Environment and Planning A, 1976, volume 8, pages 189-214

The method of residues in urban modelling

M Batty
Department of Geography, University of Reading, Whiteknights, Reading, RG6 2AF, Berkshire, England
L March
Department of Design Technology, The Open University, Walton Hall, Milton Keynes, MK7 6AA, England
Received 11 July 1975

Abstract. This paper seeks to extend the macrostatic approach to urban modelling by treating modelling problems as many-stage processes. Within such a process the early stages are concerned with explaining the relatively trivial characteristics of the phenomena of interest, and the later stages are devoted to explaining more important behavioural issues. Coleman (1964) calls this approach the 'method of residues', and its power is first demonstrated here by a reinterpretation of the well-known gravity model. An ad hoc test of the method on the Toronto-centred region serves to emphasise the need for a more formal approach, and thus an analogy between the method and the Bayesian viewpoint is introduced. A method of information minimising, more general but consistently and unambiguously related to the method of entropy maximising, is used to make the formal approach operational, and the method is used to generate an 'extended' family of spatial-interaction models. A number of spatial-interaction models are derived, and the paper is concluded by a test of two of these models on the Toronto-centred region.

1 Introduction

In the effort to design manageable models of cities and regions that are capable of use in a policymaking context, the most successful approach so far has undoubtedly been framed in terms of macroanalysis. Most urban models have been built at a highly aggregate level, in which the simplest form of urban structure has been described statistically in static terms. The appeal of this approach is easy to comprehend for the requirements of simplicity and practical applicability, combined with the somewhat primitive level of available urban theory, have made the macrostatic approach the only one feasible (Batty, 1976). Yet in its wake there has come the call for extensions to this approach, particularly through the treatment of system behaviour through time (dynamics), and through the description of more disaggregate structure (Wilson, 1974). Progress has, however, been slow. Despite the evident need for such improvements, more microdynamic modelling has been inhibited by the perennial data problem, by statistical significance problems and, more importantly, by problems of formulating hypotheses capable of being disentangled from the myriad of factors affecting urban systems and capable of being tested accordingly.

This recognisable pressure towards more microdynamic modelling may well contain the seeds of its own destruction, at least in the short term. The early history of urban modelling in the United States was beset by problems originating from efforts to model disaggregate systems, and the prospect for such models, especially in a practical context, is still not good (Lee, 1973). However, there are other possibilities for improving urban models which retain a macrostatic emphasis and, although these approaches are less easy to engender, one such approach will be advanced in this paper. The basis of this approach rests on the idea that, within any theory of urban structure, there are sets of ideas which are more appealing and more relevant than others. In one sense, it is possible to view any theory as composed of a number of distinct parts which could be ordered according to their sophistication, relevance, and certainty, or testability. For example, current urban economic theories seem to contain much stronger insights about residential location than about industrial or service location, although they attempt to explain both. The method to be outlined here is based on the premise that theories can be so ordered and that different types of models must be used to describe different parts of the theory. In another sense, this approach is one in which a hypothesis or theory is modelled sequentially in stages, each stage building on the previous one.

The simplest form for this approach involves a two-stage process of model building, and in statistical applications this process has been known for many years. A typical example might involve a linear model of some phenomenon calibrated by a technique such as least squares, and embodying some ideas about the causal structure of the system through its independent variables. This would constitute the first stage, and a second stage is often suggested in which a further explanation of the phenomenon is attempted by examining the residuals, and building another model to explain this variation. The approach just described is plagued by problems involving the separability of the phenomena into these stages, and in the past the second stage has been rarely attempted and has never been treated as a matter of course. Yet the idea seems to have some merit, especially if a convincing means for distinguishing the various stages is devised. In fact one of the starting points of this paper involves such a distinction; Coleman (1964), in his book on mathematical sociology, presents a profound example of the method which he terms the 'method of residues'. He demonstrates how a trivial aspect of spatial variation must be factored out from data in the first stage, and in the second stage the residues become the all-important phenomena to be explained. There is a subtle change in emphasis here, one which will be developed in the argument of this paper, for it will be demonstrated that a possible approach to urban modelling can be devised in which the initial stages involve modelling the more trivial characteristics and the later stages the more important ones.

Although the development of the method of residues will comprise the core of this paper, it is not proposed to develop a formal framework for applying the method at the start, for there are a number of problems to be explored first. Thus the thesis of this paper will be organised in such a way that the initial applications of the method to location in the Toronto-centred region lead naturally to problems which can be resolved by a more formal approach. The method will be described first, and its application in understanding the form of the well-known gravity model of spatial interaction will be presented. In fact the method leads to some new insights into the form of this model, and some long-standing problems, if not resolved, are at least reinterpreted. The basic method is applied in an ad hoc fashion to the Toronto-centred region, and on the basis of these applications a new framework is defined.

From these rather ad hoc attempts at making the method operational it is clear that a more consistent approach is required, and thus the new framework attempts to see the two-stage approach to modelling as one in which prior and posterior probability models are derived. In particular, a Bayesian approach to the method of residues is implied, and a formal technique for deriving single models incorporating both stages is presented. The technique is based on information theory and is one of information minimising, which is intrinsically related to the well-known technique of entropy maximising. A demonstration of the technique in terms of the gravity models introduced earlier is given, and in particular the family of spatial-interaction models due to Wilson (1971) is extended. These new models are then applied to the Toronto-centred region and comparisons are drawn between the ad hoc and formal approaches.

Like any departure from existing practice this new approach is speculative and tentative; the style chosen for presenting the approach involves a mixture of simple theory and empirical application. The basis of the ideas contained in this paper is very simple, and thus they are presented in as elementary a way as is possible. An essential feature of this approach is to contrast theory against application so that both can draw from each other. And in this context the interposing of theory and application is all the more necessary and important owing to the somewhat speculative nature of the argument. But before the method is applied to locational modelling, it is worthwhile to consider its logic in more detail for this will serve to delimit the scope of the proposed approach.

2 The method of residues

2.1 *Theory and explanation as a many-stage process*

It is generally accepted by the scientific community that theory is improved sequentially and marginally, and that this is equivalent to the theory passing through a sequence of stages which reflect different identifiable problems facing the field. This is the process which Kuhn (1962) characterises as 'normal science', and the process of theory-building is usually achieved by "conjecture and refutation", using Popper's (1965) phrase. Moreover, in the eyes of later observers certain well-established theories, or parts of a theory, might appear trivial, yet it is a feature of all knowledge that when an idea is absorbed it seems obvious. Yet seldom are theories themselves divided into distinct parts or stages which build on each other, and although this might seem odd, given the history of theory-building as one of successive improvement, it is easily explained. In science there is a widespread reticence against coupling distinct theories or ideas together, for this is likely to lead to inconsistencies and unforeseen repercussions. Such lack of consistency, however practical a coupling of theories might appear, is against the grain of science—for it has none of the conceptual and theoretical elegance which characterises great theory. In physical science, however, the desire for elegance has generally been met, for there is a major force towards ironing out inconsistencies and deriving universals through research.

But in the social sciences the situation is quite different. The social world cannot be interpreted in the same way as the physical world, and the nature of social systems, and the intrinsic difficulty in explaining their form forces the researcher into more expedient ways. Furthermore the requirement for practical methods to solve problems *now* also focuses the field on more ad hoc and pragmatic approaches, and inconsistency within and between theory is the norm. In the field of urban modelling, examples of coupling theories together or breaking models into distinct stages are plentiful. For example, the CONSAD model built for the Pittsburgh Community Renewal Program (Steger, 1965) was composed of three submodels—input–output, industrial location, and residential-service location—which were strung together in a simple linear sequence. Even in theoretical developments of urban models, researchers have been forced to couple together existing techniques. Wilson's (1974) general model is conceived in this way, and inevitably there have been major criticisms of such approaches owing to the potential propagation of errors (Alonso, 1968).

In the introduction another reason, apart from expediency, was implied in support of urban modelling as a many-stage process. Because of the nature of urban systems, different parts of any theory have different degrees of strength and significance, and thus it is quite logical to treat these different parts accordingly. And there is one particular case where to treat an urban model in any other way would be quite foolish, and this is the case where the urban model contains characteristics which are trivial in terms of their explanation. If a model contains such characteristics, then these can come to dominate the model, and this criticism has been frequently levelled against several major applications. If trivialities are identifiable the logical approach would be to filter out these characteristics in the early stages of modelling, in preparation for the true 'explanation' achieved in later stages, using a process analogous to a communication engineer's attempt to filter out noise, or a statistician's attempt to define periodicity in time series. This problem is not an easy one to evaluate, but in urban modelling it is immensely significant, as can be seen from the debate over spatial correlation and boundary effects (Curry, 1972; Cliff et al., 1974), which plagues the field of spatial analysis with nontrivial effects from trivial causes.

Coleman (1964) sums up the problem quite cogently when he states, "in considering a given complex social phenomenon, certain aspects of it are explainable by 'sociologically trivial' assumptions, or by matter irrelevant to the substantive matters under investigation. If we examine what part of the behavior can be explained by these 'trivial' factors, then the remainder stands out to be explained by less trivial factors". Coleman calls the method, which he goes on to develop, the 'method of residues', and he demonstrates how the method can be used sequentially and hierarchically to filter out and explain various characteristics of a problem. Coleman refers to such trivial characteristics as 'null hypotheses' which form some starting point or baseline on which to begin the explanation. In this paper, a proposal for applying a two-stage method of residues to urban modelling is elaborated and applied both in pragmatic and in formalised ways. The particular model which is used in an urban context is the gravity model; this example demonstrates that the general approach implied by the method of residues does not only lead to a better statistical explanation but it also leads to a more satisfying theoretical explanation, and this fact reinforces the argument that this method has much more than superficial value.

