Environment and Planning, 1971, volume 3, pages 1-32

A family of spatial interaction models, and associated developments †

A G Wilson
Department of Geography, University of Leeds
Received 28 December 1970

Abstract. This paper shows that the gravity model is not a single model but that there is a whole family of spatial interaction models. The properties of this family are outlined in some detail. Basic concepts of such models can be developed in a variety of ways, and these are illustrated. The paper then outlines a number of other theoretical developments, and is particularly concerned with the disaggregation of such models, with the incorporation of time variables, and with the relation of spatial interaction, to more general, models. Uses of spatial interaction models are outlined briefly and the final section of the paper draws a number of conclusions and presents a summary.

1 Introduction

The study of spatial interaction has been important for a long time in the social sciences, especially in geography but also in economics and sociology. Models have been built which relate to a wide range of spatial interaction phenomena: for example, all kinds of person movement (journey to work, journey to shop, journey to school, and so on through all trips for different purposes, together with associated flows of money and goods), migration, goods movement, telephone calls, marriage selection, newspaper circulation, bank cheques, and spread of innovation.

Among the models used, the gravity model has perhaps been used most of all. In general, it is probably still seen as something based on a Newtonian analogy. This paper notes that more fruitful analogies can be developed, and that then the name 'gravity model' is something of a misnomer—though we will probably keep it for historical reasons—and that a whole family of spatial interaction models can be generated as extensions of the old gravity model. Consequently, this new framework and recognition of the existence of a family of models provides a basis for future development. The paper goes on to discuss developments of the basic model concepts and other theoretical developments. The use of models is discussed briefly, followed by a summary and conclusion.

This is not a review paper in the sense that it attempts to list and review a large body of literature. The excellent reviews by Carrothers (1956) and Olsson (1965) are informative in themselves and between them they cite most of an extensive literature up to 1965. (Chapter 11 of Isard, 1960, is also useful in this respect.) This paper concentrates mostly on developments which have taken place since then. It notes a number of earlier developments in related fields—such as the model used mostly by civil engineers in transportation studies, which had not been studied much by geographers and hence did not figure in Olsson's review. The paper is a review paper, however, in that it attempts to record the present state of the art. It is an extension and elaboration of two earlier papers (Wilson, 1970a, and Cordey Hayes and Wilson, 1970). Many of the points to be raised are discussed in detail in earlier papers by the author and others. These are referred to where appropriate and detailed arguments, which are often elaborate, will not be repeated here. In effect, an attempt is being made to write mainly about wood and not about trees; but the

† Paper presented to the Quantitative Methods Study Group of the Institute of British Geographers, Sussex, January 1971.

reader is reminded that the details do exist elsewhere and that arguments presented with relatively little support in this paper are often substantiated in other papers.

2 A family of spatial interaction models

2.1 *A family of models*

Suppose we have a study area divided into zones which can be labelled i, j, and so on. Then the traditional gravity model can be written as:

$$T_{ij} = K\frac{W_i^{(1)}W_j^{(2)}}{c_{ij}^n} , \tag{1}$$

where T_{ij} is a measure of the interaction between zones i and j, $W_i^{(1)}$ is a measure of the 'mass term' associated with zone i, $W_j^{(2)}$ is a measure of the 'mass term' associated with zone j, and c_{ij} is a measure of the distance, or generalised cost of travel, between zone i and zone j. K is a constant of proportionality and n is a parameter to be estimated. This equation can be written slightly more generally as:

$$T_{ij} = KW_i^{(1)}W_j^{(2)}\mathrm{f}(c_{ij}) , \tag{2}$$

where $1/c_{ij}^n$ has been replaced by $\mathrm{f}(c_{ij})$, some function of c_{ij} which decreases as c_{ij} increases. Now, a family of models can be constructed as follows. Initially we take account of any additional knowledge which should be built into the model to constrain the possible values that the interaction variable, T_{ij}, can take. Such information usually takes the form of knowledge about the total interaction flows leaving a particular zone and/or the total interaction flows terminating in a particular zone. For example, if we know the total flow O_i originating at each zone i then we can write:

$$\sum_j T_{ij} = O_i . \tag{3}$$

Similarly, if we know the total of flows terminating at zone j we can write:

$$\sum_i T_{ij} = D_j \tag{4}$$

where D_j is the known total. We can now distinguish four cases:
(a) neither Equation (3) nor (4) holds—the *unconstrained* case;
(b) Equation (3) holds—the *production constrained* case;
(c) Equation (4) holds—the *attraction constrained* case;
(d) Equations (3) and (4) hold simultaneously—the *production-attraction constrained* case.

The model representing the unconstrained case can be taken as Equation (2). To obtain the alternative models which incorporate our additional knowledge as constraints, we must replace the constant of proportionality, K, by a set of such constants (or 'balancing factors' as they are sometimes known) appropriate to the constraints assumed: for example, K is replaced by a set A_i if Equation (3) holds, by a set B_j if Equation (4) holds, and by the product set A_iB_j if Equations (3) and (4) hold simultaneously. Further, in the case where O_i is known this acts as a mass term and replaces $W_i^{(1)}$, and in the case where D_j is known this can act as the mass term and replace $W_j^{(2)}$. If these changes are made where appropriate, we shall have generated a family of four models. The unconstrained model is still given by Equation (2). The others are:

production constrained

$$T_{ij} = A_iO_iW_j^{(2)}\mathrm{f}(c_{ij}) ; \tag{5}$$

attraction constrained

$$T_{ij} = B_j W_i^{(1)} D_j \mathrm{f}(c_{ij}) \; ; \tag{6}$$

production–attraction constrained

$$T_{ij} = A_i B_j O_i D_j \mathrm{f}(c_{ij}) \; . \tag{7}$$

The A_i's and B_j's are calculated to ensure that the constraint equations are satisfied.
Note that each model is of the form

interaction = balancing factors x mass term x mass term x distance function (8)

and the balancing factors and mass terms are specified in accordance with the nature of the known constraints. The A_i's and B_j's in the different cases are:
production constrained [with Equation (5)]

$$A_i = \frac{1}{\sum_j W_j^{(2)} \mathrm{f}(c_{ij})} \; ; \tag{9}$$

attraction constrained [with Equation (6)]

$$B_j = \frac{1}{\sum_i W_i^{(1)} \mathrm{f}(c_{ij})} \; ; \tag{10}$$

production–attraction constrained [with Equation (7)]

$$A_i = \frac{1}{\sum_j B_j D_j \mathrm{f}(c_{ij})} \, , \qquad B_j = \frac{1}{\sum_i A_i O_i \mathrm{f}(c_{ij})} \; . \tag{11}$$

The reader will be familiar with applications of unconstrained models of the form of Equation (2) from a large range of literature. It may be useful to give examples of applications of other members of the family of models.

An example of the *production–constrained model* is the model which is used to estimate retail sales in shopping centres, a distribution of population together with their purchasing power being given. Since the purchasing power of the residents of a particular zone is assumed given, this is equivalent to building in a production constraint of the form of Equation (3), where O_i would be the purchasing power of the residents of i and the interaction term T_{ij} would be the flow of money from zone i to shops in zone j. In this model, $W_j^{(2)}$ would be a measure of the attractiveness of shops in j. More typically perhaps, the interaction variable in this case would be written as S_{ij} and the purchasing power of residents of zone i could be represented as $e_i P_i$, where e_i is the mean expenditure of residents of zone i, and P_i is the population of zone i. The complete model can then be written:

$$S_{ij} = A_i (e_i P_i) W_j^{(2)} \mathrm{f}(c_{ij}) \, , \tag{12}$$

where

$$A_i = \frac{1}{\sum_j W_j^{(2)} \mathrm{f}(c_{ij})} \; . \tag{13}$$

Typically, $W_j^{(2)}$ is taken as a measure of shopping centre size on the assumption that this is a proxy for attractiveness. It can be related to the notion of scale economies (Wilson„ 1967). This model has been used to study shopping patterns by Lakshmanan and Hansen (1965), and by many other authors subsequently. It has also been used to study flows of patients to hospitals by Morrill and Kelley (1970).

It should be noted that Equations (12) and (13) can be combined and written as:

$$S_{ij} = (e_i P_i)\frac{W_j^{(2)}\mathrm{f}(c_{ij})}{\sum_j W_j^{(2)}\mathrm{f}(c_{ij})} . \tag{14}$$

This, in turn, can be written:

$$S_{ij} = (e_i P_i)\mathrm{p}_{ij} , \tag{15}$$

where

$$\mathrm{p}_{ij} = \frac{W_j^{(2)}\mathrm{f}(c_{ij})}{\sum_j W_j^{(2)}\mathrm{f}(c_{ij})} . \tag{16}$$

Huff (1964) hypothesised p_{ij} as the probability that one resident of zone i would shop in zone j. He then derived the above shopping model on the basis of this hypothesis. This kind of transformation and interpretation can be used for any singly constrained model (production-constrained or attraction-constrained). Equation (16) is a common kind of formula. More generally, we might define u_{ij} as the utility gained by an i-resident by shopping in j. Then

$$\mathrm{p}_{ij} = \frac{u_{ij}}{\sum_j u_{ij}} \tag{17}$$

and this utility notion can be used in other contexts. An equation of this type is used by Golledge (1969) in exploring a conceptualisation of the market decision process. It is related to Thurstone's principle in psychology (cf.Lambe, 1969).

