) for economic development or other purposes, or analysis of interrelations of multiple phenomena over the landscape. Additionally, they can be used effectively with dynamic models in a combined attack on problems of distributional analysis.

Dynamic models in geography are primarily concerned with how phenomena come to be sited (in a time-space rather than strictly a spatial sense, since we include time as a specific parameter of a dynamic model), and with how and in what manner and form and volume interactions between areas occur through time for a specific phenomenon or combination of phenomena. The most

---

[4]See Brian J. L. Berry, "A Method for Deriving Multifactor Uniform Regions," Przeglad Geograficzny, XXXIII (1961), pp. 263-82.

[5]Clifford E. Tiedemann, "Two Models for the Inferential Analysis of Central Place Patterns" (unpublished Ph. D. dissertation, Department of Geography, Michigan State University, 1966); also see the extensive listings in Brian J. L. Berry and A. Pred, Central Place Theory: A Bibliography of Theory and Applications (Bibliography Series No. 1; Philadelphia: Regional Science Research Institute, 1961; with supplement, 1965).

notable use of dynamic models in geography has been the Monte Carlo method for studying the path of innovation diffusion through time[6] although random methods are not the only means of following such processes.[7]

Combined models which use both the static and dynamic approaches are probably more commonly used than the pure dynamic approach. They have in fact some obvious advantages over either the entirely static or entirely dynamic models. In a study concerned primarily with comparative distributions, for example, generic information on the phenomena often proves essential for understanding the significance of the spatial pattern studied as it pertains to different places.[8] A dynamic model seeking to trace out the space-path of a process, on the other hand can scarcely be comprehended in a geographical sense without consideration of pattern, even if that be presented in a form other than a map.[9]

The ideal model or research approach, then, in geography -- and surely in many other social sciences -- is neither restricted to time-less space or space-less time but rather combines both

---

[6]Initially by Torsten Hägerstrand in, "The Propagation of Innovation Waves," No. 4, Lund Studies in Geography; Series B; Human Geography (Lund, Sweden: C. W. K. Gleerup, 1952), pp. 3-19. A fairly complete (as of 1965) list of such studies by geographers is in Olsson, op. cit.

[7]Lawrence A. Brown, "The Diffusion of Innovation: A Markov Chain-Type Approach" (Department of Geography Discussion Paper No. 3; Evanston, Illinois: Northwestern University, 1963).

[8]Leighly, loc. cit.

[9]For instance in a linear graph or in matrix representation.

aspects of reality. In actuality, of course, we may have to emphasize one or the other aspects depending on the purpose of the moment but exclusion of either can only be justified for the most restricted kinds of problems.

In accord with these sentiments, this work outlines a modeling system designed to facilitate the building of models to analyze spatial processes. Considering the importance of dynamic systems in present-day geographic research and the concurrent practical need for predictive and planning models, it is hoped that such an attempt at synthesizing into a common structure apparently disparate approaches will provide the basis for significant advances in the field.[10]

We start in the next chapter by advancing a common conceptual framework which, we feel, underlies models of spatial processes. Then, to emphasize the evolutionary rather than revolutionary nature of our proposals, we examine two traditional models used to study such processes, pointing out those facets that do not meet all our criteria. The third chapter provides a more complete explication of our conceptual structure and includes some suggestions for integrating the components of spatial process models. Finally, we use a simple example to demonstrate our model-building scheme and point out some problems yet to be solved.

---

[10]Earlier attempts at synthesis in other fields of geography have been made by B. J. L. Berry, "Approaches to Regional Analysis: A Synthesis," Annals of the Association of American Geographers, L (1964), pp. 2-12 and by R. Chorley, "Geomorphology and General Systems Theory." In F. Dohrs and L. Sommers (eds.), Introduction to Geography. Selected Readings (New York: T. Y. Crowell Co., 1967), pp. 285-301.

CHAPTER II

DEVELOPMENT OF THE MODEL FRAMEWORK

In the last chapter, we mentioned several types of models which have been or can be used to analyze spatial processes and patterns. These have been characterized as static, dynamic, or a combination of the two. The static models, whether they be maps or graphs of distributions in one or more dimensions are of most use in analyzing the spatial manifestations of one or a combination of variables at a particular point of time.

While such models may be used to give some impression of change through a series of "snapshots" taken at discrete time intervals, spatial processes are perhaps best understood through the workings of models which specifically include the time dimension as a parameter. This point has been succinctly emphasized by Rogers in a paper on simulation and diffusion processes.[11] The task of this chapter, then, is to introduce the framework for such a modeling system, and compare its assumptions to those underlying other dynamic models of spatial process used by geographers and other social scientists. In the succeeding chapters, we attempt further explication and apply the resulting model to an example.

---

[11] Everett M. Rogers, *et al.*, "Computer Simulation of Innovation Diffusion: An Illustration from a Latin American Village." Paper presented at a joint session of the American Sociological Association and the Rural Sociological Society, Chicago, 1965.

A Brief Overview of the Conceptual Framework

Geography, as that discipline which is pre-eminently concerned with the spatial-locational dimension of phenomena, should logically structure its inquiries in such a way as to emphasize the spatial aspects of relationships and of behavior. In dynamic models, we are fundamentally interested in the results of repeatedly applying some transformational function (operator) to a set of operands.[12] We can leave aside, here, the question of the characteristics of the operator except to note the necessity of the concept and to mention that its nature may change with time, with achievement of certain threshold values in the operand set, and so on. The critical problem is the definition and depiction of "geographic" operands.

These operands, from our viewpoint, must have locational and spatial attributes in addition to any intrinsic properties; furthermore, the intrinsic properties cannot be thought of in isolation from the spatial ones. As an example, if we are investigating, say, total employment trends in the United States, as geographers, we would not focus on the interpretation of national aggregate figures. Instead, we would probably take areal samples of some kind, associate the places picked with our employment information and obtain an employment surface. For convenience, we might additionally classify our results to obtain employment

---

[12]This, at least, would be the interpretation derived from cybernetic principles. See chapter 2 in W. R. Ashby, Introduction to Cybernetics (New York: John Wiley & Sons, Inc.; Science Editions, 1966).

regions. In the dynamic situation, our task would be more involved and difficult but we would essentially be trying to depict (and perhaps account for) the pulsating nature of the national employment surface. We must, then, structure our investigation in terms of spatial (place) attributes and the relative location of these places. Location is always an implicit category in our scheme.[13]

Before developing this structure more fully in the next chapter, it will be instructive to compare its assumptions with those underlying two developed dynamic models of spatial behavior. This should serve two purposes: to demonstrate the non-radical nature of our approach by tying it to a developed tradition in geographic research, and to point out places where established models do not fit our present structure. The nature of the discussion is, by choice, rather more expository than analytic.

Antecedents

Ideas seldom spring from whole cloth. The ideas and methodology underlying the modeling system presented in the next chapter build on two sources. These are:

(1) The pioneering work on the diffusion of innovations through space started by Torsten Hägerstrand in Sweden and subsequent researches on diffusion models by several American geographers.

---

[13] William Bunge, "Locations are not Unique," Annals of the Association of American Geographers, LVI (1966), pp. 375-76.

(2) The behavioral insights of Georg Karlsson, the Swedish sociologist, into the communication process and more recent work by communications researchers in the study of the diffusion of ideas.

The Hägerstrand Simulation Model and Its Successors

Torsten Hägerstrand, a geographer at the Royal University of Lund, Sweden, initiated research efforts by geographers on the simulation of spatial diffusion in the early 1950's.[14] His work was especially noteworthy for two innovations: the use of probabilistic methods (Monte-Carlo technique) based on probabilities derived from measurements on observed data and the use of a digital computer to carry out the large number of calculations inherent in dynamic probability models. Hägerstrand's ideas attracted favorable attention in the United States in the late 1950's[15] and since that time a steadily increasing number of American

---

[14] Torsten Hägerstrand, op. cit.; also Innovation Diffusion as a Spatial Process. Translated from the Swedish edition (1953) by A. Pred (Chicago: University of Chicago Press, 1967). His most influential paper so far as American geographers are concerned appeared in 1960 in mimeograph, titled "On Monte Carlo Simulation of Diffusion." It has since been reprinted as "A Monte Carlo Approach to Diffusion" in B. Berry and D. Marble, eds., Spatial Analysis (Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1968), pp. 368-84.

[15] In 1958-1959, Professor Hägerstrand was a visiting professor of geography at the University of Washington where his ideas met great acceptance by the students of mathematical geography working with William Garrison.

geographers have experimented with simulation techniques applicable to geographic problems.[16]

According to his paper, "On Monte Carlo Simulation of Diffusion," Hägerstrand appears to have arrived at his model by considering what processes could cause the nebula-like patterns characterizing many economic and cultural phenomena. These patterns consist of dense "core" areas surrounded by border zones of outwards decreasing density. Given these types of distributions, he looked for processes which could create similar patterns and settled on the diffusion of techniques and ideas through social contacts as being peculiarly suitable for investigation. As he emphasized, "there is nothing such as one single and simple explanation of the 'nebula-distribution'." Thus, the fact that his investigations have been concerned with the diffusion of innovations rather than some other process has been mostly a matter of convenience.

Hägerstrand explains the process of innovation diffusion over space in regard to the adoption of a new farming practice in Sweden:

> A start is made by a rather concentrated cluster of carriers. This cluster expands step by step in such a way that the probability of a conversion always seems to be higher among those who live near the carriers than among those who live further away. The potential carriers become 'blackened' in a spatial continuity which reminds [sic] of the development of a photographic plate seen in a microscope.

---

[16]Olsson, op. cit. Also see Lawrence Brown, "A Bibliography on Spatial Diffusion," (Department of Geography Discussion Paper No. 5, Evanston, Illinois: Northwestern University, June, 1965), and his "Models for Spatial Diffusion Research: A Review," Technical Report No. 3, Spatial Diffusion Study (Department of Geography, Evanston, Illinois: Northwestern University, 1965).

> A convenient term from [sic] the phenomena could be borrowed from this physical process: 'neighborhood effect.'[17]

The inverse relationship between distance and the probability of adoption (or more generally, interaction) is based on the well-accepted notion that there is decreasing contact or influence between people as the distance between them increases. As contacts decrease, the occasions for learning of and adopting an innovation become fewer and fewer. This effect of distance on interaction was confirmed here by analysis of telephone traffic and local migration figures. These data were used as surrogates of information flows permitting construction of a contact matrix (the Mean Information Field); in each cell of this matrix are the derived probabilities of contact at specified distances from a carrier of an innovation. The probabilities reflect: intensity of information flow which is hypothesized to be directly related to acceptance of the innovation, and the average spatial pattern of day-to-day, or short run, contacts.

A second element of the neighborhood effect developed from Hägerstrand's observation that there exists a hierarchy of innovation centers with well-defined and relatively stable communication channels connecting them. The probability of an idea spreading through a social system is greatest if it is initially propagated through the upper levels of the hierarchy and if it uses existing information channels.[18]

---

[17] Hägerstrand, "On Monte Carlo Simulation of Diffusion," op. cit., p. 3.

[18] Ibid., p. 5.

To test his ideas, Hägerstrand developed a dynamic model to simulate the diffusion of an innovation in a population through time using Monte Carlo techniques. The Monte Carlo approach in simulation implies that the interactions of individual elements are governed by probabilistic rules given in the model rather than deterministic ones. The probabilities used are derived from real-world data and serve to provide an underlying stratum of reality upon which the model operates. In Hägerstrand's example, they come from measurements of local communication patterns and are expressed in the Mean Information Field.[19]

The original model had several self-imposed limitations for the purpose of simplifying exposition. In the process, only person-to-person communication was considered. Newspapers, radio, television, books, public lectures, and other communications media were not included in the model. The following series of rules were adopted to govern the simple model's operation:[20]

(1) Only one person carries the item at the start.

(2) The item is adopted at once when heard of.

(3) Information is spread only by telling at pairwise meetings.

(4) The telling takes place only at certain times with constant intervals (generation intervals) when _every_

---

[19]Hereafter referred to as the M. I. F. For other examples of the computation and use of the M. I. F. see D. Marble and J. Nystuen, "An Approach to the Direct Measurement of Community Mean Information Fields," _Papers of the Regional Science Association_, XI (1963), pp. 99-109, and R. Morrill and F. Pitts, "Marriage, Migration and the Mean Information Field: A Study in Uniqueness and Generality," _Annals of the Association of American Geographers_, LVII (1967), pp. 401-22.

[20]Hägerstrand, "On Monte Carlo Simulation of Diffusion," _op. cit._, p. 9.

carrier tells one other person, carrier or non-carrier.

(5) The probability of being paired with a carrier depends on the geographical distance between teller and receiver in a way determined by empirical estimate.

This simplest version of the Hägerstrand simulation model works in the following way. A map of the area being studied is prepared and divided into equal-area grid cells (normally squares) with the number of individuals in each grid cell being noted (Figure 1). The contact probabilities in the Mean Information Field (Figure 2) are converted to integer form (Figure 3). The central square of the floating grid -- the integer form of the M. I. F. -- is placed over one individual who has adopted the innovation previously. The adopter then communicates the innovation to another individual as determined by the probabilities in the other cells of the floating grid. The choice of the individual to be contacted is made by picking random numbers from a rectangular distribution. Here, for example, if the knower is in cell (4, 5) and the number 58 is selected, then contact is made with cell (3, 4).

The process is repeated until all adopters have communicated with another person. After all previous adopters have communicated, one "generation" of the simulation model is completed. At first, the number of adopters is small and their number initially increases slowly. As more people adopt the innovation, there are more "tellers" to spread the information and the rate of diffusion increases until a saturation point is reached at which point it

	1	2	3	4	5	6	7	8	9
1	2	3	5	10	1	3	8	7	10
2	1	5	6	8	7	4	5	3	7
3	1	4	18	40	6	3	2	11	8
4	1	9	20	125	35	31	5	13	24
5	8	6	50	200	78	25	20	18	8
6	2	5	12	100	24	118	25	4	13
7	15	4	32	70	19	15	14	6	3
8	5	3	10	50	6	7	7	3	5
9	8	1	3	30	2	3	4	7	2

Figure 1. Example of Population Grid for a Diffusion Model

.060	.100	.040
.150	.400*	.090
.100	.050	.101

Figure 2. A Mean Information Field

*Each entry gives the probability of being contacted by the "Knower" in the central cell.

000-059	310-409	860-899
060-209	410-809	900-989
210-309	810-859	990-999

Figure 3. Floating Grid

becomes difficult to find persons who have not yet adopted the innovation. The result of the process then, is typically an S-shaped curve.

Use of the Monte Carlo simulation models for research into actual process demands the use of digital computers. This is not due to any inherent complexity in the models themselves but to the enormous number of random numbers necessary to carry the model through a number of generations over the geographic field. The random numbers are necessary to determine which of the many possible courses of action will be utilized at any one step, and the number needed in even moderate-size simulation study will run into the thousands. This characteristic of the approach is demonstrated in an example in the next chapter.

Despite the modifications made to this simple model by Hägerstrand and other geographers to fit it better to reality[21] it suffers from some basic weaknesses from our point of view. Geographically, the most serious of these is the lack of spatial differentiation. That is, each cell of the "map" represents a place wherein people are present and each cell is assumed to be like every other cell. In the work by Yuill,[22] simulating barrier effects,

---

[21]Such as the introduction of a "resistance-curve" for persons contacted based on an assumed distribution of attitudes towards new ideas. See Forrest R. Pitts, "Problems in Computer Simulation of Diffusion," Papers of the Regional Science Association, XI (1963), pp. 111-19.

[22]Robert Yuill, "A Simulation Study of Barrier Effects in Spatial Diffusion Problems," (Michigan Inter-University Community of Mathematical Geographers Discussion Paper No. 5, Ann Arbor: The University of Michigan, 1965).

a separate information vector identifies the grid locations of barriers and also notes their character (absorbing, reflecting, or permeable) and Morrill[23] utilizes a similar method to give character to space in his work on town development in Sweden. But these studies, while important advances on the original model, are only capable of expressing one differentiating characteristic.

The M. I. F. which governs the distance and direction of contact also distorts reality in that it has no provision for differential distance and directional biases. That is, the same contact probabilities are applied around each cell on the "map."

There are a number of reasons why we would expect such differentials in the real world. Most obvious among these reasons are local biases in the communications network which may be due to topographic conditions or to peculiar historical circumstances. Variation from a uniform population density surface would also produce place to place differences in the M. I. F. In terms of such interactions as are expressed by shopping trip behavior, selection of marriage partner and other types of personal contacts there are also indications that rural and urban M. I. F. 's differ substantially. Marble and Nystuen note in a study of the M. I. F. concept based on trip behavior data collected in Cedar Rapids, Iowa,[24] that distance

---

[23] Richard Morrill, Migration and the Spread and Growth of Urban Settlement (Lund Studies in Geography, Series B: Human Geography, No. 26; Lund, Sweden: C. W. K. Gleerup, 1965).

[24] D. Marble and J. Nystuen, *op. cit.*, pp. 107-08.

decay functions for such behavior were steeper in that city than in the rural Asby area of Sweden used by Hagerstrand and further note that Stouffer's measures on distance between homes of marriage partners in Cleveland tends to confirm the existence of steeper decay rates in urban areas. The distance decay exponents for Asby, Cedar Rapids and Cleveland were, respectively: -1.58, -3.035, and -2.49. An obvious first explanation for these rural-urban differences lies in the greater density of opportunities for contact in metropolitan areas which, *ceteris paribus*, would lead to more spatially restricted contact fields.[25]

Of course, we might speculate that for certain types of contacts, urban areas might well have flatter distance decay functions than rural places; for instance, in behavior affected by exposure to mass media. But regardless of which one, or combination of these factors is operating, the M.I.F. in any varied landscape should differ from place to place and at micro-scale each place, and even each individual, would have its own peculiar field. And to complicate matters further we must note that contact fields for any area may also vary over time with changes in the communications system, cultural preferences and so on.

Finally, the Hägerstrand model is deficient -- from the point of view of serving as a general approach to analyzing spatial processes -- in not including varying behavioral parameters for sub-groups in the area being studied. That is, it is strictly an

---

[25]This explanation was advanced by Professor Gerard Rushton, Department of Geography, Michigan State University.

aggregative model wherein all members of the population are assumed to behave similarly.

In the case of the diffusion of an innovation, we should be at least aware of the possibility of the existence of varying group or individual attitudes towards new ideas depending on stage of the life-cycle, economic status, educational level, ethnic or religious characteristics and even the individual's or group's historical experience.

We do not mean to imply by the above criticisms that the general Hägerstrand model (including its successors) is not perfectly adequate for describing a large variety of geographic problems involving diffusion. It has been used to provide great insight into the spatial patterns resulting from the introduction of agricultural innovations and has even served to provide experimental verification of central place theory.[26] All that is said above, therefore, merely means that the Hägerstrand model should be used only with its stated limitations in mind -- a caution that applies to all modeling systems of any kind.

The Karlsson Model of Interpersonal Communication

In his book, Social Mechanisms[27] the Swedish sociologist Georg Karlsson develops several models of interest to geographers. The one most germane to our present interests is his model of

---

[26]Morrill, Migration and the Spread and Growth of Urban Settlement, op. cit.

[27]Georg Karlsson, Social Mechanisms: Studies in Sociological Theory (New York: The Free Press of Glencoe, 1958).

interpersonal communication. It is important because it adds the element of behavioral depth to the traditional Hägerstrand simulation models and also because of its relevance to the modeling scheme developed below. This relevance is due above all to the findings that suggest that many of the decisions that affect the landscape -- migration, economic development, and so on -- are ultimately based on the communication process as it takes place on this simple level.[28]

In any communication system, four basic elements are present: the message; the communicator(s); the receiver(s), and the environment (physical or social) in which the process takes place.

We also assume that a motivation to communicate is present, this being a function of the importance of the message, the character of the communicator or of exogenous factors.[29] The simple model of interpersonal communication places restrictions on these elements in the interest of simplicity and clarity of exposition. However, it illustrates the general mechanism through which any such simulation model operates.[30]

---

[28] Ibid., p. 53. Also see Julian Wolpert, "Behavioral Aspects of the Decision to Migrate," Papers of the Regional Science Association, XV (1965), pp. 159-69.

[29] Karlsson, op. cit., p. 29.

[30] Ibid., p. 47ff. The following discussion of Karlsson's original model is based on this.