2.2 *Gravity: the physics of space*

Almost from the late 1940s, when models of spatial interaction were first formulated explicitly, there has been a debate about whether or not such models have a behavioural significance and whether or not such models simply reflect the way in which boundaries are drawn and space is partitioned. In recent years this problem has been considered by Curry (1972), who suggests that the form of the map over which spatial interaction takes place is an all-important determinant of the pattern of interaction. In other words it is space itself rather than any behavioural structure which works the model, and thus the performance of the model is an obvious consequence of the physics of the space. The uncertainties surrounding the argument have perhaps inhibited its full development, for there are few clear statements of the problem, and little formal analysis of the effect of space.

Yet there is a simple and cogent way of deriving a model which describes spatial interaction and which requires only minimal assumptions. Consider the probability of interaction p_{ij} between two points i and j separated by some distance d. Then the total interaction between i and j must be proportional to all possible pairwise connections, ϕ_i, generated by the persons living at i and, ϕ_j, by those at j. Thus p_{ij} is directly proportional to the product $\phi_i\phi_j$. If we assume equal population densities and also that a person is equally likely to travel distance d miles as $d+\Delta$ miles (in the absence of knowledge to the contrary), it can be shown that the probability p_{ij} varies in inverse proportion to distance. Consider two destinations, j_1 and j_2, such that the distances take the following order,

$$d_{ij_1} = d < d+\Delta = d_{ij_2} \ .$$

Now the destination j_1 can be located anywhere on the circumference of a circle of length $2\pi d$, whereas destination j_2 can be anywhere on the circle of length $2\pi(d+\Delta)$.

With equal population densities, there are $(d+\Delta)/d = 1+\Delta/d$ as many persons on the more distant circle, and thus the probability of interaction to all points on the nearer circle is higher by this factor. This in turn implies that the probability of interaction from a fixed origin to any destination declines proportionately with distance from that origin. Stated formally, these conditions can be combined to give

$$p_{ij} \propto \frac{\phi_i \phi_j}{d_{ij}} . \tag{1}$$

The model in equation (1) is an example of what Coleman (1964) refers to as a *null hypothesis*: it is the expected pattern of interaction, given no assumptions about the behaviour of the population apart from the assumption of travel indifference to distance. The similarity to the well-known gravity model is striking, but this is more a consequence of the physics of the situation than any behavioural postulates. In fact this model was originally derived by Zipf (1949) who used an argument similar to that above. However, Zipf never expounded his argument in detail, and in his work he simply states that equation (1) is a direct consequence of the law of precise geometrical probability. Zipf's work was swamped by the alternative derivations of similar models, particularly those by Stewart (1942; 1947), and the essential logic of Zipf's ideas appears to have been forgotten by succeeding generations of model builders. But, in this paper, equation (1) takes on a new significance because it becomes a central model in applications of the method of residues and, more important, in a reinterpretation of spatial-interaction models.

In contrast to Zipf's (1949) work, most early researchers who adopted the gravity model did so by using the analogy with the Law of three-dimensional Gravitation, which is a direct consequence of Newton's Second Law of Motion. The equation describing the gravitational force between two bodies i and j, with masses M_i and M_j respectively, is given by

$$F_{ij} = G\frac{M_i M_j}{d_{ij}^2} , \tag{2}$$

where

F_{ij} is the force between i and j, and

G is the gravitational constant.

Some researchers, for example Reilly (1931), have used equation (2) directly for estimating spatial interaction, but an alternative analogy was used by Stewart (1942). Stewart, in a series of articles, argued that the correct relationship for interaction phenomena was not one of force but one of energy. The equation for gravitational potential energy is defined in Newtonian mechanics as being proportional to the integral of equation (2), and this is given by

$$V_{(\text{ex})ij} = G\frac{M_i M_j}{d_{ij}} , \tag{3}$$

where

$V_{(\text{ex})ij}$ is the potential energy interchange between mass i and mass j.

Stewart then went on to define potential energy at a point or origin i by summing equation (3) over i and rearranging, thus

$$V_i = \frac{\sum_j V_{(\text{ex})ij}}{M_i} = G\sum_j \frac{M_j}{d_{ij}} , \tag{4}$$

where

V_i is the potential energy at i.

Stewart maintained that equation (4) was the correct one for describing demographic potential, with population density replacing M_i, and the constant G fixing the appropriate potential scale.

There are several comments to be made on these arguments. The seeming correspondence between Zipf's derivation [equation (1)] and Stewart's analogy [equation (3)] is ambiguous, to say the least. In the endless debate about whether or not distance should be raised to the second power, or whether or not the power should be a parameter to be estimated, Stewart and Warntz (1958) maintained that equation (3) is not inconsistent with equation (2) since both have their derivation from Newtonian mechanics. But equations (1) and (2) do appear inconsistent, for one is derived from two-dimensional probability considerations and the other from three-dimensional gravitational considerations. In fact, very few researchers appear to have picked up this point—that two-dimensional gravitational models are likely to be much more suitable than three-dimensional models for social systems of interest. From the literature of that time Stewart and Zipf make substantial reference to each other's work, but they do not really emphasise any of their differences, and there were several. In fact, because of the similarity of equations (1) and (2), and because these similar equations were both tested empirically by each author, results were quite consistent with the use of either equation regardless of how it was derived. But it does seem that Zipf was aware of the differences between the ways in which these equations were postulated. Yet he was able to rationalise the problem by arguing that his 'null hypothesis' contained only one of several effects which account for variations from equation (1). In this way he was able to accept the need for variation in the exponent on distance much more readily than Stewart, and his work is quite consistent with Coleman's (1964) method of residues, from which it seems to have been derived. In fact Zipf distinguished between two effects of distance: the first, an effect concerning the increasing cost or effort of overcoming distance, and the second, the declining opportunity for contact per unit area at decreasing distance. And he explicitly made tests of the second effect (the null hypothesis), leaving the first effect to be explained by other methods.

The null hypothesis contained in equation (1) is so appealing that an ad hoc test of the model as a first stage in the method of residues, and a second-stage explanation of the residuals, is attempted below. But, as a baseline to this work, an elementary residential-location model of the kind derived by Wilson (1969) will also be applied and used to engender comparisons. The applications throughout this paper are based on data taken from the Toronto-centred region which includes the cities of Hamilton, Toronto, and Oshawa, and stretches some sixty miles along the northern shore of Lake Ontario.

3 Ad hoc applications of the method: examples in the Toronto-centred region

3.1 *The singly-constrained residential-location model*

The first model to be applied here—the baseline model—is perhaps the best known and most widely used of all gravity models at the present time. It exists in various forms: in traffic-distribution modelling due to Voorhees (1955); in retail-location studies due to Casey (1955), Huff (1964), Schneider (1959), Harris (1964), and Lakshmanan and Hansen (1965); and in residential modelling due to Wilson (1969) The model is of particular significance, for although it is difficult to derive behaviourally it is widely believed that the model 'captures' actual behaviour. Thus a comparison of this model with the null hypothesis is likely to be revealing.

The model can be stated in probability terms as follows:

$$p_{ij} = K_i O_i D_j \exp(-\lambda d_{ij}) \ , \qquad (5)$$

where
p_{ij} is the probability of interaction between origin i and destination j,
O_i and D_j are the amounts of activity in origin i and destination j respectively,
λ is a parameter controlling the influence of distance d_{ij}, and
K_i is a balancing factor defined below.
λ is a parameter to be estimated by some numerical–statistical procedure and is approximately inversely related to the total distance travelled in the system. The model is subject to a constraint on the origins, which is written

$$\sum_j p_{ij} = \frac{O_i}{T} , \tag{6}$$

where T is the total activity in the region. This is defined as

$$\sum_i O_i = \gamma \sum_j D_j = T . \tag{7}$$

Note that γ is an arbitrary scaling constant and, without loss of generality, can be assumed to be equal to unity. Summing equation (6) over i and substituting from equation (7), it is clear that

$$\sum_i \sum_j p_{ij} = \frac{\sum_i O_i}{T} = 1 . \tag{8}$$

The constant K_i can be evaluated by summing equation (5) over j and rearranging by use of equation (6). Then

$$K_i = \left[T \sum_j D_j \exp(-\lambda d_{ij}) \right]^{-1} , \tag{9}$$

and the interaction model can now be written out explicitly in terms of the trips, called T_{ij}, made between i and j, thus

$$T_{ij} = Tp_{ij} = O_i \frac{D_j \exp(-\lambda d_{ij})}{\sum_j D_j \exp(-\lambda d_{ij})} . \tag{10}$$

Of particular interest in such a model is the number of trips attracted to any particular destination zone j, which can be regarded as an estimate of the location of activity in j. This prediction, called D_j^*, is calculated by summing equation (10) over i, thus

$$D_j^* = \sum_i T_{ij} = T \sum_i p_{ij} . \tag{11}$$

In the context of the residential-location model to be applied here, O_i is an estimate of the persons working in i, and D_j is a measure of the attraction of the residential zone j. Thus D_j^* is the predicted number of workers who live at j.

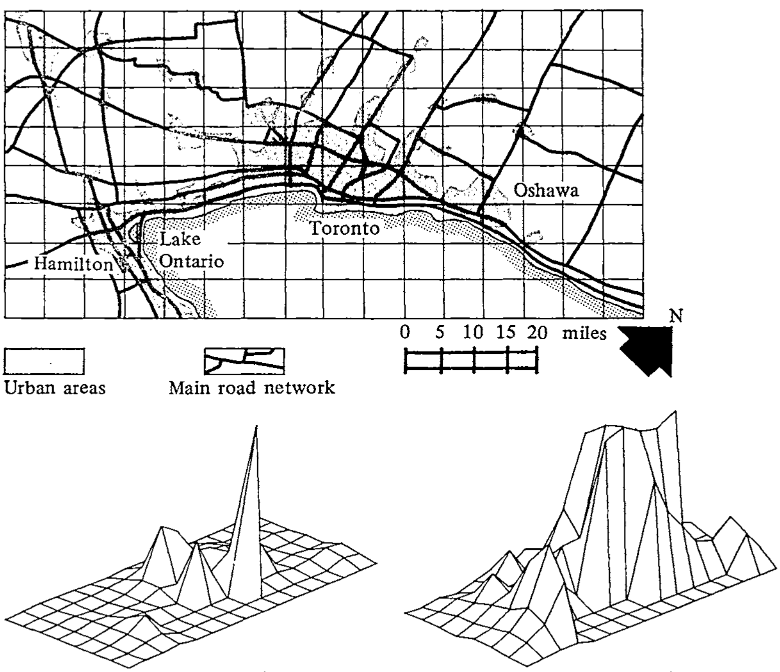
The model given in equations (10) and (11) has been calibrated to the Toronto-centred region shown in figure 1. The origin and destination distributions $\{O_i\}$ and $\{D_j\}$ are equivalent to the distribution of employment $\{E_i\}$ and the scaled distribution of population $\{P_j(E/P)\}$, where E and P are total employment and total population in the system respectively. The data is taken from the 1971 National Census, and a zoning system based on a neutral square grid has been adopted for the modelling work. Figure 1 shows the urbanised area and main transportation routes, and the isometric solids for employment and scaled population reveal that the region has the characteristic spatial form of most Western cities. The employment density is sharply peaked, in downtown Toronto and north and west of the city, in contrast to the population density which is much less peaked but nevertheless declines quite regularly with increasing distance from downtown Toronto. The region had a population of some three million people in 1971, and these characteristics make it highly suited to the application pursued here.