One other point can usefully be taken up at this stage in relation to the production-constrained model. So far, we have said relatively little about attractiveness factors, but have assumed implicitly that they are determined outside the model. One interesting possibility, which has arisen in the context of a retail sales' model, is to take

$$W_j^{(2)} = \sum_i T_{ij} . \tag{18}$$

$\sum_i T_{ij}$ can be written as T_{*j} for short—an asterisk replacing a suffix being used to denote summation. The attractiveness is now determined inside the model and the equation has to be solved iteratively for T_{ij}. The problem of the convergence of this process is explored by Eilon *et al.* (1969).

An example of the attraction-constrained model is one of the elementary residential location models. This assumes that workers can be allocated to residences in zones around their workplaces in a gravity model fashion, but that the distribution of jobs is given. This is equivalent to assuming a constraint of the form of Equation (4). $W_i^{(1)}$ in this case becomes a measure of the attractiveness of zone i as a residential zone. Thus, if T_{ij} is the number of people who live in zone i and work in zone j, and if D_j is replaced by E_j as the given employment in zone j, then this model can be written:

$$T_{ij} = B_j W_i^{(1)} E_j \mathrm{f}(c_{ij}) , \tag{19}$$

where

$$B_j = \frac{1}{\sum_i W_i^{(1)}\mathrm{f}(c_{ij})} . \tag{20}$$

The production–attraction constrained model is employed in transportation studies where it is assumed that 'trip ends', the number of origins in each zone of the given kind of trip, and a number of destinations in each zone of the same kind of trip, are given and the purpose of the model is to estimate T_{ij} given this information. Thus, we may be building a model of the journey to work, in which case O_i would be the number of workers resident in zone i and D_j would be the number of jobs in zone j; Equations (3) and (4) then hold as constraints. The appropriate model is then that given by Equations (7) and (11).

The models were used in these various ways in the early to mid-1960's without it really being recognised explicitly that they were all different members of the same family, and that their position in the family was determined by what might be called constraint information. It proves much easier to see the structure of this family if the models are viewed within an entropy maximising framework, as we shall see below. The nature of the family of models was first discussed in this way by the author in a paper on commodity flows (Wilson, 1970b) and was elaborated in a further paper (Cordey Hayes and Wilson, 1970).

One point which should be noted explicitly at this stage of the argument is that any member of the family other than the production–attraction constrained model can be, and often is, used as a *location model*. For example, in the production constrained case $\sum_i T_{ij}$ is not assumed to be known outside the model and so can be estimated inside it. In the example which was given of this model (for estimating retail sales), it was used in exactly this way to estimate the total sales from shops in each particular zone. It can also be used in this way to estimate the impact of new shopping centres. Similarly, the elementary residential location model is being used to provide an estimate of $\sum_j T_{ij}$—the distribution of population. We shall see later that it is necessary to discuss carefully whether interaction models *ought* to be used in this way. A further minor point can be noted at this stage: when a location model is being built from an interaction hypothesis—as Lowry's very elementary residential location model (Lowry, 1964)—it is best to state explicitly the interaction assumption in terms of an interaction variable as we have been doing here. This makes it easier to see how to improve the model (cf.Wilson, 1969g).

As a final point on the nature of the family of spatial interaction models, we can note that the family can be extended further for multistage interactions. Consider, for example, a set of exporting regions, i, in one country, and a set of importing regions, k, not in that country, and let the first country have a set of ports, j. Goods can flow from i, through to j, to k, and such a flow can be described by the variable T_{ijk}. If O_i is the exportable produce of i, D_k the imports of k, and X_j the capacity of j (and it is assumed for the sake of this example that all capacity is used) then

$$\sum_j \sum_k T_{ijk} = O_i \; , \tag{21}$$

$$\sum_i \sum_k T_{ij} = X_j \; , \tag{22}$$

$$\sum_i \sum_j T_{ijk} = D_k \; , \tag{23}$$

and, if the travel cost per unit of good is written c_{ijk}, then the corresponding spatial interaction model might be

$$T_{ijk} = A_i B_j C_k O_i X_j D_k \exp(-\beta c_{ijk}) \; , \tag{24}$$

where

$$A_i = \frac{1}{\sum_j \sum_k B_j C_k X_j D_k \exp(-\beta c_{ijk})} \tag{25}$$

$$B_j = \frac{1}{\sum_i \sum_k A_i C_k O_i D_k \exp(-\beta c_{ijk})} \tag{26}$$

$$C_k = \frac{1}{\sum_i \sum_j A_i B_j O_i X_j \exp(-\beta c_{ijk})} \; . \tag{27}$$

So, we now see at least the formal possibility of a spatial interaction model with three mass terms and three balancing factors (which will, incidentally, ensure that the overall dimensionality is correct).

2.2 *Entropy maximising: gravity models as statistical averages*

In the previous sub-section (2.1) a family of models has been described which could be considered to arise as modifications of the unconstrained Newtonian model. In this section, it is perhaps appropriate to explain briefly how this family of models can be derived on the basis of a different analogy. The simplest explanation of the analogy is based on notions from statistical mechanics although in many ways, this is the least rigorous way of proceeding, a point which the reader should bear in mind. Some other ways of proceeding will also be described briefly below.

The statistical mechanics analogy depends essentially on the notion of a 'state' and of a 'state description'. To fix ideas, suppose that we are interested in the journey to work problem described earlier. In principle we can label all the workers who are making trips, and so the most detailed description of the journey to work 'state' would be one which specified the origin and destination of each individual worker. Such a description can be called a micro-state description. However, we may be interested only in the total number of people travelling from one zone i to any other zone j—this is our variable T_{ij}. Such a description of the state may be known as a meso-description. There is an even more coarse level of description: the macro-state via the overall constraints on the system—the O_i's and the D_j's, and the total expenditure on the journey to work. This last constraint is being introduced for the first time; we shall find it essential to generate the appropriate models. However, the important point to note at this stage is that many micro-states give rise to one meso-state, and many meso-states give rise to one macro-state. This hierarchical arrangement is shown in Figure 1.

We can now state our problem as follows: we wish to find the most probable meso-state compatible with a set of constraints defining a given macro-state.

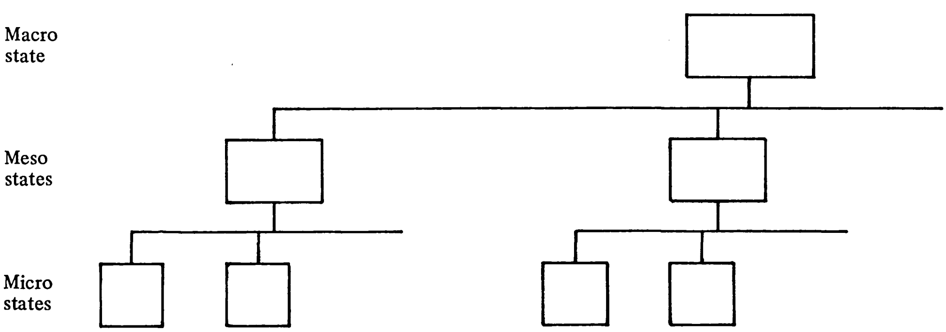

Figure 1. Hierarchical relationships between state descriptions.

Assuming that we are ignorant about micro-states, we can assume that all such states are equally probable. We can thus find the most probable meso-state by counting the number of micro-states associated with each meso-state which satisfies the constraints, and then finding the one that has the greatest number of micro-states associated with it. A meso-state is characterised by the matrix T_{ij}. It can easily be shown that the number of possible micro-states which give rise to such a meso-state is

$$W(\{T_{ij}\}) = \frac{T!}{\prod_{ij} T_{ij}!}, \tag{29}$$

where T is the total number of trips (Wilson, 1967).

So, all we now have to do is to find a set of T_{ij}'s which maximise this expression subject to whatever constraints obtain. For the journey to work, these constraints are given by Equations (3) and (4) together with a constraint on total cost which can be written as:

$$\sum_i \sum_j T_{ij} c_{ij} = C. \tag{30}$$

In fact, we maximise $\ln W$ rather than the W of Equation (29). This of course, produces the same solution for the T_{ij}. It can then be shown that the estimate of T_{ij}, which results from maximising $\ln W$ subject to constraint Equations (3), (4), and (30), is

$$T_{ij} = A_i B_j O_i D_j \exp(-\beta c_{ij}), \tag{31}$$

where

$$A_i = \frac{1}{\sum_j B_j D_j \exp(-\beta c_{ij})} \tag{32}$$

and

$$B_j = \frac{1}{\sum_i A_i O_i \exp(-\beta c_{ij})}. \tag{33}$$

This model is identical to that given by Equations (7) and (11) except that the general function $f(c_{ij})$ is replaced by the exponential function $\exp(-\beta c_{ij})$. The significance of this will be discussed later.