Our milieu is a group of _n_ persons of whom _m_ possess the information to be transmitted. The message is simple and is not changed or distorted in the process of diffusion. Knowers of the message are motivated to communicate; they do so only at pairwise meetings with _a_ contacts per time period. Since the motivation to communicate lessens over time with some empirically-determined decay function, the teller communicates only for _k_ periods -- a total of _ak_ contacts. In reality, _a_ and _k_ are probably stochastic variables but for our purposes (and for most models looking for gross explication) their means (for observed data) are used.

We are interested in observing the character of the diffusion process -- that is the spread of the idea through time -- in the subject population. Therefore, the model must include the probability of each person's receiving the message in any time period and the independent probability of his accepting the message and in turn becoming a teller in the next period.

These probabilities depend on the geographical and social distance between teller and contact. The following probabilities then must be estimated from the data for any particular project:[31]

Probabilities governing receipt of the message:

p_{gs} = Probability of a person at distance _g_ and social distance _s_ receiving the message.

p_{gs}/n_{gs} = Probability of contact for an individual if there are n_{gs} individuals in a cell.

---

[31] Ideally, in a manner similar to Hägerstrand's.

Probabilities governing acceptance of the message:

p_{ac} = Probability of accepting a message with attitude $\underline{a}$ on the side of the contact and credibility $\underline{c}$ on the side of the teller.

Further, since Karlsson saw no reason to suppose a dependency between physical and social distance, we assume $p_{gs} = p_g p_s$ and formally assess the probability of contacting someone who is already a knower as r^t_{gs}. This probability depends on what happens during the diffusion process being a function of the proportion of knowers in a particular cell; it designates the probability that a message directed to cell gs at time t hits a knower and is lost.

The probabilities of non-acceptance of the message because of the contactee's attitudes and the teller's credibility are likewise assumed to be independent so that $p_{ac} = p_a p_c$. Notationally, we designate by p_f the probability that a person in cell gs has probability p_a of acceptance; this is equal to the fraction of non-knowers in cell gs belonging to social category a.

Bringing the above together in an overall formula, we note that the probability of a message being directed to a non-knower in cell gs and being accepted by him is

$$\sum^{p} \sum^{a} p_g p^{(1-r_{gs})} p_a p_f p_c p_e$$

where r_{gs}, p_f, and p_e are parameters determined by the stochastic nature of the diffusion process. There is no explicit measure for their value as a function of time and the other parameters.

Fortunately, as Karlsson notes,[32] they are only needed for the formal development of the model.

The lack of a workable formal definition for the time-path of the message through the population makes it necessary to use Monte Carlo procedures to find an approximate distribution of the time-paths. This requires only the parameters p_s and p_g to determine the cell gs to which the information is directed -- giving each member of the cell a chance to receive the information. If he has not heard it before, he then has a chance $p_a p_c$ to believe the message based on his own attitude and the credibility of the teller. It is necessary in the actual computation to keep track of each knower and each contactee to determine the relevant attitude and credibility ratings in each case.

Karlsson provides an example of the procedure for hand computation using the data matrix in Table 1. The cell entries reflect: social class, A or B; attitude of members towards new ideas -- H=unreceptive to new ideas, and credibility -- T=low credibility. The matrix represents a group of 100 persons distributed in a square with equal distances between positions. The positions are regarded as main staying places because of the necessity for the carriers of the message to move about and meet other persons.

The following distance probabilities are assumed for the example:

---

[32]Karlsson, op. cit., p. 50.

Table 1. Data Matrix for Karlsson Simulation[33]

	a	b	c	d	e	f	g	h	i	j
1	A T	AH	B	A T	A	B	AHT	A	B	A T
2	A	BHT	A	A	B T	A	AH	B T	A	A
3	BHT	A	A	B T	A	AH	B T	A	A	BHT
4	A	A	B T	A	AH	B T	A	A	BHT	A
5	A	B T	A	AH	B T	A	A	BHT	A	A
6	B T	BH	A	B T	B	AH	B T	B	A	B T
7	BH	A	B T	B	AH	B T	B	A	B T	BH
8	A	B T	B	AH	B T	B	A	B T	BH	A
9	B T	B	AH	B T	B	A	B T	BH	A	B T
10	B	AH	B T	B	A	B T	BH	A	B T	B

$p_{g1}=0.5, \quad p_{g2}=0.4, \quad p_{g3}=0.1.$

The geographical distance cells are defined as:

Cell 1: the positions in the square of cells nearest the communicator.

Cell 2: the positions in the square of cells next to cell 1.

Cell 3: the positions in the square of cells next to cell 2.

If the communicator, then, is in position 4d, cell 1 consists of 3c, 3d, 3e, 4c, 4e, 5c, 5d, 5e: cell 2 of positions 2b-2f, 3b, 3f, 4b, 4f, 5b, 5f, and 6b-6f, etc.

The other probabilities are assumed to be $p_s=0.9$ for the communicator's own stratum and 0.1 for the other stratum; these

[33] Ibid., p. 51.

probabilities are the same for both strata. If a person has attitude h, his p_a=0. 3; if his attitude is $\overline{H}$, p_a=0. 9. A communicator marked with a T has a probability of acceptance of his information p_c=0. 7; if he is $\overline{T}$, p_c=1. 0. Since $p_{ac}=p_a p_c$, we get:

Condition	p_{ac}	Condition	p_{ac}
Non H from non T	.90	H from non T	.30
Non H from T	.63	H from T	.21

A trial run of this model was made using random numbers from a table and the following procedure:

(1) Determine the geographical region (cell 1, cell 2, cell 3) to be contacted.

(2) Find the social stratum to be contacted within the selected cell.

(3) Select a person to be contacted within the gs cell, giving each the probability $1/n_{gs}$.

(4) If the person is not a knower already, his probability of accepting the message is given by p_{ac}.

(5) Steps 1 through 4 are repeated for each knower in each time period (ten periods are used in the example -- these are what Hägerstrand calls generations).

(6) The initial and sole knower is placed in cell 5e. Each knower communicates only three times after which he is inactive. Only one contact is made by a knower in any one generation.

The results of this trial are listed in Table 2.

Table 2. Trial Run 1 of Karlsson's Simulation[34]

Step	Knowers	Hits Producing No New Knowers
0	5e	-
1	5e	3g
2	5e	8f
3	5e, 4c	-
4	(5e), 4c	3d
5	(5e), 4c, 6a	-
6	(5e), 4c, 6a, 7b	6a
7	(5e, 4c), 6a, 7b, 8a	4c
8	(5e, 4c), 6a, 7b, 8a, 8c, 8e, 6c	-
9	(5e, 4c, 6a), 7b, 8a, 8c, 8e, 6c, 5a, 5b, 5c	6a, 7f
10	(5e, 4c, 6a, 7b), 8a, 8c, 8e, 6c, 5a, 5b, 5c, 8f, 4a, 7e	10b, 5e, 5d, 6a

non-active knowers are in parentheses

Several contrasts should be pointed out between the Karlsson and the Hagerstrand-derived models. First and most important is the greater attention paid to behavioral elements by Karlsson. In essence, this difference is akin to the distinction between aggregative and disaggregative models. The great advantage of the latter, especially in geographic research, is their ability to more accurately portray the complexity of the landscape by allowing for the diversity of elements present in any area. This greater accuracy carries a penalty, of course, if pursued too far: the lack of, or great difficulty in collecting adequate data and the subsequent problem, if the first be overcome, of aggregating our results in some way so as to make them comprehensible. Nonetheless, the inclusion of several behavioral variables in each cell (place) of the

---

[34] Ibid., p. 52.

Karlsson model's data matrix gives it a great advantage over previous models of diffusion and we shall follow this approach in our own modeling system.

The greater attention paid to behavioral variables by Karlsson is counterbalanced, for geographers, by his relative neglect of spatial variation. It would be an error for us to use this model in unmodified form because it has no provision for inclusion of such physical variables as barriers (social barriers are, of course, included in the model). While not an a-spatial model, Karlsson's treatment of distance is also unsatisfactory for us. If Hägerstrand's mean information field is too aggregative for studies of large regions, it does, at least, provide for local directional biases and allows quite fine control of the distance decay function. Karlsson's distance rings recognize the attenuating effects of distance (whether measured in physical or social units) but do not allow discrimination within the rings. Both models have no means of taking into account the possibility of differing spatial preferences among groups in the population. In the next chapter, where we develop our own structure in more detail, we use what we can of the Hagerstrand and Karlsson approaches and at the same time attempt to avoid some of their deficiencies.

CHAPTER III

STRUCTURE OF THE MODEL SYSTEM

We have previously only alluded inferentially to the model system we are proposing, first in the discussion of general approaches to geographic models and then, more concretely, in the preceding discussion. The purpose of this chapter is to describe, in more detail, the common structural characteristics of our models and to indicate some important operational considerations. The example that follows in the next chapter is designed to illuminate these features.

First of all, let us again examine the characteristics one might rationally associate with or expect to find in any model of a spatial process, be it migration, information diffusion or changes in the properties of an air mass over time. We find we need be concerned with three: place attributes; place location; and the rules specifying interactions among places under given conditions. The need for a knowledge of attributes follows from the desire to examine interactions among places; since these can presumably occur only under certain conditions and are dependent on place characteristics, it is necessary to know these characteristics. Given such a description in terms of "relevant" variables, we then need to specify the permissible paths over which information, people or other phenomena can flow from or to a place; in a sense, we must specify distance-decay functions. So far, we have presented a landscape, a map as it were. To breathe life into this landscape and to model a process acting in it, it becomes necessary to provide a set of rules which directs interactions according to place characteristics and connectivities.

The following diagram (Figure 4) illustrates these considerations and integrates with them the important steps of initial question posing and model testing and verification.[35] The reverse arrows indicate in a schematic way the process of feedback (positive and negative) and also the existence of a continuum of models resulting from refinement of our questions and changes in selections of subsystems to be investigated and even our image of the "real" world.

The problems of question-posing and model testing have been extensively discussed elsewhere in the geographic literature[36] and we will not depart from our scheme to consider them here. Our attention is instead directed towards the task of expanding on problems of specifying attributes, connection patterns and rules of interaction and of integrating these into a well-ordered experimental design.

---

[35] The general form of the diagram and some of the concepts underlying it are based to an extent on the work of Richard Chorley, "Geography and Analogue Theory," Annals of the Association of American Geographers, LIV (1964), p. 129.

[36] See J.O.M. Broek, "Some Research Themes," in Geography. Its Scope and Spirit (Social Science Seminar Series; Columbus, Ohio: Charles E. Merrill Books, 1965); also "Four Problem Areas and Clusters of Research Interest," in NAS-NRC, Earth Sciences Division, The Science of Geography: Report of the Ad Hoc Committee on Geography (Publication 1277; Washington: 1965). The literature on model testing is scattered but a good discussion is available in Peter Haggett, Locational Analysis in Human Geography (London: Edward Arnold Ltd., 1965), pp. 277-310.

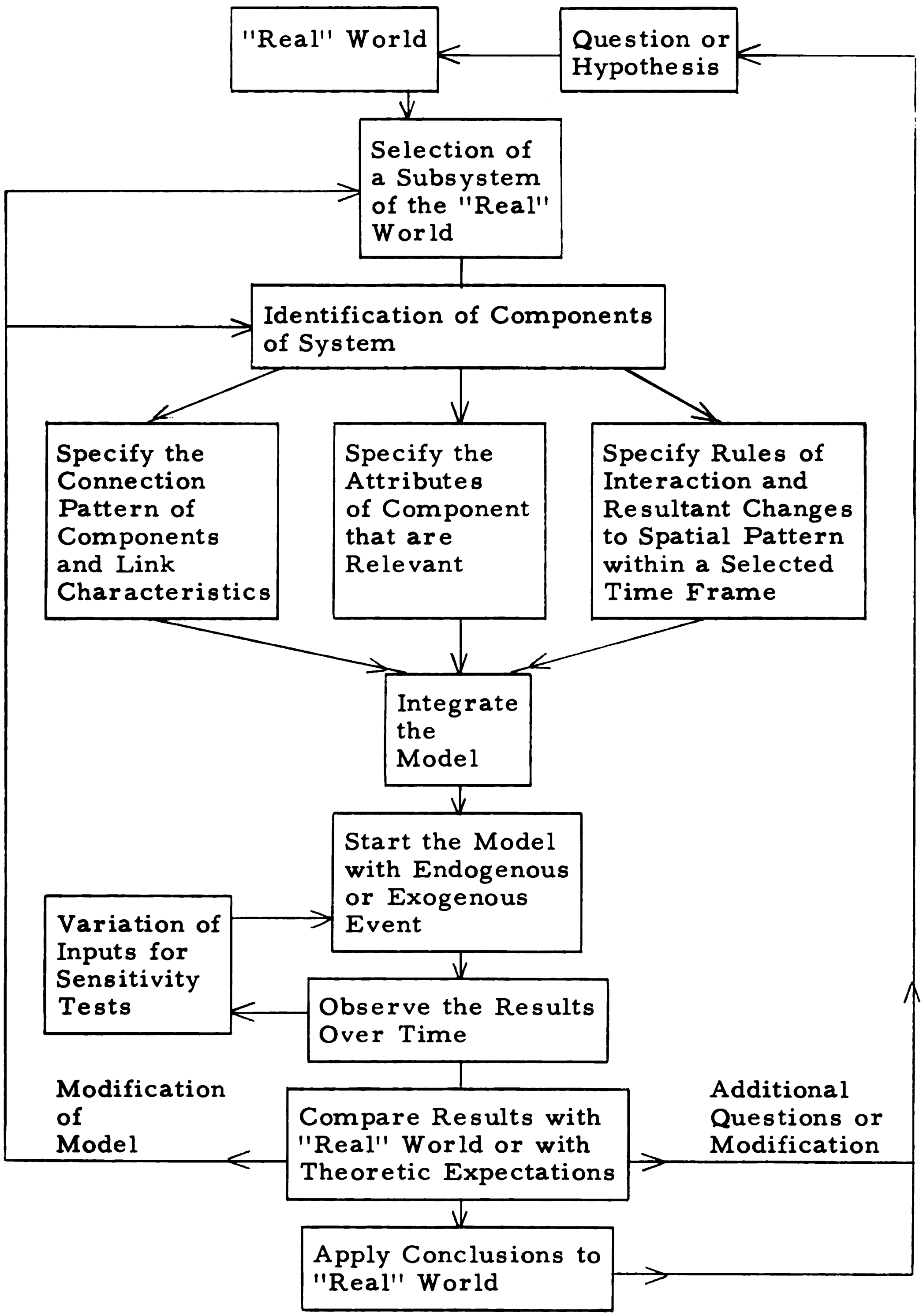


Figure 4. A Model of a Model of Spatial Process

Specification of Attributes

Once a choice of some subsystem of the real world has been made for the study and its components specified,[37] we confront the problem of selecting those attributes of the system deemed to be of significance in the process being investigated. We shall define an attribute of a place as being both a property and an associated intensity.[38]

There now exist two problems: identification of the relevant properties of the system and the choice of measures of magnitude for each. The identification problem can be handled in a number of ways. If we are posing questions in a hypothetico-deductive framework it is certainly reasonable to postulate the importance of certain key variables or properties, subject to test of course. This was done in the Karlsson model mentioned above.[39] Or we may take another well-defined system which we presume behaves similarly to the one under investigation and search for analogous variables. Ellis' study of the Michigan recreation system follows this course.[40] Another approach might make use of multivariate

---

[37]Normally these will be areal units but depending on the study scale they can also be line or point phenomena or a combination of all three.

[38]For example, a property might be manufacturing employment; then its associated intensity would be the number of workers employed.

[39]See pages 11-12ff.

[40]Jack B. Ellis, "The Description and Analysis of Socio-Economic Systems by Physical Systems Techniques," (unpublished Ph. D. dissertation, Department of Electrical Engineering, Michigan State University, 1965), pp. 9-11.

statistical techniques, typically multiple regression or factor analysis, to select a constellation of independent variables in a parsimonious fashion according to some efficiency criterion such as percent of system variance explained by the selected group of variates. Departures from linearity and the possible existence of interaction effects vitiating the assumptions of independence on which these techniques depend should of course be investigated when they are used.[41]

Any of these approaches to the selection of some fundamental set of properties may be used singly or in combination. It is vital to realize that good judgment on the part of the investigator is always required in making the selection as is an adequate knowledge of the system being studied. Where "ideal" measures are lacking we are often reduced to using some available surrogate which hopefully mirrors the behavior of the conceptually desired variable. In the case of one Hägerstrand model of information diffusion which utilized telephone messages as a substitute for some ideal measure of information flow[42] the assumption of analogous behavior can be

---

[41] A simple and clear exposition of the dangers of assuming linearity where it does not exist is in Frederick V. Waugh, Graphic Analysis: Applications in Agricultural Economics (Agricultural Handbook No. 326, United States Department of Agriculture, Economic Research Service; Washington: U.S. Government Printing Office, 1966), pp. 24-25. A discussion of interaction effects and their importance can be found in John Sonquist and James Morgan, The Detection of Interaction Effects (Institute for Social Research, Survey Research Center; Ann Arbor: The University of Michigan, Monograph No. 35, 1964).

[42] Hägerstrand, "On Monte Carlo Simulation of Diffusion," op. cit., p. 6.

easily defended. But in general the use of surrogates requires an even higher degree of subject knowledge and integrity on the investigator's part than normal.

The choice of a metric to define intensity of a property is also not a simple thing. To some extent there is a dependency on the choice of the variable to be measured. Also, we may either be limited to given measures as in the case of census data or might be so uncertain of the accuracy with which the variable was measured that we deliberately choose a measuring tool of greater "fuzziness" than may be available to blur inaccuracies. This implies that we use a weaker measurement scale than is naively indicated by our data in an attempt to balance the accuracy desired with our confidence in the measurement itself.

Using Steven's classification of measurement scales (Table 3) we can illustrate this technique with the following example. Assume we have two measures of income in dollars for each of five places:

	Place				
	Able	Baker	Charlie	Dog	Echo
Average Gross Income	12,543	4,950	6,800	10,200	4,000
Average Net Income	10,500	4,500	6,000	8,600	3,700

If we are aware that measurement errors may be present in this data (as a result of poor sampling technique, incorrect responses, coding errors and so on) we may decide not to trust the apparently exact figures above but instead to blur the assumed inaccuracy by converting the data to an ordinal scale:

Table 3. A Classification of Scales of Measurement and Examples of Statistical Measures Appropriate to Each

Scale*	Basic Empirical Operations	Measures of Location	Dispersion	Association or Correlation	Significance Tests
Nominal	Determination of equality	Mode	Information, H	Information transmitted, T Contingency correlation	Chi square
Ordinal	Determination of greater or less	Median	Percentiles	Rank-order correlation measures	Sign test Run test
Interval	Determination of the equality of intervals or of differences	Arithmetic mean	Standard deviation, Average deviation	Product-moment correlation	T test F test
Ratio	Determination of the equality of ratios	Geometric mean, Harmonic mean	Per cent variation		

Source: S. S. Stevens, "Measurement, Psychophysics, and Utility" in C. W. Churchman and P. Ratoosh (eds.), Measurement: Definitions and Theories (New York: John Wiley & Sons, 1959), pp. 25 and 27.

*The more powerful scales can be used for any operation that can be performed using a weaker scale; also, all measures available for a weak scale are available for the stronger scale.

	Place				
	Able	Baker	Charlie	Dog	Echo
Average Gross Income	1	4	3	2	5
Average Net Income	1	4	3	2	5

What we have done, here, is sacrifice some of the discriminating power inherent in the ratio scale to obtain simpler but less "noisy" information. We also find, as we might expect, that the types of statistical measures obtainable from our data are less sophisticated. But they are also less likely to lead us into making incorrect inferences since they increase the chance of rejecting a hypothesis.