The model was calibrated by finding a value for λ which reproduced the mean travel distance generated in the region; that is, λ was chosen so that

$$\frac{1}{T}\sum_i \sum_j T_{ij}^{\text{obs}} d_{ij} = \sum_i \sum_j K_i O_i D_j \exp(-\lambda d_{ij}) d_{ij} \ , \qquad (12)$$

which several authors (Hyman, 1969; Evans, 1971; and Kirby, 1974) have shown to be the maximum-likelihood estimate for λ. Equation (12) was solved by using the Newton–Raphson method (Batty, 1976), and four iterations of the method were required. In terms of the measures of fit used, the performance of this model is fair, and the statistics summarising this fit are shown in table 1. Figure 2 gives a graphical demonstration of the fit in terms of four isometric plots. The first two show the observed and predicted distributions $\{D_j^{\text{obs}}\}$ and $\{D_j^*\}$, whereas the second two are plots of the residuals. First, the actual residuals, R_j, defined as

$$R_j = D_j^* - D_j^{\text{obs}} \ , \qquad (13)$$



Distribution of employment $\{O_i\}$ Distribution of population $\{D_j\}$

Figure 1. Spatial form of the Toronto-centred region.

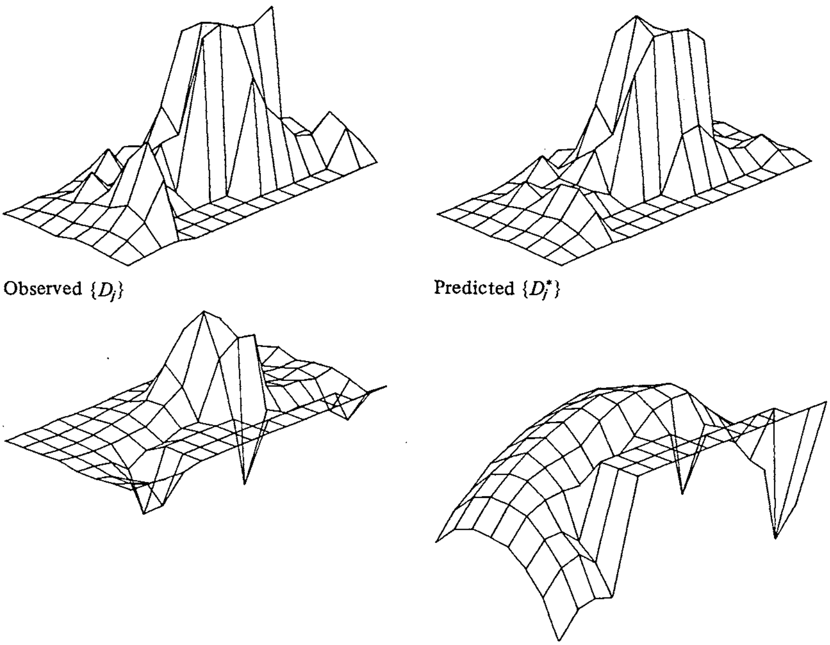
Table 1. Performance of the singly-constrained baseline model.

Observed mean travel distance	6·1000
Predicted mean travel distance	6·0997
Parameter, λ	0·2230
Correlation r of $\{D_j^{\text{obs}}\}$ and $\{D_j^*\}$	0·9757
Coefficient of determination, r^2	0·9521
Intercept of regression of predictions on observations	1784·8750
Slope of regression	0·8382
Sum of the absolute deviations between predictions and observations	295377

and then, the percentage residuals, r_j, defined as

$$r_j = \frac{D_j^* - D_j^{\mathrm{obs}}}{D_j^*} , \tag{14}$$

are plotted. The percentage residuals are perhaps the most important in evaluating whether or not any systematic pattern in the residuals is present and, from figure 2, it appears that the model underestimates population in the actual downtown zones and on the periphery of the region—at Hamilton and Oshawa. This might be corrected by finding a better distance function which gives greater estimates at the origin than the negative exponential function. However, the real interest here is in assessing how much of the explanation accounted for by this residential model can be accounted for by the null hypothesis, and this constitutes the next application.



Observed $\{D_j\}$ Predicted $\{D_j^*\}$

Absolute residuals $\{R_j\}$ Percentage residuals $\{r_j\}$

Figure 2. Spatial predictions and residuals from the singly-constrained model.

3.2 *The Coleman–Zipf model: a null hypothesis*

The model based on equation (1) can be rewritten using the notation and definitions given in the previous section. The model then becomes

$$p_{ij} = K O_i D_j d_{ij}^{-1} , \tag{15}$$

where O_i and D_j have replaced ϕ_i and ϕ_j respectively, and K is a scaling constant which is introduced to ensure that

$$\sum_i \sum_j p_{ij} = 1 . \tag{16}$$

K can easily be evaluated in the usual manner by summing equation (15) over i and j, and rearranging. Then

$$K = \left(\sum_i \sum_j O_i D_j d_{ij}^{-1} \right)^{-1} , \tag{17}$$

and the explicit trip model is thus given as

$$T_{ij} = Tp_{ij} = T\frac{O_i D_j d_{ij}^{-1}}{\sum_i \sum_j O_i D_j d_{ij}^{-1}} . \tag{18}$$

The model in equation (18) is a particular case in the class of unconstrained spatial-interaction models defined by Cordey-Hayes and Wilson (1971); in this model, we can estimate not only the trips attracted to a destination j but also the trips produced in an origin i. First, the trips produced in i, which are related to activity located at i, in this case employment, are calculated as

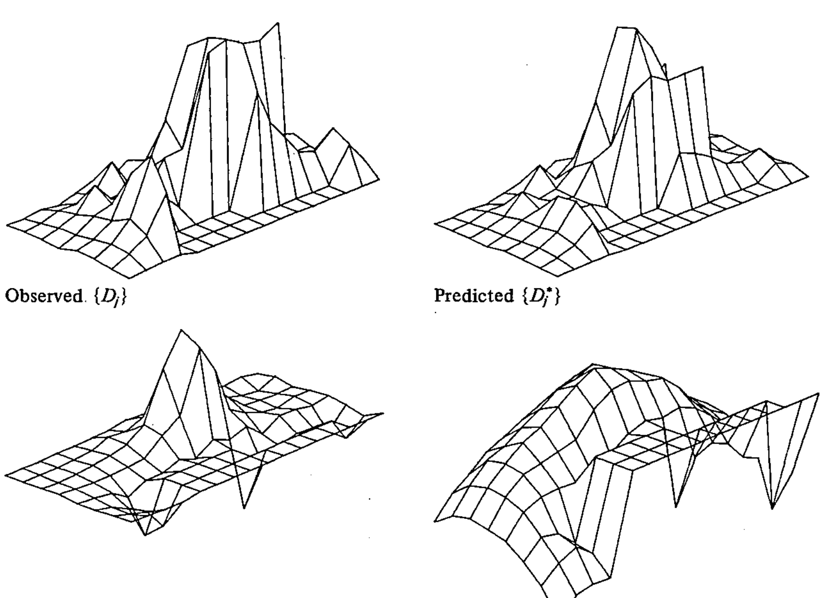
$$\begin{aligned} O_i^* &= \sum_j T_{ij} = TKO_i \sum_j D_j d_{ij}^{-1} \\ &\propto O_i V_i^{(1)} . \end{aligned} \tag{19}$$

The interpretation of equation (19) is interesting in the light of the earlier use of potential by Stewart (1942), and it is clear that the estimated activity in origin i, predicted by the model, is proportional to the actual measure weighted by its potential. The same interpretation can be made for the activity located at the destination j, which is calculated as

$$\begin{aligned} D_j^* &= \sum_i T_{ij} = TKD_j \sum_i Q_i d_{ij}^{-1} \\ &\propto D_j V_j^{(2)} . \end{aligned} \tag{20}$$

Of further interest is the relationship between this model and Hansen's original accessibility-potential model (Hansen, 1959), but this will not be explored further here.

The model defined by equations (18) and (20), which embodies the null hypothesis, has been applied directly, and its performance is presented in table 2. In figure 3, a visual presentation of the origin and destination estimates and their residuals is also made, and it is clear from these results that the model performs quite well. What is remarkable from a theoretical perspective is the huge amount of variation that is



Observed $\{D_j\}$ Predicted $\{D_j^*\}$

Absolute residuals $\{R_j\}$ Percentage residuals $\{r_j\}$

Figure 3. Spatial predictions and residuals from the Coleman-Zipf model.

accounted for by this null hypothesis: 84% of the origin activity and 93% of the destination activity are accounted for by the model; yet, this is not so remarkable if the forms of the singly-constrained baseline and the Coleman–Zipf models are compared directly. The models are similar in several senses, hence the equally creditable performance of the Coleman–Zipf model against the baseline is explicable. What is quite intriguing, however, is the fact that a large proportion of the variance in these spatial activities can be accounted for by a model based on minimal assumptions, and this leads one to consider additional assumptions which have more potential explanatory power. Some may immediately object that these models are tautological in that some variables concerning the origin and destination activity, O_i and D_j, are used to predict the same variables O_i^* and D_j^*. The argument is recurrent throughout the use of gravity models as location models, and it is not proposed to detail the arguments again here; but interested readers are referred to Batty (1976), Wilson (1970), and Echenique et al. (1969) for various viewpoints. For those who remain sceptical, note that the predicted trip distribution against the observed trip distributions give a performance similar to the performance of the marginal totals of these distributions compared here.