The model was derived essentially in this way in an earlier paper by the author (Wilson, 1967). It was also shown in that paper how the other models of the family could be derived, whilst it is shown explicitly for all the family in the other two papers cited earlier (Wilson, 1970a and Cordey Hayes and Wilson, 1970). The derivation can also be presented using the concept of entropy as it is employed in information theory (Wilson, 1969a); additionally, it can also be related to Bayesian methods of inference (Hyman, 1969). There is one further method which this author finds the most satisfying of all in terms of both rigour and intuitive understanding of what the method achieves: this is the equivalent of the Darwin–Fowler method in statistical mechanics. This method assumes, in effect, that we are calculating the statistical averages of behaviour, possibly subject to various constraints, of all the individuals in the system. Thus the entropy maximising method can be seen simply as a statistical averaging method. A recent book by the author includes a section which shows how the Darwin–Fowler method can be applied in this particular case (Wilson, 1970c). The four model building procedures which have been outlined or mentioned above—the first three essentially entropy maximising and the fourth statistical averaging—are all more or less equivalent. As it happens, they are progressively more difficult mathematically in the order in which they have been discussed above, so that, although the statistical mechanics analogy is perhaps least satisfactory from the point of view of rigour, we can use it with confidence knowing that more rigorous methods would lead to the same answers in the cases which interest us.

This author was interested to note the following remark on rereading Carrothers (1956): "As has been pointed out, the gravity and potential concepts of human interaction were developed originally from analogy to Newtonian physics of matter. The behaviour of molecules, individually, is not normally predictable, while in large numbers their behaviour is predictable on the basis of mathematical probability. Similarly, while it may not be possible to describe the action and reactions of the individual human in mathematical terms, it is quite conceivable that interactions of groups of people may be described in this way." This is a good short description of the potential of using analogies with statistical mechanics rather than Newtonian mechanics, though evidently no one at the time took up the suggestion!

The entropy maximising method is essentially a model building tool. We have seen that the family of models which can be derived in this way can also be derived by modifying the original gravity model to take account of the constraints. Thus, while it can be considered that entropy maximising is based on a more realistic analogy, we have not so far made the case that it can lead to models which could not be discovered some other way. Indeed, it is probably true, in general, that all models which can be discovered by entropy maximising methods could be discovered in other ways. However, we shall see that in more complicated situations, although this is a theoretical possibility, it would be extremely difficult to achieve in practice. However, it is perhaps appropriate to note at this point that the existence of an entropy maximising derivation of a model does not of itself give that model any validity. Once the model has been set up it has to be tested in the usual way (Clough, 1964). Further, there are many models which can be derived by methods which are not compatible with entropy maximising. Again, they can be tested in the usual way. However, there are situations where one would expect a model to be derivable by an entropy maximising method, and in this case if such a proof cannot be produced then this can be taken as evidence that the model is less than satisfactory. Thus, the main advantages of the entropy maximising method can be summarised under three headings (Wilson, 1970d):

1 *Model generation*: the method is extremely effective in certain situations for generating models of rather complex phenomena and which are internally consistent in the sense that they satisfy each of a relatively large number of constraints that have to be imposed. Usually, also, it is possible to see how to extend the models in a variety of ways. We shall see examples of this below.
2 If a model can be derived in an entropy maximising way, it is useful to note the constraints that give rise to the model. This often helps considerably in the *interpretation* of the terms of the model. For example, the balancing factors, A_i and B_j, in the journey to work model are closely related to Lagrangian multipliers associated with constraint Equations (3) and (4) respectively. The parameter β which appears in the exponential function is the Lagrangian multiplier associated with the cost constraint Equation (30). This helps considerably in the interpretation of this term. For example, rather than estimating β directly, it may be appropriate to estimate C. This also creates a potential method for estimating the change of β over time. We shall take up this point again below.
3 We shall also discuss later the treatment of time in these kinds of models and the problems of building *dynamic* models. Entropy maximising potentially helps here

because it is associated with a body of mathematics which has been applied in part to the study of dynamic systems.

2.3 *The properties of the balancing factors*

The production–attraction constrained model has always had more theoretical problems associated with it than the other models because of the nature of Equations (11) for A_i and B_j. The reader will see quickly that these equations have to be solved iteratively. There has long been a fascination with these A_i's and B_j's, their interpretation, and their other properties. Considerable advances have been made in recent years, some of which have been reported in a recent special issue of *Transportation Research*. In particular, Evans (1970a) has been able to show that A_i's and B_j's converge to a unique solution for the product A_iB_j, no matter what starting values are assumed in an iterative process. (Typically all the B_j's are taken as 1 initially and then an iterative solution is obtained.)

Several authors have produced alternative models based on empirical observations of spatial distributions of A_i and B_j values—notably Kirby (1970) and Edens (1970). In production-constrained models and attraction-constrained models, the A_i's and B_j's have long been interpreted as either competition or accessibility as appropriate. This point is taken up in section 2.5 below. Similar interpretations can be maintained (Wilson, 1967) for the A_i's and B_j's in the production–attraction constrained model, although in this case each is influenced by the other.

2.4 *Calibration methods*

A lot of work has also been done on calibration methods—these raise some general problems. This model is non-linear in a very interesting and useful way, but this means that the standard techniques for parameter estimation cannot be used. In particular, the reader should note that it is dangerous to take logarithms of a spatial interaction model equation and then to use regression analysis to estimate the parameters. This is because of the nature of the A_i and B_j terms. Suppose we take logarithms in a re-arranged Equation (1):

$$\ln\frac{T_{ij}}{W_i^{(1)}W_j^{(2)}} = \ln K - n\ln c_{ij} \,. \tag{34}$$

We can estimate n by regression analysis provided that we can treat K as a constant. In practice, in most spatial interaction models, K is more likely to be A_i, B_j or A_iB_j (or at least to have the properties of these terms even if the model has been mis-specified as unconstrained), and a glance at A_i, B_j or A_iB_j equations will show that these expressions *are functions of* n. Hence regression analysis of (34) as it stands will lead to a biassed estimate of n. Experience of this author and various colleagues has shown this result to be true.

It is useful to note at this point that if the model to be calibrated involves attractiveness factors, then such factors may be assumed to be a 'size' variable raised to some power which has to be estimated. Thus, a production constrained model might be

$$T_{ij} = A_iO_iW_j^{(2)\alpha}c_{ij}^{-n} \,, \tag{35}$$

where α and n are parameters to be estimated.

Typically, it is necessary to define a set of goodness-of-fit statistics which can be calculated for each (α, n) pair. Values of α and n can then be chosen which maximise one or more of these statistics. Unfortunately, it is often the case that not all the goodness-of-fit statistics are maximised at the same (α, n) values.

A good example of this kind of calibration procedure is provided by the Government paper, *Portbury* (Ministry of Transport, 1966). If we use the superscript 'obs' to denote an observed value, then the goodness-of-fit statistics used are:

1. $\dfrac{\sum_i \sum_j T_{ij} c_{ij}}{\sum_i \sum_j T_{ij}}$ mean trip cost, to be equated with observed trip cost,

2. $\sum_j \left\{ \sum_i T_{ij} - \sum_i T_{ij}^{\text{obs}} \right\}^2$ the sum of squares of differences between trip attractions, should be maximised,

3. $\sum_i \sum_j (T_{ij} - T_{ij}^{\text{obs}})^2$ the sum of squares of differences between model and observed interactions, should be minimized.

'3' is obviously better than '2', but '2' sometimes has to be used because observed interaction data are not always available. These statistics should always be used in conjunction with '1'—there are good theoretical reasons for this associated with the trip cost constraint (Hyman, 1969; Evans, 1970b)—especially where '2' is being used without '3'. This last point is vital, since the sum in '2' is identically zero when $\alpha = 1$, $n = 0$, and can bias the results. The point is made very clear by the example in the *Portbury* paper. Furthermore, '2' is insufficiently sensitive, as shown by Batty (1970).

One other point can be usefully noted while the *Portbury* paper is being considered: there are situations where different models of the family can be used for the one situation. In the *Portbury* case, it was possible to use either an attraction-constrained model or a production–attraction-constrained model. Both models were calibrated and gave the *same n* values. It may be thought a reasonable conjecture that this result is true generally.

At this point it is appropriate to make a further comment: it would be very useful if parameters resulting from calibrations of different models in different situations could be compared. Should this be attempted, however, great care must be exercised. If a power function is used, as c_{ij}^{-n}, then n is dimensionless; but if an exponential function, as $\exp(-\beta c_{ij})$, is used, β is *not* dimensionless, and so its value depends on the units of c_{ij}.

2.5 *Associated concepts*

The most famous concept which is associated (at least to some extent) with gravity models is that of *accessibility*. For example, if we follow Hansen (1959) in relation to the shopping model defined by Equations (12) and (13), we can define the accessibility of residents of zone i to shops in zone j to be $X_{i \to j}$ where

$$X_{i \to j} = W_j^{(2)} f(c_{ij}) . \tag{36}$$

We can sum the expression on the right hand side of Equation (36) over all zones j, and define X_i to be the total accessibility to shops of residents of zone i

$$X_i = \sum_j W_j^{(2)} f(c_{ij}) . \tag{37}$$

(Note that X_i is equal to $1/A_i$; so this gives us an interpretation of A_i.)

Many social scientists, and geographers in particular, have defined similar concepts of accessibility in the past. The advantage of this one is that it contains a distance function, $f(c_{ij})$, which appears in a model that reproduces interaction behaviour—hopefully reasonably well. In previous attempts, the function used has been, say, a power function with a guessed exponent value—not one which has been calibrated within a model. This concept of accessibility is well known and we need not dwell on it in this paper. We can note in passing, however, that two terms contribute to accessibility multiplicatively: one associated with attractiveness, the other with travel impedance.