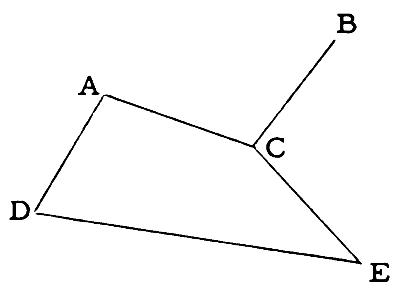
We may also make such a measurement scale transformation if the relationship of interest can be perfectly well expressed using a weaker scale. That is, if we are only interested in the presence or absence of a phenomenon, then a binary nominal scale is quite adequate. Similarly, an interest in the ordering of places (A is greater than B) need not call for the use of a scale any stronger than the ordinal. In our original data collection, however, it is wisest to start with the most discriminating measure possible since we can always coarsen measurements but cannot read into them any more power than is originally there.[43]

---

[43] A strong argument for this approach is made by S. Goldberg in his Probability: An Introduction (Englewood-Cliffs, N.J.: Prentice-Hall, Inc., 1960), pp. 45-46. For an interesting example of the use of scale transformation see J. Nystuen and M. Dacey, "A Graph Theory Interpretation of Nodal Regions," Papers and Proceedings of the Regional Science Association, VII (1961), pp. 29-42.

Specification of Connection or Position

If we view the attribute space of a spatial system as corresponding to the classical idea of site, then position space corresponds to location. If we are only interested in the property of connectivity, then a simple linear graph (Figure 5A) suffices to portray this. The equivalent matrix representation (Figure 5B) is more convenient for arithmetic manipulation but fails to give the visual image geographers find useful. In theory, the quality of a connection channel can also be considered in these representations by specifying an unambiguous scale based on the capacity of the channel.



(A)

$$\begin{array}{c|ccccc} & A & B & C & D & E \\ \hline A & 0 & 0 & 1 & 1 & 0 \\ B & 0 & 0 & 1 & 0 & 0 \\ C & 1 & 1 & 0 & 0 & 1 \\ D & 1 & 0 & 0 & 0 & 1 \\ E & 0 & 0 & 1 & 1 & 0 \end{array}$$

(B)

Figure 5. Linear Graph and Matrix Representations of a Connected Network

The above representation seems to indicate that specifying connectivity is simpler than of specifying attributes. But this is not the case, for we encounter several basic problems. In some cases, these may stem from a disparity between technical channel capacity and actual utility -- a situation most often found in underdeveloped or authoritarian countries where such phenomena as "showpiece" roads are not uncommon. The use of estimates of traffic generating capacity of terminals achieved by gravity model techniques or analogue methods is one way of attacking this type of problem. But we encounter more substantial difficulties when we consider the hierarchical nature of many communication flows in the real world. As Hägerstrand noted in connection with the diffusion of ideas through pairwise meetings, some individuals operate on a local level only while others operate on a regional or even international level as well. Analogous points have been made by Tiedemann and Van Doren in their work on the diffusion of hybrid corn in Iowa.[44]

When dealing with a hierarchic connectivity net such as the transportation system pictured (Figure 6), normal concepts of distance cease to give adequate explanations of observed interactions over space. Here, the major, "A," centers are connected on one distance continuum while the subsidiary centers, "B," interact on an entirely different distance scale. Even though the subsidiary centers are closer to each other in a "pure" distance sense than the major centers, they are further apart and relatively isolated if we measure distance on an access-time scale or in terms of "effective" connectivity.

The most familiar examples of this phenomenon are found in communication and transportation systems with high-quality

---

[44]Hägerstrand, "On Monte Carlo Simulation of Diffusion, op. cit., p. 5. C. E. Tiedemann and C. S. Van Doren, "The Diffusion of Hybrid Seed Corn in Iowa: A Spatial Simulation Model (Technical Bulletin B-44, Institute for Community Development and Services; East Lansing: Michigan State University, December, 1964).

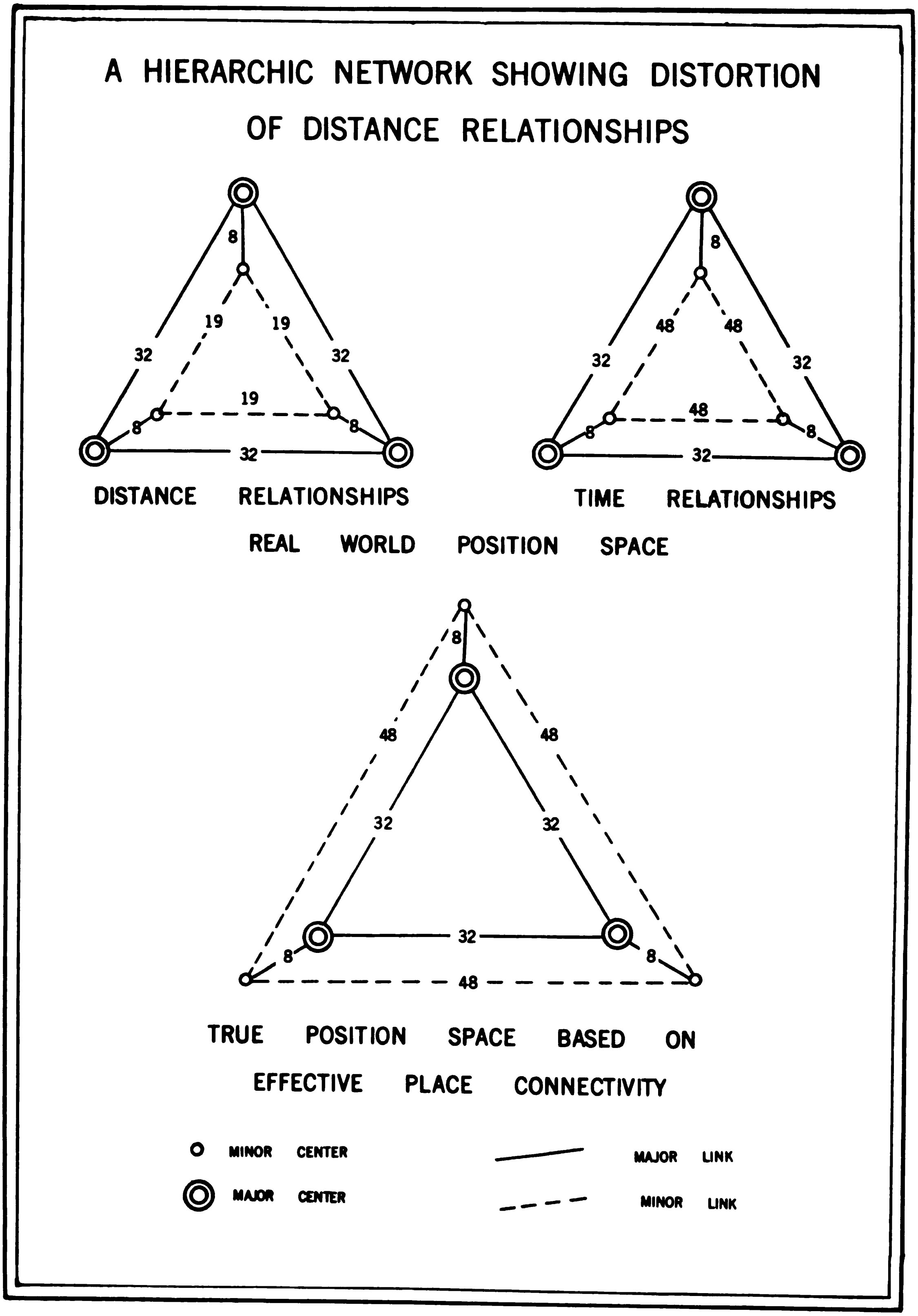

A HIERARCHIC NETWORK SHOWING DISTORTION
OF DISTANCE RELATIONSHIPS
8
19
19
32
32
19
8
8
32
DISTANCE RELATIONSHIPS
8
48
48
32
32
48
8
8
32
TIME RELATIONSHIPS
REAL WORLD POSITION SPACE
8
48
48
32
32
32
8
8
48
TRUE POSITION SPACE BASED ON
EFFECTIVE PLACE CONNECTIVITY
MINOR CENTER
MAJOR CENTER
MAJOR LINK
MINOR LINK

Figure 6.

connections functioning primarily as links between major centers, leaving subsidiary points in the network to make do as well as possible with less efficient links. Apart from obvious effects on such processes as industrialization and the diffusion of innovations which depend heavily on efficient communications, the existence in spatial networks of such hierarchical structures poses some interesting modeling problems.

Pictorial inversion of the system to produce true relations in position space as demonstrated in the second part of Figure 6 is obviously a limited solution to be applied only to topologically simple networks.[45] A more logical approach in the general case would be to divide the system studied into homogeneous groups based on efficiency of connections and treat each group separately while assuring that correct inter-group linkages are maintained.

Specification of Interaction Rules

While many useful geographic analyses can be performed using the static structure of position and attribute-space, it seems doubtful that an adequate knowledge of spatial processes can be attained without the employment of dynamic models having time as an inherent parameter. An essential prerequisite for such models ~~is~~ the construction of a set of interaction rules. These serve to link the static tableau of the landscape with the investigator's ideas about the mechanisms of the process studied.

---

[45]But the technique is appealing pedagogically and might be used, for instance, in an attempt to clarify problems of the isolation of certain districts in a country or a city from the "mainstream," in presentations to legislative groups or planners.

The rules have to meet two conditions. First, they must be capable of being expressed in the form of a statement calculus[46] of the form

IF A THEN B ELSE . . .

where the tested condition may be a string of propositions joined by logical operators. The condition must have a truth value in each instance. If the condition is true, then "B" results, otherwise we proceed to an alternate result or to a further series of test statements. The result need not be deterministic since it may consist of a vector of outcomes, exact choice of which would be determined by probability rules. Obviously, these, too, must be specified as part of the model. Rules that can be expressed in a statement calculus are unambiguous; for our purposes in constructing a logical model they must also be complete. That is, no logical possibility can be overlooked in our specifications. Unfortunately, we cannot be sure that a set of rules meeting these two requirements will give "correct" answers. The format does, however, facilitate testing through a stage process to help insure that our logic is tight and can point up those propositions needing further research.

Once we go beyond these basic requirements, the task of establishing modeling rules becomes more complex as they are guided by the purpose of the research and some attitudes of the investigator. A basic dichotomy must be faced at the outset; this

---

[46]For a readable exposition of the basic requirements of a statement calculus see J. Kemeny, L. Snell, and G. Thompson, Introduction to Finite Mathematics (2nd edition; Englewood-Cliffs, N.J.: Prentice-Hall, Inc., 1967), pp. 1-52.

is a choice between specifying a deterministic or a probabilistic model. The great appeal of deterministic models lies in their conceptual simplicity. They also have intuitive appeal for those who value conciseness in statement and dislike introducing the concept of chance into explanatory models. Where a deterministic model produces adequate results (in a predictive sense), these virtues of simplicity must recommend it.[47] But if we agree with Saushkin that

> geography is the science dealing with complex dynamic spatial systems that develop on the earth's surface as a result of interplay between nature and society. . . .[48]

then we can make a strong case for making probabilistic models the norm in geographic research. A practical reason for doing so is that much of our raw data is sampled, formally or otherwise, making our conclusions statistically rather than absolutely valid. Nagel also makes the point that laws in the social sciences are (most likely to be) statistical in nature because they are stated as though applicable to the real world rather than to an ideal state.[49] When our models are focused on examining human actions on a

---

[47] In fact, in a recent article that recast the gravity model to accord with a probabilistic philosophy, the author implies that the added complications did not produce significantly better results. See B. Harris, "Probability of Interaction at a Distance," Journal of Regional Science, V (1964), pp. 31-35.

[48] Y. G. Saushkin, "An Introductory Lecture to First-Year Geography Students," Soviet Geography: Review and Translation, VII (1966), p. 59.

[49] Ernest Nagel, The Structure of Science (New York: Harcourt, Brace, and World, Inc., 1961), pp. 507-09.

less than universal scale, we should also be aware that individual responses to stimuli are governed by interpretations of external conditions rather than the conditions themselves and that the resulting uncertainty in the model must be handled probabilistically.[50]

Assuming that one chooses to construct a probabilistic model, there are a number of possibilities. Two types that have been proposed for and used in geography are Markov chain models and variations on the Monte Carlo method.[51] The Hägerstrand and Karlsson models which were discussed in the last chapter are examples of the latter approach; enough has been written about Monte Carlo models in the geographic literature to make a discussion of underlying assumptions superfluous here. But this is not the case with Markov models and a brief account of the basic finite model,[52] its assumptions and its limitations for research into spatial processes might be useful.

We assume a set of experiments having the following properties. The result of each experiment is one of a finite number of outcomes $[x_1, x_2, \ldots x_k]$. The probability of any outcome x_j is

---

[50]At the small, similar conditions are encountered in the physical sciences. See Werner Heisenberg, Physics and Philosophy (New York: Harper & Brothers, 1958).

[51]I. Lowry, "A Short Course in Model Design," Journal of the American Institute of Planners, XXXI (1965), pp. 158-65; W. Garrison, "Towards Simulation Models of Urban Growth and Development," No. 24, Lund Studies in Geography, Series B; Human Geography (Lund, Sweden: C. W. K. Gleerup, 1962), pp. 91-108. One of the few examples of the use of Markov chain model by a geographer is found in Brown, op. cit.

[52]Adapted from E. Parzen, Stochastic Processes (San Francisco: Holden-Day, Inc., 1962), pp. 188-306.

not necessarily independent of previous outcomes; but at most it depends on the outcome of the immediately preceding experiment. The probability of outcome x_j given that x_i occurred on the previous experiment is given by p_{ij}. The p_{ij}'s are termed <u>transition probabilities</u> and the set of outcomes $[x_1, x_2, \ldots x_k]$ are called <u>states</u>. The p's are calculated experimentally or in social science applications are frequencies reduced to probabilities. If we know that a process begins in a particular state, then given the transition probabilities, we have sufficient information to calculate the probabilities of the experiment ending in any given outcome at any future time.

A process combining the transitional probabilities and states is most conveniently represented in matrix form

$$x_t = p\, x_{t-1},$$

where the x_t represents the state vector at time [t]; its values are the result of the interaction of the transition probability matrix p and the state vector at time [t-1]. Operationally, the interaction can be computed as a succession of matrix vector multiplications but it is easy to prove that given the vector x at time t = 0, then x_n is simply $p^n x_0$.

The superficial simplicity of the approach makes it tempting to use as a predictive model. But there are some drawbacks we should be aware of. Where there is no absorbing state present (a state from which there can be no transition) the state vector reaches equilibrium after a number of cycles and thereafter stays constant. It is possible to obtain this equilibrium solution of the state vector analytically, that is, without carrying out "n" matrix multiplications. The number of cycles taken to reach equilibrium is then unknown. Where cycle times are very short as in many physical science processes this is often not critical. But where they are each a year or longer which is common if we are using social science data (censuses or surveys) it becomes misleading to consider the equilibrium situation without knowledge of the time taken to reach it. For instance, an otherwise excellent recent paper predicted interregional migration and used equilibrium values of the state vector as the basis for discussion of policy questions.[53] In checking this model using Roger's initial state vector and transition probabilities which were based on five-year migration frequencies, the author attempted to ascertain, using a computer for performing the matrix multiplications, the time required for the system to reach equilibrium. After 2,000 years of simulated time this had still not occurred.

Even over a shorter time period, the transitional probabilities in the real world system are likely to change, especially if we are working with a social or economic system. But the transition probabilities in the Markov chain model are fixed initially and never change.

Olsson and Gale have recently[54] proposed that these and other objections can be met by using an n-dimensional Markov process

---

[53]Andrei Rogers, "A Markovian Policy Model of Interregional Migration," Papers of the Regional Science Association, XVII (1966), pp. 205-24.

[54]G. Olsson and S. Gale, "Spatial Theory and Human Behavior: Human Behavior and Anarchistic Vector Spaces." Paper presented before the Regional Science Association, Boston, 1967.

model. The multi-dimensional feature would allow the researcher to consider more than one variable operating at a place. This would, of course, let us come much closer to the multi-factor real world problems we are interested in exploring. Adoption of the more general Markov process also allows the transitional probabilities to change over time as a function of the immediately preceding transition matrix and state condition and the place of the sequence in a set "T" of sequence conditions.

It is difficult to quarrel with this ambitious proposal, especially since all reasonable-appearing models should be explored at this stage of geographic methodology. Two caveats might, however, be raised. First, the model may be too sophisticated, for our present ability to specify interactions and transition probabilities in the detail required by the model is limited. Second, the Markovian model is linear and may not "fit" many spatial processes. Olsson and Gale speculate that the traditional linear operator in the Markov process could be replaced by non-linear operators; but, as they admit it is not at all clear in general what form such operators might take.

Roger's migration model and the proposals advanced by Olsson and Gale point up two critical problems in devising an adequate set of interaction rules for spatial processes: the selection of reasonable time parameters and the necessity for including provision in the model for changes in position and attribute space as a result of the operation of the system over time. Both are rather difficult problems and no easy answers are available.

In the case of the time parameters, the investigator should specify a "working time" for the simulated system which covers the period for which he thinks his rules hold true. At the end of that time, the system should be examined for "reasonableness" and tested against expected outcomes. He should also be aware that, in theory, each component of the system might have its own time response function. We can all think of certain attributes of an urban system, say, that respond quickly to changes in the external environment whereas others respond more slowly or not at all in the short run.[55] A good set of rules for a spatial process should reflect these differentials where they exist in the system.

It is also necessary to make provision for structural changes in the system resulting from operation of endogenous or exogenous forces. Most of the process models used to date in geography have this facility only to a limited extent. The difficulty in extending this extremely important quality results from our ignorance and inability to specify exactly all forces likely to impinge on a system. In this case, we are in a position similar to an analyst who was asked why he had not predicted present United States troop levels in Vietnam three years ago. Our own answers are not likely to be much better but we have the obligation to try.

Integrating the Model Structure

Thus far, we have tried to lay bare the elements of a model system for the analysis of spatial process. We have attempted to

---

[55]Private employment volume and welfare payments are examples of rapidly responding attributes and government employment an example of a more stable one.

justify the use of a framework that considers the location of components of a system, the attributes of these components, and have also indicated some principles that should underly the rules which make the system operate through time. Some consideration must now be given to the problem of integrating these elements into a well-ordered model. This can be attacked in several ways; we choose to do so, here, from an operational point of view.

For most spatial systems of interest -- apart from pedagogical examples -- analysis and experimentation require the use of computing machines. This need not be due to the intellectual complexity of the system but may simply be the result of the presence of a large data base or a set of rules requiring a large number of operations. Once the decision is made to computerize a fully thought-out model, fairly standard generalized techniques are available for making it operational.

We can think of a digital computer as a machine for processing information under the control of a set of instructions written by the user. In our case, the instructions are mainly composed of our modeling rules recast into a computer language. The latter, unlike natural languages, have a logical structure and a helpful by-product of the transformation from rules to computer instructions is that it helps the researcher to spot mistakes in his original formulations. While most geographers have used standard languages such as Fortran for all their work, it is usually possible to find a language particularly well-suited to processing a particular type of information (see Appendix A).

Regardless of the language used, the conceptual structure of the string of instructions, usually called a program, is invariant. It is most conveniently described in terms of a series of blocks, or subroutines, each devoted to performing a single task (Figure 7). As indicated in the diagram, the sequence of control in most large programs is not linear; in general it is helpful to have a control block which calls on the use of sub-blocks where needed. This may be done several times in the program for the same sub-block; differences in the basic task, which might, for instance, be an operation performed on several different data structures during the course of the program, are indicated by supplying controlling parameters in the calling statement. Since control always returns to the calling procedure after a task is executed, the procedure allows a smooth flow of command. The building block approach to designing also facilitates testing and replacement of sub-blocks and the insertion of extra blocks when the requirements of the model expand.

