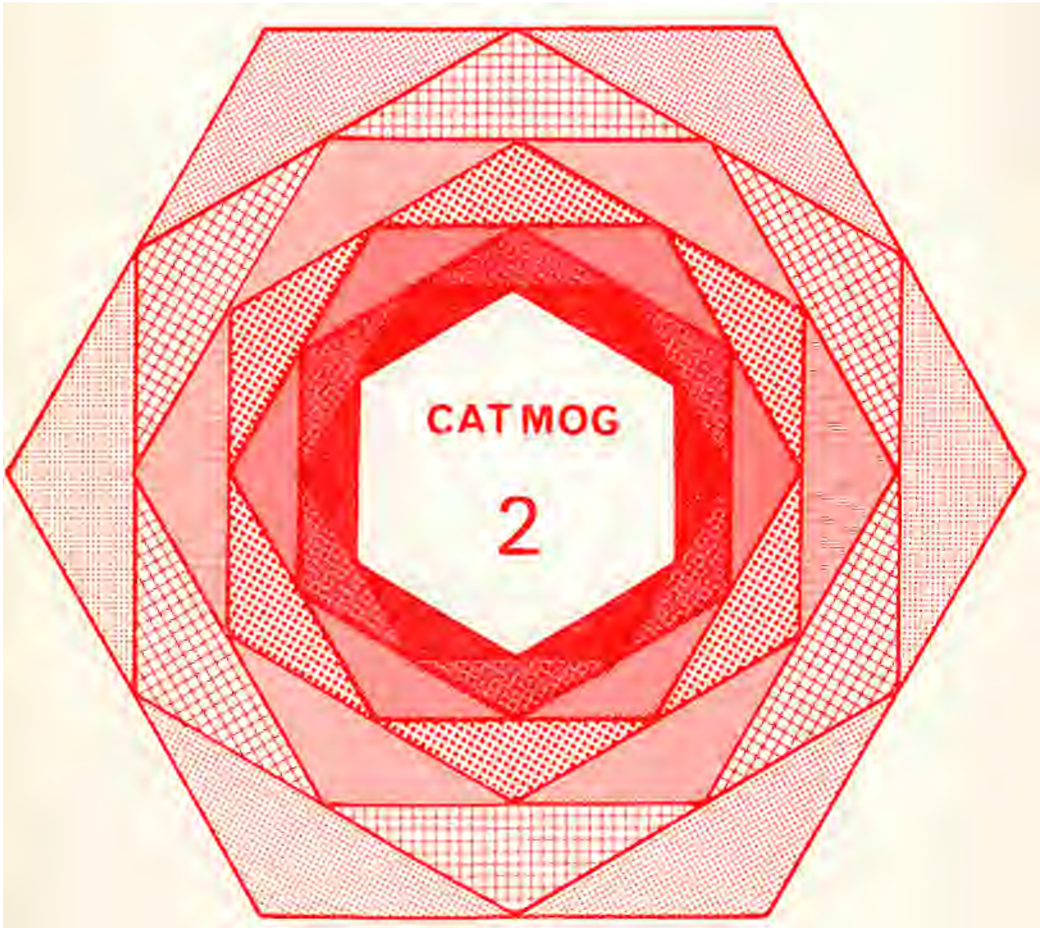
DISTANCE DECAY IN SPATIAL INTERACTIONS

Peter J. Taylor



ISBN 0 86094 090 X
ISSN 0306 6142

CATMOG - Concepts and Techniques in Modern Geography

CATMOG has been created to fill in a teaching need in the field of quantitative methods in undergraduate geography courses. These texts are admirable guides for teachers, yet cheap enough for student purchase as the basis of classwork. Each book is written by an author currently working with the technique or concept he describes.

I. Introduction to Markov chain analysis - L. Collins
2. Distance decay in spatial interactions - P.J. Taylor
3. Understanding canonical correlation analysis - D. Clark
4. Some theoretical and applied aspects of spatial interaction shopping models - S. Openshaw
5. An introduction to trend surface analysis - D. Unwin
6. Classification in geography - R.J. Johnston
7. An introduction to factor analysis - J.B. Goddard & A. Kirby
8. Principal components analysis - S. Daultrey
9. Causal inferences from dichotomous variables - N. Davidson
0. Introduction to the use of logit models in geography - N. Wrigley
I. Linear programming: elementary geographical applications of the transportation problem - A. Hay
2. An introduction to quadrat analysis (2nd edition) - R.W. Thomas
3. An introduction to time-geography - N.J. Thrift
4. An introduction to graph theoretical methods in geography - K.J. Tinkler
5. Linear regression in geography - R. Ferguson
6. Probability surface mapping. An introduction with examples and FORTRAN programs - N. Wrigley
7. Sampling methods for geographical research - C.B. Dixon & B. Leach
8. Questionnaires and interviews in geographical research- C.B. Dixon & B. Leach
9. Analysis of frequency distributions - V. Gardiner & G. Gardiner
20. Analysis of covariance and comparison of regression lines - J. Silk
21. An introduction to the use of simultaneous-equation regression analysis in geography - D. Todd
22. Transfer function modelling: relationship between time series variables - Pong-wai Lai
23. Stochastic processes in one dimensional series: an introduction - K.S. Richards
24. Linear programming: the Simplex method with geographical applications - James E. Killen
25. Directional statistics - G.L. Gaile & J.E. Burt
26. Potential models in human geography - D.C. Rich
27. Causal modelling: the Simon-Blalock approach - D.G. Pringle
28. Statistical forecasting - R.J. Bennett
29. The British Census - J.C. Dewdney
30. The analysis of variance - J. Silk
31. Information statistics in geography - R.W. Thomas
32. Centrographic measures in geography - Aharon Kellerman
33. An introduction to dimensional analysis for geographers - Robin Haynes
34. An introduction to Q-analysis -J. Beaumont & A Gatrell
35. The agricultural census - United Kingdom and United States G. Clark
36. Order-neighbour analysis -Graeme Aplin

This series is produced by the Study Group in Quantitative methods, of the Institute of British Geographers.'
For details of membership of the Study Group, write to the Institute of British Geographers, I Kensington Gore, London SW7.

The series is published by:
Geo Books, Regency House, 34 Duke Street, Norwich NR3 3AP, England, to whom all other enquiries should be addressed.

DISTANCE DECAY IN SPATIAL INTERACTIONS

by

Peter J. Taylor
(University of Newcastle upon Tyne)

CONTENTS

Acknowledgement

We thank T. Hagerstrand for permission to reproduce three figures and table from Innovation diffusion as a spatial process. 1953 (English translation 1967).

DISTANCE DECAY IN SPATIAL INTERACTIONS

I INTRODUCTION

(i) Prerequisities and Purpose

In this monograph we deal with the analysis of flows of phenomena over the landscape. These flows may be of goods, people, conversation or ideas. We shall use the term spatial interaction or just simply interaction to denote any one of these flows over some study area. Data on such interaction may be available as either individual movements or aggregated flows between sub-areas within a study area. The former can be mapped simply as a collection of lines joining the origin of the interaction to its destination. Aggregated data are usually mapped using choropleth maps to denote varying densities of flow from a single origin or to a single destination. In both these cartographic forms the geographer has long been familiar with working with interaction data. Our purpose here is to seek out the general patterns and regularities that exist within such data which are not always evident from the cartographic portrayal.

The techniques we use for this task are all various applications of standard regression analyses which produce best fit (i.e. 'least squares') functions fora set of data. The resulting function then becomes a simple mathematical model for representing the original data. In this monograph we use this approach to model the concept of distance decay as observed in interaction data. Several different models are presented and we show how interaction data can be modified to fit these models' needs. Thus the only prerequisities for our discussion are a general knowledge of mathematical functions with their associated statistical regression analyses. Our discussions below illustrate how this knowledge can be used very flexibly in numerous practical geographical research situations.

(ii) Spatial Interaction Patterns

Let us consider the sort of basic data we will be dealing with.

Figure 1 shows migration 'paths' for one year in the Asby district of Sweden. The map has been constructed by joining the initial home of each migrant (the origin) with his new home (the destination). The result is a criss-cross pattern of lines over the map. This reflects the individual migrants making their own separate decisions. It is certainly very difficult to see any 'pattern' in this data. In contrast consider Figure 2. In this case we are dealing with a much larger scale showing migrants to London as recorded in the 1971 census for English counties. This map does seem to exhibit a certain pattern. Clearly, in general at least, more migrants to London have originated in the counties around London than in counties further away. We might suggest that there is some evidence in Figure 2 for a decay effect on migration numbers away from the destination. This decay effect has been normally interpreted as reflecting a migration field - the area about some destination from which migrants are drawn. Figure 2 is a cartographic portrayal of London's migration field among English counties and thus illustrates a very general pattern. In fact we could have literally taken any town or city in the world for which such data is available and produced a similar migration field map showing a distance decay effect. Thus whereas Figure 1 seems to typify uniqueness,

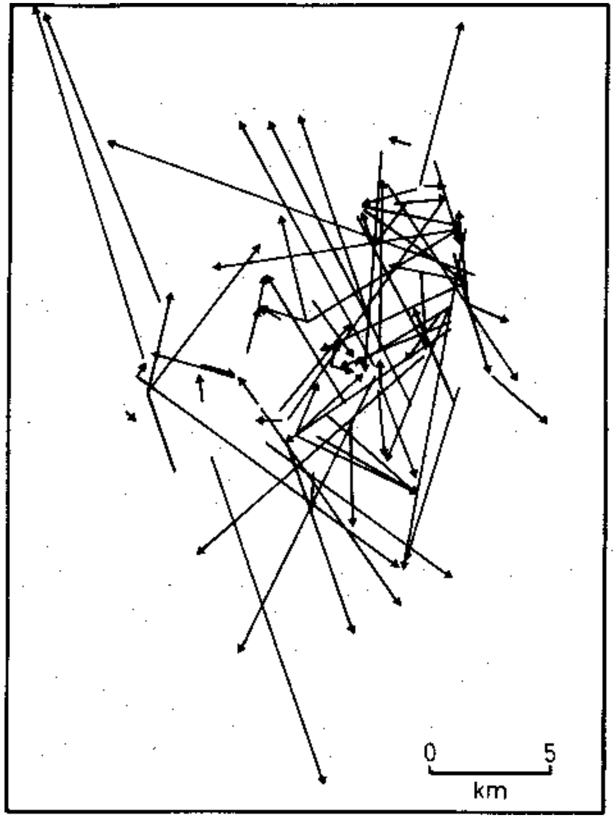


Fig. 1 Local Migrants in Asby, Sweden

Figure 2 seems to represent <u>generality.</u> Let us investigate these apparently important- differences.