We noted that an interpretation of the term A_i which appears in the model has been offered us and is given by Equation (13). In fact, within the context of the shopping model, it is probably more normal to interpret A_i, which is of course the inverse of what we have defined to be the accessibility, as a term which represents the *competition* of other shopping centres for the trade of residents of zone i as perceived by shopkeepers in zone j.

Note that the argument can be turned the other way about. We can define the term $Y_{j \to i}$ as being the potential purchasing power of shops in j from residents in zone i, and write

$$Y_{j \to i} = (e_i P_i)\mathrm{f}(c_{ij}) , \tag{38}$$

and then we can sum the expression on the right hand side of Equation (38) over i to give

$$Y_j = \sum_i (e_i P_i)\mathrm{f}(c_{ij}) , \tag{39}$$

where Y_j now represents something like *population potential*—in this case measured in terms of spending power—for shopkeepers in zone j. Note, however, that this term does not play a direct role in the model as X_i did. The shopkeepers in zone j, therefore, could only treat this as a rough guide, as it does not take account explicitly of the competition of other shopping centres. Any shopkeepers planning on using quantitative methods would be better advised to use the full shopping model rather than this concept of potential! Note also that this concept, and related concepts, of population potential should not be confused with that introduced by Stewart (1947) by analogy with Newtonian physics. A careful reading shows that they are not the same. The concept of 'potential' is not being used here as it is in physics—that is, the interaction is not the gradient of a potential field.

Another associated concept can be mentioned at this point: namely that of *market area*. One of the most attractive features of the gravity model when used to measure market phenomena, as in the shopping model, is that market areas overlap. This seems to give the model a strong edge over more geometrically based models, such as those of central place theory. (It would be an interesting research exercise nonetheless to attempt to reconcile the two: that is to rewrite the central place theory in such a way that some of the most important relationships connected with hierarchies are preserved, but market areas still overlap in a gravity model-like way.) If it was useful to define market areas based on gravity model concepts, then this could be done in a variety of ways: we could say, for example, that residential zone i was in the market area of shops in zone j, if one of a number of possible conditions were satisfied. For example:

(a) if $S_{ij} > A$ (40)

where A is some given number. Alternatively

(b) if $\dfrac{S_{ij}}{e_i P_i} > B$ (41)

where B is some given number. It might be interesting to construct maps of market areas defined in such ways for various alternative values of A and B, recognising of course that the boundaries constructed are not absolute boundaries. An alternative approach, based on probability contours, is given by Huff (1964).

The concepts discussed in this subsection all have one feature in common: they are mostly defined for convenience of presentation and model interpretation. They do not have to have, in any inherent sense, these meanings associated with them within the model itself. Sometimes, however, it would be useful to define some of these concepts in a model-based way for use in other models. Accessibility is a case in point, for one often needs a measure of accessibility to use in such a way.

2.6 *Other approaches to the construction of spatial interaction models*
There are, of course, many alternative approaches to the construction of spatial interaction models, one of the most famous being the intervening opportunities model. This paper is not intended to argue that the gravity model is the only possible model even at the scale of analysis (that is, resolution level) at which gravity models work best. It has been shown, however, that the family of interaction models outlined in this section do arise in a particularly natural way. It has also been shown (Wilson, 1967) that the intervening opportunities model can be derived by an entropy maximising method, but that some rather unusual assumptions have to be made about the form of the cost constraint—the equivalent of Equation (30).

It is hoped that many of the points raised so far in this section, and to be raised in subsequent sections of the paper, are general points which apply to any spatial interaction models at this scale of analysis. A particular family of models has been used to illustrate all the points raised, but any enthusiast of, for example, the intervening opportunities model could tackle some or most of the problems raised in similar ways, and modify and develop the model in ways in which the gravity model has been developed. It is hoped therefore that the paper is of more general applicability than may appear at first sight.

The intervening opportunities model has been mentioned as one alternative approach. A number of others can be mentioned. The gravity model and intervening opportunities model can be combined to some extent if intervening opportunities are employed as a measure of cost in an otherwise gravity model framework (Wilson, 1967, 1969d). This idea has been applied by Cripps and Foot (1969) in a residential location model. The notion of *competing opportunities* was introduced by Tomazinis (1962) to offer another kind of spatial interaction model. Yet another approach, using a kind of entropy maximising method, has been developed by Loubal (1968), and this has been generalised by Ferragu and Sakarovitch (1970). Most of these models are essentially multiplicative in the various terms which are introduced to explain interaction. Additive models have been attempted (Osofsky, 1958, as described by Bruton, 1970), but one would not expect them to be very successful. Alternatively, there is the other extreme position of *log* linearity of the independent variables—but no balancing factors. An example of this approach is the so-called abstract mode model of Quandt and Baumol (1966).

3 Developments of basic model concepts

3.1 *Introduction*
Within this section it is hoped to distinguish between developments up to the present time and possibilities of future development. However, the two are inextricably interwoven. Many of the ideas to be described here have been developed theoretically, a few of them have been tested empirically, others are in the process of being tested, and yet others still have to be tested. The status of the various ideas to be discussed will be pointed out where possible.

The main concepts of a gravity-type spatial interaction model can still be described as the masses, the cost, and the nature of the cost function. We might add to this list the nature of the balancing factors, the A_i's and/or the B_j's, to be used in the model. In effect, this last item means 'which member of the family should be chosen for the particular purpose in hand?' There had been much discussion many times in the past about the possible development of these various concepts.

There has been a reluctance in the past to make the *mass terms* some kind of composite index, and even to measure travel cost in a composite kind of way. At times there was also a feeling that because the model was essentially a gravity model, the cost function ought to be a power function. Further, in the early days of gravity models, it was felt strongly that the exponent ought to be 2 as it is in physics. Until relatively recently, before the impact of the large urban transportation studies, there have been few discussions of the role of the balancing factors and the family of models. With the framework that has been developed and outlined in this paper, we are now in a much stronger position to decide what we can do and what we cannot do in terms of extending and modifying various concepts. These are discussed in turn below.

3.2 *The mass terms*

We have seen that the mass terms can each be one of two kinds: either some measure of total interaction flow out of a zone or into a zone, or some kind of attraction factor. In the first case, the mass obviously has a clearly defined dimension—as a number of trips or whatever; however, this number may be estimated in another sub-model, for example, in the way in which trip ends in transportation studies are estimated in trip generation sub-models, and so even in this case the mass term may be a function of a number of other variables. In the second case, where the mass term is a measure of attractiveness, it is difficult to associate a dimension with the term, and from the point of view of the model unnecessary, as the reader will note that any attractiveness mass term appears in both numerator and denominator (i.e. in the A_i's and/or B_j's) of the expression on the right hand side of the main model equation. This creates all kinds of possibilities for forming composite indices for this type of mass term. It has been argued in other papers (Wilson, 1967, for example) that this type of mass term can be associated with various kinds of scale economies. This sometimes helps us to interpret the term, and also to decide what form it should take.

It is also important, especially for attractiveness-type mass terms, to try to ensure that the models do not make estimates and predictions which are dependent on zone size: if some measure of 'size' is being used as a proxy for attractiveness, this can easily happen if all the zones are not of equal size (as they would be if a grid was being used for example). If the zones are of unequal size, then the definitions of the attractiveness terms should be carefully scrutinized and possibly amended, for example by dividing by the area of the zone, in order to be made dimensionally correct.

So far, we have assumed mostly that mass terms can be estimated outside the interaction model (except when we considered, in section 2.1, that $W_j^{(2)}$ might be taken as T_{*j}). In particular we have at least implicitly assumed that the mass terms (the *total* interaction generated) are independent of the ease of interaction. We have also assumed that total interaction is inelastic with respect to ease of interaction. We might remedy this by assuming a relationship of the form

$$O_i = O_i^0 A_i^\alpha , \tag{42}$$

where

$$A_i = \sum_j D_j \exp(-\beta c_{ij}) \tag{43}$$

is an accessibility measure, and α is then a measure of generation elasticity. This form of trip generation sub-model uses interaction model independent variables as its own independent variables, but it is still exogenous to the interaction model. A situation could be envisaged where D_j in Equation (43), for example, was itself a function of the interaction. The trip generation estimates would then become partly endogenous. Relatively little progress has been made in the development of this kind of model, though common sense suggests that it is badly needed, because of the stringent time series data requirements of the related calibration process.

3.3 *Travel cost*

It seems intuitively clear that there are several components to what we call *travel cost*. If any one of us is making a journey, for example, then we might well take account of the money cost, the different kinds of time expended (travelling time, waiting time, and possibly transfer time), and possibly other features such as comfort and convenience. It seems reasonable, therefore, to try to develop some composite measure of travel cost and to estimate the weight of the various components so that they represent people's actual perception of these weights. An example of this kind of development can be seen in the SELNEC Transportation Study, where the following measure of travel cost was used:

$$c_{ij}^k = a_1 t_{ij}^k + a_2 e_{ij}^k + a_3 d_{ij}^k + p_j^k + \delta^k \,, \tag{44}$$

where

c_{ij}^k = travel cost from i to j by mode k;
t_{ij}^k = travel time from i to j by mode k, in minutes (which included car parking time and walking time on the public transport system);
e_{ij}^k = excess time in the journey from i to j by mode k (waiting time in the public transport system);
d_{ij}^k = distance from i to j in miles (used for estimating perceived operating costs for car drivers and fares for public transport passengers, which are assumed proportional to distance);
p_j^k = terminal cost at j for mode k (car parking charges);
δ^k = modal penalty;
a_1, a_2, a_3 = parameters.