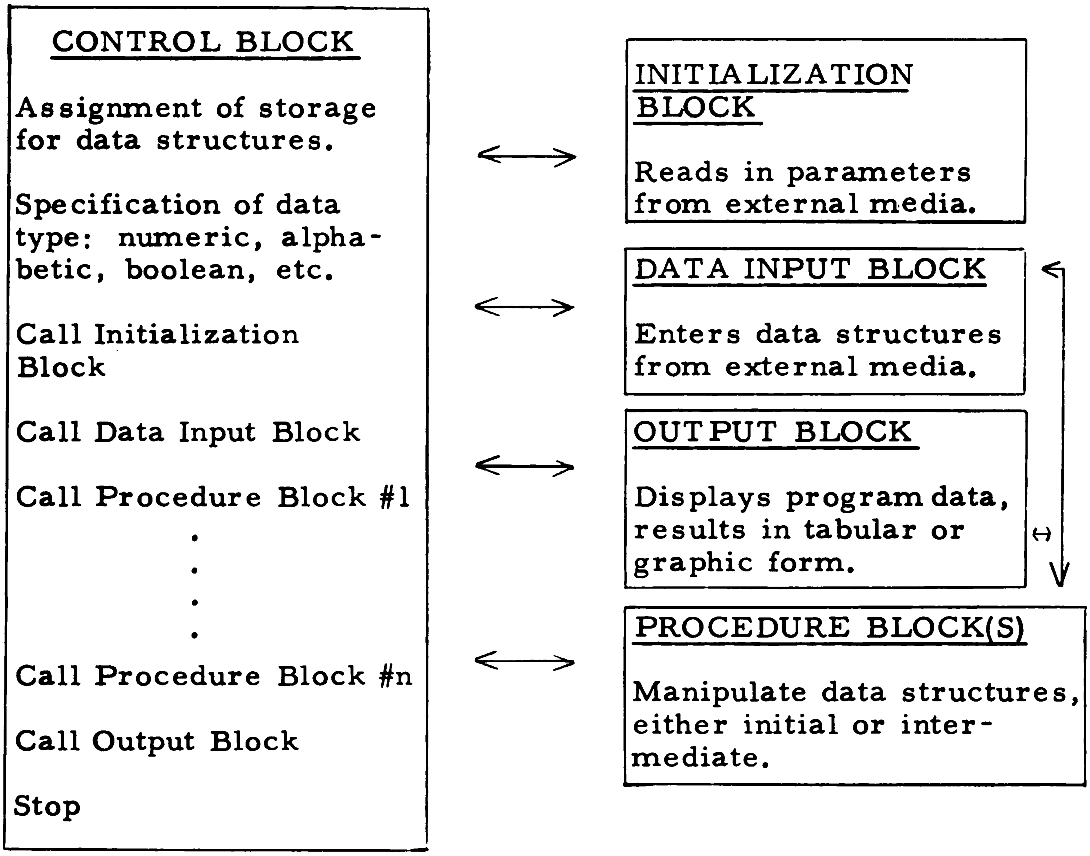


Figure 7. A Schematic Diagram of a Computer Program

CHAPTER IV

A WORKED EXAMPLE

Now that we have discussed the general structural characteristics of models of spatial processes, it would be helpful to see them applied to an example. The system we will examine is chosen from the field of medical geography, and is artificially created to dramatize only the essential elements of the process. The advantages of this abstract approach are simplicity and clarity. Only those spatial attributes deemed necessary to the process are present; they are not obscured by the multiplicity of variables present in the real world with the associated complexity of chains of cause(s) and effect(s), nor is the possibility of having left out important explanatory variables present. The process itself is also controlled and specified solely by our interaction rules; but, as we shall see, this need not mean our results are predictable or uninteresting.

The Spread of a Disease in an Isolated Region

The field of medicine contains many problems of interest to geographers. Among the most "geographic" of these is the study of the spread of infectious diseases over space. Thorough understanding of these processes requires skills of a high order. First, the geographer must understand the functional causes of contagion for any disease studied to a degree sufficient to make predictions about its spread. This will necessarily include knowledge of incubation time and the period of infectiousness if these are known or the ability to make reasonable estimates about them when they are not. He must then relate the characteristics of the disease agent to those of the environment, which requires all of the geographer's skills in classification techniques and model building. Finally, the results of the study have to be presented in a form that is understandable and which emphasizes the spatial pattern of the disease as it changes through time.

One might think that the field of epidemiology would provide models of such processes as they operate over space. But this is not the case since as Bailey notes[56] the major methodological work has been devoted to studies of infection rates and removal rates in a homogeneously mixed sample in a spaceless environment. The deterministic model of the geographic spread of a disease also presented by Bailey,[57] while a good basis for further development is conceptually quite sparse as it is limited to consideration of the spread of an infection over an infinite uniform plain with even population density. The population is assumed to be uniform in all respects but for the presence or absence of the disease. Coleman's presentation of a model of infection for incompletely mixed (i. e., socially stratified) populations[58] while extending previous models

---

[56]N. T. J. Bailey, The Mathematical Theory of Epidemics (London: Charles Griffin & Company Limited, 1957), chapters 1 and 2.

[57]*Ibid.*, p. 32ff.

[58]See J. S. Coleman, "Diffusion in Incomplete Social Structures," in F. Massarik and P. Ratoosh, eds., Mathematical Explorations in Behavioral Science (Homewood, Illinois: Richard D. Irwin, Inc., 1965), pp. 214-32.

to a more complicated universe does not pay attention to differentiation over space; that is, changes in the attribute space from place to place. So while the system modeled below is certainly too over-simplified for real world application it does involve more considerations of real conditions than is common in this area.

Statement of the Problem

In this example, we wish to investigate theoretically the spatial patterns resulting from the spread of an infectious viral disease through an isolated area. We are particularly interested in the spatial progress of contagion through the first several weeks of the outbreak since the strain mutates and dies out in a closed system after that time. The results of the analysis are to be used to guide further research by physicians into producing an effective vaccine for the prevention of the disease. An experimental vaccine has been given to a randomly selected group of villages in the study area and its effects on limiting the spread of infection will also be studied by physicians based on our results.

Construction of a model to investigate this problem requires information about the disease agent and about the environment in which it is assumed to operate. From this basic structure we then have to construct a set of rules that will allow us to monitor the dynamics of the process. The important problems involved in testing the results of models similar to this simple example are discussed at the end of the chapter but are not applied here because we lack a template against which to match our results.

The Disease Agent

Our agent is a virus which is spread by contact and only affects or is transmitted by humans. Continuous contact over a period of several days always results in transmission of the virus and subsequent development of symptoms which are unpleasant but not disabling or offensive to other persons. The incubation period of the disease is constant at six days; thereafter, the host is a potential source of infection to those he contacts for three weeks, after which time there is spontaneous remission. No subsequent reinfection occurs for at least a year.

Apart from total quarantine, two factors can hinder the spread of this virus. Assuming the contact period between a carrier and a contact is short, a day or less, a moderately high level of sanitation on the latter's part offers a 70 percent chance of catching the disease if exposed. A vaccine has also been developed which retards the spread of infection, not so much by preventing a vaccinated person's contracting the disease, although the symptoms are ameliorated, as by cutting down on the probability of his transmitting it to another person. Without the vaccine, an infected person will invariably transmit the virus; with it there is a 70 percent chance of his doing so.

The Environment

We assume our environment consists of a gently hilly region essentially closed to the outside world (Figure 8). The region contains 25 regularly shaped small village communities of about

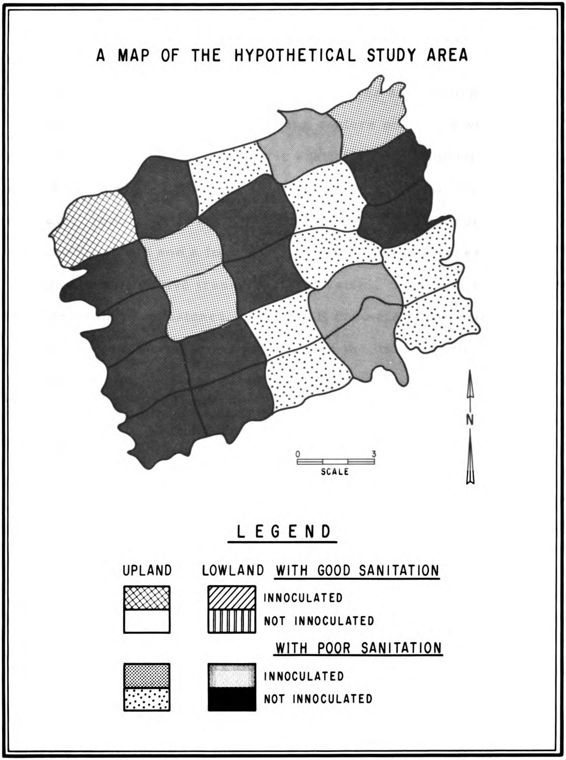

A MAP OF THE HYPOTHETICAL STUDY AREA
N
0
3
SCALE
LEGEND
UPLAND
LOWLAND
WITH GOOD SANITATION
INNOCULATED
NOT INNOCULATED
WITH POOR SANITATION
INNOCULATED
NOT INNOCULATED

Figure 8.

the same size, each with its surrounding agricultural and forest lands. Eleven communities are located on uplands with the remainder in the broad valley bottoms. Each community is basically self-sufficient but for religious purposes and trading, representatives regularly visit neighboring centers one day a week. The pattern of intercommunity visits is spatially quite random but the upland and lowland communities do have a strong tendency not to interact. In fact, only about one time in ten is there a contact between an upland and a lowland community, or vice versa. Because of a primitive transportation system, there is a rapid drop in interaction with distance. Half of all contacts occur between adjacent communities; about two fifths of the time a contact occurs between communities separated by another center. Only about once in ten times is there a contact with a center as much as three villages away. Finally, some communities are a bit more advanced than others in that they have piped water and primitive but adequately hygienic waste disposal systems.

Interaction Rules

The lack of a workable analytic formulation for this problem makes it necessary to use a probabilistic approach in our model to find an approximate spatial distribution over time. The fact that our basic information about both the disease and the environment is based quite largely on estimates of frequency reinforces the decision to formulate the model in a Monte Carlo framework.

Our rules then must include the probability of each community's receiving the virus in each week and the independent probability of being infected and in turn passing on the disease in the succeeding time periods. These probabilities depend on the geographical distance between the transmitting and receiving communities (in location space) and on their respective attributes. The following probabilities must be calculated from our data:

p_{dt} = the probability of a community at distance d and of type t receiving a contact from an infected community, and

p_{si} = the probability of successful transmission of the viral disease to a community with a sanitation standard s from a community with innoculation status i.

Since there is no dependency between physical distance and community type in our region we assume that $p_{dt} = p_d p_t$. Because the disease invariably affects everyone in a community once it is transmitted we need not consider the number of persons in each community nor the characteristics of those who travel to other communities. Furthermore, since the assignment of vaccine to communities was random and independent of local sanitary conditions we can also assume that $p_{si} = p_s p_i$.

The actual probabilities based on our raw data are as follows:

for contact at distances d_1, d_2, and d_3 the probabilities are respectively 0.5, 0.4, and 0.1 where the d_i's refer to the successive rings of communities around the transmitting center since there is no directional bias in the system;

the probability for contact to occur between communities of the same type (upland or lowland) is 0.9 and the probability of contact between different types is 0.1;

the probability of contracting the disease in a community with adequate sanitary facilities is 0.3 and in the absence of these facilities, it is 0.9; and

the probability that a person who has not been innoculated will transmit the virus is 1.0, whereas if he has been innoculated the transmittal probability drops to 0.7.

We indicate the possible combinations using a table:

Table 4. Combined Probability of Infection in a Contacted Community

Contacted Community	Transmitting Community	
	Innoculated	Not Innoculated
Adequate sanitary facilities	0.21	0.30
Inadequate sanitary facilities	0.63	0.90

Since our region is isolated, we can treat its boundaries as reflecting barriers which simply means that instead of contact circles around each potential transmitting community we may have to assume contact arcs. Also, to simplify our process somewhat we make the assumption that we are only interested in contacts from infected villages to other communities. This could result in the real world from a situation where visitors would not approach an infected community but where the latter's inhabitants would continue their usual visiting habits since they would realize catching the disease twice is not possible.

Our time periods (generations) consist of weeks since this is the visiting pattern and we restrict our time frame to a period of five weeks because of the mutation characteristics of the virus. The initial appearance of the virus is assumed to occur in the central community. We will run the model twice to obtain a visual impression of the pattern of spatial spread and to examine the possible influences of distance, village type and sanitation. Then, in order to average out the run-to-run variation inherent in Monte Carlo processes we will examine the aggregate results from 100 simulations.

Making the Model Operational

Two simulations of the system were made using a specially written computer program (see Appendix B) incorporating a pseudo-random number generator to emulate the probabilistic characteristics of the model. The program was written using the building block approach indicated in Figure 7.

The initial program block was used to describe the structure of the data and to enter as program parameters the probabilities mentioned above, the number of system time units to be run, and the number of simulation runs to be made. If more than one simulation run is requested, the initial block also sets up conditions for repeating the model as many times as required. Because of the arrangement of the sub-cells of the region, the communities, it proved possible to describe (and later enter) the locational and attribute data in the form of a square 5 x 5 matrix with each cell representing a community and containing four bits of binary information indicating whether it was:

upland (0) or lowland (1);

innoculated (0) or not innoculated (1);

possessed of adequate sanitary facilities (0) or not (1), and

free of infection (0) or infected (1).

The second block of the program was used to enter the data, store it, and also to key the middle cell with an indication that it was infected with the disease. The map (Figure 8) was transformed to coded form on cards which are interpreted by the program as represented below:

0000	1110	0110	1010	0010
1110	0010	1110	0110	1110
1110	0010	1111	0110	1110
1110	1110	0110	1010	0110
1110	1110	0110	1010	0110

The several procedure blocks in the program had the following tasks in order:

1. Select an infected community from the previous time period and initiate the process of contact for each of them repeating the following procedure blocks for each such community. In essence, step one is called by the main control block and then serves as a master for each of the following procedure blocks.

2. Select one of the three possible rings of surrounding communities to contact by drawing a random number from 0 to 99. If the number is between 0-49 contact is made in the closest ring, between 50-89 in the next closest ring, between 90-99 in the third ring.

3. Drawing a random number between 0-99, determine whether contact is to be made with a community of the same (0-89) or different (90-99) type. Find one such community in the ring with equal preference given to each.

4. If the community is not already infected, its probability of becoming so depends on whether it has adequate sanitary facilities and whether the transmitter is from an innoculated community.

The job of the output block was to produce a simplified pictorial representation of the spread of the disease for each of five generations and an indication of the pattern of spatial contact, both those which result in the spread of the infection and those which do not. Additionally, for the purpose of testing the randomness of the process (if this is desired as a check), the random digits generated can be printed. For the case where such output would be too voluminous, as when several hundred simulations of a process are desired, the output block can be instructed to provide only overall frequencies of contact and other summary statistics.

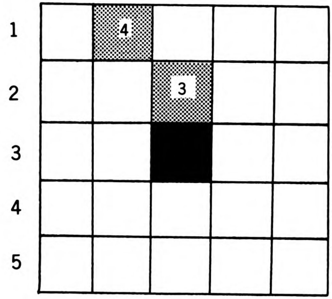
Results of the Model Runs

The results of the two simulations of five generations each are presented below (Figure 9 and Table 5).[59] The most significant

---

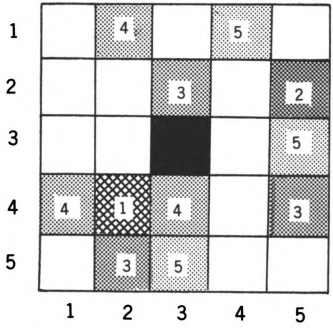
[59]The original map of the area has been transformed to a cartogram with grid references for convenience.

SPATIAL PROGRESS OF THE DISEASE



SIMULATION 1

4 Indicates infection in fourth week



SIMULATION 2

Figure 9.

Table 5. Results of Two Simulations of Five Weeks Each

Week	Carrier Village	Contacted Villages	Infected Villages	Sequence of Random Numbers
Simulation 1				
1	3, 3	1, 1	--	98, 88, 1, 53
2	3, 3	4, 2	--	35, 4, 7, 95
3	3, 3	2, 3	2, 3	19, 48, 2, 33
4	2, 3	1, 2	1, 2	11, 26, 1, 33
5	1, 2; 2, 3	2, 3; 3, 3	--	35, 49, 3, 14, 8, 6
Simulation 2				
1	3, 3	4, 2	4, 2	61, 45, 7, 68
2	3, 3; 4, 2	2, 5; 1, 2	2, 5	45, 75, 5, 10, 37, 91, 2, 95
3	2, 5; 3, 3; 4, 2	4, 5; 2, 3; 5, 2	4, 5; 2, 3; 5, 2	94, 96, 6, 89, 20, 11, 2, 86, 16, 8, 6, 6
4	2, 3; 2, 5; 4, 2; 4, 5; 5, 2	1, 2; 2, 5; 4, 1; 4, 2; 4, 3	1, 2; 4, 1; 4, 3	75, 65, 2, 9, 89, 44, 2, 40, 87, 11, 70, 95, 93, 12, 91, 14, 2, 66
5	1, 2; 2, 3; 2, 5; 4, 1; 4, 3; 4, 5; 5, 2	3, 5; 1, 4; 3, 5; 5, 3; 5, 3; 4, 5; 5, 2	3, 5; 1, 4; 5, 3	71, 93, 7, 27, 36, 3, 23, 46, 12, 4, 97, 59, 6, 70, 34, 39, 6, 18, 40, 2, 11, 15, 3

impression received from the output is the totally different pattern presented by each simulation. While such large run to run (or more precisely, sample to sample) differences are most apparent in processes with stringent criteria for success, they are characteristic of all Monte Carlo procedures.[60] This is one reason why we should interpret results presented on the basis of one or two simulations[61] with caution. To find "expected" distributions, we must typically run several dozen experiments, in this case simulations of our process, under identical initial conditions but with different selections of random numbers.[62]

In this case, we have run the model 100 times and present the results in frequency form (Table 6). Here, the cell entries represent number of times we would expect a given village to become infected were we to replicate our process this many times. It would also be possible to convert the frequencies to relative frequencies by dividing each cell entry by the total number of infecting contacts summed over the 100 samples. To emphasize this element of variability in small samples, we also generated and present (Figure 10) the expected number of new infections per generation for 2, 50, and 100 simulations of a similar process occurring in a larger area over 30 generations. The point to make here is that

---

[60] J.M. Hammersley and D. C. Handscomb, Monte Carlo Methods (London: Methuen & Co. Ltd., 1964), chapter 1 and pp. 55-74.

[61] See, for instance, R. Morrill, *op. cit.*

[62] For an example in a slightly different context see J. Herniter, A. Williams, and J. Wolpert, "Learning to Cooperate," Papers of the Peace Research Society, VII (1967), pp. 67-82.

Table 6. Frequency of Infection Summed over 100 Simulations Using the Sample Area*

1	5	44	13	44	11
2	17	10	45	13	49
3	29	21		16	44
4	30	44	14	74	17
5	27	54	20	60	7
	1	2	3	4	5

*Total frequency sums to 708.

For identification of cells by characteristics see Figure 8.