Table 2. Performance of the Coleman–Zipf model.

	Origin activity $\{O_i^*\}$	Destination activity $\{D_j^*\}$
Correlation r of predicted and observed activities	0·9171	0·9653
Coefficient of determination r^2	0·8410	0·9318
Intercept of regression of predictions on observations	−839·5391	1534·8516
Slope of regression	1·0754	0·8604
Sum of the absolute deviations between predictions and observations	368921	293325

Predicted mean travel distance for the system = 6·1684

3.3 *Models of residues*

The method of residues is based on the idea that once the relatively straightforward characteristics, such as those reflected in the Coleman–Zipf model, have been filtered from the data, then the task to explain the resulting data begins in earnest. In this context the filtering out of these characteristics has already been achieved by the application of the model in the previous section, and the actual residuals provide the new set of observations on which the second stage of modelling is based. This division of the process into two stages is based on nothing more than common sense, for no effort has yet been made to model both stages as one; in fact, this commonsense approach is a pragmatic one, hence the term ad hoc applications.

The search for pattern or order in the residuals can embody a variety of approaches: the search could proceed inductively by using factor analysis, it could proceed by using an elementary theory of linear additive causality, or it could proceed by using stronger deductive theory, thus giving rise to integrated nonlinear models such as those used in spatial interaction. The variables used in the first stage can be used again if it is felt that certain effects related to these variables have not yet been captured, or they might be excluded on grounds of duplication; whatever the mix of variables adopted, their ultimate significance will depend upon the plausibility and strength of the theory postulated to explain the second stage and, perhaps more importantly, on the relationships between the first- and second-stage theories. In this example, two models of the residues will be advanced; the first is a linear model which contains a variety of variables whose inclusion depends upon their performance, and the second is a spatial-interaction model of the kind given in equation (10), which includes a hypothesis concerning the additional effect of distance on the spatial form of the system.

The linear model used to explain the pattern of residuals, R_j, has been applied both to the origin and the destination residuals that are computed from the predicted origin and destination activities given in equations (19) and (20). The model has the general form

$$R_j = \sum_l \alpha_l X_{jl} , \qquad l = 1, 2, ..., L , \tag{21}$$

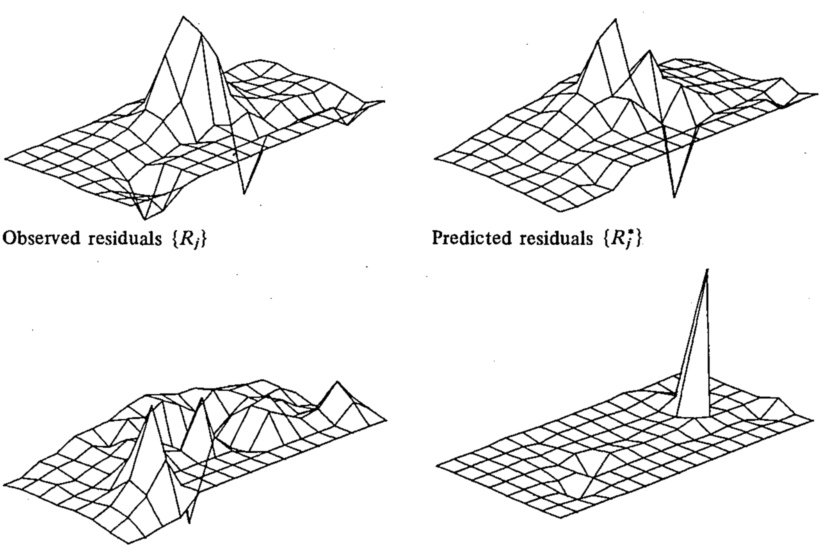
where

X_{jl} is the value of the independent variable l in zone j, and

α_l is the coefficient describing the effect of X_{jl} in the regression.

The model has been calibrated using least squares, and from an initial set of L independent variables the set has been narrowed by using a stepwise regression of the kind developed by Efroymson (1962). Four runs were made; the origin and destination residuals were modelled, first using the partial-correlation criteria of entry into the regression, and second using the stricter criteria based on the partial F-test (for details see Draper and Smith, 1966). These results are given in table 3, and it is immediately clear that a good proportion of the residual variance is accounted for by this procedure. However, a more dramatic demonstration of the way in which pattern can be accounted for is given in figure 4 where the destination residuals and residuals of residuals are plotted; these diagrams reveal that the previous pattern in the residuals is almost entirely 'explained' by the regression, and the remaining effect is quite random, apart from the anomalous case of downtown Toronto.

The second model used in attempting to explain the residuals was a version of the singly-constrained residential-location model already given in equations (10) and (11).



Observed residuals $\{R_j\}$ — Predicted residuals $\{R_j^*\}$

Absolute residuals of residuals — Percentage residuals of residuals

Figure 4. Spatial predictions and residuals from the linear model of destination residuals.

The use of this model was based on the attempt to incorporate explicitly an additional factor of distance into the analysis by following Zipf's argument that apart from the null hypothesis effect of distance, there is another effect which reflects the declining propensity to travel with increasing distance, due to the additional effort and cost required. However, because interaction models of the type developed here are unable to account directly for negative quantities, the residuals were arbitrarily scaled as follows:

$$\overline{O}_i = (O_i^* - O_i) + |\min_i (O_i^* - O_i)| + 1 , \tag{22}$$

and

$$\overline{D}_j = (D_j^* - D_j) + |\min_j (D_j^* - D_j)| + 1 , \tag{23}$$

Table 3. Estimation of the linear models of residuals.

	Origin residuals		Destination residuals	
	with partial F-tests	without partial F-tests	with partial F-tests	without partial F-tests
Multiple correlation R^2	0·9493	0·9507	0·8340	0·8491
0 Constant	−6949	−10506	−5969	−11167
1 Population	0·0523 (3·2285)	0·0539 (3·0325)	–	−0·0134 (−1·0163)[a]
2 Services	−1·9237 (−16·2511)	−1·9236 (−13·6908)	−0·7000 (−7·8329)	−0·7494 (−7·1557)
3 Basic employment	–	–	–	–
4 Total employment	0·6107 (9·8638)	0·6144 (8·2875)	0·3111 (7·1109)	0·3552 (6·4293)
5 Access to population	–	−0·0033 (−0·3225)[a]	–	−0·0017 (−0·2301)[a]
6 Access to services	0·1117 (6·2711)	0·1565 (1·7029)	–	0·0584 (1·2586)[a]
7 Access to basic employment	–	–	0·0567 (7·7865)	0·0392 (7·2596)
8 Access to total employment	–	−0·0100 (−0·2407)[a]	–	–
9 Available land	–	26·3427 (0·2847)[a]	–	114·0006 (1·6583)
10 Yearly income	–	3·0620 (0·9565)[a]	–	−0·9583 (−0·4027)[a]
11 Net migration	–	−0·3064 (−0·0417)[a]	–	−0·5056 (−0·0926)[a]
12 Weekly rent	–	−5·8604 (−0·3029)[a]	–	−20·5372 (−1·4284)[a]

(1) The first number alongside each variable is the α_l coefficient, the second number in parenthesis is the Student t statistic.
(2) [a] indicates that the α_l coefficient is not significantly different from zero in terms of the t statistic set at the 95% confidence level.
(3) The F values for entry of a variable in the regression were set at 3·5, and for deletion at 3·0.
(4) Accessibilities were all calculated from the potential formula V_i; see equation (4) in the text.

and the explicit interaction model had the following form:

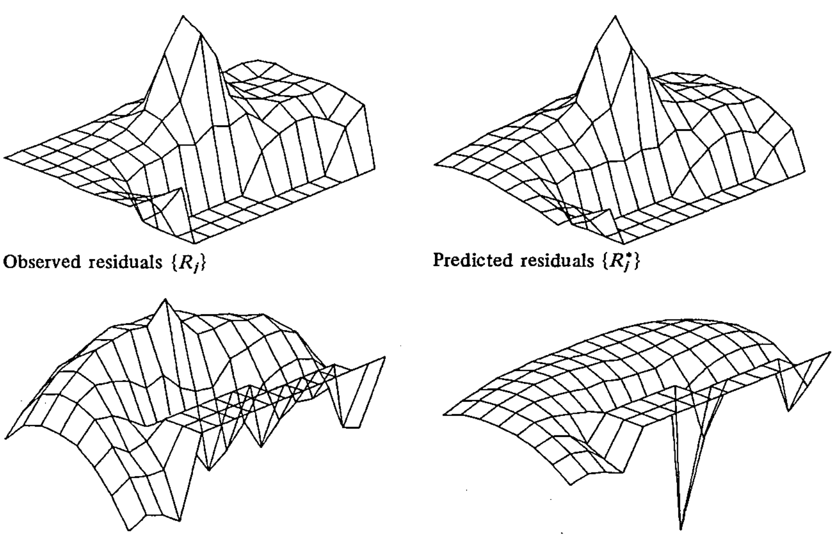
$$T_{ij} = T \frac{\overline{O}_i}{\sum\limits_i \overline{O}_i} \frac{\overline{D}_j \exp(-\lambda d_{ij})}{\sum\limits_j \overline{D}_j \exp(-\lambda d_{ij})} . \tag{24}$$

This model was recalibrated to the existing travel pattern by use of the Newton–Raphson method, and its performance seemed reasonable. The results are given (see table 4 and figure 5) in terms of the predicted residential locations of employees, and a cursory evaluation of these results reveals that most of the residual variance, except for downtown Toronto, is accounted for. And of additional significance is the fact that the variables used in the linear model of destination residuals are similar to those used in this interaction model of the same phenomena.