This turns out to be rather a sensitive measure of cost and can reflect network congestion, for example, through the travel time variables. When distance or travel time is used as a measure of travel cost, then such measures are special cases of this kind of generalised cost.

Formulation of this sort raises the problem of estimating the coefficients in equations such as (44). Quarmby (1967) did this using discriminant analysis, for example.

Note also that one of the costs in Equation (44), p_j^k, is a terminal cost. Evans (1970a) has an interesting result in this context: if costs c_i or c_j are added to columns or rows of the cost matrix, and if constraints of the form (3) and (4) are imposed, then the predicted value of T_{ij} is unaffected by the new costs. In the SELNEC Study, from which Equation (44) was taken, this does not apply because the model is a multi-modal model, and an equivalent of Evans' row in the cost matrix runs over j *and* modes k, so the same term is not being added to each element of a row. The significance of this result to the reader who is unacquainted with these models may be more clear after section 4.2 in which a disaggregated transport model is outlined.

One more point on cost measurement can usefully be noted here. In the spatial interaction models outlined so far, we have used a term c_{ij} to represent travel cost. However, Equation (44) represented *modal* cost, c_{ij}^k, and it is clear that when more than one mode is available, it is modal costs which are actually measured. c_{ij} is some kind of *average* of a set of c_{ij}^k's, and so we have to suggest how it should be obtained. This argument can be taken a step further: costs are actually obtained from detailed analysis of modal networks, and in fact, measured costs are costs on *routes* through networks. Thus, even modal costs, c_{ij}^k, are averages over route costs, say γ_{ij}. Usually, of course, the minimum route cost within a mode is taken as the modal cost, but this is only one kind of averaging, and because everyday experience tells us that not all people travel on minimum cost routes, it may not be the most satisfactory form of averaging. This point will be taken up again in section 4.2 where we will have more tools available to take it further forward.

3.4 *The cost function*
With the *cost function*, there seems to be a good case for trying to find empirically the shape which gives the best fit to the observations. (The best known procedure for doing this has been given by the Bureau of Public Roads, 1965.) However, this is also a point where the entropy maximising method can help us interpret any such empirical results rather usefully. The reader will recall that the cost constraint, Equation (30), gives rise to the negative exponential function within the spatial interaction model and this is true for any member of the family of models. Suppose now that a power function fits better empirically. What does this imply? We can note that if c_{ij} is replaced by $\ln c_{ij}$ in the cost constraint, Equation (30), then $\exp(-\beta c_{ij})$ is replaced by $c_{ij}^{-\beta}$, and we could argue in this case that the traveller is *perceiving* travel cost subjectively, in a way which is more like the logarithm of the travel cost measured objectively. This is likely to be true for longer trips. For example, consider the following two situations: traveller A can either travel one mile or 51 miles, while traveller B can travel 300 miles or 350 miles; traveller B, or traveller A in traveller B's situation, is likely to value his additional miles as being rather less expensive than traveller A values his additional 50 miles. The situation represents a logarithmic-like perception of travel cost (that is, travel cost is perceived to increase less than linearly relative to the objectively measured cost). Indeed, it has been remarked by various researchers that power functions often fit better in interurban studies, whereas the negative exponential function often fits best in intraurban studies. Obviously, in the first situation trips are much longer on average than in the second. This kind of argument can be applied to any other form of cost function which turns up empirically. In other words, we are saying that the calibration of the cost function helps us *measure perceived cost.* (See also, section 4.2 below on transport model disaggregation.)

3.5 *The balancing factors*
We have already illustrated, in section 2, models being used for different purposes within which different sets of *balancing factors* are employed. This point will be taken up again below, where we shall see that the choice of model from the family might be dependent on the resolution level at which it is being applied.

3.6 *Hybrid models: policy constraints*
The entropy maximising methodology facilitates the development of hybrid models. This is a general notion but will be illustrated here by a consideration of the problem of imposing policy constraints in an elementary residential location model. The model given by Equations (19) and (20) is used to estimate $\sum_j T_{ij}$, the residential distribution of the population. As a matter of planning policy, upper bounds may be set on this figure for some zones, and these upper bounds represent constraints which have to be imposed in the model. This leads to a hybrid model which is singly constrained with respect to unconstrained population zones and doubly constrained otherwise. Suppose the constraints can be written as equalities (as can be assumed without loss of generality), and suppose Z is the set of all zones, Z_1 the set of population constrained zones, and Z_2 the remainder, so

$$Z = Z_1 \cup Z_2 \,. \tag{45}$$

Then the model to represent this situation, which can be derived by an entropy maximising procedure (Wilson, 1969d) is

$$T_{ij} = A_i B_j O_i E_j \mathrm{f}(c_{ij}) \,, \qquad i \in Z_1 \tag{46}$$

where

$$A_i = \frac{1}{\sum\limits_{j \in Z} B_j E_j \mathrm{f}(c_{ij})} \,, \qquad i \in Z_1 \tag{47}$$

$$T_{ij} = B_j W_i^{(1)} E_j \mathrm{f}(c_{ij}) \,, \qquad i \in Z_2 \tag{48}$$

and

$$B_j = \frac{1}{\sum\limits_{i \in Z_1} A_i O_i \mathrm{f}(c_{ij}) + \sum\limits_{i \in Z_2} W_i^{(1)} \mathrm{f}(c_{ij})} \,, \qquad j \in Z \,; \tag{49}$$

the population constraint has been assumed to be

$$\sum_j T_{ij} = O_i, \quad i \in Z_1 \,. \tag{50}$$

The model has been employed in this way by Batty (1970).

3.7 *External zones*

So far, we have assumed implicitly that we are modelling spatial interaction within a closed system. In practice, we usually have to incorporate a representation of interaction across the system boundary [though the ports' model in the *Portbury* paper (Ministry of Transport, 1966), is one exception to this].

To illustrate the theoretically best way of closing a system, suppose the study area is a region S which has a circular boundary of radius a. Let R be the distance within which x% (say 90%) of the interaction occurs; let S_1 be an annular region of radii a, $a+R$; and let S_2 be a second annular region of radii $a+R$, $a+2R$. Then, S_1 contains zones which will generate some interaction which will terminate in S, and which will be attracted out of S. S_2 has to be added, as S_1 generated interaction which might go to S could alternatively be attracted to zones in S_2. A little thought shows that this closed system is an approximation for all interaction affecting S, but a good approximation.

Because R is often large, this method of system closure is usually a luxury which we cannot afford. It is more usual to designate a number of zones outside the study area as external zones. The entropy maximising method helps us again here: we want to represent the interaction between zones of the study area and external zones, but not between external zones. The appropriate methods are explained in Chapter 5 of a book by the author (Wilson, 1970c).

4 Further theoretical developments

4.1 *Introduction*

As noted earlier, many authors have scrutinized the basic concepts of the spatial interaction model and suggested improvements. Such improvements, as outlined in the previous sub-sections, might be described as traditional. In this section we want to consider a number of less traditional improvements and ways in which they can arise. The topics to be examined here will be discussed under three headings: firstly we will examine the implications of reviewing the level of description we adopt (resolution level) in relation to a system we are attempting to model; secondly we will consider how to treat time; and thirdly we will note that spatial interaction models are often partial models of systems of interest, and we will examine the implications of completing the models.

4.2 *Level of description: systematic disaggregation*

It has long been recognised that some kind of sector disaggregation is desirable for the kinds of models discussed so far in this paper. For example, we travel by particular modes and one of the interesting questions in transport modelling is to predict the number of travellers on each mode. So far we have attempted to predict only total numbers of trips. If we now change the level of description in our approach to the system and admit the possibility of several transport modes, how do we proceed? The traditional way of proceeding in this kind of situation has been to write down spatial interaction model equations for each mode, and then to run the model separately for each mode. This is often unsatisfactory for a number of reasons. For example, if a gravity model is run for trips for each mode separately it is then necessary to predict trip ends by mode. This may be feasible for the residential end of a trip, but seems unreasonable for the destination end: we probably wish to hypothesize for example that all workers are competing for the same jobs irrespective of mode of transport. The entropy maximising method enables us to handle this kind of situation and by proceeding to what might be called systematic disaggregation and writing down the constraint equations on the disaggregated variables, it is possible to develop a set of models at the disaggregated level which are interlinked in an appropriate way. We can also note at this point that disaggregation with respect to one kind of index—in the case of our example the introduction of several modes—may force disaggregation in other ways. For example, it is only sensible to build a transport model covering several modes if we also recognise that particular subsets of the population have only subsets of all modes available to them; in particular, non-car owners do not have the option of travelling by car. It is then possible to handle this problem also. The key point becomes the definition of the disaggregated spatial interaction variable. In the example we have been considering, we started with the aggregate variable T_{ij} representing total trips from zone i to zone j. We wish to disaggregate by mode and also by person type—where the person type index must tell us which modes are available to different people. It will suffice for illustrative purposes simply to distinguish between car owners and non-car owners, and to model a situation with two modes only—private and public transport. In this case, we can introduce, as additional superscripts, k to represent mode, and n to represent person type. Although we have suggested two modes and two person types to fix ideas, the equations to be presented as illustrations below are perfectly general. Note also, as a further preliminary, that any k summation always follows either an n summation or is in an equation where n is a free subscript; it is implicit that the k summation is only over modes which are available to the corresponding person type, and it is left implicit here for convenience. In earlier papers (for example, Wilson, 1967) a notation has been developed to represent this explicitly.