SMOOTHING OF MEAN VALUES WITH MORE SAMPLES

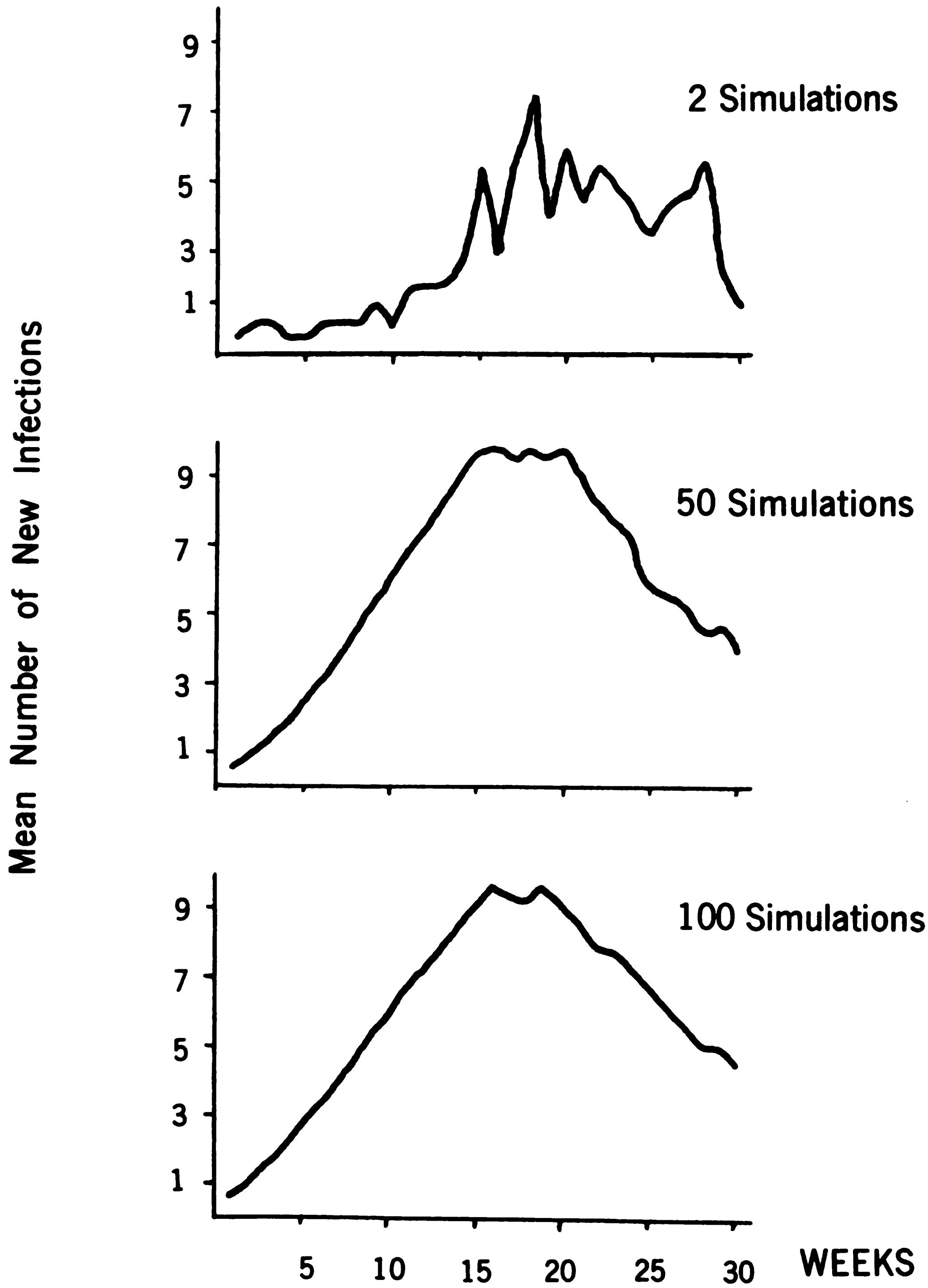


Figure 10.

as we accumulate more information about the process under investigation by obtaining more samples, we come closer and closer to some bounding values and begin to detect true regularities rather than the random aberrations which can occur with only two samples. If our distribution of rates of infection over time is normal[63] then we can also indicate, using confidence bands based on the standard deviation, the confidence we have that any specific value at a point in time will occur with a certain margin of error. Of course, since the pattern we find in real life is essentially one sample drawn from a large or infinite number of possible ones, it may well turn out that our expected or average value (of, say, the number of villages infected in a given week) is quite different from the actual one. But in terms of predicting the likelihood of a pattern, the expected value is the best guess we can make in advance.

Returning to the results of the two simulations, we find the following patterns (Table 7) of infection as related to three variables of interest: type of village, sanitary facilities, and distance from the original source of infection. We can interpret this information in a variety of ways given our initial conditions and the characteristics of the original source of infection. The importance of the barriers to inter-group communication is evident in the small number of upland centers infected. Given the almost equal representation of the two types, 14 lowland villages and 11 upland villages, we might expect a more even distribution of infection; but

---

[63]We can test this by using the X^2 statistic or the Kolmogorov-Smirnov test.

Table 7. Two Simulations: Classification of Infected Villages

Simulation	Village Type		Sanitary Facilities		Distance from Origin	
	Lowland	Upland	Present	Absent	1	2
1	2	0	0	0	1	1
2	8	3	0	11	3	8

the low interaction rate mitigates against this. In the context of diffusion processes where the interest is in means of accelerating the spread of a phenomenon, we have an indication that information on intergroup interaction rates may be critical. With a situation similar to this, vis-à-vis interaction rates, it would seem almost mandatory to initially "plant" the thing to be spread in each of the groups involved. Where we are interested in retarding the spatial spread of a process as here, an efficient strategy would be to attempt to intensify the effect of existing barriers to communication.

At first glance, it would seem that the effects of sanitation are even more pronounced than those of social group. This may be so, but our evidence cannot be conclusive in this environment because of two confounding effects: first, the village having sanitary facilities is of the upland type which has fewer transmittals anyway, and second, the location of the village is peripheral which lessens the probability of its being contacted. The effects of distance in the model are also somewhat ambivalent. While we note the presence of an edge effect as would be predictable from general diffusion theory, there are also voids -- villages which escape infection -- close to the original source of infection and to subsequently infected villages. The theoretical role of distance in models of similar processes, then, should not be assumed to be significant *a priori*. While a distance decay factor may be of general importance in spatial models, in any particular case we may find it of minor significance or even, given a particular distribution of other attributes over the area being investigated, find an inverse distance effect.

Turning to the larger sample of 100 simulations, we find a total of 708 successful transmissions of the disease, an average for each simulation of 7.08. This gives some perspective to the totals for the individual simulations described above of 2 and 11 transmissions. In terms of the same variables we used above, the following pattern emerges:

Table 8. 100 Simulations: Classification of Infected Villages

	Village Type		Sanitary Facilities		Distance from Origin	
	Lowland	Upland	Present	Absent	1	2
Actual Number	561	147	5	703	237	471
Percent of all villages	56.0	44.0	4.0	96.0	32.0	64.0
Percent of all Infections	79.23	20.77	0.71	99.29	33.47	66.53*

*The average number of infections per cell in the inner ring is 29.5. In the second ring the average number is 29.

There are a number of ways of testing the importance of each of these variables as contributors to the pattern of spread found by our model. With the three variables above and under our assumption of their independence we might find in a practical case some use in constructing a multiple regression model to test the importance of each of the independent variables as predictors of infection in a village.[64] Effects of different origins of the infection could also be tested by comparing the standardized beta weights on each variable. For this simple example it was decided not to follow such a course; instead, a X^2 statistic was constructed for each variable to test whether the difference between the observed frequency (actual infections) and expected frequency under a hypothesis of independence (expected frequency being equal to the proportion of villages represented by the trait examined) was significant. The only variable that proved to be in the critical range of the test (at the .05, .01, and .001 levels) was village type. Differences between observed and expected values for distance and the presence of sanitation facilities were not significant at any level.

So with the greater amount of information given by 100 simulations, we are essentially in the position of being able to say that under the conditions of our experimental environment, the only variable which has a consistent effect on the spread of our disease

---

[64]Of course, to make any probabilistic or inferential statements about our results, the residuals should be tested for normality, independence, and an expected mean of zero. See K. W. Smillie, An Introduction to Regression and Correlation (New York: Academic Press Inc., 1966), especially chapter 1 and pp. 72-75 and 91-96.

is village type. The edge effects and those internal villages without infection in our first two simulations and the villages with very low frequencies of infection in the large sample of 100 simulations are seen to be functions of the different environments, lowland or upland and more importantly of low contact frequencies between the two types.

What would be the next step in using this simple model to explore the effects of those variables of interest on the spread of this disease? We can easily investigate the sensitivity of the model by altering our basic probabilities; for instance, by changing the probability of intergroup communication from .10 to say .30 we could get a good indication of the effects of freer movement in our experimental area. Other probabilities, for distance, or those related to sanitation could also be changed; if we wished to investigate the changes resulting from the development of a vaccine offering some protection against infection this would also be easily done. Beyond changes of this type which are easily made, it would be conceptually simple (but would offer some programming difficulties) to add more groups to our population, to enlarge the size of the area under investigation, to vary the period during which a person would be a carrier of the disease, and so on.

Problems of Testing and Verification

So far, we have not discussed in any detail the substantial problems encountered in testing or verifying models of spatial processes. In large part this is because ". . . the problem of verifying simulation models remains today perhaps the most elusive of all the unresolved problems associated with computer simulation techniques."[65] While we cannot, here, present definitive solutions to these problems we can identify them, point out some of their principal causes, and suggest some possible approaches towards their solution.

The difficulties encountered are both general ones that apply to all problems of testing and specific ones of particular moment to the types of models we have been discussing. At the heart of the general problem is the implication of the concept of verification itself, for:

> To verify or validate any kind of model means to prove the model to be true. But to prove that a model is 'true' implies (1) that we have established a set of criteria for differentiating between those models which are 'true' and those which are not 'true' and (2) that we have the ability to readily apply these criteria to any given model. Yet the concept of 'truth' has successfully eluded philosophers and theologians since the history of mankind [sic]. To decide upon a particular set of criteria that must be satisfied before we can have 'truth' suggests that we must choose a subset of rules (truth rules) from an infinite set of rules handed down by philosophers, theologians, and metaphysicians. When placed in this perspective, the problem of verification is completely overwhelming because it may well be argued that man is incapable of recognizing 'truth' at all, even if 'truth' exists.[66]

The selection of a set of criteria for determining the adequacy of our models is another source of difficulty. Here we may adopt any of a number of philosophical approaches ranging from a position

---

[65] Thomas Naylor, Joseph Balintfy, Donald Burdick, and Kong Chu, Computer Simulation Techniques (New York: John Wiley & Sons, Inc., 1966), p. 310.

[66] Ibid.

that a satisfactory model is one that is logically deducible from a series of "self-evident" premises or axioms to a crude empiricism that accepts any model that can produce satisfactory predictions of the behavior of the system being studied.

In the particular case of models of spatial processes there are at least two major problems. One of these is the selection of a "norm" against which the model's results can be compared. In the case of a Monte Carlo model of diffusion, for example, it is not at all obvious as to how the researcher can conclude that the spatial pattern output from the model is "good." The second problem is encountered when the researcher considers the use of statistical tests to evaluate, for instance, the effects of certain variables on the system under investigation. In this case we must consider the fact that models of any spatial process except purely random ones embody both time and space dependence while statistical tests, both parametric and nonparametric, are based on the assumption that observations must be independent in the sense that the value of a variable for one observation should not bias the value assigned to any other observation.[67]

A pragmatic solution to the general problems mentioned above has been suggested by Naylor and Balintfy.[68] They advocate the use of a rather eclectic multi-stage procedure for verification consisting of: 1) the formulation of a set of postulates or hypotheses describing

---

[67] Sidney Siegel, Nonparametric Statistics for the Behavioral Sciences (New York: McGraw-Hill Book Company, Inc., 1956), p. 19.

[68] Naylor, et al., op. cit., pp. 316-19.

the behavior of the system of interest; 2) an attempt to "verify" the soundness of these postulates using applicable parametric and non-parametric statistical techniques, and 3) the testing of the model's ability to predict behavior of the system over time. Possible approaches to such testing include matching the model's output, generated from historical data, with actual values. Goodness of fit tests would be used to compare generated and actual time series for timing and amplitude. But the ultimate test of any simulation model, in their view, is its ability to predict the future behavior of the actual system being studied. The general approach of these authors seems reasonable, if perhaps difficult to accomplish in all details. One modification that would have to be made for spatial models would be to devise a goodness of fit test that would compare spatio-temporal series.

One obvious solution to the problem of selecting a standard against which to compare the results of spatial models is to initially operate with historical data and visually compare generated and actual maps. If satisfactory calibration is achieved, the model could then be used as a predictive tool for similar systems. This procedure, while conceptually and operationally simple, has several drawbacks. First, patterns similar to those of the real system may be generated in a particular case even though the rules of the model may bear little resemblance to those of the real world. This would not be a cause for anxiety if similar patterns were always generated by the model and the real system. But this might not be the case and one would certainly be ill-advised to use a model tested in this manner for prediction. Second, there is the problem of deciding when two patterns are "close" enough to be considered similar. Visual comparison, even of patterns of numbers, is usually inaccurate and difficulties are compounded with the two-dimensional patterns with which geographers are concerned.[69] There are, of course, various statistical techniques that are in practice used to compare such patterns but it is at this point that we encounter the basic problem that the surfaces being compared will usually embody spatial and temporal dependence.

Except for those cases where the spatial process itself (as opposed to any patterns it generates) is fundamentally random, we should, then, exercise considerable caution in making inferences based on the application of standard statistical techniques. This would be so even when our data is selected through the use of some suitable sampling technique[70] and the use of inferential statistics without sampling is, of course, not to be countenanced.

An indirect approach to testing models of diffusion has recently been advanced by Harvey;[71] it appears to be promising and can also be used with models of other spatial processes. The

---

[69]Harold H. McCarty and Neil E. Salisbury, "Visual Comparison of Isopleth Maps as a Means of Determining Correlations Between Spatially Distributed Phenomena" (Iowa City, Iowa: Department of Geography, The University of Iowa, mimeographed, 1961).

[70]Brian J. L. Berry and Alan M. Baker, "Geographic Sampling," in Brian J. L. Berry and Duane Marble, eds., Spatial Analysis (Englewood, N.J.: Prentice-Hall, Inc., 1968), pp. 91-100.

[71]David Harvey, "The Analysis of Point Patterns," Transactions of The Institute of British Geographers, XL (1966), pp. 81-95.

technique is based on the analysis of patterns resulting from the operation of the system. Elements of the pattern are sampled using the quadrat sampling techniques developed by ecologists. The resulting distribution is then compared with some standard generating function such as the Poisson or negative binomial using some distribution-free test such as X^2. If the fit is satisfactory then the process can be classified as belonging to a particular group of models that generate such patterns. The advantages of this method include its relative objectivity and the opportunity it gives the researcher to compare patterns produced by the same model in different regions or with changes in various parameter values. In the latter case the method provides a means for testing sensitivity of the model (and by implication, the system), to changes in these values. Also of considerable importance, if the model matches some standard generating function it is possible to utilize knowledge about analogous models in explaining the system being studied. Still, neither this procedure nor the other approaches mentioned are optimum since what is needed is some means of testing the dynamic aspect of our models and not only the patterns generated by them. Unhappily, this type of test or verification appears to be beyond our reach at this moment.

CHAPTER V

CONCLUSIONS

We have attempted to explicate and demonstrate the existence of a common structure underlying models of spatial processes. It is argued that this structure consists of three major components: an attribute space, a location space, and a set of rules which link these components and simulate the dynamics of the system. An examination of models of spatial processes developed by Georg Karlsson and by Torsten Hägerstrand and other geographers indicated that our framework is basically consistent with and a logical development of their (often implicit) assumptions about the nature of these processes.

After a discussion of some of the more important elements of each of these components, we turned to a consideration of the problem of their integration into functioning models of the system being studied. For pragmatic reasons we advocated a block structure approach which is based on techniques that have proved successful in the construction of many computerized models. Since many, if not all, analytical studies of real world spatial processes will, in the future even more than at present, depend on efficient utilization of computers it seems reasonable to take this factor into consideration at the very beginning.

To demonstrate in a simplified way the process of disaggregating a complex system into the above components, we constructed an abstract example based on considerations relative to the spread of a disease in an isolated region. The succeeding steps of integrating these components into a working model of this process, then presenting the effects of its operation over time and space were also shown. Because of the nature of the data available, the rules in our example were cast in a Monte Carlo framework. Although this probabilistic approach is not necessary in all cases, we feel that because of the complex nature of the processes of interest to geographers and because of the form of most of our data it is likely to become the most common.

Our exposition has not touched on some important points. The most critical of these is the general problem of framing the initial questions asked about some process in such a manner as to guide our initial selection of a study area and to suggest the particular sub-components of our model. If we can use the experience of the physical sciences as an analogue, it seems reasonable to think that to a considerable degree future progress in geographical research will be intimately dependent on developing a standardized and effective approach to answering the problem of what questions should be asked and how they should be posed. Schemes, such as the one advocated here, B. J. L. Berry's geographic matrix, or Chorley's model of models can hopefully provide some guidelines. But we are, at this moment, rather far away from developing answers to such question-posing choices.

There are also severe problems in geographical work connected with questions of measurement of data and tests of the results of our models. Some general principles to guide our choice of measurement were discussed above (in Chapter III) but here again we need to develop standards. Most especially, it would be useful and may indeed be essential to fundamental progress to agree on definitions of fundamental units of measurement for spatial processes. There is some reason to feel that the physical systems approach developed by H. Koenig and his associates[72] which classifies fundamental variables as either flow (through variables) or potential (across variables) can provide us with the guidelines we need in this area.

In presenting the results of the simulation runs on the example in the last chapter, the problem of testing was mentioned. Noteworthy advances have been made in the past decade in attacking problems of testing geographic models. Most of them, however, are most specifically applicable to the analysis of patterns. The methods used include classificatory statistical models such as factor analysis and numerical taxonomy, analysis of variance or covariance for comparing populations or regions on the basis of several variables and, of course, the general multivariate regression model. Some of these techniques are useful also in testing the results of dynamic models of processes. We could for instance simulate the performance and behavior of a system using different estimates of certain attribute values. The areal patterns resulting from each set of assumed conditions could then be factored and

---

[72] J. B. Ellis, op. cit., chapter 1. Also H. Koenig, Y. Tokad, and H. Kesavan, Analysis of Discrete Physical Systems (New York: McGraw-Hill Book Company, 1967), p. 7.

scored and the resulting scores compared using regression techniques or an analysis of variance model. It would then be possible, in theory at least, to evaluate the importance of initial assumptions as they affect the behavior of a system over space.

In many instances, though, we find the standard tests are not optimum (or even adequate) for dynamic processes. This is largely a result of their having been developed for problems in other disciplines where the spatial factor is not assumed to be important. In using them in geographical research we may not only be employing a technique that is inefficient in some way but may also be guilty of ignoring fundamental assumptions built into the test model. For the regression model the major assumption is that the value of each observation on the dependent variable is one random observation from _n_ different normally distributed variates. This implies that there is no serial correlation between adjacent observations. Where our observations are geographical units, this assumption is often not met as there _is_ dependence between adjacent regions. Any inferential use of results from regression models, then, may be invalid although the model is still useful as a descriptive device. The solution to these testing problems must depend on more geographers turning their attention to them and lessening slavish dependence on the work of others whose models may give erroneous results when applied to our data.

Leaving these, as yet unsolved, problems we must finally ask ourselves what is the utility of the conceptual framework herein advocated for the study of spatial processes. If we accept the value of retaining a certain amount of continuity in the traditions of a discipline then we find that our scheme is evolutionary and does no violence to these traditions. What it does do is lay out explicitly the nature and the fundamental structural building blocks of spatial processes.

The advantages of such a conceptualization are both theoretical and practical. In the first instance, the acceptance of a common structure places under the same rubric processes which at a superficial level appear disparate. This encourages a search for fundamental general rules of behavior applicable to all or to major subclasses of spatial processes. It also would have the effect of countering trends towards particularity that tend to afflict fields such as geography that are undergoing revolutionary changes in methods and to a lesser extent in theory. On the practical side we would expect to encourage the search for fruitful analogies whereby results from processes that are relatively well understood would be tentatively used to throw light on aspects of other processes that are mysterious in the original. If we look at the history of other sciences, such analogies have been of the utmost importance in their development.

APPENDIX A

SOME COMMENTS ON COMPUTER LANGUAGES

Geographers, in seeking to solve problems, gain insights into spatial processes and develop theory, are increasingly turning to digital computers as a useful and sometimes indispensable tool.[73] They are especially valuable as an aid in understanding and manipulating systems with large numbers of variables, poorly defined processes and either too many or too few observations. The advantages that computing machines have over more traditional tools are increased speed and accuracy, the ability to store data and intermediate results and instructions, and very importantly the ability to make logical decisions about data handling based on programmed logic.

The rub in all this is the fact that computers must be fed not only raw information, but instructions as to what to do with it. These instructions must be internally consistent, provide for all possibilities inherent in the data (check for errors and so on) and be written not in the natural language of the researcher but in a language acceptable to the machine. This imposes a new requirement on the scholar wishing to use the power provided by computers: learning one or more computer languages. It is, of course, often possible to use service programs (standard statistical "packages"

---

[73] Analog computers have also been used for special purposes but the present discussion deliberately concentrates on digital machines because of their wider applicability.

provided by the manufacturer or the computing center) or employ a professional programmer; even in these instances, though, some knowledge of computers and their languages is useful. The choice of which computer language(s) to learn is based on two considerations: what is available at the local computing center, and what is the best language for the application. The remainder of this appendix is devoted to the second consideration.

COMPUTER LANGUAGES. There are presently more than 200 computer languages. At first glance this seems to make a choice among them somewhat difficult, but fortunately the babel of tongues can be resolved into four reasonably self-contained groups: (1) machine and assembly languages; (2) algorithmic languages; (3) list processing and simulation languages, and (4) others. The basis for the choice of the tool language(s) is presented through the description of the features of each group.