There are a number of related problems which immediately arise when models of residuals are built along the lines sketched above. Perhaps the major problem revolves around the difficulty of evaluating how the two stages interact with each other, but more important is the impossibility of separating effects, such as those of distance, due to two or more causes. But within the framework used here, it is not feasible to resolve these problems satisfactorily, for what is required is a stronger theory capable of accounting for the relationships between the stages. Given such a theory, it should be possible to build integrated single models in which all the stages are weighted appropriately, and it is to tackle this problem that the theory proposed in the following sections is addressed. First, an appropriate theory will be suggested and, in the rest of the paper, it will be elaborated and tested empirically.

Table 4. Performance of the singly-constrained model of destination residuals.

Observed mean travel distance	6·1000
Predicted mean travel distance	6·0999
Parameter λ	0·1563
Correlation r of $\{\overline{D}_j^{\text{obs}}\}$ and $\{\overline{D}_j^*\}$	0·9769
Coefficient determination r^2	0·9543
Intercept of regression of predictions on observations	936·0703
Slope of regression	0·9151
Sum of the absolute deviations between predictions and observations	148916



Observed residuals $\{R_j\}$ Predicted residuals $\{R_j^*\}$

Absolute residuals of residuals Percentage residuals of residuals

Figure 5. Spatial predictions and residuals from the singly-constrained model of destination residuals.

4 An information-minimising approach to the method of residues

4.1 *Prior and posterior probabilities: the Bayesian viewpoint*

One way of interpreting a many-stage modelling process is through information theory, in which each stage gives rise to additional information which needs to be incorporated within the model. In the special case of a two-stage procedure, the first stage can be likened to designing a model of *prior* probabilities which are transformed into *posterior* probabilities by the extra information generated at the second stage. There is, in fact, a very well-defined technique for making such a transformation in the case of the two-stage process, and one way of writing the formula for achieving such a transformation is as follows:

$$p_{ij} = \frac{p_{ij}^{\mathrm{o}} L_{ij}^{\mathrm{p}}}{\sum_i \sum_j p_{ij}^{\mathrm{o}} L_{ij}^{\mathrm{p}}} \tag{25}$$

where

p_{ij}^{o} is the prior probability assigned to the event ij,

L_{ij}^{p} is the likelihood of the event ij given this prior, and

p_{ij} is the posterior probability, which is clearly the product of the prior and the likelihood.

Equation (25) is of universal significance and is known as Bayes' equation; indeed the whole logic surrounding the Bayesian approach to statistical reasoning rests on the acceptance of prior probabilities, and much effort has been devoted to methods for assigning appropriate priors. In the context of the urban-modelling problems pursued here, the prior might relate to the Coleman–Zipf model, the likelihood to additional effects over and above the effects incorporated in the prior, and the posterior to an integrated probability distribution combining both prior and additional effects, or first- and second-stage information.

Problems of assigning prior probabilities have plagued the school of Bayesian statisticians since the time of Laplace, but quite recently a number of theorists, namely Cox (1961), Jaynes (1968), and Tribus (1969), have suggested that a consistent approach to assigning priors is through the technique of maximum entropy. The technique consists of deriving probabilities which are maximally unprejudiced but subject to certain information known about the context; the principle of maximising entropy involves encoding this information into a probability distribution which contains the greatest amount of uncertainty with regard to possible outcomes. It is, in this sense, 'maximally unprejudiced'. The technique will be outlined in the following section, but many readers will already be familiar with the method in urban modelling due to the pioneering work of Wilson (1970), who has used it extensively in deriving spatial-interaction models. In the work of Jaynes (1968) and Tribus (1969), in particular, it is suggested explicitly that the technique should be used to generate p_{ij}^{o}, but it is unclear how the resulting prior is to be extended within Bayesian analysis. Indeed Jaynes and Tribus would probably argue that the derivation of posteriors from priors is outside the scope of their immediate interest although, in this context, it is critical.

Therefore, the technique to be proposed here is somewhat different from the common use of maximum entropy to assign prior probabilities, for it is based on the notion that a technique of encoding a sequence of information into a probability distribution is more appropriate than a once-for-all encoding. In the case of the two-stage process it is proposed to apply a technique to the problem referred to as 'information-minimising', but intrinsically and unambiguously related to entropy-maximising; first, the information concerning the prior will be encoded implicitly, and second, additional information, such as that referred to by Zipf, will be encoded, and this will lead to a model of posterior probabilities identical to Bayes' equation [equation (25)]. Before proceeding to detail these developments it is of interest to note that Stewart (1950), some twenty-five years ago in a review of social physics, forecasted a similar line of development to that pursued here, which was started in the field of urban modelling by Wilson (1970).

4.2 *Entropy maximising and information minimising*

The principle of entropy maximising is based on maximising a quantity H which measures the expected amount of uncertainty or missing information in a probability distribution $\{p_{ij}\}$. This quantity H is the measure of entropy defined by Shannon (1948) in a probability context but previously used in statistical thermodynamics as a measure for statistical disorder. Shannon (1948) and others have shown that a measure for H which meets certain conditions or desiderata can be defined as

$$H = -\sum_i \sum_j p_{ij} \ln p_{ij} , \qquad (26)$$

where p_{ij} is now being used as the general probability for event ij without any prior or posterior connotation. It is clearly a measure of uncertainty, for it takes on a maximum value when $\{p_{ij}\}$ is a uniform distribution, that is, when there is no way of distinguishing between the probability of any particular outcome occurring, and a minimum value when $\{p_{ij}\}$ is unity for one event kl, and zero for all other events $ij \neq kl$, namely, when there is complete certainty in the outcome. The principle of maximum entropy involves maximising H subject to a series of constraints which reflect information known about the probability set, information which must be taken into account by any assignment of probabilities.

The formal problem as stated by a number of theorists can be presented as follows: maximise H, subject to the following set of constraints:

$$\sum_i \sum_j p_{ij} = \langle 1 \rangle , \qquad (27)$$

and

$$\sum_i \sum_j p_{ij} \mathrm{f}^k(x_{ij}) = \langle x_k \rangle , \qquad k = 1, 2, ..., K , \qquad (28)$$

where

$\mathrm{f}^k(x_{ij})$ is some function f^k of the information x_{ij}, and

$\langle x_k \rangle$ is the expected value of that information.

There are K constraints of this form, embodying information about K functions of (x_{ij}), and the additional constraint in equation (27) ensures that the probability set sums to unity. The method of constrained maximisation is well-known and can be found in the works of Jaynes (1957; 1968), Tribus (1959; 1969), Wilson (1970), and Hobson (1971). The method will not be presented here, but the maximisation always leads to a solution for p_{ij} of the following form:

$$p_{ij} = \frac{1}{Z(\lambda_1, \lambda_2, ..., \lambda_K)} \exp\{-\lambda_1 \mathrm{f}^1(x_{ij}) - \lambda_2 \mathrm{f}^2(x_{ij}) - ... - \lambda_k \mathrm{f}^K(x_{ij})\} , \qquad (29)$$

where $\lambda_1, \lambda_2, ..., \lambda_K$ are parameters of the distribution related to the expected values $\langle x_1 \rangle, \langle x_2 \rangle, ..., \langle x_K \rangle$ respectively, and $Z(\lambda_1, \lambda_2, ..., \lambda_K)$ is a partition function related to the normalisation constraint in equation (27), and is given by

$$Z(\lambda_1, \lambda_2, ..., \lambda_K) = \sum_i \sum_j \exp\{-\lambda_1 \mathrm{f}^1(x_{ij}) - \lambda_2 \mathrm{f}^2(x_{ij}) - ... - \lambda_k \mathrm{f}^K(x_{ij})\} . \qquad (30)$$

If the constraints in equations (28) are all redundant, then equation (29) implies a uniform or rectangular distribution, and if there are as many constraints as events in equations (28), the model reproduces the data or information. By using this technique it is quite possible to derive the singly-constrained model in equations (10) and (11) if H is maximised subject to

$$\sum_i \sum_j p_{ij} = \left\langle \frac{1}{T} \sum_i O_i \right\rangle \quad \text{and} \quad \sum_i \sum_j p_{ij} d_{ij} = \langle \bar{d} \rangle \; ;$$

thus it should be possible to extend this technique to deal with additional information such as that encoded in the Coleman–Zipf model.

The technique, as it stands, relates to only one probability distribution $\{p_{ij}\}$, but there is a continuing debate concerning this question. Many information theorists argue that information can only be measured relative to two or more distributions and that in the case where there appears to be only one distribution, there is always a concealed distribution. Thus information is a relative, not an absolute, concept and must be defined accordingly. The argument is involved and it relates to discourse on subjective probability. It has been examined by the authors in a related paper (March and Batty, 1975), and it is sufficient to note here that a long line of probability theorists including Fisher, Turing, Good, Lindley, and Kullback have suggested that relative information be measured by the formula

$$I(p^{\text{¤}} : p^{\text{¤o}}) = \sum_i \sum_j p_{ij} \ln \frac{p_{ij}}{p_{ij}^{\text{o}}} \, , \tag{31}$$

where
p_{ij}^{o} is the prior probability of event ij,
p_{ij} is the posterior probability of event ij, and
$p^{\text{¤}} = \{p_{ij}\}$ and $p^{\text{¤o}} = \{p_{ij}^{\text{o}}\}$.
Equation (31) measures the information difference or gain generated by moving from the prior to the posterior. When the posterior is equal to the prior, no information is gained, and equation (31) is zero. Whenever prior and posterior are different, the information gain is positive. Another advantage in using equation (31) is that it is dimensionless; it does not suffer from the difficulties associated with moving from discrete to continuous distributions, as does H (see Batty, 1974), and it is never negative.