The next step is to write down the constraint equations on the interaction variable. The implicit assumption to be incorporated here is that origin trip ends can be estimated by person type, while destination trip ends cannot; it is also assumed that the cost constraint should be written down for each person type separately, so that different overall patterns of travel behaviour with respect to average length of trip can be predicted for different person types. The constraint equations are then

$$\sum_j \sum_k T_{ij}^{kn} = O_i^n \,, \tag{51}$$

$$\sum_i \sum_n \sum_k T_{ij}^{kn} = D_j \,, \tag{52}$$

$$\sum_i \sum_j \sum_k T_{ij}^{kn} c_{ij}^k = C^n \qquad \text{(where } c_{ij}^k \text{ is the cost of travel by mode } k\text{).} \tag{53}$$

The reader will note how these are obvious disaggregations of Equations (3), (4), and (17). The entropy term is now $\ln T!/\prod_i \prod_j \prod_n \prod_k T_{ij}^{kn}!$; this can be maximised subject to the constraints in the usual way, and leads to the following set of equations

$$T_{ij}^{kn} = A_i^n B_j O_i^n D_j \exp(-\beta^n c_{ij}^k) , \tag{54}$$

where

$$A_i^n = \frac{1}{\sum_j \sum_k B_i D_j \exp(-\beta^n c_{ij}^k)} , \tag{55}$$

and

$$B_j = \frac{1}{\sum_i \sum_n \sum_k A_i^n O_i^n \exp(-\beta^n c_{ij}^k)} . \tag{56}$$

It can be seen that we now have a spatial interaction model for each k–n combination. However, the models are linked in various ways through the A_i^n's and the B_j's. The reader can easily check that for any person type with a choice of mode, this set of equations also leads directly to an estimate of modal split, which can be written in the form

$$\frac{T_{ij}^{kn}}{\sum_k T_{ij}^{kn}} = \frac{\exp(-\beta^n c_{ij}^k)}{\sum_k \exp(-\beta^n c_{ij}^k)} . \tag{57}$$

This model, or set of models, has been tested in the context of the SELNEC Transportation Study and seems to work reasonably well (Wilson *et al.*, 1969).

We can now consider some further aspects of the disaggregation of the transport model. Much of the earlier work has concentrated on modification of the cost function, and we can now see this as a disaggregation. In particular, it has often been noted empirically that better fits of model against observations could be obtained, if a different function was used according to the spatial location of the trip origin. In cities, if a negative exponential function was being employed, it usually meant a decrease in β as the distance of the trip origin from the centre increased. The reader will probably have already realised that this kind of disaggregation can be achieved by stating the cost constraint equation by origin zone:

$$\sum_j T_{ij} c_{ij} = C_i , \tag{58}$$

when the cost function becomes $\exp(-\beta_i c_{ij})$. However, there are no sound *theoretical* reasons as yet for doing so. This can be remedied by hypothesising that trip-making behaviour (with respect to average length of trip, which is what we are mainly discussing) is a function of income. We would expect β to decrease with income, and as average income in fact increases with distance from a city centre in general, this would explain the empirical results referred to earlier. The analysis is made explicit in a paper by Hyman (1970).

A further and most ingenious extension of this kind of idea has been provided by Edens (1970), where he introduces a cost function with both origin and destination variation—in effect (Wilson, 1970a)—by having two cost constraints of the form

$$\sum_j T_{ij} c_{ij} = C_i^{(1)} , \tag{59}$$

and

$$\sum_i T_{ij} c_{ij} = C_j^{(2)} \, . \tag{60}$$

The model is more difficult to interpret, however, and is similar to another approach used by Halder (1970).

Finally, it is useful to note the version of the cost function suggested by Archer (1970). He points out that an alternative way of handling the empirical results mentioned earlier is to make β itself a function of distance. Thus, for a cost function $\exp(-\beta c_{ij})$, he takes

$$\beta = -ac_{ij} + b \tag{61}$$

(expecting a and b to be positive) so that β is less for people making longer trips. This could be interpreted as another way of handling the income disaggregation, or, since we have a new time function of the form $\exp(-a_1 c_{ij} + a_2 c_{ij}^2)$, it could be interpreted in terms of the way people *perceive* costs of longer journeys. It is interesting that we can find two possible interpretations of the theory underpinning a model in this way.

Archer's cost function seemed to be particularly helpful in improving the fit of the modal split function of the form of Equation (57). [Note, incidentally, that the same kind of modal split function would be obtained if we decided to estimate modal split directly using *logit regression* analysis—see Theil (1965)—and also from discriminant analysis—see Quarmby (1967).]

We noted earlier, in section 3.4, that costs were actually measured on *routes*—as γ_{ij}, say, and that we have to construct modal cost c_{ij}^k out of these. This shows also that there is a *route split* problem—the allocation of trips among routes—as well as a *mode split* problem. In fact, we can *define* a mode as being a bundle of routes with similar characteristics: road, rail, bus and so on. It turns out (Wilson, 1969e) that an equation can be developed for route split which is similar in form to the mode split equation. However, a choice then arises:

1 We can perform the route split calculations and then find mode split by summing over modes within a route, or
2 we can perform mode split calculations, and then route split calculations within each mode.

In practice it is found that these procedures give different answers unless there is a particular relationship between modal costs and route costs. This takes the form

$$\exp(-\beta c_{ij}^k) = X \sum_{r \in k} \exp(-\lambda \gamma_{ij}^r) \, , \tag{62}$$

where r represents a route and X and β can be chosen arbitrarily—they determine the zero and units of the modal costs' scale—and the summation is over routes which constitute the mode k. Equation (62) offers a form of 'averaging' over the γ_{ij}^r's to form c_{ij}, and if this form is adopted, then no difference results from the choice of mode split/route split mechanism. However, there is no *a priori* reason why this should hold; if it did not, then we would have to choose between mechanisms '1' and '2' and we could deduce that it implied something about travellers' perception of modes and routes—in case '1' routes being perceived in a primary way, in case '2', modes. There is scope for much empirical work in this field.

It is important to realise that, if it was possible to build a route split model, it would be a substantial contribution to the assignment part of the traditional transport model. We appear to be well on the way to a complete integration of the distribution, modal split and assignment parts of such models already, and if the potential developments in estimating trip ends in section 3.2 materialised, the whole of the transport model will have been integrated.

One final topic on the disaggregation of transport models: clearly, the main parts of the model have to be run for each *trip purpose*, and we have to decide the level of detail at which they should be specified. (This applies to person-trip transport models, and to models such as the retail sales model, where a disaggregation by type of good is equivalent.) One way of pursuing this question is to group together types of trips with similar travel time characteristics, and a systematic way of doing this is offered by Dickey and Hunter (1970).

As a second example, let us consider the disaggregation of the elementary location model used earlier as an example. The model to be presented here is a simplification (and probably a useful development) of a model given in an earlier paper (Wilson, 1969b). In the earlier model, the spatial interaction variable was T_{ij}—the number of people who worked in zone j and who lived in zone i. This model makes no reference to the fact that households have different amounts of income available for house purchase and foı the journey to work, nor does it recognise that there are many types of house available. We can disaggregate now by defining the appropriate spatial interaction variable, which we do by recognising that most income is derived from work; it is also useful to be able to build into the model the possibility that wages vary by job location. Again, it is useful to permit house prices to vary by location, but having built this in, we then assume that price and location between them form a good index of house type. Thus, we can now define our new spatial interaction variable, T_{ij}^{pw} as representing the number of income w households with a member working in zone j, living in a price p house in zone i. (Note that for the time being we are assuming one worker per household.) We can now write down the appropriate constraints on this variable; in this case the addition of a constraint which relates income and price of the house bought, and leads to another interesting feature of the model. The reader will be able to see how to do this by modifying the constraint equations presented in the paper mentioned earlier (Wilson, 1969b). The resulting model equation is

$$T_{ij}^{pw} = A_i^p B_j^w H_i^p E_j^w \exp(-\beta^w c_{ij}) \exp\{-\mu^w [p - q^w(w - c_{ij}^1)]^2\}, \tag{63}$$

where

$$A_i^p = \frac{1}{\sum\limits_j \sum\limits_w B_j^w E_j^w \exp(-\beta^w c_{ij}) \exp\{-\mu^w [p - q^w(w - c_{ij}^1)]^2\}}, \tag{64}$$

and

$$B_j^w = \frac{1}{\sum\limits_i \sum\limits_p A_i^p H_i^p \exp(-\beta^w c_{ij}) \exp\{-\mu^w [p - q^w(w - c_{ij}^1)]^2\}}. \tag{65}$$

H_i^p is the number of households in price p houses living in zone i. Once again we have a *set* of spatial interaction models, one for each p–w combination, and again interlinked through the balancing factors (which are themselves disaggregated by p and w respectively). The model given by Equation (63) is still recognisably a spatial interaction model. The first two terms are balancing factors, then follow a mass term, then another mass term, and then the distance function.