MACHINE AND ASSEMBLY LANGUAGES. Each computer model has its own unique instruction set based upon the logic elements wired into the machine and the purpose for which it was designed. For digital computers the generally used dichotomy is between machines designed for data-processing and those designed for scientific computing. The former emphasize the processing of records (e. g. , a personnel file) which requires efficient and extensive input-output instruction sets and good capability for logical operations. The latter are designed to operate on sets of variables (e. g. , sample survey data, readouts from remote sensing apparatus) which requires a powerful arithmetic instruction set and efficient indexing instructions but places less weight on sophisticated input and output capabilities. However, these distinctions are blurring and the more modern machines can be and are used for both purposes. The one-to-one coupling that exists between the machine-language and its computer has several important programming implications, notably:

(a) Machine-dependence. This means that, in general, a program written in machine-language for one computer will not run on another model. For the researcher, it also means that knowledge of a machine-language is basically a non-transferable asset although it provides insights into computer structure.

(b) Program complexity. Since the basic instruction set of a computer is made up of so-called elementary operations, instructing the machine to perform even simple tasks requires the programmer to string together a number of these operations. For instance, the statement A = B - C would require three instructions:

load B into the arithmetic register;

subtract C from the contents of the arithmetic register, and

store the contents of the arithmetic register in A.

Additional statements would be required if there were a mix of integer and decimal values since computers store them differently. Of course, it is initially necessary to learn all machine codes for the operations; these codes can be octal, decimal, or hexadecimal depending on the machine used. Setting up a "filing system" (assignment of internal storage space) for data and instructions is also part of the programmer's job when operating at this language level.

(c) Optimal machine efficiency. Machine execution time for a program can be optimized by using machine-language. As an example, a single instruction on the Control Data Corporation 3600 can search a list to find a threshold value or one equal to some preset quantity; the same operation in an algorithmic language such as Fortran would take at least four statements and generate 20-30 machine instructions.

(d) Difficulty in correcting errors or modifying the program. Since the computer storage allocation for instructions and data are made by the programmer and are usually assigned sequentially, changes to the program (such as adding a new block of instructions) can be made only with some difficulty. "Debugging" or correcting programs written in machine-language is also not an enviable task because of the awkward coding format and the high chance of mechanical error inherent in writing numeric strings of instructions.

In practice, the theoretical advantage of machine-language in optimizing computer execution times is negated by the accompanying increase in time required to write and debug the program. To avoid some of the programming problems associated with machine-languages without losing their real efficiencies, a class of languages known as assemblers were developed starting in the early 1950's.

An assembler is actually a computer program provided by the manufacturer of the machine whose basic task is to translate other programs written in the assembly-language into the language of the machine. Like machine-language, assembly-languages are different for each computer, although these differences may be negligible within any one series of models such as the Control Data Corporation's 3000 range. They also allow a programmer to use any peculiar feature built into the computer that might serve to increase program efficiency.

While assembly-language programs generally consist of the same strings of elementary operations that characterize machine-language programs, pre-written blocks of instructions that perform certain commonly-needed tasks (finding square roots, logs, etc.) are provided in the manufacturer's assembler program; these blocks of code can be included in a user's program by writing a single instruction called a "macro" or "pseudo" instruction. Also, operation codes and variables can be referred to symbolically rather than numerically (e.g., ADD A rather than 51 073251) which eases programming and tends to minimize coding errors. Other substantial advantages of assembly over machine-languages include the automation of most "book-keeping" tasks (automation of storage assignments for data files and single variables) and the provision of automatic debugging aids for program testing and correction.

ALGORITHMIC LANGUAGES. A fundamental difference between the machine and assembly-languages and the languages described below is that the latter are problem oriented. This implies that the user can communicate with the computer in his rather than its language. Although complete naturalness has not been achieved, some of the problem oriented languages come very close to eliminating the language barrier to communication. Problem oriented languages have one other great advantage for the user or programmer; a program written in such a language can be run on a different computer with few modifications. Because of this, it has been possible to disseminate widely used types of programs (especially statistical programs for such techniques as factor analysis, regression, analysis of variance, etc.) throughout the computing field.

Many, if not all, statistical and numerical problems can be solved by methods consisting of several expressions or formulas to be evaluated using a given set of data. Algorithmic (or algebraic) languages take advantage of this fact and a programmer working with one of them writes his program as a series of equations including where necessary data input and output commands, decision statements, and control statements. Pre-programmed statement blocks (or sub-programs as they are usually called) to evaluate roots, trigonometric and logarithmic functions, to generate random numbers and so on are embedded in these languages and can be used with great facility.

The two best known algorithmic languages are Fortran[74] and Algol.[75] It is safe to say that more persons have been introduced to programming by these languages and their variants than by any other and this dominance seems likely to continue since most high school and college courses in computer programming are taught in them. A somewhat unfortunate result of this dominance is that programs are written in Fortran, for example, that could be programmed more easily and more efficiently in some other class of language (see under Simscript and COBOL below).

As a consequence of the wide range of applications to which they have been applied, Fortran and Algol have undergone continual modification. This has proceeded in two directions: revisions of the "standard" language and development of variants to suit local needs. In some instances, the resulting languages are non-compatible.[76] Fortran especially has been extensively modified and is a much more powerful and sophisticated language than it was when developed by the IBM Corporation in the middle fifties. In fact, the deficiencies of the early Fortran systems -- slow translating times from the source language to machine-language;

---

[74]Especially good introductions are found in Donald Dimitry and Thomas Mott, Jr., Introduction to Fortran IV Programming (New York: Holt, Rinehart and Winston, 1966), and in Donald J. Veldman, Fortran Programming for the Behavioral Sciences (New York: Holt, Rinehart and Winston, 1967).

[75]See "The Algol Programming Language," and "Revised Report on the Algorithmic Language -- Algol 60," in Saul Rosen, ed., Programming Systems and Languages (New York: McGraw-Hill Book Company, 1967), pp. 48-117.

[76]Elliott I. Organick, A MAD Primer, privately printed, Houston, 1964.

inefficient and restricted input and output facilities and heavy dependence on IBM hardware design features -- led to the development of the other major algorithmic language, Algol, in 1958 by an international group of computer experts. The current version of that language, Algol 60, in fact is <u>the</u> medium for communication of programming information among the computing fraternity. Two theoretically good features of Algol make it relatively more difficult to learn than Fortran: a rigidly defined syntax which has to be memorized, and the existence of three subsets of the language: a reference language very close to mathematics; a publication language, and several hardware languages.

In most cases the user benefits from continual improvement in computer languages. But for those engaged in exchanging programs to build "libraries" or those moving to places with different computer systems, the blessings are mixed because of the resultant increase in communication problems.

LIST PROCESSING AND SIMULATION LANGUAGES. Certain kinds of information processing problems, notably in the areas of artificial intelligence, simulation of thought processes, mechanical translation, information retrieval and operations research cannot be handled easily or efficiently by the previously mentioned types of programming languages. Two types of operations characterize these problems: manipulation of symbols rather than numbers, where information is carried by the relational structure as well as symbol content, and unpredictable storage requirements which require addition or deletion of storage cells as a program is executed. Several types of geographical applications share these characteristics. Probabilistic models of regional development where places add or delete functions or facilities, become more or less accessible with changes in the transport network or change socially or morphologically in some random manner are a case in point.[77]

There are several well-known list processing languages which have potential geographic applicability. Among them are: IPL-V, COMIT, LISP, and SLIP. One simulation language, Simscript, has several features which should commend it to those working with spatial processes. SLIP is and Simscript has embedded in a Fortran compiler (translator) allowing them to use all the strengths of the algorithmic language. The others are independent languages. These latter suffer uniformly from rather poor arithmetic capabilities and have some peculiar input and output restrictions. Rather than go through an exhaustive description of each of the above languages which is available elsewhere,[78] we will concentrate on IPL-V and Simscript which are quite widely available and are good examples of the two types of languages described in this section.

Information Processing Language, Version Five. Development of IPL began in 1955 by researchers, notably at the RAND

---

[77]Morrill, Migration and the Spread and Growth of Urban Settlement, op. cit.

[78]Rosen, op. cit.

Corporation, who were interested in developing computer programs for the simulation of cognitive processes and the study of artificial intelligence. Green[79] gives a cogent description of IPL from which the following account is adapted.

The language is designed to be interpreted by a special computer program which turns the machine being used into an IPL computer with a flexible memory structure and the capacity for executing recursive subroutines (sub-programs defined in terms of themselves). Programming is relatively simple because a few instructions can accomplish a lot. The language is hierarchical which leads naturally to building a program out of small routines; combining these into larger routines is simple because little must be done to assure communication. Special push-down storage lists which operate like plate-trays in automats and cafeterias make communication simple and debugging is aided by trace, snap-shot and post-mortem routines which print out information in the symbols used by the programmer wherever possible.

The IPL system processes information by manipulating lists of symbols. Every item of information is represented by either a symbol or a list. Each symbol or list may likewise name further lists, so a single item may be represented by a hierarchy of lists called a list structure. The lists and list structures are manipulated by IPL instructions which are themselves represented by lists with branched decision-points as needed. There are only seven

---

[79] Bert F. Green, Jr., "IPL-V: The Newell-Shaw-Simon Programming Language," Behavioral Science, V (January, 1960), pp. 94-98.

instructions in the language but additionally there are about 200 pre-programmed routines in IPL itself and in assembly language. Thus, a list processing program is set up by combining the seven instructions with these basic processes. The following example may make some of these points clearer:

> A region with four centers could be represented as a main list with four sublists. Each sublist might contain a list of facilities and characteristics of one center while the main list would consist of the names of the centers used as sublist names. The order of the names in the main list might be arbitrary or could represent relative importance of the centers. Interchanging the position of center names in the main list would, in the latter case, represent changing their ranks. If the computer were programmed to simulate intra-regional competition, for instance, it would have to compute characteristics of each center, such as the number of facilities, their magnitude, the relative importance of each facility in the center's employment structure, and perhaps others.
> To store these quantities, a special attribute list would be associated with the standard list. The attribute list would contain a pair of symbols for each attribute, the first describing the particular attribute (manufacturing employment, say) followed by the value of that attribute for the particular center being described. In some cases the value of the attribute might be the name of another attribute list. If a major attribute was manufacturing employment, the

sub-attribute list could include information on the number of supervisors, clerks, skilled workers, production workers, and so on. The list structure representing the region now has one main list, four sublists each with an attribute list, and other lists associated with these attributes. IPL can operate with these data at any level of detail; its unit in any one instruction may be a symbol, a sublist and attribute list, a main list, or the entire list structure.

Self-modifying programs call for a different use of lists. In the example above, the program might have a large number of rules to indicate development strategies for combinations of characteristics. As the program played, its experience would indicate which strategies were good. The list might indicate priorities with rules moved up the list when successful and demoted when they lead to disaster. A sophisticated program would keep several priority lists with descriptions of the situations in which to use each.

A major feature of IPL (and other) lists is that items can be added or deleted at any place on the list without disturbing anything else. The potential length of each list need not be specified in advance. The new cells are linked into the list by means of its address. Each cell contains <u>two</u> items of information: an IPL symbol and the location of the cell containing the next symbol on the list. Other list processing languages also use similar conventions but may use two computer words for each symbol/link pair rather than one as in IPL. Variable length lists are useful in the regional growth example, because the play of the game can involve addition or deletion of facilities for any center in each time period.

Simscript. This simulation language shares several features with list processing languages, notably the flexible use of computer memory, provision for attribute lists, and powerful symbol manipulation features. However, its arithmetic capabilities are more comprehensive as it was originally integrated with a Fortran translator.

Data for a Simscript program is described in terms of STATUS, while the operation of a program is in terms of EVENTS. These concepts may be defined as follows:

STATUS:	Entities Temporary Permanent Attributes Sets	EVENTS:	Exogenous Endogenous

By letting events happen to the data, simulation is set in motion. A feature of major advantage is that time can be controlled by the programmer by specifying timing of events in terms of days, hours, and minutes. This makes it possible to achieve relatively close accord with events in the real process being simulated.

Despite their many theoretical advantages for a large class of interesting applications, simulation and list-processing languages have seldom been used by geographers. This has been due largely to the overwhelming acceptance of Fortran, the reluctance to learn more than one computer language, and to the fact that the best language for a particular application is often not available at the researcher's institution. But as geographers investigate more complex systems and our models become more sophisticated we can anticipate that their use will increase.

OTHER LANGUAGES. Languages other than those described above are probably of less general use to social scientists. Thus, they are mentioned only briefly in the interests of completeness. First is the class of languages used for data processing in the business sense; these include COBOL and Autocoder. These are basically English language systems designed to be readily taught to persons with some background in business or industry. COBOL, Common Business Oriented Language, has some potential for wider use since it has powerful logical facilities and excellent and sophisticated input and output structures. But improvements in recent versions of Fortran have made it more satisfactory in these areas and its superiority in arithmetic and indexing facilities coupled with language simplicity have probably foreclosed the potential market for data-processing languages among social scientists.

With the introduction of its 360 series, IBM has proposed and developed a language, PL/1, which is intended to combine data-processing, algorithmic and simulation facilities as subsets. Adequate experience to comment on its usefulness has not been accumulated by the author. It also seems probable at this moment that the language will not be widely available on other than IBM equipment.

Among the many specialized languages developed for limited purposes, two are of special interest: linear programming packages and PERT systems. The former has been used by many geographers in the past several years and their use requires no more than the ability to specify parameters in some preset order (and, of course, the ability to interpret the program's output). The Program Evaluation and Review Technique (PERT) is a device useful in planning, monitoring and evaluating projects and programs. As such, its potential is primarily in applied geography and planning where it can serve as a yardstick for judging alternate approaches to problem solutions.

Finally, the increasing use of time-sharing systems and graphic display consoles will impose on any user the requirement to learn a job control or monitoring language. These are designed to provide information to the computer system about the programming language used, amount and type of data expected and its location. These languages are usually uniform within any one manufacturer's equipment and at most installations a simple subset of control statements suffices for normal applications.

CONCLUSION. We have attempted to bring programming languages into perspective from the viewpoint of a prospective user in geography or any social science. Detailed information on particular computers is available from makers or at computing centers and there are well-organized bibliographies available for more information on particular languages or computer techniques.[80]

[80] Aaron Finerman and Lee Revens, eds., Permuted (KWIC) Index to Computing Reviews (1960-1963) (New York: Association for Computing Machinery, 1964); also their Permuted and Subject Index to Computing Reviews (1964-1965), and Comprehensive Bibliography of Computing Literature, 1966.

APPENDIX B

PROGRAM KARLSSON

The computer program in this appendix was used to perform the simulations described in Chapter IV. It is originally based on Georg Karlsson's model of interpersonal communication but can be used to simulate any system with up to three variables in each geographical cell and with probabilistic rules of interaction. To adapt codes to the program other than those used as standard (see the program writeup following the listing of the computer program) it is necessary to revise the input subroutine. The output subroutine can similarly be changed to provide for nonstandard forms of printed or punched output.

Comments cards within the program provide information on the sequence of particular operations; they are phrased in terms of Karlsson's model but are readily interpretable for other problems of the same type. The program writeup provides complete information regarding input data formats and the various options available to the user as well as a summary of restrictions in the program as written.