The debate between uniqueness and generality has been an important one in geography. If we adopt a strictly uniqueness standpoint (i.e. all migration decisions are different from one another) then any theoretical geography is not possible and there can be no 'science' of geography. Rejection of the uniqueness doctrine has opened up geography to the world of models and theories thus ushering in the 'new geography'. Has Figure 1 no place in this new geography? What are the basic differences between Figures 1 and 2? The most obvious difference is the change in geographical scale. However this hides the more fundamental difference in terms of 'statistical scale' or aggregation. Figure 1 deals with individual migrants while Figure 2 portrays large numbers of migrants considered together i.e. all the migrants from one county, to London. It is when we aggregate migrants into groups that apparent common tendencies begin to arise. We shall see in the subsequent discussion that a simple aggregation of the data in Figure 1 yields results which produce findings closely related to the traditional concept of a migration field. Thus

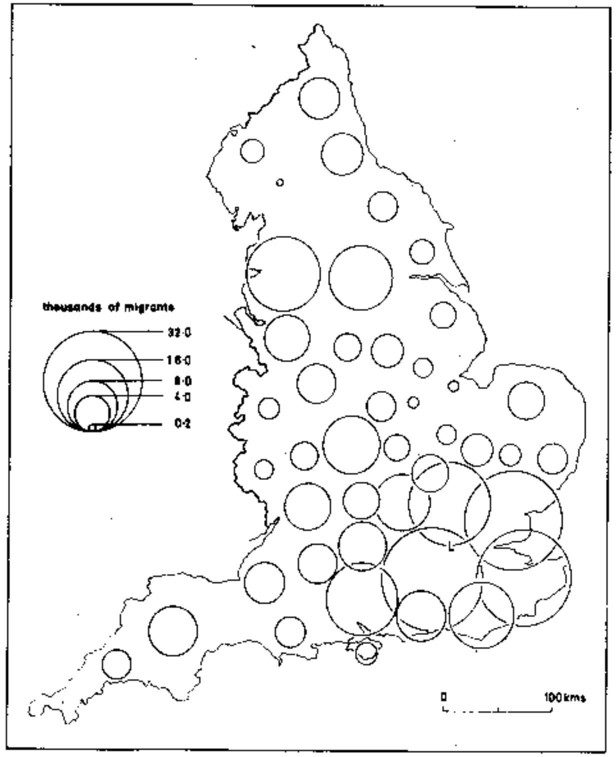


Fig. 2 Migrants to London, 1966-71

a distance decay effect is operating in Figure 1, just as in Figure 2, although it is obscured by the disaggregated level of the data. This is not to deny the 'individuality' of our migrants. We accept that they do indeed each make their own separate decisions. However each of these decisions is made within a framework of constraints. For many migrants a major aspect of these constraints is represented simply by distance. Quite simply,however much our philosophy may like to emphasise the free will of the individual, in practice migrants are not 'free' to live anywhere in the world. Their cultural preferences and their economic condition typically results in specific migration fields. We can begin to understand the working of these constraints by measuring the distance decay effect exhibited in aggregate migration flows.

The discussion above has been wholly related to one type of spatial interaction-migration. However what we have said equally applies to other interactions. Thus the migration field is merely a special type of the more general concept of a <u>functional region.</u> All functional regions are defined by flows which typically incorporate a distance decay effect. Commuter zones based on journey to work, shopping trade areas based on journeys to shop and port hinterlands based on transfer of imported goods are all other familiar examples of functional regions, studied by geographers and which will typically involve distance decay effects. Thus this effect has come to play an important role in very many models in geography. In fact it is not overstating the case to suggest that the relationship between interaction and distance is one of the most fundamental relationships in theoretical geography.

(iii) Mechanisms Behind Distance

Before we begin our discussion of how to model the distance decay effect we need to inquire as to the role that distance plays in the processes that produce the decay effect. Distance has long been used as an independent variable by social scientists other than geographers. Economists have employed distance as it relates to transport costs in their economic location theories. Sociologists have employed distance as a general determinant of friendship patterns and other social relations. In certain situations political scientists have even used distance as a variable to help explain voting patterns! However despite this widespread application, it must be conceded at once that distance is not a 'dynamic' process variable but remains a simple structural characteristic of any situation. Nonetheless, as a structural element it directly influences many of the dynamic processes. It is, in short, a very general surrogate which has the basic advantage that it can be easily measured and incorporated into our analyses. What are the processes at work which result in this very useful surrogate variable?

We can begin to answer this question by looking at what is probably the most unlikely type of interaction to be affected by distance. It has been shown through analysis of marriage records that the distances between the homes of the bride and groom typically show a definite decay pattern - there are many short 'marriage distances' but far fewer longer 'marriage distances'. Why on earth should distance affect our choice of mate? This is in fact a classic example of contraints operating in what our society designates as an area of free-will decision making. Thus there are several reasons that can be postulated which make it more likely that we choose a partner from near our home rather than from further distances away. The first is simply the probability of unintentional contacts. A person's activities take place around his home base which is the beginning and end of his day to day movements. Thus two people living near one another are more likely to meet unintentionally during their daily activities than two people living far apart. One of these chance meetings may of course be the first step on the road to marriage. However these are not the only ways in which couples first meet. Many more intentional meetings occur at entertainment centres for young people such as cinemas, dances and discotechques. Since each such centre draws its customers from its surrounding area, contacts made at such places are likely to be local rather than distant and so ultimately this favours short marriage distances. This begs a further question - why do these entertainment centres draw customers from just a local area? Clearly travel time and costs may often be important in this context. With two equivalent entertainment centres it seems reasonable that young people should patronise the one nearest to their home. There will normally need to be some special reason for them to travel further than necessary and thus waste both time and money. Hence entertainment centres have 'functional regions' about them just like shopping centres.

The movements referred to above are about the home base. When this home base is itself changed we talk of migration. In our example (Figure 2) of a migration field, migrants moved from all over England although most came from near London. All but the shortest migrations involve a change in job. Potential

migrants are more likely to know of vacant homes and jobs near their original home base than further away. Therefore his knowledge of opportunities is greatly biased in terms of local vacancies and hence there is a disposition to more short rather than long distances. In the case of intra-urban migrations the above argument holds for housing vacancies although most of such moves do not include a change of job. However at this scale other factors are often at work which also favour very short distance moves. Cities are divided up into residential areas in terms of social class, cultural groups and race. Very many potential migrants do not wish to move out of their own familiar social environment. Hence many moves are restricted to their own residential area and are thus over very short distances. This argument also applies to marriage distances where social, cultural and racial ties have a similar strong influence on the decision being made.

Thus here we have several reasons why we should expect distance to affect social types of spatial interaction. In the case of economic spatial interaction - flows of goods - it is usually hypothesized that it is the cost factor that leads to shorter distances being favoured over larger distances. Obviously a factory owner will try and obtain his raw material as cheaply as possible and this will often be from the nearest source with its lower transport costs. Similarly the factory owner will usually supply his finished articles to wholesalers and retailers from the factory nearest the outlet to avoid unnecessary transport costs.

All the mechanisms and, processes we have considered are reflections of, and are reflected in, communications. These can be directly measured using telephone data. Such data will include phone calls concerning social contacts as well as business calls arranging economic interactions. The combined result for any single telephone exchange is a communication functional region. It is the communication aspect of spatial interactions that may be reflected in social and political attitudes. Dodd (1951) has very explicitly illustrated this point in the context of racial intolerance. In this study 171 white people were interviewed in Seattle in 1948 a few weeks after a white woman had been raped in her home by a black person. The resulting racial tension, as measured by the average number of anti-black comments, has a definite spatial bias (Figure 3). It seems that local communications had led to variations in racial attitudes that can be displayed through distance zones about the site of the crime. A follow-up survey one year later shows the temporary nature of this particular pattern of racial tension (Figure 3). Thus once again we have a situation where distance is an indirect but clear determinant, this time on attitudes rather than the interactions themselves.

We conclude that there are several reasons why spatial interaction, and reflections of this interaction, should be related to distance. Not all of these mechanisms need be at work at the same time and indeed we might expect some types of processes to be dominant in the case of specific interactions - flows of goods and transport costs being the obvious example. Why, then, do we not use measurements of these mechanisms as our independent variables in spatial interaction studies? Quite simply distance is widely available from maps and is easily measured from this source. This seems to have been sufficient reason for distance to be widely used as an independent variable in spatial interaction research. Two basic types of models have been used to accommodate distance in this role and we devote a section to each of them.