It has also been shown by Hobson and Cheng (1973) that the formula for information I has much more general properties and significance than H; indeed, H is a special case of I, and to demonstrate this consider the amount of missing information, H^*, defined using equation (31). This information can be calculated as the difference between the maximum obtainable information between the prior $\{p_{ij}^{\text{o}}\}$ and some distribution $\{p_{ij}^{\text{m}}\}$, and the prior $\{p_{ij}^{\text{o}}\}$ and the actual distribution $\{p_{ij}\}$. Then

$$H^* = I(p^{\text{¤m}} : p^{\text{¤o}}) - I(p^{\text{¤}} : p^{\text{¤o}})$$

$$= \sum_i \sum_j p_{ij}^{\text{m}} \ln \frac{p_{ij}^{\text{m}}}{p_{ij}^{\text{o}}} - \sum_i \sum_j p_{ij} \ln \frac{p_{ij}}{p_{ij}^{\text{o}}} \; . \tag{32}$$

Now if the prior $\{p_{ij}^{\text{o}}\}$ is the uniform distribution, and the distribution $\{p_{ij}^{\text{m}}\}$ is the completely certain outcome in which $p_{kl} = 1$, and $p_{ij} = 0$, all $i \neq k$, and $j \neq l$, then it is easily demonstrated that the first term in equation (32) is zero, and that H^* collapses to H. Formally,

$$H^* = -\sum_i \sum_j p_{ij} \ln p_{ij} = H \, . \tag{33}$$

Therefore it seems more appropriate to use H^* rather than H in the principle of maximum entropy, for this more general formula is able to encompass the notion of both prior and posterior probabilities.

Maximisation of H^*, subject to known information expressed as constraints, proceeds in exactly the same manner as the previous maximisation of H. The formal problem can be set out as follows:

$$\max H^* = \max\left\{\sum_i \sum_j p_{ij}^{\text{m}} \ln \frac{p_{ij}^{\text{m}}}{p_{ij}^{\text{o}}} - \sum_i \sum_j p_{ij} \ln \frac{p_{ij}}{p_{ij}^{\text{o}}}\right\}, \tag{34}$$

subject to constraints on the posterior distribution, $\{p_{ij}\}$, which have the same form as those in equations (27) and (28). Equation (34), however, can be simplified because the maximisation of H^* is with respect to the posterior probability p_{ij} only, and thus the first term is constant. Then

$$\max H^* = \text{constant} + \max\left\{-\sum_i \sum_j p_{ij} \ln \frac{p_{ij}}{p_{ij}^{\text{o}}}\right\}, \tag{35}$$

and it is clear from equation (35) that the maximisation of H^* is equivalent to the minimisation of $I(p^{\text{m}}:p^{\text{mo}})$. Thus entropy maximising in this case is also information minimising, although not in the sense used by Evans (1969), where the term is used as an alternative for entropy maximising. The formal problem can now be written as follows:

$$\min I(p^{\text{m}}:p^{\text{mo}}) = \min \sum_i \sum_j p_{ij} \ln \frac{p_{ij}}{p_{ij}^{\text{o}}}, \tag{36}$$

subject to $\sum_i \sum_j p_{ij} = \langle 1 \rangle$, (27)

and

$$\sum_i \sum_j p_{ij} \mathrm{f}^k(x_{ij}) = \langle x_k \rangle, \qquad k = 1, 2, ..., K. \tag{28}$$

The minimisation is accomplished using the same technique as is used to maximise entropy, and the general solution to this class of problem is given as

$$p_{ij} = \frac{p_{ij}^{\text{o}}}{Z(\lambda_1, \lambda_2, ..., \lambda_K)} \exp\{-\lambda_1 \mathrm{f}^1(x_{ij}) - \lambda_2 \mathrm{f}^2(x_{ij}) - ... - \lambda_K \mathrm{f}^K(x_{ij})\}, \tag{37}$$

where $Z(\lambda_1, \lambda_2, ..., \lambda_K)$ is a partition function defined as

$$Z(\lambda_1, \lambda_2, ..., \lambda_K) = \sum_i \sum_j p_{ij}^{\text{o}} \exp\{-\lambda_1 \mathrm{f}^1(x_{ij}) - \lambda_2 \mathrm{f}^2(x_{ij}) - ... - \lambda_K \mathrm{f}^K(x_{ij})\}. \tag{38}$$

There are several comments pertinent to these questions of two-stage modelling by information minimising. Perhaps the most important is the fact that Bayes' equation is a direct consequence of the procedure, as is clear by comparing the equation for $\{p_{ij}\}$, equation (37), by information minimising with equation (25); these equations have the same form. The second point of note is the fact that the maximum-entropy distribution of Shannon, Jaynes, and Tribus results if a uniform distribution is followed by the prior. A third point concerns similar derivations in the literature; Charnes et al. (1972) have taken a similar approach to deriving the gravity model by using Kullback's statistic and minimising information gain. Their treatment is less general than the one given here, but is nevertheless motivated by similar considerations. Moreover, Morphet (1975) has pursued a similar course in interpreting the Furness time-function iteration procedure for the gravity distribution model as a case of information minimising. And finally there is the question of extending the method to deal with a sequence of information gains. An example suffices to illustrate the point: first the prior could be fixed as a uniform distribution by maximising entropy according to Jaynes, subject only to the normalisation constraint, then a first posterior could be derived by minimising information between the prior and the posterior and deriving a Coleman-Zipf type model. This model could then be treated as a new prior, and further information could be added by minimising information gain once more. This process could be continued indefinitely, each stage corresponding to one of the stages in the many-stage modelling process, and such a procedure would ensure that a consistent model would result. In this way it is possible to build up related families of models and, in the following section, an attempt will be made to extend the family of spatial-interaction models due to Cordey-Hayes and Wilson (1971).

4.3 *An extended family of spatial-interaction models*

All the models derived by Cordey-Hayes and Wilson (1971) using the technique of entropy maximising can be rederived with a variety of possible priors by using information minimising. The list of possible models is quite extensive and will be discussed in more detail in a subsequent paper. But to give a flavour to the wealth of possible models in the 'extended' family, three models of particular significance will be examined. First, consider the prior probability distribution based on the Coleman-Zipf model; this seems a reasonable prior, given the previous argument, and it can be stated as

$$p_{ij}^{\circ} = KO_iD_jd_{ij}^{-1} = \frac{O_iD_jd_{ij}^{-1}}{\sum_i\sum_j O_iD_jd_{ij}^{-1}} . \tag{39}$$

Now consider a traffic-distribution model in which origins and destinations are balanced, and in which there is a finite amount spent on travel in the region. Then a model encompassing this information and incorporating the prior can be derived by minimising $I(p^{\text{\tiny{H}}}:p^{\text{\tiny{H}}\circ})$ subject to

$$\sum_j p_{ij} = \left\langle \frac{O_i}{T} \right\rangle , \tag{40}$$

$$\sum_i p_{ij} = \left\langle \frac{D_j}{T} \right\rangle , \tag{41}$$

and

$$\sum_i\sum_j p_{ij}d_{ij} = \langle \bar{d} \rangle , \tag{42}$$

where $\langle \bar{d} \rangle$ is the mean travel distance in the region. The solution to the problem from equation (37) is given as

$$p_{ij} = p_{ij}^{\circ}\exp(-\lambda_{i1}-\lambda_{j2}-\lambda_3 d_{ij}) , \tag{43}$$

which can be written as

$$p_{ij} = p_{ij}^{\circ}K_iK_j\exp(-\lambda d_{ij}) , \tag{44}$$

where

$$K_i = \exp(-\lambda_{i1}) , \qquad K_j = \exp(-\lambda_{j2}) , \qquad \text{and} \qquad \lambda = \lambda_3 .$$

The factors K_i and K_j can be defined explicitly by summing equation (44) over i and j, respectively, and substituting from equations (40) and (41). Then

$$K_i = \frac{O_i}{T\sum_j p_{ij}^{\circ}K_j\exp(-\lambda d_{ij})} , \tag{45}$$

and

$$K_j = \frac{D_j}{T\sum_i p_{ij}^{\circ}K_i\exp(-\lambda d_{ij})} . \tag{46}$$

By some manipulation and redefinition of constants—by absorbing them into other terms—the model in equation (44) can be rewritten as follows:

$$p_{ij} = \alpha_i O_i \beta_j D_j \exp(-\lambda d_{ij}) d_{ij}^{-1} , \qquad (47)$$

$$\alpha_i = \left[T \sum_j \beta_j D_j \exp(-\lambda d_{ij}) d_{ij}^{-1}\right]^{-1} , \qquad (48)$$

and

$$\beta_j = \left[T \sum_i \alpha_i O_i \exp(-\lambda d_{ij}) d_{ij}^{-1}\right]^{-1} . \qquad (49)$$

A number of comments which refer to all the models in the extended family can be made with regard to this model. First, it is clear that any information already included in the prior, such as $\{O_i\}$ and $\{D_j\}$, is not duplicated in the posterior if it occurs again in constraint equations (40) and (41). Thus information which is redundant drops out, but note that the term d_{ij}^{-1} is still quite distinct in the posterior. Second, it is quite simple to derive the singly-constrained and unconstrained models with the Coleman–Zipf priors from equation (47). For the singly-constrained models, set either α_i or β_j equal to unity, and for the unconstrained model set both these terms equal to unity; as in the original Cordey-Hayes–Wilson family, these models can be seen as special cases of the doubly-constrained model.

A particularly interesting case occurs when equation (42) is replaced by the geometric mean for travel distance, that is

$$\sum_i \sum_j p_{ij} \ln d_{ij} = \langle \overline{\ln d} \rangle . \qquad (50)$$

Now minimise $I(p^{\pi}:p^{\pi o})$, subject to equations (40), (41), and (50) with the Coleman–Zipf prior. This gives a model of the form

$$p_{ij} = \alpha_i O_i \beta_j D_j d_{ij}^{(-\lambda-1)} , \qquad (51)$$

where α_i and β_j are defined to ensure that equations (40) and (41) are satisfied. The model in equation (51) is the usual traffic-distribution model with an inverse power function which is separable into two components: one arising from the prior, the other from the additional effects of travel distance. If α_i and β_j are set equal to unity, the unconstrained model which was extensively applied by sociologists during the early 1950s is derived, and the separation of exponents reveals how the two effects can be consistently combined within the same model. Note that the use of constraint equation (50) implies that travellers perceive the effect of distance logarithmically, a possibility consistent with certain observed behaviours. And note also that the argument about the power to which distance might be raised in gravity models becomes manageable and comprehensible within this framework.