APPENDIX B. 1

```
 PROGRAM KARLSSON
 DIMENSION IDATE(50,50), IPLUS(50),KRING(64),IFMT(10),
1TITLE(10),
1 DATE(50,50),KNOWER(500)
 COMMON IDATA,IPLUS,M,N,IGEN,KY,IGENK,NUMSIM,MSIM,KTAPE
1, NOTEL,
1IFMT,TITLE, KRINGA,KRINGB,KRINGC,KONA,KHT,KH,KT,KNULL
1,MSWICH,
2DATA,KNOWER,KRING
COMMENT
C       THIS IS A SIMULATION MODELING PROGRAM BASED ON THE SIMPLE
C       MODEL IF INTERPERSONAL COMMUNICATION DEVELOPED BY GEORG
C       KARLSSON.  WHILE THE PROGRAM IS WRITTEN USING THE VARIABLE
C       NAMES FROM KARLSSON'S MODEL, IT CAN BE USED TO SIMULATE ANY
C       SYSTEM WITH THREE CHARACTERISTICS IN EACH CELL AND USING
C       SIMILAR RULES OF INTERACTION.  IF DESIRED, THE PRINTOUT CAN
C       BE ADJUSTED FOR SUCH MODELS.
C       THE INPUT SUBROUTINE--READIN--GIVES COMPLETE DETAIL AS TO
C       THE FORM OF THE INPUT DECK.  THERE IS ALSO A WRITEUP
C       AVAILABLE FROM A. WILLIAMS.
C       WRITTEN IN CDC 3600 FORTRAN.
COMMENT---CDC3600 FORTRAN IS SIMILAR TO FORTRAN IV.
```

```
C HOWEVER, TO RUN THIS PROGRAM ON ANOTHER MACHINE IT MAY BE
C  NECESSARY TO CHECK LENGTH OF VARIABLE NAMES, ALPHA FIELD LENGTH
C  (8 IN THIS PROGRAM), AND ALSO TO BE SURE THAT THE MASKING STATE-
C  MENTS ARE LEGITIMATE.  ON IBM 360 SERIES COMPUTERS, FOR EXAMPLE,
C  IT MAY BE NECESSARY TO USE ASSEMBLY-LANGUAGE SUBROUTINES, OR EVEN
C  TO REWRITE THE PROGRAM IN SOME LANGUAGE LIKE PL/1.
C
C        MAIN PROGRAM STARTS HERE.
      ACTIVE = 0.0
      REWIND 54
      PRINT 3 $IZERO=0   $ MSKA= 0077000000000000B
    3 FORMAT (*1KARLSSON SIMULATION*//*  REFERENCE--GEORGE KARLSSON
     1(1958)00000003
     1, PAGE 45 ET SEQ*//*0PROGRAMMED BY A V WILLIAMS, GEOGRAPHY
     2DEPT, MICHIGAN STATE UNIVERSITY*/)
      DO 4 I = 1,50
    4 IPLUS(I) = 2H+
      BB = 100.0
      GO TO 424
  420 NUMSIM = NUMSIM -1
      WRITE (54) (KNOWER(I),I=1,IGENK)
      IF (NUMSIM) 1424,1424,425
 1424 CALL FIGURE
      GO TO 424
  425 READ (53,IFMT) ((IDATA(I,J),J=1,N),I=1,M)
      REWIND 53
```

```
      IGEN = IGENK   $ IGENK = 0
      MSIM = MSIM + 1
      GO TO (427,428,428) MSWICH
  427 PRINT 426,MSIM,TITLE
  428 DO 429 I = 1,IGEN
      KNOWER(I) = 0
  429 CONTINUE
  426 FORMAT (*2SIMULATION RUN NUMBER*I5,* FOR THIS DATA*//110A8)
      GO TO 430
  424 CALL READIN
      MSIM = 1
COMMENT--SEARCH FOR TELLERS FOLLOWS
COMMENT-------MAIN PROGRAM LOOP
  430 IGENK = IGENK + 1
      GO TO (432,431,431) MSWICH
  432 PRINT 6,IGENK
  431 CONTINUE
      DO 50 I = 1,M $ DO 50 J = 1,N
      KMARK = IDATA(I,J)
      MASKA = KMARK .AND. MSKA
      IF (MASKA .EQ. 8H0X000000) 110,50
  110 KMARK = IDATA(I,J) .AND. 77B
      IF (KMARK .NE. 0) GO TO 115
      KMARK = KMARK + 1 $ IDATA(I,J) = IDATA(I,J) .OR. KMARK $ GO TO 50
  115 IF (KMARK .GT.NOTEL) 117,114
  117 GO TO (112,50,50) MSWICH
```

```
  112 PRINT 111,I,J $ GO TO 50
  114 KMARK = KMARK + 1 $ IDATA(I,J) = IDATA(I,J).AND.777777
     17777777700B
      IDATA(I,J) = IDATA(I,J) .OR. KMARK
      ACTIVE = ACTIVE + 1.0
      GO TO (118,120,120) MSWICH
  118 PRINT 113, I,J
COMMENT--FORMAT FOR INACTIVE KNOWER
  111 FORMAT (17X,I2,1H,I2)
COMMENT--FORMAT FOR ACTIVE KNOWER
  113 FORMAT (34X,I2,1H,I2)
COMMENT--MASKB WILL CONTAIN SOCIAL CLASS TO BE CONTACTED
  120 MASKB=IDATA(I,J) .AND. 77000000B
      KRA = KRANDF(1.0,BB ,KY)-1
      GO TO (123,124,124) MSWICH
  123 PRINT 500,KRA
  500 FORMAT (1H+98X  I4)
COMMENT--THE FOLLOWING IF STATEMENT CAN BE CHANGED TO ALLOW A
C       THIRD, FOURTH, ETC. GROUP
  124 IF (KRA .LE. KONA) GO TO 130
  125 MASKB = MASKB .AND. 01000000B
      IF (MASKB) 1125,1120,1125
 1120 MASKB = 21000000B  $ GO TO 130
 1125 MASKB = 22000000B
COMMENT---GET RING AND CELL FOR INTERACTION
  130 IRING = KRANDF(1.0,BB ,KY) -1
```

```
  501 FORMAT (1H+106X  I4)
      GO TO (132,133,133) MSWICH
  132 PRINT 501,IRING
  133 IF (IRING .LE. KRINGA) 135,140
  135 IRING = 1  $ GO TO 150
  140 IF (IRING .LE. KRINBG) 145,142
  142 IRING = 3  $ GO TO 150
  145 IRING = 2
  150 ILFT = J-IRING $IRT = J+IRING $ JUP = I-IRING $JDN = I + IRING
      IF (ILFT-1) 155,160,160
  155 ILFT=1
160   IF (IRT-N) 170,170,165
  165 IRT = N
  170 IF (JUP) 175,175,180
  175 JUP = 1
  180 IF (JDN-M) 190,190,185
  185 JDN=M
  190 CELLS = (2*(IRT-ILFT) + 2*(JDN-JUP))
      KNUM = KRANDF (1.0,CELLS,KY)
  502 FORMAT (1H+ 114X  I4)
      GO TO (192,194,194) MSWICH
  192 PRINT 502,KNUM
  194 LA = 0
      DO 200 KI = ILFT,IRT
      LA = LA+1
  200 KRING(LA) = IDATA(JUP,KI)
```

```
      KIA = JUP+1
      DO 205 KI = KIA,JDN
      LA = LA+1
  205 KRING(LA) = IDATA(KI,IRT)
      KIA = IRT-1  $KIB = IRT
      DO 210 KI = ILFT,KIA
      LA = LA+1  $ KIB = KIB-1
  210 KRING(LA) = IDATA(JDN,KIB)
      KIA = JDN-1  $KIB=JDN  $  KIC = JUP + 1
      DO 215 KI = KIC,KIA
      LA = LA +1
      KIB = KIB - 1
  215 KRING(LA) = IDATA(KIB,ILFT)
      KNUMA = KNUM -1
COMMENT
C         CONTACT RING NOW IN VECTOR STARTING AT RANDOM NUMBER KNUM
      DO 225 KI = KNUM,LA
      MASKZ = KRING(KI) .AND. 77000000B
      IF (MASKB .EQ. MASKZ) 235,225
  225 CONTINUE
      DO 226 KI = 1,KNUMA
      MASKZ = KRING(KI) .AND. 77000000B
      IF (MASKB . EQ. MASKZ) 235,226
  226 CONTINUE
COMMENT---NO CONTACT MADE--NO SOCIAL CLASS MATCH IN RING
      GO TO (228,50,50) MSWICH
```

```
  228 PRINT 230
COMMENT   NOT COUNTED AS A CONTACT FOR STATISTICAL PURPOSES
    6 FORMAT ( *1GENERATION*I5//10X*KNOWERS*30X*CONTACTS*55X*RANDOM
     1NUMBERS*//15X *INACTIVE*
     1  10X*ACTIVE*10X*ACCEPTORS*10X*REJECTORS*20X*SOC.TYPE
     2 RING   NUMBER ACCEPT*//)
  230 FORMAT (1H+67X*NO CONTACT*/)
      GO TO 50
  235 MASKG = KRING(KI) .AND.MSKA
      IF (MASKG .EQ. 8H0X000000) GO TO 270
      MASKD = KRING(KI) .AND. 770000B
      NUMBER = KRANF(1.0,BB   ,KY)      - 1
  503 FORMAT (1H+ 122X  I4)
      GO TO (237,238,238) MSWICH
  237 PRINT 503,NUMBER
  238 MASKE = IDATA(I,J) .AND. 7700B
CCCCCCCCC        MASKING TYPE IF STATEMENT
C
      IF (MASKD .EQ. 300000B) 240,250
  240 IF (MASKE .EQ. 6300B .AND. NUMBER .LE. KHT) 260,245
  245 IF (MASKE .NE. 6300B .AND. NUMBER .LE. KH) 260,270
  250 IF (MASKE .EQ. 6300B .AND. NUMBER .LE. KT) 260,255
  255 IF (MASKE .NE. 6300B .AND. NUMBER .LE. KNULL) 260,270
C ACCEPT INNOVATION --KI CONTAINS INDEX OF LINEAR POSITION IN RING
  260 KRING(KI) = KRING(KI) .OR. 0067000000000000B
      ASSIGN 300 TO KSWICH $ GO TO 280
```

```
  270 ASSIGN 310 to KSWICH
  280 LA = 0
C           REPLACE NEW KNOWER IN MATRIX BY FINDING CARTESIAN
C         COORDINATES
      DO 282 KM = ILFT , IRT
      LA = LA + 1
      IF (LA .EQ. KI) 281,282
  281 KROW = JUP  $ KCOL = KM
      GO TO 350
  282 CONTINUE
      KIA = JUP + 1
      DO 285 KM = KIA,JDN
      LA = LA + 1
      IF (LA .EQ. KI) 283,285
  283 KROW = KM $ KCOL = IRT
      GO TO 350
  285 CONTINUE
      KIA = IRT -1   $ KIB = IRT
      DO 287 KM = ILFT,KIA
      LA = LA +1     $ KIB = KIB - 1
      IF (LA .EQ. KI) 286,287
  286 KROW = JDN  $ KCOL = KIB  $ GO TO 350
  287 CONTINUE
      KIA = JDN -1   $ KIB = JDN   $ KIC = JUP + 1
      DO 290 KM = KIC,KIA
      LA = LA + 1  $ KIB = KIB - 1
```

```
    IF (LA .EQ. KI) 288,290
288 KCOL = ILFT   $ KROW = KIB
    GO TO 350
290 CONTINUE
350 GO TO KSWICH
300 IDATA(KROW,KCOL)      = KRING(KI)
    KNOWER(IGENK) = KNOWER(IGENK) = 1
    IF (KROW .LT. I) 301,302
302 IF (KROW .EQ. I . AND. KCOL .LT. J) 301,307
301 IDATA(KROW,KCOL) = IDATA(KROW,KCOL) .OR. 1
305 FORMAT (1H+50X I2,1H,I2/)
315 FORMAT (1H+69X I2,1H,I2/)
307 GO TO (317,318,318) MSWICH
317 PRINT 305,KROW,KCOL
318 DATA(KROW,KCOL) = DATA(KROW,KCOL) + 1.0
    GO TO 50
310 GO TO (320,321,321) MSWICH
320 PRINT 315,KROW,KCOL
321 DATA(KROW,KCOL) = DATA(KROW,KCOL) + 1.0
 50 CONTINUE
    IGEN = IGEN - 1
    GO TO (325,327,325) MSWICH
325 JACK = 2  $ CALL MPRINT(JACK)
327 IF (ACTIVE) 330,330,329
    GO TO 420
329 ACTIVE = 0.0
```

```
      IF (IGEN) 420,420,430
      END
      SUBROUTINE MPRB
      DIMENSION IDATA(50,50),IPLUS(50),KRING(64),IFMT(10),TITLE(10),
     1 DATA(50,50),KNOWER(500)
      COMMON IDATA,IPLUS,M,N,IGEN,KY,IGENK,NUMSIM,MSIM,KTAPE, NOTEL,
     1IFMT,TITLE, KRINGA,KRINGB,KRINGC,KONA,KHT,KH,KT,KNULL,MSWICH,
     2DATA,KNOWER, KRING
   45 FORMAT (5X,10I9)
   55 FORMAT(/I5,5X,10(A8,1X))
      IOUT = 61
      KM = (((N-1)/10 + 1)*10)
      DO 30 L = 10,KM,10
      NN = L - 9     $ IF (L-KM) 10,5,10
    5 L = N
   10 WRITE (IOUT,45) (I,I=NN,L)
      DO 20 I = 1,M
   20 WRITE (IOUT,55) I, (IDATA(I,J),J=NN,L)
   30 WRITE (IOUT,35)
   35 FORMAT (1H1)
      RETURN
      END
      SUBROUTINE PRINTP(JACK,SUM)
      DIMENSION IDATA(50,50),IPLUS(50),KRING(64),IFMT(10),TITLE(10),
     1 DATA(50,50),KNOWER(500)
      DIMENSION ROW(50),COLUMN(50)
```

```
      COMMON IDATA,IPLUS,M,N,IGEN,KY,IGENK,NUMSIM,MSIM,KTAPE, NOTEL,
     1IFMT,TITLE, KRINGA,KRINGB,KRINGC,KONA,KHT,KH,KT,KNULL,MSWICH ,
     2DATA,KNOWER, KRING
      EQUIVALENCE (ROW(1),IPLUS(1)), (COLUMN(1),KRING(1))
COMMENT--PRINTS OUT PROBABILITY MATRIX
  335 FORMAT (1H2)
  340 FORMAT (/* COLUMN*/* TOTALS   *10F10.5)
  345 FORMAT (5X 10I10)
  347 FORMAT (1H+109X F10.5)
  355 FORMAT(/I5,5X,10F10.5)
  360 FORMAT (1H+109X* ROW TOTAL*)
  365 FORMAT (*0SUM OF MATRIX ELEMENTS=*F12.5/1H1)
  140 FORMAT (/* COLUMN*/* TOTALS   * 10F10.0)
  147 FORMAT (1H+109X F10.0)
  155 FORMAT (/I5,5X 10F10.0)
  165 FORMAT (*0SUM OF MATRIX ELEMENTS=* F12.0/1H1)
      IOUT = 61
      KM = (((N-1)/10 + 1) * 10)
      GO TO (100,303) JACK
  303 DO 330 L = 10,KM,10
      NN = L-9  $ IF (L-KM) 310,305,310
  305 L = N
  310 WRITE (IOUT,345) (I,I=NN,L)
      WRITE (IOUT,360)
      DO 320 I = 1,M
      WRITE (IOUT,355) I, (DATA(I,J),J=NN,L)
```

```
  320 WRITE (IOUT,347) ROW(I)
      WRITE (IOUT,340) (COLUMN(J),J=NN,L)
  330 WRITE (IOUT,335 )
      WRITE (IOUT,365) SUM
      RETURN
  100 DO 130 L = 10,KM,10
      NN = L-9
      IF (L-KM)110  ,105,110
  105 L = N
  110 WRITE (IOUT,345) (I,I=NN,L)
      WRITE (IOUT,360)
      DO 120 I = 1,M
      WRITE (IOUT,155) I, (DATA(I,J),J=NN,L)
  120 WRITE (IOUT,147) ROW(I)
      WRITE (IOUT,140) (COLUMN(J),J=NN,L)
  130 WRITE (IOUT,335)
      WRITE (IOUT,165) SUM
      RETURN
      END
      FUNCTION KRANDF(A,B,KY)
C     PSEUDO-RANDOM-NUMBER GENERATOR
C     FROM PIKE AND HILL, ALGORITHM 266
C     COMMUNICATIONS OF THE ACM, 8(OCT,65), 605
      ENTRY KRANF
      KY = 3125 * KY
      KY = KY-(KY/67108864) * 67108864
```

```
      KRANDF = KY/67108864.0 * (B-A) + A  + 0.5
      END
      SUBROUTINE READIN
      DIMENSION IDATA(50,50),IPLUS(50),KRING(64),IFMT(10),TITLE(10),
     1 DATA(50,50),KNOWER(500)
      COMMON IDATA,IPLUS,M,N,IGEN,KY,IGENK,NUMSIM,MSIM,KTAPE, NOTEL,
     1IFMT,TITLE, KRINGA,KRINGB,KRINGC,KONA,KHT,KH,KT,KNULL, MSWICH,
     2DATA,KNOWER, KRING
C     READS IN PARAMETERS AND DATA IN THIS ORDER--
C PARAMETERS
C             NUMBER OF ROWS AND COLUMNS IN DATA MATRIX, NUMBER OF
C             GENERATIONS TO BE RUN, RANDOM START NUMBER--ODD AND
C             LESS THAN 67 MILLION
C             NUMBER OF SIMULATIONS TO BE RUN ON THIS DATA,
C             TAPE CONTAINING DATA.
C             STATISTICAL OPTION. 1=PRINTOUT + STATISTICS
C                                 2=ONLY STATISTICS
C                                 3=STATISTICS+LAST MATRIX
C  SECOND PARAMETER CARD
C             NUMBER OF TELLINGS ALLOWED,RING CONTACT PROBABILITIES
C             A,B,C
C             PROBABILITY OF CONTACTING OWN SOCIAL GROUP
C             PROBABILITIES OF ACCEPTANCE
C                   H FROM T, H FROM NON-T, NON-H FROM T, NON-H
C             FROM NON-T
C  FORMAT CARD DESCRIBING DATA PLACEMENT--USE R OR A FIELD
```

```
C              DESCRIPTION
C  TITLE CARD WILL BE PRINTED BEFORE EACH RUN ON THIS DATA
C     DATA ARE INSERTED HERE IF THEY ARE ON CARDS
C              KNOWER CARD(S)
C               PUNCH ROW AND COLUMN OF EACH KNOWER SEQUENTIALLY
C              IN 3-COLUMN FIELDS
C               ON THIS CARD(S).  USE AS MANY CARDS AS YOU LIKE.
C              READING WILL STOP ON ENCOUNTERING BLANK OR ZERO-
C              FILLED COLUMNS.
C              NOTE--12 KNOWERS PER CARD IS THE MAXIMUM THAT IS
C              ALLOWED
    5 FORMAT (10A8)
    6 FORMAT (9I4)
      READ 10,M,N,IGEN,KY, NUMSIM,KTAPE,MSWICH
   10 FORMAT (3I5, I10,3I5)
   42 FORMAT (*1END OF SIMULATION*)
      IF (M) 40,40,13
   40 PRINT 42
      STOP
   13 READ 6, NOTEL,KRINGA,KRINGB,KRINGC,KONA,KHT,KH,KT,KNULL
      READ 5,IFMT
      READ 5,TITLE
      PRINT 5,TITLE
      PRINT 15, M , N , IGEN, NUMSIM, KTAPE $ IGENK = 0
      PRINT 73,KY,MSWICH
   73 FORMAT (//* RANDOM START NUMBER=*I10 //* STATISTICAL OPTION
```

```
      1*I2//)
       PRINT7,NOTEL,KRINGA,KRINGB,KRINGC,KONA,KHT,KH,KT,KNULL
     7 FORMAT(*  NUMBER OF TELLINGS=*I4/*  RING THRESHOLDS (0-99)-
      1RING 1
      1=*I4,* RING 2=*I4,* RING 3=*I4/*  SOCIAL CONTACTS (0-99)-
      1OWN GROUP=*
      2*I4/*   ACCEPTANCE THRESHOLDS--*/4X*H FROM T=*I4,4X*H FROM
      1NON T=*
      3I4,4X*NOT H FROM T=*I4,4X*NOT H FROM NOT T=* I4//)
    15 FORMAT (*OPARAMETERS*//*  INPUT MATRIX IS *I4,* BY*I5/I6 *
      1GENERATIONS TO BE RUN*/*ONUMBER OF SIMULATIONS ON THIS DATA=
      1*I5//* DATA
      2TAPE=(I5//)
       DO 20 I = 1,50 $ DO 20 J = 1,50
       DATA (I,J) = 0.0
    20 IDATA(I,J) = 0
       IF (KTAPE .NE. 60) REWIND KTAPE
       READ (KTAPE,IFMT) ((IDATA(I,J),J=1,N),I=1,M)
       REWIND 53
       DO 130 I = 1,M  $ DO 130 J = 1,N
   130 IDATA(I,J) = IDATA(I,J) .AND. 60606060777777700B
C      INSERT KNOWERS IN MATRIX FROM CARD
    99 READ 100,(KRING(I),I=1,24)
   100 FORMAT (24I3)
       DO 110 I = 1,24,2
       J = I + 1
```

```
      IF (KRING(I)) 115,115,105
  105 IDATA(KRING(I),KRING(J)) = IDATA(KRING(I),KRING(J)) .OR.
     10067000000000001B
  110 CONTINUE
      GO TO 99
  115 CONTINUE
      WRITE (53,IFMT) ((IDATA(I,J),J=1,N),I=1,M)
      REWIND 53
      DO 120 I = 1,500
  120 KNOWER (I) = 0
      PRINT 25
   25 FORMAT (*1INPUT MATRIX*//)
      CALL MPRINT(1)
      RETURN
      END
      SUBROUTINE MPRINT(I)
      DIMENSION IDATA(50,50),IPLUS(50),KRING(64),IFMT(10),TITLE(10),
     1 DATA(50,50),KNOWER(500)
      COMMON IDATA,IPLUS,M,N,IGEN,KY,IGENK,NUMSIM,MSIM,KTAPE, NOTEL,
     1IFMT,TITLE, KRINGA,KRINGB,KRINGC,KONA,KHT,KH,KT,KNULL,MSWICH,
     2DATA,KNOWER, KRING
COMMENT---PRINTS SQUARE ALPHA MATRIX
      IF (I.EQ.2) GO TO 50
      CALL MPRB
   50 PRINT 70,IGENK
   70 FORMAT (////* MATRIX OF KNOWERS IN GENERATION * I5  //)
```

```
      PRINT 52, (I,I=1,M)
   52 FORMAT (8X,40I3)
      NN = N+1
      DO 160 I = 1,M
      PRINT 55,I,(IDATA(I,J),J=1,N)
   55 FORMAT(/I3,5X,50(1X,A2))
  160 CONTINUE
      PRINT 65
   65 FORMAT (1H2)
      RETURN
      END
      SUBROUTINE FIGURE
      DIMENSION IDATA(50,50),IPLUS(50),KRING(64),IFMT(10),TITLE(10),
     1 DATA(50,50),KNOWER(500)
      DIMENSION SUMM(500),SUMSQ(500),STDEV(500),DDATA(50,50),AVE(500)
      DIMENSION ROW(50),COLUMN(50)
      EQUIVALENCE (IPLUS(1),ROW(1)), (KRING(1),COLUMN(1))
      EQUIVALENCE (SUMM(1),IDATA(1)),(SUMSQ(1),IDATA(501)),(STDEV
     1(1),IDA
     1TA(1001)), (DDATA(1),IDATA(1)),(AVE(1),IDATA(1501))
      COMMON IDATA,IPLUS,M,N,IGEN,KY,IGENK,NUMSIM,MSIM,KTAPE, NOTEL,
     1IFMT,TITLE, KRINGA,KRINGB,KRINGC,KONA,KHT,KH,KT,KNULL,MSWICH,
     2DATA,KNOWER, KRING
      PRINT 5,MSIM  $ SUM = 0.0
    5 FORMAT (*1STATISTICS FOR*I5,* SIMULATIONS ON DATA*//)
      DO 10 I = 1,M  $ DO 10 J = 1,N
```

```
      ROW (I) = 0.0
      COLUMN(J) = 0.0
      SUM = SUM + DATA(I,J)
   10 CONTINUE
      PRINT 6
    6 FORMAT (*OCONTACT FREQUENCIES FOR CELLS, ROWS, AND COLUMNS*//)
      DO 100 I = 1,N
      DO 100 J = 1,M
  100 COLUMN(I) = COLUMN(I) + DATA(J,I)
      DO 110 I = 1,M
      DO 110 J = 1,N
  110 ROW(I) = ROW(I) + DATA(I,J)
      JACK = 1
      CALL PRINTP(JACK,SUM)
      DO 2- I = 1,M  $ DO 20 J = 1,N
      ROW(I) = 0.0
      COLUMN (J) = 0.0
      DDATA(I,J) = 0.0
      DATA(I,J) = DATA(I,J) / SUM
   20 CONTINUE
      DO 200 I = 1,N
      DO 200 J = 1,M
  200 COLUMN(I) = COLUMN(I) + DATA(J,I)
      DO 210 I = 1,M
      DO 210 J = 1,N
  210 ROW(I) = ROW(I) + DATA(I,J)
```

```
      SUM = 0.0
      DO 310 I = 1,M
      SUM = SUM + ROW(I)
  310 CONTINUE
      JACK = 2
      PRINT 7
    7 FORMAT (*1CONTACT PROBABILITIES FOR CELLS, ROWS, AND COLUMNS*//)
      CALL PRINTP(JACK,SUM)
      REWIND 54
      DO 34 I = 1,MSIM
      READ (54) (KNOWER(K), K = 1,IGENK)
      DO 34 J = 1,IGENK
      SUMM(J) = SUMM(J) + KNOWER(J)
      SUMSQ(J) = SUMSQ(J) + KNOWER(J) *KNOWER(J)
   34 CONTINUE
      GENK = MSIM
      DO 40 I = 1,IGENK
      AVE(I) = SUMM(I)/MSIM
      STDEV(I) = SQRTF((SUMSQ(I)-(SUMM(I)*SUMM(I))/GENK) /(GENK-1.0))
   40 CONTINUE
      PRINT 45
   45 FORMAT (*1NUMBER OF NEW KNOWERS/GENERATION*//10X* GENERATION
     1*10X
     1*MEAN*10X*STD DEV.*//)
      DO 50 I = 1,IGENK
      PRINT 55,I,AVE(I),STDEV(I)
```

```
   55 FORMAT (1H014X I4,7X F10.5,7X F10.5)
   50 CONTINUE
      DO 60 I = 1,500
   60 KNOWER(I) = 0
      DO 47 I = 1,50  $ DO 47 J = 1.50
      IDATA(I,J) = 0
      DATA(I,J) = 0.0
   47 CONTINUE
      REWIND 54
      RETURN
      END
0000500005000050000300001000020000600000l
000300490089009900890021002000630090
    (5A8)
1EXAMPLE. 5X5 MATRIX, 2 SIMULATIONS, 5 GENERATIONS EACH.
    AHTO    B   0    A   0    B  TO    A  TO
    B   0    A  TO    B   0    A   0    B   0
    B   0    A  TO    B   0    A   0    B   0
    B   0    B   0    A   0    B  TO    A   0
    B   0    B   0    A   0    B  TO    A   0
003003
```

APPENDIX B. 2

OPERATING CHARACTERISTICS OF THE COMPUTER MODEL

KARLSSON: Computer Simulation of the Diffusion of Innovations Using Georg Karlsson's Simple Model of Interpersonal Communication

Language: CDC 3600 FORTRAN (FORTRAN IV)

Programmer: A. V. Williams, Department of Geography, Michigan State University

Description:

The program carries out a Karlsson simulation on a population arranged in an m x n matrix where m, n 50. The characteristics of each member are punched on cards -- social class, attitude toward new ideas, trustworthiness -- and this data along with certain parameters to govern the process make up the input to the program.