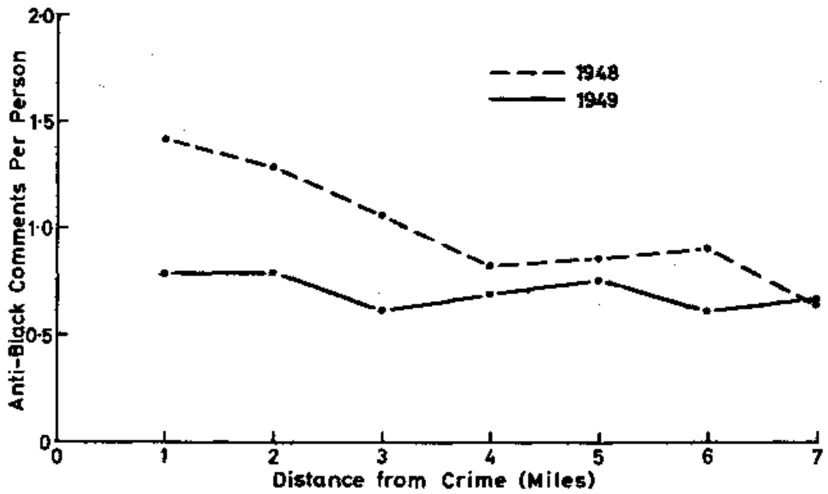


Fig. 3 Spatial Bias in a Racial 'Wave'

II ELEMENTARY SOCIAL PHYSICS FORMULATIONS

(i) The Gravity Model Analogy

It is perhaps inevitable that relatively undeveloped sciences, such as the social sciences, should cast envious eyes upon the more spectacular achievements of the physical sciences. This sometimes leads to ambitious attempts to emulate physical studies by using analogies. Probably the most famous of these physics-social science analogies has been the attempt to develop a social physics by application of Newton's laws of gravity to spatial interactions. Such attempts have a long history going back into the last century. Thus Ravenstein's celebrated 'first law of migration' that most migrants only travel short distances is often quoted as an early social physics study. However the American social scientist H.C. Carey had been much more explicit even earlier when he referred to the 'great law of molecular gravitation' for man. In a more practical context the transport engineer, Dionysius Lardner, was writing of a 'principle of high generality' that most train passengers only travel short distances. Such a simple finding was of obvious direct relevance in an age of railway building and speculation. Many years later, in the 1920s and 30s, three American researchers independently 'rediscovered' the gravity model for describing movements of farm population (Young), shopping travel (Reilly), and marriage distances (Bossard). Thus there has been quite an impressive parade of scientific discovery of the distance decay effect which ante-dates the more well known social physics of recent years (Carruthers, 1956).

Most modern interest on the social gravity model has stemmed from two independent sources - George K. Zipf's 'Principle of Least Effort' (Zipf, 1949) and John Q. Stewart's most explicit 'social physics'. The former researcher developed what he termed the "P_1P_2/D hypothesis" (Zipf, 1946) to describe flows of goods and people between cities in any single country where P_1 and P_2 are the populations of two cities and D is the distance between them. This was hypothesized as the mechanism at work in balancing the forces of diversification and unification within a country. Quite simply the larger the cities the larger the flow between them, while conversely the larger the distance separating them, the smaller the flow ('the principle of least effort'). Zipf's approach to testing these ideas has been only rarely employed and so we concentrate on Stewart's work in this discussion.

The most explicit drawing of the social physics analogy is found in the work of the Princeton astronomer Stewart. His initial interest was stimulated by his observation that students at Princeton came mainly from the local region with progressively less as distances from Princeton increased (Stewart, 1941). This led him to propose that by replacing the physicist's 'masses' by demographic masses - i.e. populations - we can derive demographic laws of gravity (Olsson, 1965). Thus gravitation force (F) between two masses (m_1, m_2) is given by

$$F = g \frac{m_1 m_2}{d_{12}^2} \qquad (1)$$

where d_{12} is the distance separating the masses and g is the gravitational constant. By analogy we have demographic force (DF) between two populations (P_1, P_2)

$$DF = k \frac{P_1 P_2}{d_{12}^2} \qquad (2)$$

where k is a constant to be calibrated. Similarly gravitation energy and demographic energy are given by

$$E = g \frac{m_1 m_2}{d_{12}} \qquad (3)$$

and

$$DE = k . \frac{P_1 P_2}{d_{12}} \qquad (4)$$

respectively. (Notice that demographic energy is equivalent to Zipf's hypothesis.) In physics the total energy at a point relative to all other points in a set is known at gravitational potential so that

$$\text{potential at point } i = g \, \Sigma \frac{M_j}{d_{ij}} \qquad (5)$$

where we sum over all j points. Total demographic energy therefore leads to population potential

$$V_i = \Sigma \frac{P_j}{d_{ij}} \qquad (6)$$

which has been commonly used as an aggregate measure of relative position in economic geography. However here we will concentrate on the two concepts of demographic force and energy. Notice that they only differ with respect to the exponent used for the distance variable.
Thus we can generalise to the single equation

$$X_{12} = k \frac{P_1 P_2}{d_{12}^{\;b}} \tag{7}$$

so that when b = 1, X is demographic energy and when b = 2, X is demographic force.

Thus we have two theoretical concepts relating pairs of settlements. What use are they? In the original studies it was suggested that the force and energy between two settlements would influence the amount of interaction between them. Thus Stewart (1947) postulates that the number of students from an area attending a university is related to the 'energy' between the university and the student's home area. However most subsequent researchers have not found it necessary to restrict themselves to the theoretical concepts derived from the analogy. In this situation we argue that b = 1 and b = 2 in our general equation are simply two special cases among an infinite number of possible values that b can take. When viewed in this light we soon begin to ask the question - what b value is most appropriate for any particular set of data? The solutions to this problem have used regression analysis.

(ii) The Linear Regression Model with Three Independent Variables

For a regression analysis we need to deal with several pairs of settlements for which we have locations, populations and some measures of interactions or flows between the settlements. We shall use the artificial data in Table 1 which gives migration from 2 villages to 3 country towns. Let us consider how we can use this data and our 'general' gravity equation in a regression analysis.

Table 1 Some Hypothetical Data for Gravity Modelling

X_{ij}	P_i	P_j	d_{ij}
9	500	5,000	2
74	500	10,000	1
23	500	15,000	3
6	1,000	5,000	5
30	1,000	10,000	4
89	1,000	15,000	3

In our general equation we have written down no exponents for the two populations P_1 and P_2. This amounts to an assumption that both exponents are 1. in our subsequent regression analysis we make no assumptions about constants so that we need to generalise further and allocate exponents to our populations.

Thus we produce

$$X_{ij} = k \frac{P_i^{\,b_1} P_j^{\,b_2}}{d_{ij}^{\,b_3}} \tag{8}$$

where i and j are any pair of settlements. (Now when $b_1 = 1$, $b_2 = 1$ and $b_3 = 1$ we have demographic energy.) When we log each variable on both sides of this revised gravity model the following equation results:

$$\log X_{ij} = \pm a \pm b_1 \log P_i \pm b_2 \log P_j \pm b_3 \log d_{ij} \tag{9}$$

where a = log k. Since this equation has a linear form the constants a, b_1, b_2 and b_3 can be easily estimated from data such as in Table 1. All that is required is that each variable is logged before entering our multiple regression analysis. When this is done for the data in Table 1 the following equation results:

$$\log \hat{X}_{ij} = -9.4452 + 1.8690 \log P_i + 1.5738 \log P_j - 1.6691 \log d_{ij} \tag{10}$$

where $\hat{X}_{ij}$ is estimated migration from i to j.

We can easily transform this linear equation into a gravity model type format by anti-logging to produce

$$\hat{X}_{ij} = -(2.787 \times 10^{-9}) \frac{P_i^{\,1.8690} P_j^{\,1.5738}}{d_{ij}^{\,1.6691}} \tag{11}$$

This equation defines neither demographic energy nor force but represents a best fit description of the interaction flows between the settlements in terms of their populations and the distances between them.

This form of gravity model has been used in a study of journey to work among local authority areas on Merseyside (Rahmatullah and O'Sullivan, 1968). In this case it was found that

$$\log \hat{X}_{ij} = -1.476 + 0.5057 \log P_i + 0.7786 \log P_j - 1.8066 \log d_{ij} \tag{12}$$

where $\hat{X}_{ij}$ is the estimated journeys to work from area i to area j, P_i is the number of economically active population in area i and P_j is the number of jobs in area j. This modification of the 'mass' variables into populations directly relevant to the interaction under consideration is a very common strategy in gravity modelling.

(iii) The Linear Regression Model with Two Independent Variables

Sometimes it has been considered unnecessary to allocate separate exponents to the two population variables. Instead they can be considered one variable - the size effect - given by the product (P_iP_j). By treating the populations as a single variable we produce the linear equation (by logging)

$$\log X_{ij} = \pm a \pm b_1 \log (P_i P_j) \pm b_2 \log d_{ij} \tag{13}$$

which represents the gravity model

$$X_{ij} = k \frac{(P_i P_j)^{b_1}}{d_{ij}^{\,b_2}} \qquad (14)$$

where a = log k. This model can be fitted simply by multiplying together populations and logging the product, and then logging the interactions and distances before fitting a multiple regression equation. When we do this for the data in Table 1 we produce

$$\log \hat{X}_{ij} = -9.1242 + 1.6428 \log (P_i P_j) - 1.5525 \log d_{ij} \qquad (15)$$

Notice that our constants are different from the previous case with three independent variables. This is simply because of the fact that we are fitting a slightly different form of model. However "notice that the constant for distance is only very slightly different. This reflects the fact that we have not appreciably changed the role of distance in the model.

This type of gravity model has been employed in a study of 116 migration streams between selected major cities in the U.S.A. (Galle and Taueber, 1966). In this case the model was calibrated as

$$\log \hat{X}_{ij} = 1.60 + 0.96 \log (P_i P_j) - 0.42 \log d_{ij} \qquad (16)$$

where X_{ij} is the estimated migrants from city i to city j, P_i is the total out-migrants from city i and P_j is the total in-migrants to city j. Notice that the populations have been defined in terms of 'migrant-populations' which incorporate such features as the 'attractiveness' of a city as well as simply total population size.

(iv) The Linear Regression Model with One Independent Variable

In the previous models absolute levels of interaction have been related to populations and distances. An alternative approach is to use relative levels of interaction as the dependent variable by dividing through by the populations. Thus the original general gravity model

$$X_{ij} = k \frac{P_i P_j}{d_{ij}^{\,b}} \qquad (17)$$

becomes

$$\frac{X_{ij}}{P_i P_j} = k \frac{1}{d_{ij}^{\,b}} = k d_{ij}^{\,-b} \qquad (18)$$

When we log the equation we produce a simple bi-variate linear function

$$\log \frac{X_{ij}}{P_i P_j} = a - b \log d_{ij} \qquad (19)$$

This equation simply has the logarithm of relative interaction as a linear function of the logarithm of distance. The equation can be calibrated by first generating the dependent variable by dividing the product of the populations into the interactions. This dependent variable and the independent variable, distance, can then both be logged and a simple bi-variate regression analysis applied. In the case of the data in Table 1 we produce the following equation

$$\log \left(\frac{X_{ij}}{P_i P_j} \right)' = -4.8584 - 1.2936 \log d_{ij} \qquad (20)$$

This type of equation was used in the Chicago Area Transportation Study (Carroll and Bevis, 1957). In this case traffic flows between pairs of zones in the metropolitan area were related to the product of the number of trips generated by the two zones as a proportion of all trips recorded in the study's origin and destination survey. From the data the following equation was calibrated

$$\log \left(\frac{X_{ij}}{P_i P_j / P_T} \right)' = 12.68 - 1.63 \log d_{ij} \qquad (21)$$

where X_{ij} is the traffic flow between i and j, P_i is the total flow generated by i, P_j the total flow generated by j and P_T the total flow generated by all zones combined.