The final example worth detailing here is called Morphet's case, after Morphet (1975), who was the first to discuss it. It consists of taking a prior distribution as being proportional to the observed distribution. In the case of observed trips, T_{ij}^{obs}, the prior is given as

$$p_{ij}^{o} = \frac{T_{ij}^{\text{obs}}}{\sum_i \sum_j T_{ij}^{\text{obs}}} = \frac{T_{ij}}{T} . \qquad (52)$$

Now if information is minimised, subject only to equations (40) and (41), a model of the form

$$p_{ij} = p_{ij}^{o} K_i K_j \qquad (53)$$

results, which Morphet shows to be equivalent to the Furness time-function iteration in which an observed trip matrix, $\{T_{ij}^{\text{obs}}\}$, is transformed into a predicted matrix by row and column balancing operations; that is

$$T_{ij} = T_{ij}^{\text{obs}} \alpha_i \beta_j \ . \tag{54}$$

This kind of process has also been discussed by Theil (1972) and Bacharach (1970) in the context of input-output analysis. By using this framework, many other models are possible, ranging from extremes involving uniform priors, priors based on zone size, and priors based on data, to conventional spatial-interaction models which were originally derived quite differently. The power of the method is demonstrated by this richness of interpretation, but to relate this more formal technique to the earlier ad hoc applications, and to provide some measure of closure to the argument, it is now necessary to present empirical applications of posterior probability models in the Toronto-centred region.

5 Formal applications of the method: examples in the Toronto-centred region

5.1 *A negative exponential model with Coleman-Zipf prior*

One of the ad hoc applications of the method of residues consisted in first applying the Coleman-Zipf model, and then applying a singly-constrained model to the residuals based on the notion that additional effects of distance required explanation. The first model applied here treats the Coleman-Zipf model as the prior and includes these additional effects as an integral part of the posterior model. The prior is set up as in equation (39), and the model is derived by minimising $I(p^{\text{п}}:p^{\text{пo}})$ subject to equations (40) and (42). The model has the following form:

$$p_{ij} = p_{ij}^{\text{o}} K_i \exp(-\lambda d_{ij}) \ , \tag{55}$$

and K_i is easily evaluated by summing equation (55) over j and substituting from equation (40). Then

$$K_i = \frac{O_i}{T \sum_j p_{ij}^{\text{o}} \exp(-\lambda d_{ij})} \ , \tag{56}$$

and the model can be written in more familiar form as

$$T_{ij} = Tp_{ij} = \alpha_i O_i D_j \exp(-\lambda d_{ij}) d_{ij}^{-1} \ , \tag{57}$$

where

$$\alpha_i = \left[\sum_j D_j \exp(-\lambda d_{ij}) d_{ij}^{-1} \right]^{-1} \ . \tag{58}$$

The predicted number of employees living in j is calculated by summing equation (57) over i in the usual way.

The model was calibrated to the observed mean $\bar{d}$ by solving

$$\frac{1}{T} \sum_i \sum_j T_{ij}^{\text{obs}} d_{ij} = \frac{1}{T} \sum_i \sum_j O_i \alpha_i D_j \exp(-\lambda d_{ij}) \tag{59}$$

for λ by means of the Newton-Raphson method. Furthermore the additional information gained between prior and posterior was also calculated from the formula for expected information, that is

$$I(p^{\text{п}}:p^{\text{пo}}) = \sum_i \sum_j p_{ij} \ln \frac{p_{ij}}{p_{ij}^{\text{o}}} \ , \tag{31}$$

thus giving another measure of the difference between prior and posterior models. The performance of the integrated model is shown in table 5, and isometric plots of the predictions and residuals are illustrated in figure 6. From table 5, it is clear that the performance of the posterior is better than the prior, but in terms of the r^2 statistic there is little to choose between the posterior and the baseline model. Some might take this as evidence that the approach is of limited applicability since, in this case, results, as good if not better, can be obtained without the prior. Yet, in one sense, the examples reported here are likely to be ambiguous in quantitative terms, for the emphasis is on the effect of a single factor occurring both in the prior and posterior models. The more dramatic performance of the ad hoc models was due to the duplication of variables in both stages, an effect of somewhat dubious significance. Thus this empirical demonstration is of less dramatic proportions than it could be if very different information were to be incorporated between prior and posterior. Nevertheless, this application serves to demonstrate the approach.

Table 5. A comparison of model performance: prior, posterior, and baseline models.

	Posterior model: equation (57)	Coleman–Zipf prior: equation (18)	Singly-constrained model: equation (10)
Observed mean travel distance	6·1000	6·1000	6·1000
Predicted mean travel distance	6·1001	6·1684	6·0997
Parameter λ	0·0596	predetermined	0·2230
Correlation r	0·9732	0·9653	0·9757
Coefficient of determination, r^2	0·9471	0·9318	0·9521
Intercept of regression	1536·4531	1534·8516	1784·8750
Slope of regression	0·8607	0·8604	0·8382
Sum of absolute deviations	278355	293325	295377
Information gain	0·0734	not relevant	not relevant

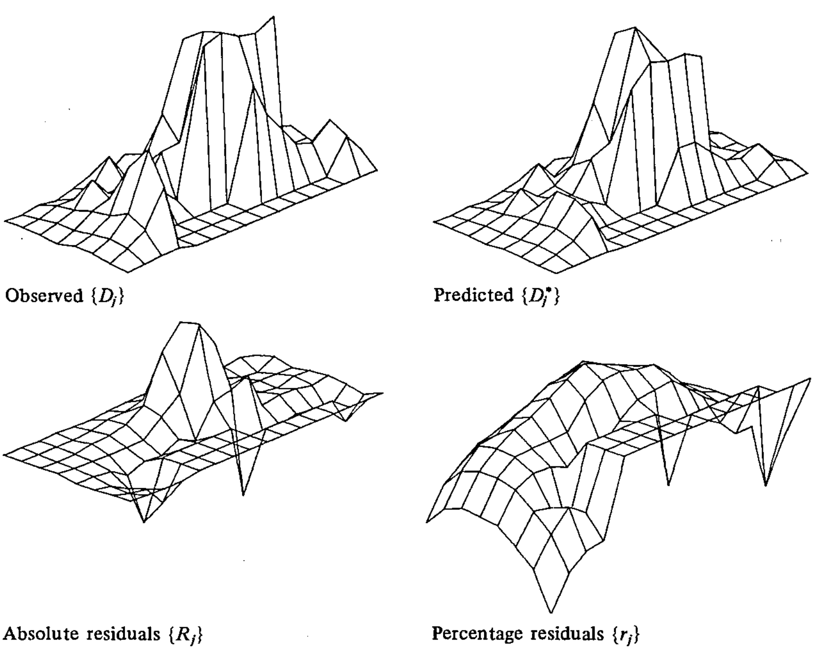


Figure 6. Spatial predictions and residuals from the negative exponential posterior model.

5.2 *A normally distributed model with Coleman–Zipf prior*

The final model applied here was based on a singly-constrained model incorporating information about the mean travel distance and the variance of the trip frequency distribution. The model included a Coleman–Zipf prior and information contained in constraint equations (40) and (42), together with the constraint

$$\sum_i \sum_j p_{ij}(d_{ij} - \bar{d})^2 = \langle \sigma^2 \rangle , \tag{60}$$

where σ^2 is the variance of the trip frequency distribution. Information is minimised subject to the constraints in equations (40), (42), and (60), and this gives rise to a model of the following form:

$$p_{ij} = p_{ij}^{\circ} K_i \exp(-\lambda d_{ij} - \mu d_{ij}^2) , \tag{61}$$

where K_i is a partition function or normalising factor defined as

$$K_i = \left[\sum_j p_{ij}^{\circ} \exp(-\lambda d_{ij} - \mu d_{ij}^2) \right]^{-1} . \tag{62}$$

This model can be rewritten in more familiar terms as

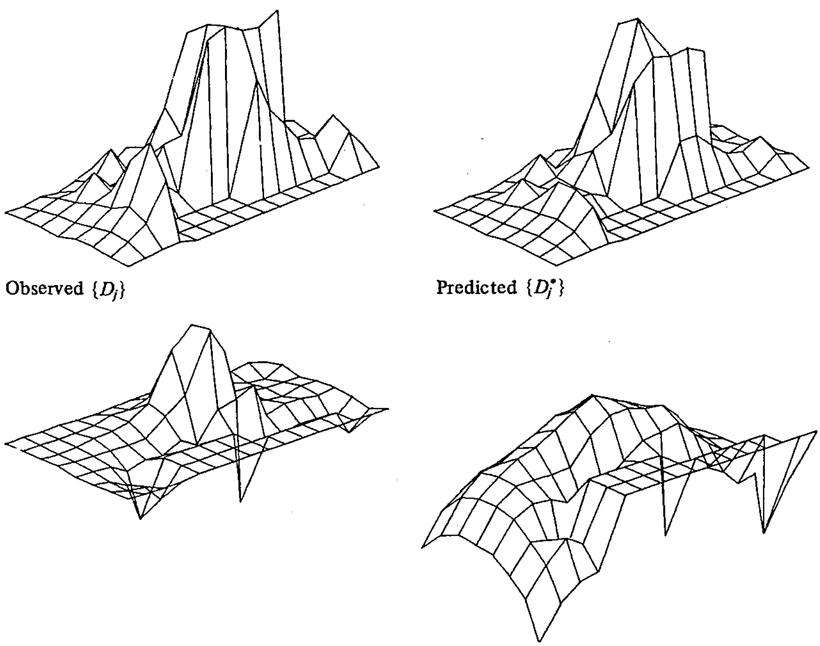
$$T_{ij} = Tp_{ij} = \alpha_i O_i D_j \exp(-\lambda d_{ij} - \mu d_{ij}^2) d_{ij}^{-1} , \tag{63}$$

where

$$\alpha_i = \left[\sum_j D_j \exp(-\lambda d_{ij} - \mu d_{ij}^2) d_{ij}^{-1} \right]^{-1} . \tag{64}$$

The exponential function in this model is proportional to the truncated normal distribution, and a derivation from this discrete form is given by Tribus (1969). The model has been calibrated in the usual manner, using the Newton–Raphson method to find λ and μ from the simultaneous solution of constraint equations (40) and (60).