However, our additional constraint has generated a further term which appears as a factor at the end of the expression on the right hand side of Equation (63). In this term q^w is the average expenditure of a w-income household on housing after journey to work costs have been deducted. In the expression, journey to work costs are written as c_{ij}^1 to distinguish between the money cost of the journey to work only as compared to the generalised cost c_{ij}. It can easily be seen that this term permits a distribution of house prices for w-income households around some mean, and the width of the distribution is determined by the parameter μ^w. It is easy to see, on reflection, that this model is capable of reproducing patterns of development in a way in which the very elementary spatial interaction model is not. It is clear though, that the model needs to operate on a quasi-dynamic basis (see section 4.3 below); then it could explain patterns which are the kinds of mixtures of Burgess–Hoyt–Harris–Ullman patterns which appear in real life.

This is an appropriate point to explain how a particular conundrum is resolved. All the residential models discussed in this paper predict, *in passing*, the journey to work; and yet the reader can easily check that these models (and hence the predictions) are different from the journey to work model used in transportation studies. The two kinds of model, however, refer to different time periods. The quasi-dynamic residential location model is applied to *potential movers* (see section 4.3 below) only, while the journey to work model explains the whole pattern. Thus, we have discovered something fairly obvious in relation to present practice: if we want a prediction of the journey to work, then the best models used in the transportation studies *are* the best currently available; residential location models which predict the journey to work will not be quasi-dynamic and will be unreliable in most respects, but certainly so in their journey to work predictions.

The third example, of a rather different kind, relates to spatial interaction models of interregional commodity flows. Attempts have been made in the past to build interaction models of such flows, usually using an unconstrained Reilly-type of model. Suppose, however, that we define a spatial interaction variable x_{ij}^m as the flow of good m from i to j—assuming production in zone i and use (including both final and intermediate use) in zone j. Then we could apply a simple spatial interaction model to each good. However, we know from our regional economic theory that the quantities of goods produced and used in various ways are related through some form of input–output model. In building a spatial interaction model of flows for a multiregion system, it turns out that we can use the input–output equations for this system (as defined by Leontief and Strout, 1963) as constraint equations. If there are no interaction-end constraints, then the only constraint equations are

$$\sum\sum x_{ij}^m c_{ij}^m = C^m \ , \tag{66}$$

$$\sum_j x_{ji}^m = \sum_n a_{mn}^i \sum_j x_{ij}^n + y_i^m \ . \tag{67}$$

This leads to a model of the form

$$x_{ij}^m = \delta_i^m \epsilon_j^m \exp(-\mu^m c_{ij}^m) \ , \tag{68}$$

where

$$\delta_i^m = \prod_n (\epsilon_i^n)^{-a_{nm}^i} \ , \tag{69}$$

and

$$\epsilon_i^m = \frac{y_i^m + \sum\limits_n a_{mn}^i \delta_i^n \sum\limits_j \epsilon_j^n \exp(-\mu^n c_{ij}^n)}{\sum\limits_j \delta_j^m \exp(-\mu^m c_{ji}^m)} \ . \tag{70}$$

The model Equation (68) has two related balancing factors which ensure that the input–output constraints are satisfied, and it has the usual kind of distance function.

It would be possible to incorporate attractiveness factors in the usual way and make it look more like an interaction-end unconstrained model. In the paper in which this work was developed (Wilson, 1970b), a whole family of these kinds of gravity-input-output models were developed to represent various situations. Once again, we have a set of interlocking models.

Note that in the case of the disaggregated residential location model, and the commodity flow model, the nature of the balancing factors has changed with disaggregation. The residential location model now has double balancing factors, $A_i^p B_j^w$, whereas earlier the equivalent aggregated model has only an A_i balancing factor. This means that when the resolutional level changes, we may have to select another model from the family. In the case of the commodity flow model the balancing factors are of a different type altogether because of the nature of the input-output constraint.

At this stage it is worth noting a perhaps obvious point, that as disaggregation continues the model may become very difficult to handle. For example, with the disaggregated residential location model, if we have say 100 zones in our study area, a number of price ranges, and a number of income groups then this leads to a very large number of 'cells' of the array T_{ij}^{pw}, which makes for difficulties in both computer programming and sampling for the purposes of model testing. One way out of this difficulty is to consider using continuous variation for some of the subscripts which we have considered hitherto to be discrete, such as that referring to income. Then, instead of generating a large number of parameters, such as β^w and μ^w, we can generate a smaller number of parameters that relate to the shape of the actual income distribution. This possibility is explored for transport models in a paper by Hyman (1970).

4.3 *Time*

The kind of spatial interaction model outlined in this paper is often described as a comparative static equilibrium model. Few people, however, have actually stated clearly just what 'equilibrium' means in this context. In this sub-section we shall discuss firstly the behaviour of the parameters of the gravity model over time, secondly how to make some kinds of model quasi-dynamic, and thirdly, how to approach the problems of building a fully dynamic model.

Forecasts of future interactions are made using gravity type models in the following way: assumptions are made about future values of the mass terms and the travel cost matrix; these are fed into the model usually with an additional assumption of the constancy of the parameters of the model. The model will then predict future interaction. The cost constraint equation, the equivalent in the particular model of Equation (30), enables us in principle to do rather better than this. If we can work out how total expenditure on transport is likely to change overall—and in principle we might predict this with some kind of economic model—then we can use the cost constraint equation to estimate β, since β is the Lagrangian multiplier associated with this parameter. A number of alternative ways of predicting change in β are discussed in a paper by Hyman and Wilson (1969).

To see the simplest way of dynamising models let us consider again the residential location model. All the residential location models discussed so far allocate the whole population to residences. It is almost certainly much more realistic to assume that in any one time period, one year say, some subset of the population forms the set of potential movers estimated for the next time period, and so on. In such a model, time does not appear as a variable explicitly and so we can call the model quasi-dynamic. A method for doing this is described in a paper by the author (Wilson, 1969b).

Finally we consider briefly the possibility of building truly dynamic models. We noted in an earlier section that the statistical mechanics analogy may be used for this purpose since, in physics, dynamic models have been built for possibly analogous systems. The problem was explored on the basis of the statistical mechanics' analogy by the author (Wilson, 1969a) and has also been approached on the basis of kinetic theory by Tomlin (1969, 1970).

When we study system dynamics in this way, it is interesting to ask whether any social science equivalents of the laws of thermodynamics hold, and in particular, whether any kind of Second Law holds. This is also discussed in an earlier paper (Wilson, 1969a). It does appear that there are situations where some kind of Second Law applies, though this also implies something about the way we attach 'values' to certain variables. In the social sciences by artificially associating certain values with certain variables, the Second Law could be infringed.

4.4 *Spatial interaction models in general frameworks: sub-systems and boundaries*
Clearly, spatial interaction models are really parts of more general models. Often there will be several spatial interaction models within the framework of a single, more general model. The relationship between several models can often be clarified by using an accounting framework (Wilson, 1969c), but this point will not be pursued further here.

The main point to note about the spatial interaction models discussed in this paper is that they are mostly in the nature of demand models: that is they represent person demand for trips, person demand for residences, and so on, and although this is not strictly demand in the economist's sense, it is something approaching it. The reader can easily see that some of the mass terms in our models represent what the supply side is offering; and so we can often complete our models by adding supply side models which predict the mass terms.

This point can be illustrated using the disaggregated residential location model (more strictly, a residential-workplace allocation model) given by Equations (63)-(65). In the model, H_i^p represents the supply of housing, and E_j^w the supply of jobs. It is interesting that in this context, one of the simplest 'old' residential location models (due to Hansen, and described in Swerdloff and Stowers, 1966) can be reinterpreted as a simple *supply side* model. In this model,

$$G_i = G_t \frac{V_i A_i^\alpha}{\sum_i V_i A_i^\alpha} , \tag{71}$$

where

G_i	= housing growth in zone i	A_i	= 'accessibility' of zone i
G_t	= total growth to be allocated	α	= parameter.
V_i	= vacant land in zone i		

For example, we might have

$$A_i = \sum_j E_j c_{ij}^{-n} , \tag{72}$$

where E_j is jobs in zone j, in the usual way. To make it fit our model we could define a coefficient a_i^p as the proportion of new houses in zone i which will sell for price p, and then take

$$H_i^p = a_i^p G_i . \tag{73}$$

The residential-workplace allocation model will then predict *who* will live in these houses. Of course, the supply side mechanism is much more complicated, but this serves to show how we must be careful in interpreting models.

4.5 *Final comments: some other possible theoretical developments*

The kind of spatial interaction models described in this paper work reasonably well at a particular scale of analysis. They work for systems which contain large numbers of decision making units, like a large population each deciding where to travel, where to purchase a house, and so on. The entropy maximising method tells us this, and indeed can be used formally to tell us whether the population in our system is large enough for a statistical averaging technique to be effective. As a result it means for example that this kind of model is not likely to be useful in the study of industrial location where the decisions are being made by a relatively smaller number of people. Having said all this, we can note that there are many other types of development in relation to the kinds of problems that we are studying and briefly it is worth relating some of these to the models which have been outlined.