We have now presented three linear equations for fitting the gravity model and, as we have seen, all three have been employed in geographical or related research. However notice that as we have reduced the number of independent variables in our models, the number of constants fitted has automatically declined so that the models progressively become less general. In fact our final bi-variate model can be easily shown to be the equivalent of the original generalisation of Stewart's gravity model analogy presented above (equation 7). In both cases we have just two constants a = log k and b:

$$X_{ij} = k \frac{P_i P_j}{d_{ij}^{\,b}} \qquad (7)$$

$$\log \frac{X_{ij}}{P_i P_j} = a - b \log d_{ij} \qquad (19)$$

Thus in effect our bi-variate model assumes $b_1 = b_2 = 1$ in the most general gravity model we considered (equation 8):

$$X_{ij} = k \frac{P_i^{\,b_1} P_j^{\,b_2}}{d_{ij}^{\,b_3}} \qquad (8)$$

Why should researchers bother to make this assumption when the more general regression models can be just as easily fitted? The answer seems to be two-fold. In much research the main interest in the modelling procedure has been in ascertaining the exponent for distance. This exponent can be directly interpreted as a measure of the rate of decline in (log) interaction with increasing (log) distance. It is generally referred to as the distance decay ex onent. The other two exponents relating to the populations are generally cons, ered to be much less interesting - it is just a matter of 'common sense' (or probability theory - Isard (1960)) that the amount of interaction between two places increases with their respective sizes. Finally the bi-variate model has the advantage of being easily presented graphically by plotting the relative interaction against distance on double-log graph paper and with the model illustrated by means of a straight line through the data points. This is shown for the data derived from Table 1 in Figure 4. With the model in this form the distance decay is directly represented as the gradient or slope of the line. Thus comparisons can be easily illustrated for different types of interaction. For instance in the Chicago study a graph comparing school, shop, work and recreation trips is shown with the slopes of the regression lines occurring in that order. This reflects neighbourhood school and shopping trips as against more metropolitan-wide work and recreation trips. We discuss applications of these equations a little more fully in the final section of this monograph. However work on such distance decay comparisons has often been carried out outside a strict gravity model formulation and we turn to this alternative modelling tradition next.

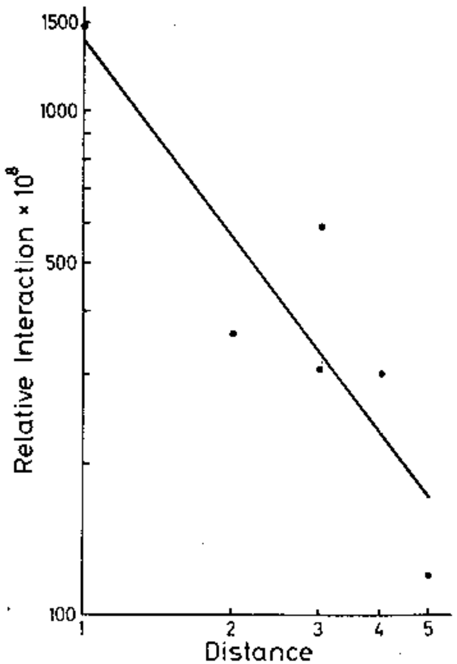


Fig. 4 A Bi-variate Function for the Data in Table 1

III DISTRIBUTIONS OF INTERACTION DISTANCES

(i) Real and Abstracted Interaction Fields

Interaction distance distributions are frequency distributions of distances over which a sample of interactions have taken place. Thus a typical example would be the number of migrants moving between 0 and 4.9 miles; the number between 5 and 9.9 miles; the number between 10 and 14.9 miles and so on. Such distributions can be constructed from either real or abstract interaction fields.

The date portrayed in Figure 2 presents little or no problems in constructing a distance distribution table. Thus we can measure the distance from London to each county on the map and allocate the observed number of migrants to the appropriate distance class. In this case we have a single specified centre, London, with a gradually declining set of interactions around it. We shall term this a real interaction field. The only problems that may occur in this case are due to the initial data aggregation and the resulting imprecision in allocating interaction to distance classes in 'bundles' rather than individually. This problem is overcome where we collect our own data on real interaction fields rather than relying on archival sources. For instance Claeson (1964) has collected data on the home locations of people visiting cinemas in various Swedish towns. The home-cinema distances were measured from this data and allocated to distance classed to form an accurate frequency distribution of these interaction distances.

In the original example of interaction we considered in Figure 1, the individual interactions are located all over the study area with no visible order. This situation seems a far cry from the real interaction fields dealt with above. In practice, however, it has been found relatively easy to convert maps such as in Figure 1 to neat orderly tables of interaction distances. All that is required is that we think in terms of abstraction of the data. This requires that we just remove from the map that information with which we are interested, ignoring the remainder. In Figure 1 the feature that is confusing the situation is the fact that the lines begin and end all over the map. This should not worry us here since we are not concerned with the relative positions of the interactions but just their individual spatial dimensions. This fact allows us to find order in the map through abstracting distances and directions from the pattern. We include directions here simply to keep a record of the individual vectors since we may later wish to consider direction biases in interaction patterns.

Once we have abstracted the vectors from the map we can re-arrange them so that they form a pattern not unlike the real interaction field. Every vector is given the same central origin, and then drawn from this perspective. We shall term such a diagram an abstract centred interaction field. With our data in this form we can see that we can produce an interaction distance table in exactly the same way as for the real interaction field.

Figure 5 illustrates this procedure. Figure 5(a) shows four flow lines that are 'centred' in Figure 5(b). The respective distances are 0.4, 1.1, 2.6 and 3.1 kilometres. These are allocated to the frequency table for distance classes in Table 2. A similar interaction frequency table can be produced however complicated the original line map. Figure 6 shows the centred interaction field for the Asby data from Figure 1 as a series of dots representing the

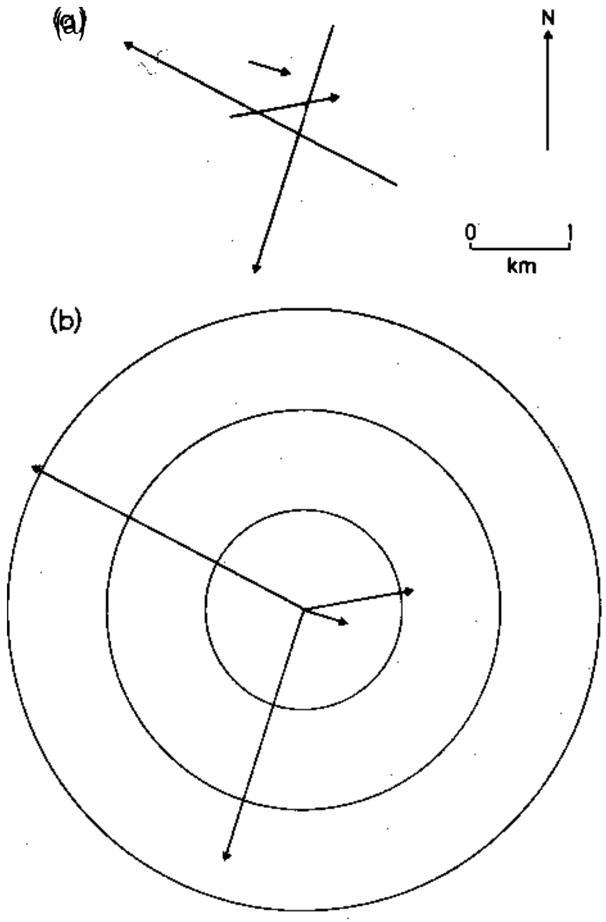


Fig. 5 Constructing an Abstract Centred Interaction Field

destinations and their distances to the origin are tabulated in Table 3. These interaction frequency tables constitute the first stage in the data preparation for modelling a distance distribution.

Why can't we use these crude frequency figures as our measures of interaction levels? In effect what we have produced are absolute frequencies which take no account of the spatial structure in which the interactions take place. In other contexts the influence of this spatial structure may be of interest although in distance decay research it is typically regarded as a disturbing influence to be 'accounted for' or held constant while we examine the interaction: distance relationship. Thus we need to convert our absolute frequencies

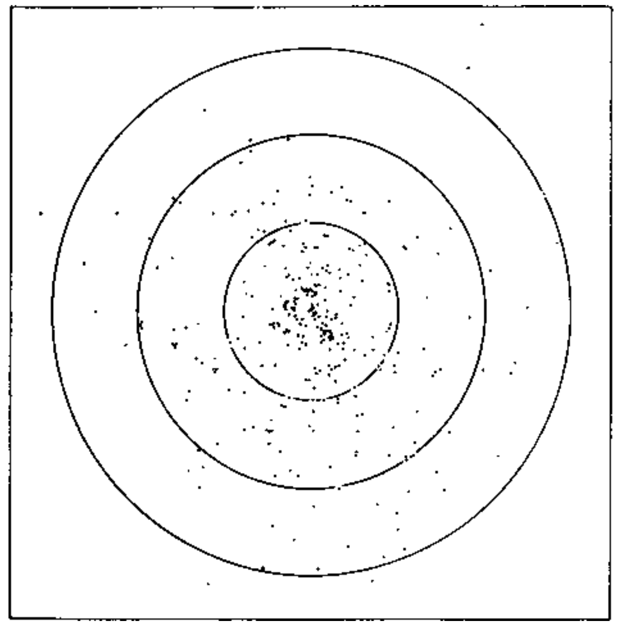


Fig. 6 The Abstract Centred Migration Field for Asby, Sweden

into relative measures related to the spatial context. How we achieve this end depends primarily on the amount of information we have about the spatial structure and upon the form that our interaction frequencies take. These considerations have led to several approaches to producing relative measures of interaction levels. Let us begin with a simple example which will illustrate the need for this manipulation of the data.