Printed output includes the job title as supplied by the user, a listing of the parameters specified, and one of the following options:

1. For each generation of each simulation:
   1a. Coordinates of knowers, active and inactive.
   1b. Coordinates of persons contacted by each knower, whether they accept the innovation or not.
   1c. Matrix of knowers at the end of each generation.

   For all simulations on a particular set of data and parameters:
   2a. A contact frequency matrix with row and column marginals and total frequency.
   2b. A contact probability matrix with row and column marginals and total probability (which will be unity if we neglect possible rounding errors).
   2c. A table giving the mean number of new knowers for each generation along with the standard deviation.

2. 2a, 2b, and 2c above.

3. 2a, 2b, and 2c plus the matrix of knowers at the end of each simulation.

Considerable output is generated by option #1 so it ought not be used except for initial experimenting with a small number of generations and simulations.

Job Deck

Explanation of cards in job deck in the order in which they appear.

PNC card -- upon being assigned a problem number by the Computer Laboratory, the user is given several of these cards. They are placed in front of each deck submitted to the computer.

Job card -- an accounting card containing problem number, job title, estimated total running time, and name.

Fortran card -- punch $^{7}_{9}$ FORTRAN, X, * starting in column 1.

Program deck -- a copy can be obtained from A. Williams, Geography Department, Michigan State University.

Run card -- punch $^{7}_{9}$ RUN, tt, pp where tt = estimated running time, pp = estimated number of lines to be printed. These parameters vary with the job, of course, but for a 10 x 10 data matrix, 20 generations, and 2 simulations with output option 1 a time limit of 3 minutes and a print limit of 5000 lines should be adequate.

Param 1 card --

Columns	Punch
1 through 5	number of rows in matrix
6 through 10	number of columns in matrix
11 through 15	number of generations
16 through 25	odd random start number less than 67 million
26 through 30	number of simulations
31 through 35	tape where data is stored -- if the data are on cards this is 60. If using previously read data use 53
36 through 40	statistical option: 1-output option 1 2-output option 2 3-output option 3

Param 2 card --

1 through 4	number of tellings allowed
5 through 8	probability of contacting ring 1(0000-0099)
9 through 12	probability of contacting ring 2(0000-0099)
17 through 20	probability of contacting own group (0000-0099)

	probability of accepting innovation (0000-0099)
21 through 24	H from T
25 through 28	H from non-T
29 through 32	non-H from T
33 through 36	non-H from non-T

Format card -- describes placement of data on cards using "A" or "R" field description with nX used to describe those card columns skipped. Assuming we are reading in a 5 x 5 data matrix with each row punched on a separate card thus:

bbbbAHT0bbbbBbb0bbbbAbb0bbbbBHT0bbbbAbT0

where b = blank. The zero is an integral part of the matrix for the program and must be included.

The format card for the above data would be either

(5A8) or (5R8).

If we want to eliminate the leading blanks, making each data cell consist of four characters (e.g. ABT0 Bbb0, etc.) then we would use the format card

(5R4)

although in this case the matrix of knowers (if we selected this option) would be printed out with a leading zero (0X 0X) rather than the more readable form (X X). Using compressed fields is most reasonable when print option 2 is selected. Otherwise, using an 8-column field for each cell of the matrix is best.

Title card -- whatever is punched on this card will be printed at the head of the results. If you wish the title to start on a new page, a 1 should be punched in column 1 (the 1 will not be printed).

Data cards -- each person in the data matrix is described in terms of three characteristics: his social class (A or B); his attitude towards new ideas (H or blank), and his trustworthiness (T or blank). In addition, for program purposes, each person has a zero punched after his T position.

Knower card(s) -- the row and column of each knower is punched in sequential 3-column fields with leading zeros as required. Punching is allowed in columns 1 through 72; 24 knowers can thus be defined in a single card. As many cards as necessary may be used. The computer will stop reading cards when it encounters a blank column.

Job Example:

To help make the preceding explanations clearer, a simple job deck is described below. We are given a 5 x 5 data matrix of persons and wish to simulate five generations of activity and to do this two times. Each teller can tell three times and the following contact and acceptance probabilities are used:

Probability of contacting cell 1 = .5
cell 2 = .4
cell 3 = .1
(note: since probabilities are computed 0-99, ring 1 probability will be entered as 49)
Probability of contacting own group = .9
Probabilities of acceptance:

H from T	= .21
H from non-T	= .30
non-H from T	= .63
non-H from non-T	= .90

The initial knower is located in row 3, column 3 (the center of the matrix). We select output option 3 to print only the final knower matrix for each of the two simulation runs plus the statistics.

Job Deck

```
PNC card

7/9JOB,998877,SIM,3.  DOE, JOHN.

7/9FORTRAN, X,*

      KARLSSON program deck

7/9RUN,3,2500

000050000500005000030000100001000600003
0003004900890099008900210030000630090
   (5A8)
1  KARLSSON MODEL FOR CHECKING STATISTICS, 5x5, 2 Sim, 5 Gen

    AHTO    B  0    A   0    B TO    A TO
    B  0    A TO    B   0    A  0    B  0
    B  0    A TO    B   0    A  0    B  0
    B  0    B  0    A   0    B TO    A  0
    B  0    B  0    A   0    B TO    A  0
003003
000000000000000000000000000000000000000000000000000000000000000000000
```

This example produces the following output (some spaces between lines left out to compress the typing):

```
KARLSSON SIMULATION
 REFERENCE--GEORGE KARLSSON (1958), PAGE 45 ET SEQ
PROGRAMMED BY A V WILLIAMS, GEOGRAPHY DEPT, MICHIGAN STATE
KARLSSON MODEL FOR CHECKING STATISTICS, 5x5 2 SIM, 5 GEN
 INPUT MATRIX IS 5 BY 5
 5 GENERATIONS TO BE RUN
```

```
 NUMBER OF SIMULATIONS ON THIS DATA =   2
 DATA TAPE = 60
 RANDOM START NUMBER =   300001
 STATISTICAL OPTION 3
 NUMBER OF TELLINGS =   3
 RING THRESHHOLDS (0-99)-Ring 1=49 Ring 2=89 Ring 3=99
 SOCIAL CONTACTS(0-99)-OWN GROUP = 89
 ACCEPTANCE THRESHHOLDS--
  H FROM T = 21 H FROM NON T = 30  NON H FROM T = 63 NON H FROM
  NON T = 90

INPUT MATRIX
       1        2        3        4        5
1      AHTO     B  0     A   0    B TO     A TO
2      B  0     A TO     B   0    A  0     B  0
3      B  0     A TO X   B   1    A  0     B  0
4      B  0     B  0     A   0    B TO     A  0
5      B  0     B  0     A   0    B TO     A  0

MATRIX OF KNOWERS IN GENERATION 0

       1        2        3        4        5
1
2
3                        X
4
5

MATRIX OF KNOWERS IN GENERATION 5

       1        2        3        4        5
1               X
2                        X
3                        X
4
5

SIMULATION RUN NUMBER 2 FOR THIS DATA
KARLSSON MODEL FOR CHECKING STATISTICS, 5x5, 2 SIM, 5 GEN

MATRIX OF KNOWERS IN GENERATION 5

       1        2        3        4        5
1               X                 X
2                        X                 X
3                        X                 X
4      X        X        X                 X
5               X        X
```

STATISTICS FOR 2 SIMULATIONS ON DATA

CONTACT FREQUENCIES FOR CELLS, ROWS, AND COLUMNS

	1	2	3	4	5	ROW TOTAL
1	1	3	0	1	0	5
2	0	0	3	0	2	5
3	0	0	1	0	2	3
4	1	3	1	0	2	7
5	0	2	2	0	0	4
COLUMN TOTALS	2	8	7	1	6	

SUM OF MATRIX ELEMENTS = 24

CONTACT PROBABILITIES FOR CELLS, ROWS, AND COLUMNS

	1	2	3	4	5	ROW TOTAL
1	0.04167	0.12500	0.00000	0.04167	0.0000	0.20833
2	0.00000	0.00000	0.12500	0.00000	0.08333	0.20833
3	0.00000	0.00000	0.04167	0.00000	0.08333	0.12500
4	0.04167	0.12500	0.04167	0.00000	0.08333	0.29167
5	0.00000	0.08333	0.08333	0.00000	0.00000	0.;6667

SUM OF MATRIX ELEMENTS = 1.00000

NUMBER OF NEW KNOWERS/GENERATION

GENERATION	MEAN
1	0.5
2	0.5
3	2.0
4	2.0
5	1.5

END OF SIMULATION

BIBLIOGRAPHY

Ackerman, Edward. Geography as a Fundamental Research Discipline. Department of Geography Research Paper No. 53. Chicago: University of Chicago, 1958.

Ashby, W. Ross. Introduction to Cybernetics. New York: John Wiley & Sons, Inc., Science Editions, 1966.

Bailey, N. T. J. The Mathematical Theory of Epidemics. London: Charles Griffin & Company Limited, 1957.

Bartlett, M. S. Stochastic Population Models in Ecology and Epidemiology. London: Methuen & Co. Ltd., 1960.

Berry, Brian J. L. "A Method for Deriving Multi-factor Uniform Regions," Przeglad Geograficzny, XXXIII (1961), pp. 263-82.

________. "Approaches to Regional Analysis: A Synthesis," Annals of the Association of American Geographers, L (1964), pp. 2-12.

Berry, Brian J. L., and Marble, D. (eds.) Spatial Analysis. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1968.

Berry, Brian J. L., and Pred, Allan. Central Place Theory. A Bibliography of Theory and Applications. Bibliography Series, No. 1. Philadelphia: Regional Science Research Institute, 1961; with supplement, 1965.

Broek, Jan O. M. Geography. Its Scope and Spirit. Social Science Seminar Series. Columbus, Ohio: Charles E. Merrill Books, 1965.

Brown, Lawrence. "The Diffusion of Innovation: A Markov Chain-Type Approach," Department of Geography Discussion Paper No. 3. Evanston, Illinois: Northwestern University, 1963.

________. "A Bibliography on Spatial Diffusion," Department of Geography Discussion Paper No. 5. Evanston, Illinois: Northwestern University, June, 1965.

Bunge, William. Theoretical Geography. Lund Studies in Geography, Series C: General and Mathematical Geography, No. 1. Lund, Sweden: C. W. K. Gleerup, 1962.

Bunge, William. "Locations are Not Unique," Annals of the Association of American Geographers, LVI (1966), pp. 375-76.

Chorley, Richard. "Geography and Analogue Theory," Annals of the Association of American Geographers, LVI (1964), pp. 127-37.

________. "Geomorphology and General Systems Theory," in F. Dohrs and L. Sommers (eds.), Introduction to Geography: Selected Readings. New York: T. Y. Crowell Co., 1967.

Churchman, C. W., and Ratoosh, P. (eds.). Measurement: Definitions and Theories. New York: John Wiley & Sons, Inc., 1959.

Coleman, James S. Introduction to Mathematical Sociology. New York: The Free Press of Glencoe, 1964.

Cox, D. R., and Lewis, P. A. The Statistical Analysis of Series of Events. London: Methuen & Co., Ltd., 1966.

Dimitry, Donald, and Mott, Thomas Jr. Introduction to Fortran IV Programming. New York: Holt, Rinehart and Winston, 1966.

Dohrs, Fred, and Sommers, Lawrence (eds.) Introduction to Geography: Selected Readings. New York: T. Y. Crowell Co., 1967.

Ellis, Jack B. "The Description and Analysis of Socio-Economic Systems by Physical Systems Techniques." Unpublished Ph. D. dissertation, Department of Electrical Engineering, Michigan State University, 1965.

Finerman, Aaron, and Revens, Lee (eds.) Permuted (KWIC) Index to Computing Reviews (1960-1963) (series). New York: Association for Computing Machinery, 1964.

Goldberg, Samuel. Probability: An Introduction. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1960.

Green, Bert F. Jr. "IPL-V: The Newell-Shaw-Simon Programming Language," Behavioral Science, V (January, 1960).

Greig-Smith, P. Quantitative Plant Ecology. London: Butterworths, 1964.

Hägerstrand, Torsten. Innovation Diffusion as a Spatial Process. Translated from the Swedish edition (1953), by A. Pred. Chicago: University of Chicago Press, 1967.

Hägerstrand, Torsten. "The Propagation of Innovation Waves," Lund Studies in Geography. Series B, Human Geography, No. 4. Lund, Sweden: C. W. K. Gleerup, (1952), pp. 3-19.

Haggett, Peter. Locational Analysis in Human Geography. London: Edward Arnold, 1965.

Hammersley, J. M., and Handscomb, D. C. Monte Carlo Methods. London: Methuen & Co., Ltd., 1964.

Harris, Britten. "Probability of Interaction at a Distance," Journal of Regional Science, V (1964), pp. 31-35.

Hartshorne, Richard. Perspective on the Nature of Geography. Association of American Geographers, Monograph Series. Chicago: Rand McNally, 1959.

Harvey, David. "The Analysis of Point Patterns," Transactions of The Institute of British Geographers, XL (1966), pp. 81-95.

Heisenberg, Werner. Physics and Philosophy. New York: Harper & Brothers, 1958.

Herniter, J., Wolpert, J., and Williams, A. "Learning to Cooperate," Papers of the Peace Research Society, VII (1967), pp. 67-82.

Karlsson, Georg. Social Mechanisms: Studies in Sociological Theory. New York: The Free Press of Glencoe, 1958.

Kemeny, J., and Snell, L. Mathematical Models in the Social Sciences. New York: Blaisdell Publishing Company, 1962.

Kemeny, J., Snell, L., and Thompson, G. Introduction to Finite Mathematics. 2nd ed. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1967.

Koenig, H., Tokad, Y., and Kesavan, H. Analysis of Discrete Physical Systems. New York: McGraw-Hill Book Company, 1967.

Lazarsfeld, P., and Henry, N. (eds.) Readings in Mathematical Social Science. Chicago: Science Research Associates, Inc., 1966.

Leighly, John. "What has Happened to Physical Geography?", Annals of the Association of American Geographers, VL (1955), pp. 309-18.

Lowry, Ira. "A Short Course in Model Design," Journal of the American Institute of Planners, XXXI (1965), pp. 158-65.

Massarik, Fred, and Ratoosh, P. (eds.) Mathematical Explorations in Behavioral Science. Homewood, Illinois: Richard D. Irwin, Inc., 1965.

Morrill, Richard. Migration and the Spread and Growth of Urban Settlement. Lund Studies in Geography, Series B: Human Geography, No. 26. Lund, Sweden: C.W.K. Gleerup, 1965.

Nagel, Ernest. The Structure of Science. New York: Harcourt, Brace and World, Inc., 1961.

National Academy of Sciences-National Research Council. The Science of Geography: Report of the Ad Hoc Committee on Geography. Publication 1277. Washington, D.C., 1965.

Naylor, Thomas, Balintfy, Joseph, Burdick, Donald, and Chu, Kong. Computer Simulation Techniques. New York: John Wiley & Sons, Inc., 1966.

Nystuen, John, and Dacey, Michael. "A Graph Theory Interpretation of Nodal Regions," Papers and Proceedings of the Regional Science Association, VII (1961), pp. 29-42.

Olsson, Gunnar. Distance and Human Interaction. A Review and Bibliography. Bibliography Series; 2. Philadelphia: Regional Science Research Institute, 1965.

Organick, Elliott I. A MAD Primer. Houston: Privately Printed, 1964.

Parzen, Emmanuel. Stochastic Processes. San Francisco: Holden-Day, Inc., 1962.

Pitts, Forrest R. "Problems in Computer Simulation of Diffusion," Papers of the Regional Science Association, XI (1963), pp. 111-19.

Rogers, Andrei. "A Markovian Policy Model of Interregional Migration," Papers of the Regional Science Association, XVII (1966), pp. 205-24.

________. Matrix Analysis of Interregional Population Growth and Distribution. Berkeley: University of California Press, 1968.

Rogers, Everett. The Diffusion of Innovation. New York: The Free Press of Glencoe, 1962.

Rogers, Everett, et al. "Computer Simulation of Information Diffusion: An Illustration from a Latin American Village." Paper presented at a joint session of the American Sociological Association and the Rural Sociological Society, Chicago, 1965.

Rosen, Saul (ed.) "Revised Report on the Algorithmic Language--Algol 60," Programming Systems and Languages. New York: McGraw-Hill Book Company, 1967.

________. "The Algol Programming Language," Programming Systems and Languages. New York: McGraw-Hill Book Company, 1967.

Siegel, Sidney. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill Book Company, 1956.

Smillie, K. W. An Introduction to Regression and Correlation. New York: Academic Press, Inc., 1966.

Sonquist, John, and Morgan, James. The Detection of Interaction Effects. Institute for Social Research, Survey Research Center, Monograph No. 35. Ann Arbor: The University of Michigan, 1964.

Tiedemann, Clifford. "Two Models for the Inferential Analysis of Central Place Patterns." Unpublished Ph.D. dissertation, Department of Geography, Michigan State University, 1966.

Tiedemann, Clifford, and Van Doren, Carleton. "The Diffusion of Hybrid Seed Corn in Iowa: A Spatial Simulation Model." Institute for Community Development and Services, Technical Bulletin B-44. East Lansing: Michigan State University, December, 1964.

Veldman, Donald J. Fortran Programming for the Behavioral Sciences. New York: Holt, Rinehart and Winston, 1967.

Waugh, Frederick V. Graphic Analysis: Applications in Agricultural Economics. United States Department of Agriculture, Economic Service, Agricultural Handbook No. 326. Washington, D.C.: USGPO, 1966.

Wolpert, Julian. "Behavioral Aspects of the Decision to Migrate," Papers of the Regional Science Association, XV (1965), pp. 159-69.

Yuill, Robert. "A Simulation Study of Barrier Effects in Spatial Diffusion Problems," Michigan Inter-University Community of Mathematical Geographers, Discussion Paper No. 5. Ann Arbor: The University of Michigan, 1965.