The performance of this model is presented in table 6 and figure 7. In this case the performance is similar to the previous model, although an examination of the residuals reveals that this model produces a pattern of residuals that is much more random than that of all the previous interaction models. The only anomaly is downtown Toronto which, in all the models throughout this paper, has been difficult



Observed $\{D_j\}$ Predicted $\{D_j^*\}$

Absolute residuals $\{R_j\}$ Percentage residuals $\{r_j\}$

Figure 7. Spatial predictions and residuals from the normally distributed posterior model.

to predict. However, one fairly encouraging point with regard to the performance of these posterior models over the baseline model is the fact that their overall pattern of deviations from reality is less than the baseline prediction. So many effects are captured by these various models that a definitive empirical demonstration of their value would be unlikely in most applications. Indeed, their value lies more in establishing a point of view which might lead to new insights than in improving model performance in narrow statistical terms. And it is with this prospect in mind that this paper is concluded.

Table 6. Performance of the normally distributed posterior model.

Observed mean travel distance	6·1000	Correlation r of $\{D_j^{obs}\}$ and $\{D_j^*\}$	0·9725
Predicted mean travel distance	6·1003	Coefficient of determination, r^2	0·9459
Parameter λ	0·0944	Intercept of regression	1498·1015
Observed variance	18·3000	Slope of regression	0·8642
Predicted variance	18·3013	Sum of the absolute deviations	274838
Parameter μ	−0·0019	Information gain	0·0283

6 Conclusions

This paper has sought to extend the macroanalytic approach to urban modelling by treating the process of model design in distinct stages. Coleman's method of residues provides a useful way of breaking the process up into stages, although the method has hitherto only been applied in an ad hoc fashion. As this type of analysis fits easily into the Bayesian viewpoint, it can be interpreted in terms of the assignment of prior and posterior probability sets, and this leads naturally to an information-theoretic principle for making such assignments. In the authors' view, there are three significant interpretations fostered by this approach; first, the method of residues itself helps to emphasise the need for model builders to clarify various parts of their work with regard to trivial effects and irrelevant variables. Second, and related to this point, is the clarification of the gravity model and arguments about its form through the use of this approach. Last, and perhaps most important, there is the development of the information-minimising approach which appears to have greater generality than entropy-maximising, as is evidenced by the treatment both of priors and posteriors, and the extended family of spatial interaction models.

Doubtless there are many possible extensions to this framework, and some have already been hinted at here. For example, the extension of the information-minimising approach to a sequence or hierarchy of stages, each adding new information and building on the previous priors, could be explicitly demonstrated and empirically tested. The development of different priors could lead to new model forms; for example, the use of zone size as a prior would lead directly to models in which such effects appeared explicitly, thus demonstrating the link between information-gain formulae and Shannon's continuous entropy (Batty, 1974). Comprehensive urban models such as those of the Lowry genus could be reworked as two-stage models; in this context, it seems that concepts such as the economic-base relationship could also be treated as relatively trivial characteristics to be incorporated as priors, thus shifting the emphasis onto explanation of variations from such baselines.

It might also be possible to extend this approach within the context of probability theory by treating dependent and conditional-marginal probability explicitly. In entropy-maximising studies this has not been formally begun, but it does appear that added insights could be generated by such analysis. And this could also open the way to assessing the relevance of building interaction models which, like the ones here, are tested on their marginal totals. However, the purpose of this paper has not been to provide a cut-and-dried argument, but rather to establish a line of inquiry which, it is hoped, others might take up, alter, or reject; for it is only in this way that a progressive approach to urban modelling can be fostered. By applying these ideas to other situations, by testing new priors, and by taking a closer look at model performance, this approach could be enriched in the hope that it would lead to greater understanding of how to model the urban system and, ultimately, to a greater understanding of how to generate urban plans.

Acknowledgements. The authors wish to thank Bruce Hutchinson and John Shortreed of the University of Waterloo for providing the data and computer time for this study.

References

Alonso W, 1968 "Predicting best with imperfect data" *Journal of the American Institute of Planners* **34** 248-255

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

Batty M, 1974 "Urban density and entropy functions" *Journal of Cybernetics* **4** 41- 55

Batty M, 1976 *Urban Modelling* (Cambridge University Press, London) in press

Casey H J, 1955 "The law of retail gravitation applied to traffic engineering" *Traffic Quarterly* **9** 313-321

Charnes A, Raikes W M, Bettinger C O, 1972 "An extremal and information-theoretic characterisations of some interzonal transfer models" *Socio-Economic Planning Sciences* **6** 531-537

Cliff A D, Martin R L, Ord J K, 1974 "Evaluating the friction of distance parameter in gravity models" *Regional Studies* **8** 281-286

Coleman J S, 1964 *Introduction to Mathematical Sociology* (Free Press of Glencoe, New York)

Cordey-Hayes M, Wilson A G, 1971 "Spatial interaction" *Socio-Economic Planning Sciences* **5** 73-95

Cox R T, 1961 *The Algebra of Probable Inference* (Johns Hopkins University Press, Baltimore)

Curry L, 1972 "A spatial analysis of gravity flows" *Regional Studies* **6** 131-147

Draper N R, Smith H, 1966 *Applied Regression Analysis* (John Wiley, New York)

Echenique M, Crowther D, Lindsay W, 1969 "A spatial model of urban stock and activity" *Regional Studies* **3** 281-312

Efroymson M A, 1962 "Multiple regression analysis" *Mathematical Methods for Digital Computers* Eds A Ralston, H S Wilf (John Wiley, New York)

Evans A W, 1971 "The calibration of trip distribution models with exponential or similar cost functions" *Transportation Research* **5** 15-38

Evans R A, 1969 "The principle of minimum information" *IEEE Transactions on Reliability* R-18, 87-90

Hansen W G, 1959 "How accessibility shapes land use" *Journal of the American Institute of Planners* **25** 72-76

Harris B, 1964 "A note on the probability of interaction at a distance" *Journal of Regional Science* **5** 31-35

Hobson A, 1971 *Concepts in Statistical Mechanics* (Gordon and Breach, New York)

Hobson A, Cheng B K, 1973 "A comparison of the Shannon and Kullback information measures" *Journal of Statistical Physics* **7** 301-310

Huff D L, 1964 "Defining and estimating a trading area" *Journal of Marketing* **28** 34-38

Hyman G M, 1969 "The calibration of trip distribution models" *Environment and Planning* **1** 105-112

Jaynes E T, 1957 "Information theory and statistical mechanics" *Physics Review* **106** 620-630, **108** 171-190

Jaynes E T, 1968 "Prior probabilities" *IEEE Transactions on Systems Science and Cybernetics* SSC-4, 227-241

Kirby H R, 1974 "Theoretical requirements for calibrating gravity models" *Transportation Research* **8** 97-104

Kuhn T S, 1962 *The Structure of Scientific Revolutions* (University of Chicago Press, Chicago)

Lakshmanan T R, Hansen W G, 1965 "A retail market potential model" *Journal of the American Institute of Planners* **31** 134-143

Lee D B, 1973 "Requiem for large-scale models" *Journal of the American Institute of Planners* **39** 163-178

March L, Batty M, 1975 "Generalised measures of information, Bayes' likelihood ratio and Jaynes' formalism" *Environment and Planning B* **2** 99-105

Morphet R, 1975 "A note on the calculation and calibration of doubly-constrained trip distribution models" *Transportation* **4** 43-53

Popper K R, 1965 *Conjectures and Refutations: The Growth of Scientific Knowledge* (Basic Books, New York)
Reilly W J, 1931 *The Law of Retail Gravitation* (The Knickerbocker Press, New York)
Schneider M, 1959 "Gravity models and trip distribution theory" *Papers of the Regional Science Association* **5** 51-66
Shannon C E, 1948 "The mathematical theory of communication" *Bell System Technical Journal* **27** 379-423, 623-656
Steger W A, 1965 "Pittsburgh urban renewal simulation model" *Journal of the American Institute of Planners* **31** 144-150
Stewart J Q, 1942 "A measure of the influence of population at a distance" *Sociometry* **5** 63-71
Stewart J Q, 1947 "Empirical mathematical rules concerning the distribution and equilibrium of population" *Geographical Review* **37** 461-485
Stewart J Q, 1950 "The development of social physics" *American Journal of Physics* **18** 239-253
Stewart J Q, Warntz W, 1958 "The physics of population distribution" *Journal of Regional Science* **1** 99-123
Theil H, 1972 *Statistical Decomposition Analysis* (North Holland, Amsterdam)
Tribus M, 1959 *Thermostatics and Thermodynamics* (Van Nostrand, Princeton, NJ)
Tribus M, 1969 *Rational Descriptions, Decisions and Designs* (Pergamon Press, New York)
Voorhees A, 1955 "A general theory of traffic movement" *Proceedings of the Institute of Traffic Engineers* **26** 46-56
Wilson A G, 1969 "Development of some elementary residential location models" *Journal of Regional Science* **9** 377-385
Wilson A G, 1970 *Entropy in Urban and Regional Modelling* (Pion, London)
Wilson A G, 1971 "A family of spatial interaction models and associated development" *Environment and Planning* **3** 1-32
Wilson A G, 1974 *Urban and Regional Models in Geography and Planning* (John Wiley, London)
Zipf G K, 1949 *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Cambridge, Mass.)