Consider the third column in Table 2 produced from the four individual flows in Figure 5(a). Let us assume that a larger sample is taken but that the same frequency pattern emerges so that with four hundred distances measured we have a frequency of 100 moves in each of the four distance classes. At first such a set of data might be considered to show no influence of distance. However such a conclusion would be premature. Let us make some assumptions concerning this example. First of all we will assume that the movements are those of migrants. Secondly we assume an isotropic plane so that opportunities for migrants are evenly spread over the land. Thirdly let us assume that distance has no influence on the migrant's choice of destination. This leads to the proposition that the probability of a migrant moving a specific distance class interval is equal to the area of a circular band defining that class interval around the abstract or real centre. This area can be easily found by subtracting the area of an inner circle defined by the lower end of the class interval from the area of an outer circle defined by the higher end of the

Table 2 Hypothetical Interaction:distance data from Figure 5

Distance:		Interaction:	
Class Interval:	Mid Point:	Frequency (X_j)	Intensity (X_j/P_j)
0-1	0.5	1(x100)	6400
1-2	1.5	1(x100)	1067
2-3	2.5	1(x100)	480
3-4	3.5	1(x100)	229

class interval. Thus for the class interval r_{j-1} to r_j, the formula is

$$\text{Area of ring} = \Pi r_j^2 - \Pi r_{j-1}^2 \qquad (22)$$

where r_j is the outer circumference of the class interval and r_{j-1} the inner circumference. The area of the total zone is Πr_n^2 where r_n is the final outer class interval boundary for n zones so that the probability of a move to distance band j, P_j, is given by

$$P_j = \frac{\Pi r_j^2 - \Pi r_{j-1}^2}{\Pi r_n^2} \qquad (23)$$

In our hypothetical example this produces:

$$\begin{aligned} P_1 \; (r_j = 1,\ r_{j-1} = 0) &= 0.0625 \\ P_2 \; (r_j = 2,\ r_{j-1} = 1) &= 0.1875 \\ P_3 \; (r_j = 3,\ r_{j-1} = 2) &= 0.3125 \\ P_4 \; (r_j = 4,\ r_{j-1} = 3) &= 0.4375 \end{aligned} \qquad (24)$$

These probabilities for movements can now be compared with the actual interaction frequency proportions X_j. The usual method is to produce a ratio X_j/P_j. These ratios are interaction levels relative to those predicted by the spatial structure and we can term them measures of interaction intensity. In the hypothetical example the absolute proportions are constant at one quarter while our probability model reflects the two dimensional nature of our space by predicting more interaction as distance increases. In terms of interaction intensity therefore interaction declines quite rapidly with distance. Thus we can interpret our original data as displaying no distance effects due to the disturbing influence of the spatial structure. With this influence allowed for in the X_j/P_j ratios, the influence of distance is revealed as shown in the last column in Table 2.

The assumption of the isotropic plane may seem rather unrealistic although it

has, in fact, been commonly used where we have no information on the spatial structure. Given this lack of knowledge it seems a reasonable assumption that 'opportunities' will increase with distance and so interaction frequencies have been standardized accordingly. We may assume errors resulting from this assumption are not important in rural areas with a largely even distribution of population. This seems to have been the case in the original work in this field carried out by Hagerstrand (1968) on local migrants in central Sweden. The final column in Table 3 shows the interaction intensities for the Asby data. It is these relative figures which exhibit the distance decay effect which we will subsequently measure.

Table 3 Migration Distances for Asby

Distance Band (Km)	Migrant Units	Units per Square Km
0.0 - 0.5	9	11.39
0.5 - 1.5	45	7.17
1.5 - 2.5	45	3.58
2.5 - 3.5	26	1.38
3.5 - 4.5	28	1.11
4.5 - 5.5	25	0.80
5.5 - 6.5	20	0.53
6.5 - 7.5	23	0.52
7.5 - 8.5	18	0.36
8.5 - 9.5	10	0.18
9.5 -10.5	17	0.27
10.5 -11.5	7	0.10
11.5 -12.5	11	0.15
12.5 -13.5	6	0.07
13.5 -14.5	2	0.02
14.5 -15.5	5	0.05

Where we have a real interaction field such as the London migration example (Figure 2) standardization is a relatively straightforward matter. Figure 2 suggests a gradual decline in level of migration out from London although some specific exceptions do exist. In the north of England some counties sent large numbers of migrants to London because their large population size inevitably leads to their having large numbers of potential migrants. This is particularly the case for Lancashire and the West Riding of Yorkshire. Similarly all the circles on Figure 2 will be influenced by the population size of the county producing the flow. We can allow for this factor by recomputing our migration flows as migrations per standard population size. On Figure 7 the migration field has been redrawn in terms of migration per thousand population. This results in the importance of Lancashire disappearing and the local counties around London become pre-eminent. The effect of distance is now more clearly illustrated with the disturbing influence of population size allowed for.

The use of standardized 'rates' of interaction and interaction intensity are two methods of defining a relative level of interaction to enable us to utilize a simple bi-variate function: As such we have reached the same stage as at the end of the last section when we defined a bi-variate version of the gravity model. In that case, by arguing from analogy we specified a particular

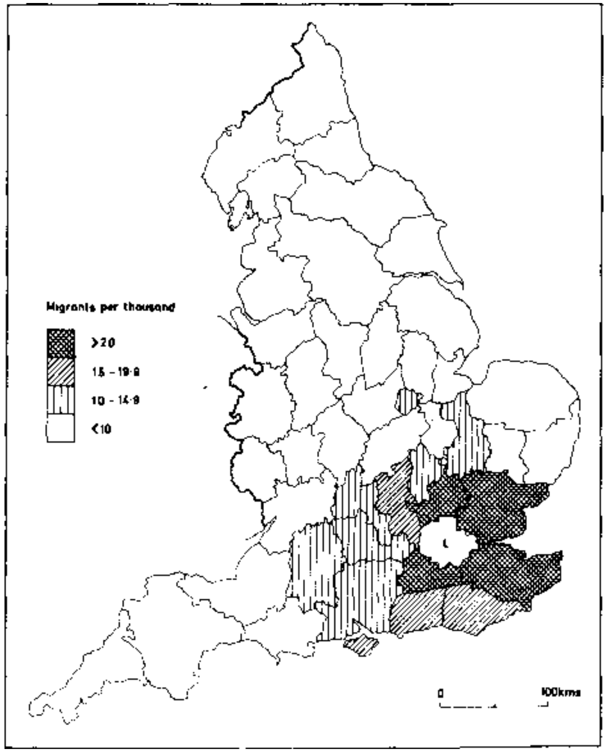


Fig. 7 Migration Rates to London, 1966-71

type of function logging both variables. In the present case we need only specify the general function

$$I_j = f(d_j) \tag{25}$$

which simply reads that interaction intensity to or from j to the centre is a function of the distance to the centre. A gravity type function is simply one of several interpretations of this general model.

(ii) Calibrating a Distance Decay Function

If the distance-interaction intensity data from Table 3 is plotted on graph paper a definite non-linear pattern emerges. An extreme concave upwards pattern is typical of much interaction data (Taylor, 1971a). Thus a simple fitting of a linear regression line is not usually appropriate so that we need to consider other approaches to the description of the data.

Faced with non-linear data a researcher can employ one of three simple strategies - he can break up his data into parts and fit several straight lines, he can fit a smooth polynomial curve or he can transform his data to make it linear. Let us briefly consider each in turn.

The simplest approach is to fit several straight lines to separate parts of the data. Figure 8 (a) shows how distance-interaction data might be treated in this way. Thus we have one function that applies from zero distance to some threshold and another function that applies from this threshold onwards. This approach may be criticised since it is cumbersome to work with several equations - it is clearly not the most succinct description of the data. Nonetheless if a clear threshold is apparent in the data then this would seem to be the most preferable method to employ. This approach infers separate processes operating over different ranges which require separate modelling. However this inference may be misleading where there is no obvious 'break' in patterns and the data portrays a 'smooth' curve. In this case there would seem to be a single process which requires a single non-linear expression.

The simple linear function we fit to any straight line data can be interpreted as the first of a whole family of curves. These 'polynomials' are identified by the number of turning points they incorporate. The linear equation has no turning points whereas the 'second order polynomial' (the quadratic function) has one turning point. Applied to our distance-interaction data this type of function has the form

$$I = a - bd + cd^2 \tag{26}$$

which is illustrated in Figure 8(b). Notice that the turning point eventually leads to a situation where interaction increases with distance so that, in order to be used in the present context,we need to truncate it at its turning point. However this approach does involve other problems. When fitting such a curve to interaction data the curve will often fall below the y axis so that 'negative' interaction intensity results. Thus the equation for the Asby data is

$$I = 8.62 - 1.87d + 0.09d^2 \tag{27}$$

which predicts an interaction intensity level of - 1.08 at 10 kilometres. Furthermore the interpretation of the additional c constant is not apparent in this migration context. Hence we can conclude that this second approach to fitting distance-interaction data is generally not appropriate.

The final approach is to transform the data so that it has a linear pattern. We then fit a simple linear regression line to the transformed data. This function will then give a smooth curve when transformed back into the original data. The problem is, of course, to choose the appropriate transformation. Both variables may be transformed and we consider the various alternatives for the independent variable distance below. However let us begin by considering the dependent variable, interaction intensity. From our previous discussion we can conclude that we do not want to 'allow' any negative values for interaction intensity. We can achieve this end simply by using the logarithm of the variable, which, by definition, can never be less than zero. Thus in all the functions that follow we will employ log I as the dependent variable.

Having chosen our transformed dependent variable and assuming we also have

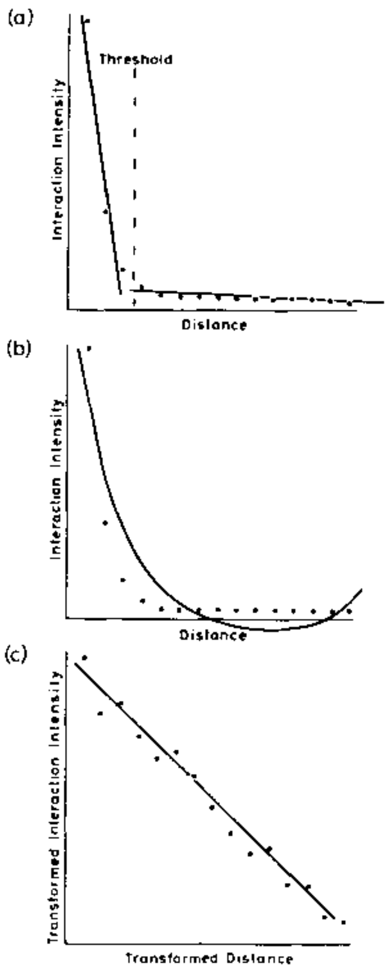


Fig. 8 Three Approaches to Fitting Interaction-Distance Data

adequately transformed distance we should produce a simple function such as that portrayed in Figure 8(c). The advantages of this approach are twofold. First we have a single function over all the range that does not need truncating and secondly the function itself is easily interpreted - the slope of the line is a distance decay gradient in terms of the transformed variables. Let us consider some examples of these types of functions.

(iii) <u>Distance Transformations</u>

We have so far decided that we are going to employ functions of the form

$$\log I = a - b\, f(d) \qquad (28)$$

where f(d) is some transformation of distance. The French geneticist J.M. Goux (1962) has proposed a useful typology of such functions. He treats these functions as a family of exponential curves of the form

$$I = ke^{-bf(d)} \qquad (29)$$

where e is the exponential constant (= 2.7183). We can transform distance by using various exponents, m. With m = 1 we actually leave distance untransformed so that a simple <u>exponential model</u> is produced

$$I = ke^{-bd} \qquad (30)$$

which has a linear form

$$\log I = a - bd \qquad \text{(where } a = \log k) \qquad (31)$$

if we use natural logs (since log e = 1). Alternatively we can transform distance by squaring (m = 2) so that

$$I = ke^{-bd^2} \qquad (32)$$

which has a linear form

$$\log I = a - bd^2 \qquad (33)$$

and can be shown to be a form of <u>normal model.</u> However geneticists have favoured decay curves that use the square root of distance (i.e. m = 0.5)

$$I = ke^{-bd^{0.5}} \qquad (34)$$

which has a linear form

$$\log I = a - bd^{0.5} \qquad (35)$$

and is termed the <u>square root exponential model</u> (Bateman 1947).

The above models can be termed 'single-log models' since a logarithmic transformation is only employed once (Taylor 1971a). Geographers have typically employed 'double-log models' where distance is also logged. In its simplest form this produces the most common function, the <u>Pareto model</u>

$$I = ke^{-b \log d} \qquad (36)$$

which has the linear form

$$tog\ I = a - b \log d \qquad (37)$$

This is the equivalent of the bi-variate gravity model described above (equation 19). Finally we should note that exponents can be added to the logarithm of distance. One example of this is a form of log-normal model where the transformed distance is the square of its logarithm

$$I = ke^{-b\,(\log d)^2} \tag{38}$$

which has the linear form

$$\log I = a - b\,(\log d)^2 \tag{39}$$

(In this case m = 2 compared to the Pareto model's m = 1).

We have now presented five different distance decay functions which have been employed in spatial interaction research. They are brought together in a diagrammatic representation of Goux's typology in Figure 9. In each case we fit the function by transforming the data first and then fitting a linear regression line to the transformed data. However so far we have not mentioned the basic reason for the transformation - the resulting linearity. How well

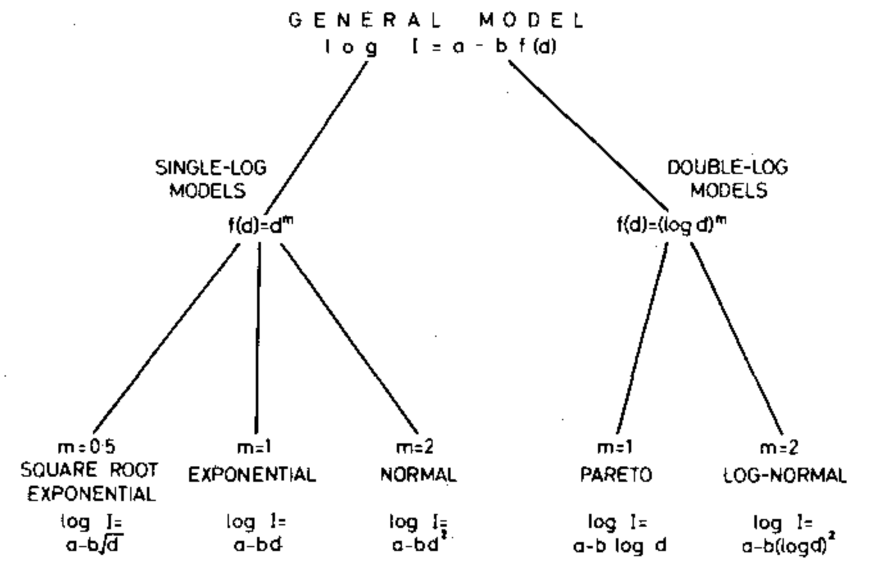


Fig. 9 The Goux Typology of Distance Decay Functions

our transformation has performed in making our data linear obviously controls the utility of the function. Furthermore, it is the basic criteria for choosing between functions in any single empirical application.

(iv) Choosing a Function

Figure 10 shows the five functions of the Goux typology fitted to the Asby data. In each case the y axis is calibrated in terms of the logarithms of interaction intensity and the x axis in terms of the particular transformation of distance under consideration. The regression lines that are fitted are as follows:

$$\left.\begin{array}{ll}\log I = 1.92 - 0.365d & \text{(Exponential Model)}\\ \log I = 0.87 - 0.219d^2 & \text{(Normal Model)}\\ \log I = 3.58 - 1.720d^{0.5} & \text{(Square Rt. Exp. Model)}\\ \log I = 1.62 - 1.478\log d & \text{(Pareto Model)}\\ \log I = 1.91 - 0.701\,(\log d)^2 & \text{(Log-Normal Model)}\end{array}\right\} \tag{40}$$

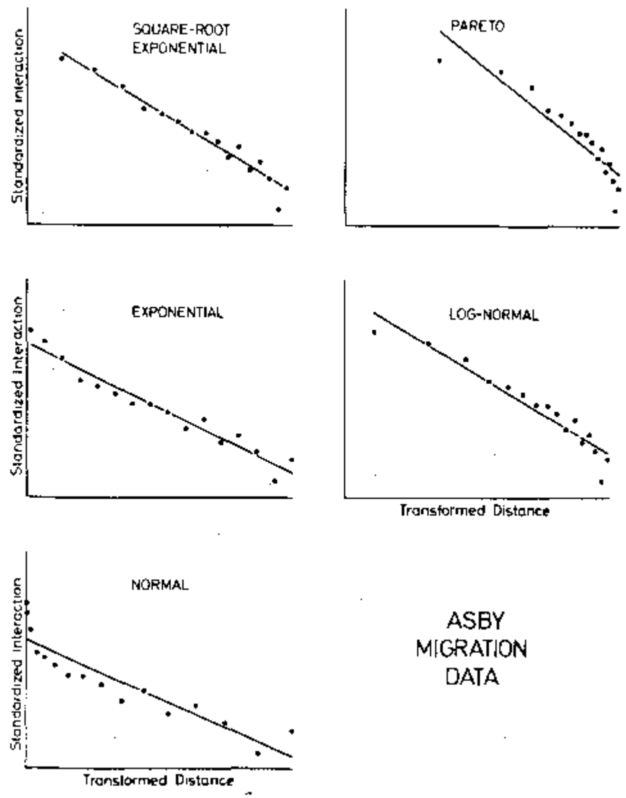


Fig.10 Fitting Five Functions to the Asby Data

All five models can be seen to give reasonable fits although some fits are clearly better than others. The simplest way to assess the validity of a regression model is in terms of its standard error of estimate. This is simply the standard deviation of the residuals about the regression line. The standard errors for the models fitted to the Asby data in Figure 10 are as follows

1. Square Rt. Exp. Model . 0.36
2. Exponential Model . . . 0.49
3. Log-Normal Model . . . 0.49
4. Pareto Model 0.67
5. Normal Model 0.71

This quantitative assessment of the models enables us to rank them in terms of goodness of fit and choose the Square Root Exponential Model as empirically the best model in the case of the Asby data.

The standard error of estimate is an a re ate measure of the empirical validity of our models. In order to understan w at the distance transformations are achieving it is useful to consider the <u>pattern</u> of the individual residuals. Consider the Pareto fit on Figure 10 fro In this case the first residual is negative, followed by a series of positive residuals with finally some further negative residuals. This pattern of residuals suggests that the transformations employed have not made the data linear but have produced a slightly 'concave-downwards' relationship . This feature of the Pareto model is a very common phenomena in published distance decay curves (Taylor, 1971(a)). In a similar way the normal model tends to produce a 'concave-upwards' relationship rather than the linear pattern expected. Figure 11 is a schematic diagram of all five models which illustrates the transforming process as it seems to work for the Asby data and for several over data sets that have been examined in this way (Taylor 1971(a)). In effect by logging distance, the Pareto model over transforms the data past linearity to a concave-downwards relationship. The log-normal model reverses this trend. In contrast the normal model under-transforms and in this case no transformation of distance (the exponential model) and the square root of distance progressively make the data more linear.

This schematic diagram seems to be a reasonable guide to the use of distance decay functions with interaction data of the local migration and marriage distances type. However,it should not be considered as a general rule. For instance, in a study of journey to work patterns to Newcastle upon Tyne (Gleave, 1971) the lowest standard errors of estimate were found when using the Pareto model. Clearly each set of distance-interaction data should be treated empirically as suggested above before the final decision as to what function to employ is made.

(v) <u>Some Outstanding Problems</u>

Before we leave our discussion of distance decay functions we should consider some problems that arise from this approach. The disadvantages of the transformations stem from the basic fact that we have not calibrated equations in terms of the variables in which we are ultimately interested but have used some transformation of them. Thus the best fitting curves are defined in terms of minimizing the squares of the <u>logarithms</u> of the dependent variable. This

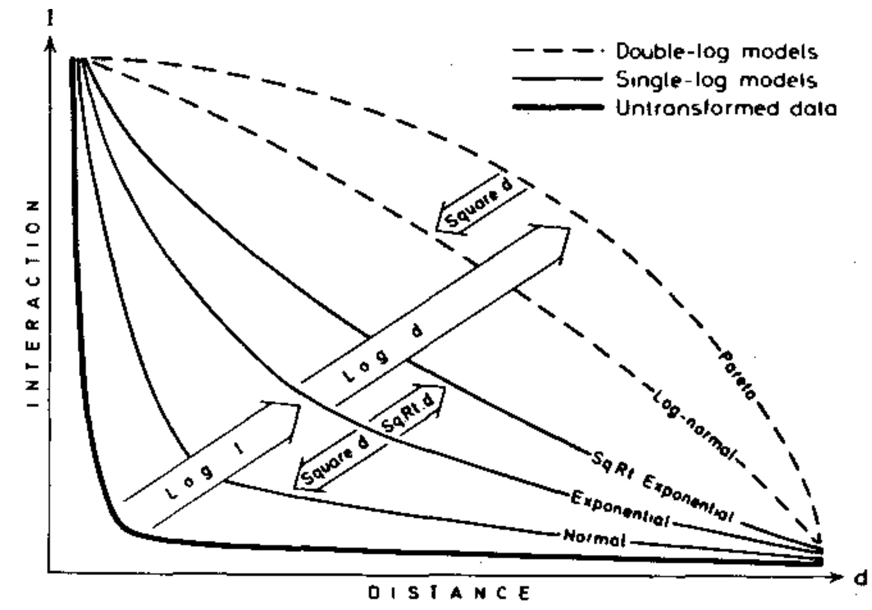


Fig. 11 A Schematic Diagram of Distance Transformations

has the effect of using 'relative' errors rather than 'absolute' errors. Consider the logs (to base 10) 0, 1, 2, and 3. If an observation is 0 and the model predicts 1 the error is just 1. Similarly the same error results with an observation of 2 and prediction of 3. However when we anti-log the first residual equals 9 and the other 900: Nonetheless fitting a least squares line to logarithms results in these two residuals being treated equally. A further implication of this situation is that the interaction levels for longer distances need to be particularly accurate. An absolute change of 2 when X 3 has a very different effect than when X = 300. In the first case log X changes from 1.099 to 1.609 whereas in the second case the change is minimal - from 5.705. to 5.710. However, since the data is derived from raw frequencies it is just these longer distance observations based on only a few interactions that are the most unstable. In the Asby data, for instance, the number of migrants observed travelling between 13.5 and 14.5 kilometres is only two (Table 3). This problem is discussed in relation to the early distance decay research of Ilke (1954) by Anderson (1955).

A further problem relating to the use of relative errors in the curve fitting is that this results in difficulties in actual testing of whether the model is a good fit to the data using inferential statistics. Such testing can usually be achieved by using one of several goodness of fit tests that are available. The most popular is the Kolmogorov-Smirnov one sample test which concentrates on the largest <u>absolute</u> deviation when the observed and predicted distributions are cumulated (Morrill and Pitts, 1964). The point that arises in this context is whether it is a fair test of the model to assess it in terms of absolute values when it has been calibrated to minimise squares of logarithmic values. Not surprisingly when such tests have been carried out the results have not indicated particularly close fits. Unfortunately no goodness of fit test exists for assessing relative errors. However in practice the lack of rigorous statistical assessment at this stage of the analysis has not seriously hampered applications of distance decay functions.

A final problem that we should mention here is the role of the spatial structure - the stage upon which the interactions take place - on the resulting distance: interaction intensity data. Although we usually take into account the probability of interaction taking place in terms of population or area in defining our relative interaction levels, we do not directly consider the relative position of the interaction within the system. Several researchers have shown that this will have an effect on the resulting decay gradients (Goux, 1962; Curry, 1972) although suggested solutions have varied widely (Rushton, 1971; Taylor, 1971b; Johnston, T973). (here nas recently been some further debate on this matter and reference should be made to Cliff et.al. (1974,1975) Johnston (1975,1976) and Curry et.al. (1975).

IV CONCLUDING REMARKS

(1) Applications with Individual Equations

Despite the various problems described above, the fact remains that distance decay function give very succinct descriptions of interaction patterns. Furthermore the basic parameter of the functions, the regression coefficients, are easily interpretable in terms of interaction. They are in fact the distance gradients - the amount of (logged) interaction intensity that falls with one unit of (transformed) distance. It is thus a simple measure of the decline of interaction with distance i.e. the distance decay. As such it is easily interpreted when we revert back to the two dimensional map of an interaction field (real or abstract). We can envisage replacing the x axis of graph paper with a two dimensional base about the origin. If we now rotate the curve about the origin our straight line on the graph describes a cone over the new two dimensional base (Figure 12). In our data preparation before calibrating our distance decay functions, we in effect 'wrapped up' our original two dimensional interaction field into a one dimensional set of data. Thus what we have done in Figure 12 is to 'unwrap' the decay function to produce a new description of the original interaction field in terms of a cone. The advantage over the earlier cartographic portrayal is an explicit description by a simple mathematical formula. Huff and Jenks (1968) give a good discussion of 'unwrapped' decay functions and Morrill (1963) and Olsson (1967) have also considered this unwrapping.

Application of this succinct description of an interaction field can be found in an area of research which we may term 'comparative distance decay'. Torsten Hagerstrand has pioneered this approach as part of his Swedish migration studies (Hagerstrand, 1957). Part of this work involved comparing migration fields by using the distance decay exponent in the Pareto model. It is best illustrated in terms of changes in migration fields over time.

We are all familiar with the notion that the development of modern industrial societies have led to a sort of 'spatial liberation' of man. The horizons of

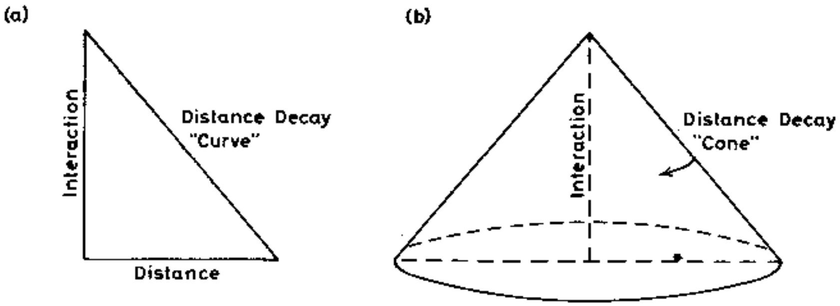


Fig. 12 'Unwrapping' a Distance Decay Curve

modern man are far wider than those of his traditional predecessor. The whole scale of society has changed and this should be reflected in larger interaction fields. Changes in decay gradients represent a specific measurement of these fundamental changes in the condition of mankind. Let us consider the phenomena of out-migration from selected Swedish settlements. This should be a very good indicator of people's widening horizons. In the century between 1850 and 1950 we would expect these migration fields to have spread out from local fields to cover much more of the area of Sweden. Figure 13 illustrates trends in the value of the decay gradients for a number of Swedish settlements over this time period. A clear pattern emerges with gradients becoming gentler indicating widening interaction fields over time. This pattern is consistent for all the settlements although differences remain between them. The three settlements with the largest gradients are rural settlements, Kavlinge and Askersund are small towns and Lund, which has the smallest gradients, is by far the largest town. This pattern is consistent with our notions of modernization and spatial horizons since it is the towns, especially the large towns, that are furthest removed from traditional societies.

Thus the very extensive field of Lund, especially of the middle class, contrasting with the restricted field of Lyrestad, complements the temporal pattern in monitoring the effects of distance on interaction in a changing society.

We can now consider exactly what these changing gradients mean. They are direct measures of the pattern of interaction fields. Thus steep gradients represent very restricted migration fields; gentle gradients represent much wider fields

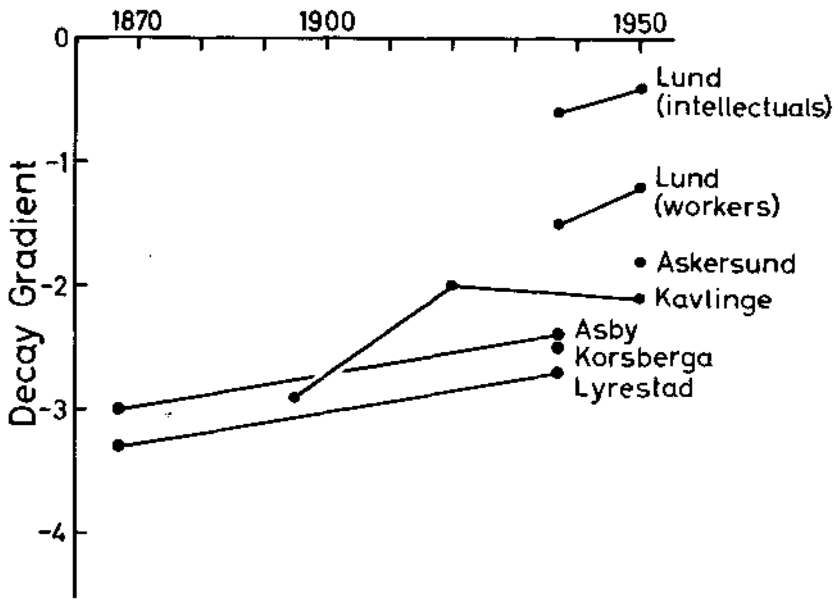


Fig. 13 Widening Horizons in Sweden, 1850-1950

perhaps covering all of Sweden. Hence Figure 13 illustrates migration fields ranging from essentially local patterns (pointed cones) to more nationwide patterns (much flatter cones). In terms of the mechanisms behind distance described at the beginning of this monograph, we can hypothesize increasing wider knowledge of opportunities leading to more migrants travelling longer distances beyond the local area.

(ii) Applications within More General Models

It is very common for applications of the distance decay concept to be part of a more general model. This is its role in most gravity model type applications. However once again Hagerstrand has been a pioneer in utilizing distance decay functions within a broader model context. Thus in his well-known Monte Carlo simulation model for the diffusion of innovations he employs distance decay functions as surrogates for the contact pattern of farmers. This relates to our original discussion of mechanisms behind distance where it was suggested that migration distances tend to be smaller rather than larger because of the pattern of local contacts. Hagerstrand turns this argument around to use migration patterns to represent the pattern of underlying contact, or as he terms it, the 'mean information field'. Thus the distance decay function plays a key role in the operation of this model since it incorporates the underlying process that generates the neighbourhood effect in patterns of innovation

acceptance in both real and simulated diffusion patterns. (Hagerstrand, 1968; Gould, 1969).

However the most widespread use of the distance decay concept in modelling has been in urban studies. The early land use-transportation studies explicitly involved linking different land uses with traffic generation. This required prediction of sizes of flows between land use zones and gravity model concepts were brought into play as the basic flow predictors in many of these studies. Thus distance decay functions typically abound in these models - the Chicago Area Transportation Study experimented with different decay gradients for different trip purposes (Carroll and Bevis, 1957) and the Pittsburgh Area Transportation Study used different decay exponents for different social groups (Lowry, 1963). It has been in this area of modelling that major advances in models incorporating the distance decay concept have occurred in recent years.

(iii) Recent Advances in Social Physics

In transportation studies the gravity model has usually employed trip origins and trip destinations as the two 'masses' of the original analogy. This model can be written

$$I_{ij} = k \frac{O_i D_j}{d_{ij}^{\,b}}$$

where Oi is the number of trips leaving zone i and D• the number of trips terminating at zone j. Let us assume two zones, a central area (C) with all of the jobs (1000) and a periphery (P) with all of the workers' residences (1000). We will further assume that the model has been calibrated so that b = 2 and k = 1/250. The distance from C to P is 2 miles. Hence journeys to work from P to C are

$$I_{PC} = \frac{1}{250} \times \frac{1000 \times 1000}{2^2} = 1000$$

$$\text{and} \quad I_{CP} = \frac{1}{250} \times \frac{0 \times 0}{2^2} = 0$$

The model is obviously working quite logically with work places and residences completely separated between zones. However let us now assume some expansion so that the jobs in C are doubled with the result that workers living in P are also doubled. What happens to the flow from P to C? This should obviously also double. However using our model above we now predict

$$I_{PC} = \frac{1}{250} \times \frac{2000 \times 2000}{2^2} = 4{,}000$$

The flow seems to have quadrupled! The reason for this is that there is a basic flaw in this transportation model. The empirical data on flows occurs in three places in the model as interaction, origins and destinations. It is therefore necessary to ensure that these three elements maintain a consistent relationship. This is exactly what we have failed to achieve above. Such a gravity model can be considered an "unconstrained" version of little utility in transportation modelling. Alan Wilson (1967) has shown that it can be replaced by a constrained model of the form

$$I_{ij} = A_i B_j O_i D_j f(d_{ij})$$

where K is replaced with two sets of constants A• and B.; linked to origins and destinations respectively. These 'balancing factors" are calibrated to ensure the predicted flows correspond to the number of origins and destinations in each zone. The method of calibrating these new constants is described in Masser (1972). However the main advance that Wilson's work has brought goes beyond solving the simple practical problem posed above. From this beginning he has been able to build a new approach to gravity modelling using an entropy maximising approach to give the model a much more rigorous theoretical foundation. This approach is introduced by Gould (1972) and widely illustrated by Wilson (1970). In terms of our original mechanisms behind distance, the decay effect is developed as a result of setting a limit on the amount of money spent on journey to work in any urban system. The gravity model has quite literally been taken out of the world of a Newtonian mechanics analogy to the sounder basis of the techniques from modern statistical mechanics.

BIBLIOGRAPHY

A. Distance Decay Modelling

1. Anderson, T.R. (1956) "Reply to Ikle", American Sociological Review, 21, 714-5.

2. Bateman, A.J. (1947) "Contamination in seed crops III Relation with isolation distance", Heredity, 1, 303-36.

3. Carrothers, G.A.P. (1956) "An historical review of the gravity and potential concepts of human interaction", Journal of the American Institute of Planners, 22, 94-102.

4. Goux, J.M. (1962) "Structure de l'espace et migration", in Sutter, J. (ed.) Human Displacements (Entretiens de Monaco en Sciences Humaines: Premiere Session).

5. Huff D.L. and G.F. Jenks (1968) "A graphic interpretation of the friction of distance in gravity models", Annals, Association of American Geographers, 58, 814-24.

6. Ikle, F.C. (1954) "Sociological relationship of traffic to population and distance", Traffic Quarterly, 8, 123-36.

7. Isard, W. (1960) Methods of Regional Analysis (M.I.T: Cambridge, Mass).

8. Marble, D.F. and J.D. Nystuen (1962) "An approach to the direct measurement of community mean information fields", Papers, Regional Science Association, 11, 99-109.

9. Morrill, R.L. (1963) "The distribution of migration distances", Papers, Regional Science Association, 11, 75-84.

10. Morrill, R.L. and Pitts F.R. (1967) "Marriage, migration and the mean information field: a study in uniqueness and generality", Annals, Association of American Geographers, 57, 401-22.

11. Olsson, G. (1965) Distance and Human Interaction. A Review and Bibliography(Regional Science Research Institute, Bibliographic Series, No. 2).

12. Stewart, J.Q. (1941) "An inverse distance variation for certain social influences", Science, 93, 89-90.

13. Stewart, J.Q. (1947) "Empirical mathematical rules concerning the distribution and equilibrium of population", Geographical Review, 37, 461-85.

14. Taylor, P.J. (1971a) "Distance transformation and distance decay functions: Geographical Analysis, 3, 221-38.

15. Zipf, G.K. (1946) "The P1P2/D hypothesis: on the inter-city movements of persons", American Sociological Review, 11, 677-86.

16. Zipf, G.K. (1949) Human Behaviour and the Principle of Least Effort (Cambridge, Mass.).

B. Applications of Distance Decay Concepts

17. Boalt, G. and Janson C.G. (1957) "Distance and social relations Acta Sociologica, 2, 73-96.

18. Carroll, J.D. and Bevis H.B. (1957) "Predicting local travel in urban regions," Papers, Regional Science Association, 3, 183-97.

19. Claeson C.F. (1964) "A regional chorological analysis of the cinema audience," (English summary) Geografiska Annaler, 46, 121-7.

20. Dodd, S.C. (1951) "A measured wave of inter-racial tension", Social Forces, 29, 281-9.

21. Galle, O.R. and Taueber K.E. (1966) "Metropolitan migration and intervening opportunities", American Sociological Review, 31, 5-13.

22. Gleave, D.J. (1970) "Trends in journey to work to Newcastle 1921-1966," Department of Geography, University of Newcastle upon Tyne, Seminar Papers, No. 13.

23. Gould, P.R. (1969) "Spatial diffusion", Resource Paper, No. 9 (Association of American Geographers).

24. Hagerstrand, T. (1957) "Migration and area", Lund Studies in Geography, B13, 27-158.

25. Hagerstrand, T. (1968) Innovation Diffusion as a Spatial Process (Translated by A. Pred, Chicago University Press).

26. Lowry, I.S. (1963) "Location parameters in the Pittsburgh model", Papers, Regional Science Association, 11, 145-66.

27. Olsson, G. (1967) "Central place systems, spatial interaction and stocastic processes", Papers, Regional Science Association, 18, 13-45.

28. Rahmatullah, Md. and O'Sullivan P. (1968) "The journey to work between local authorities on Merseyside from 1921 to 1966", London School of Economics, Graduate School of Geography, Discussion Paper, No. 12.

29. Starkie, D.N.M. (1967) "Traffic and industry. A study in traffic generation and spatial interaction", London School of Economics, Geographical Papers, No.3.

C. Further Reading

30. Cesario, F.J. (1975) "Linear and non-linear regression models of spatial interaction",Economic Geography, 51, 69-77.

31. Cliff, A.D., Martin, R.L. and Ord, J.K. (1974) "Evaluating the friction of distance parameter in gravity models", Regional Studies, 8, 281-6.

32. Cliff, A.D., Martin, R.L. and Ord, J.K. (1975) "Map pattern and friction of distance parameters: reply to comments", Regional Studies, 9, 285-7.

33. Curry, L. (1972) "A spatial analysis of gravity flows", Regional Studies, 6, 131-47.

34. Curry, L., Griffith, D.A. and Sheppard, D.S. (1975) "Those gravity parameters again", Regional Studies, 9, 289-96.

35. Gould, P.R. (1972) "A pedagogic review", Annals, Association of American Geographers, 62,.689-700.

36. Hyman, G.M. (1969) "The calibration of trip distribution models", Environment and Planning, 1, 105-12.

37. Johnston, R.J. (1973) "On frictions of distance and regression coefficients", Area, 5, 187-91.

38. Johnston,R.J. (1975) "Map pattern and friction of distance parameters: comment" Regional Studies, 9, 281-4.

39. Johnston,R.J. (1976) "On regression coefficients in comparative studies of friction of distance", Tidjschrift voor Economische en Sociale Geografie, 67, 15-28.

40. Masser, I. (1972) Analytical Models for Urban and Regional Planning (Newton Abbot: David & Charles).

41. Olsson G. (1970) "Explanation, prediction and meaning variance: an assessment of distance interaction models", Economic Geography, 46, 223-31.

42. Rushton, G. (1969) "Analysis of spatial behaviour by revealed space preferences" Annals, Association of American Geographers, 59, 391-400.

43. Taylor, P.J. (1971b) "Distances within shapes: an introduction to a new family of finite frequency distributions", Geografiska Annaler, 53B, 40-53.

44. Wilson, A.G. (1967) "A statistical theory of spatial distribution models", Transportation Research, 1, 253-69.

45. Wilson, A.G. (1970) Entropy in Urban and Regional Modelling (London:Pion).