It is not clear to what extent spatial interaction models imply an assumption about perfect information being available to people who form the components of the system being modelled. I suspect that they do not imply this. Nonetheless, it may be useful to consider explicitly the problem of using mental map type information, which is being obtained in perception studies, and integrating it with gravity models. This should not be difficult. One possibility is to incorporate a terminal 'cost', c_j, with possible destinations j, in the appropriate model, and to estimate c_j from variables which occur in mental map-type studies. Various interpretations could then be offered of the resulting term.

It is also interesting to try to relate the spatial interaction kind of model with the theory of consumer behaviour. This theory is usually based on some kind of utility maximising principle, where a utility function is maximised subject to a number of constraints. Chapter 6 of a recent book by the author discusses the role of entropy maximising methods in the analysis of utility maximising systems (Wilson, 1970c). As usual, any related entropy maximising model arises from a statistical averaging process, and in this case, it is sometimes useful to imagine that some of our models result from this kind of averaging over the preferences of the population. The preferences, expressed perhaps as indifference curves, are not known for each individual, but some kind of overall constraints may be known. Some authors, notably Neidercorn and Bechdolt (1969), have derived the simple gravity model from a maximising procedure—which, naturally, involves maximising entropy—and have then interpreted the maximand as utility! Of course, there is almost never any justification for doing this.

It is interesting to explore further the economic status of spatial interaction model equations. We have seen that some such equations have similar status to demand equations, and this fact will be used in section 5 below to develop a concept of consumers' surplus for use with spatial interaction models. However, this is not strictly accurate: the disaggregated residential location model as presented in Equation (63) is more like an equation which describes the intersection of supply and demand curves at a price, p. This makes its usage, along with the concept of consumers' surplus, still essentially sound.

However, such a model is often an anathema to the economist, who expects to build models of rational behaviour at the individual level. We have already noted that the spatial interaction model represents a statistical average over a wide range of such behaviour, involving many different utility functions. Indeed, the economist is forced to aggregate from individual behaviour to get models he can think of using in practice. This can lead to an interesting alternative approach to some of the problems for which we have suggested disaggregated spatial interaction models. It is interesting to note, for example, that the residential location model of section 4.2 (a disaggregated spatial interaction model) operates at about the same level of aggregation as the Herbert and Stevens (1960) model (an aggregated utility maximising model). A comparison of the properties of these two models will be given elsewhere by the author. (Incidentally, this example illustrates the relationship between spatial interaction approach and the linear programming models.)

Of course, there are many other alternative approaches: the economics-trained model builder might construct modal split models with parameters which can be interpreted directly as elasticities; a paper by Hyman and Wilson (1969) calculates elasticities associated with corresponding spatial interaction models to facilitate comparison. This list could be extended in many ways, but in the last analysis, we can expect different approaches to problems to be fruitful both in similar and in different ways. The only claim for the spatial interaction models made in this paper is that they represent a fruitful approach to one class of problems at a certain level of aggregation. Directions of change in urban and regional modelling are discussed more broadly in another paper (Wilson, 1970e).

5 The uses of spatial interaction models

5.1 *Introduction*

It has been argued in an earlier paper (Wilson, 1968) that there are three main kinds of task in planning, which are concerned with policy, design and analysis. *Policy* is concerned with goal formulation and the development of associated evaluation criteria so that the best plan out of a set can be chosen. *Design* is concerned with the generation of alternative plans. *Analysis* is concerned with the use of models to make predictions; this can help the designer in an obvious way to explore the impact of each alternative and can be used to help produce quantitative evaluation criteria. In this section, we are concerned to illustrate briefly how spatial interaction models can help in this kind of process. The reader is referred to various references already cited or to be cited for a more detailed account of both the framework and of particular applications.

5.2 *Transport*

The main tasks of the transport planner are to decide rights of way in the transport system, and then to decide on the amounts of investment to be made in particular time periods in each modal network. If we consider that there are only two modes, involving investment in something called public transport and something called roads, and that the planner has been told that his provisional budget of £X is fixed, then he is considering possible investments at points along the line shown in Figure 2.

Thus, the transport planner will need some measure of rate of return on investment and how it changes along the line (that is between different relative investments, in public transport and private car transport) and across the line (that is how it changes with the changing budget).

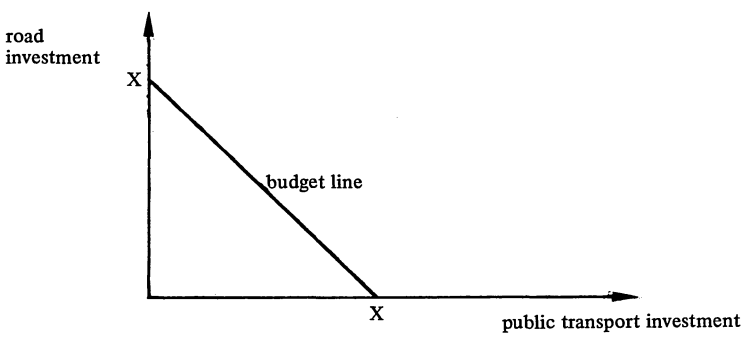

Figure 2. Budget line in transport planning.

In this case it is possible to establish some kind of evaluation criterion based on the economists' notion of consumers' surplus. Assume that we are considering all trips between zone i and zone j, T_{ij} at a cost c_{ij}, and then assume that the relationship between T_{ij} and c_{ij} is like a demand-curve as shown in Figure 3.

The shaded area in Figure 3 is the change in consumers' surplus, resulting from a cost change from c_{ij}^1 to c_{ij}^2 (for example, Wilson and Kirwan, 1969). These changes in consumers' surplus (c.s.) for one origin-destination pair can be summed to estimate the change in c.s. for a change in the whole network:

$$\Delta\text{c.s.} = \tfrac{1}{2}\sum\sum (T_{ij}^{(1)} + T_{ij}^{(2)})(c_{ij}^{(1)} - c_{ij}^{(2)}) . \tag{74}$$

A formula of this kind has been used in many transportation studies (cf.Tressider *et al.*, 1968).

Often it is much more interesting to attempt to measure the incidence of costs and benefits—the kind of disaggregated model described in section 4.2 enables us to do this. In the model the superscript n represents person type—in the SELNEC case it represented car owner or non-car owner—and we can apply the formula for changing consumer surplus to each person type separately. This then shows how the benefits accrue to different groups for particular types of network change (for example, more investment in roads rather than public transport generates proportionately more benefits for car owners).

5.3 *Retail facilities*

Equation (14) given earlier can be interpreted as a model of usage of retail facilities by a given distribution of residents. The attractiveness of the retail facilities is expressed through the set of terms $W_j^{(2)}$. As noted previously, this is usually taken to be some size variable such as floorspace, or perhaps turnover itself. The equation is repeated here for convenience:

$$S_{ij} = (e_i P_i)\frac{W_j^{(2)}\exp(-\beta c_{ij})}{\sum\limits_j W_j^{(2)}\exp(-\beta c_{ij})} . \tag{75}$$

A 'plan', in this case, will usually consist of a specification of a set of $W_j^{(2)}$'s, as floorspace or whatever. The model can then be used to predict the resulting flows, S_{ij}, and the turnover in particular shopping centres, S_{*j} (recall that an asterisk replacing a suffix denotes summation). The plan which produces a satisfactory turnover in relation to floorspace (or more generally, in relation to running and capital costs) can then be chosen. This is evaluation from the viewpoint of the developer. It would also be possible to develop criteria based on the viewpoint of the consumer—noting that the 'developers' solution' and the 'consumers' solution' for the distribution of facilities are likely to be different (Hotelling, 1929).

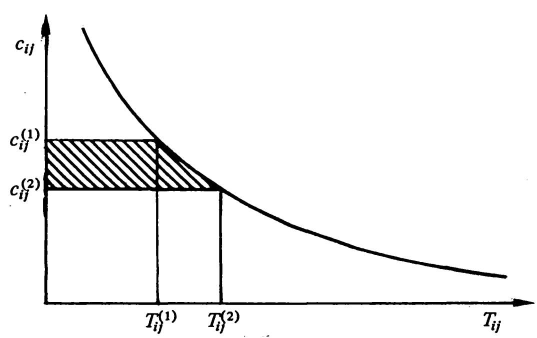

Figure 3. Change in consumers' surplus.

The model could also be used, in a similar way to the transport model, to investigate the impacts of network (c_{ij}) changes. Ideally it should be disaggregated by mode for this purpose.

5.4 *Residential location and housing policy*
It is probably clear already to the reader that the most elementary residential location models are not particularly helpful except to give general indications of pressure for residential development. This is because there is no income/house-type variable in such models. It was for this reason, namely to make such models more effective for their uses in planning, that the disaggregated residential location model outlined in section 4.2 was constructed. Equation (63) is repeated here for convenience:

$$T_{ij}^{pw} = A_i^p B_j^w H_i^p E_j^w \exp(-\beta^w c_{ij}) \exp\{-\mu^w [p - q^w (w - c_{ij}^1)]^2\} . \qquad (63)$$

The variable H_i^p which appears on the right hand side of this equation represents a statement of housing policy: it is the number of houses in each zone of the area available at a certain price. The model then predicts, in relation to a given distribution of jobs by wage, which people will live in which houses. This enables the most important problem areas to be thoroughly explored: in particular, in many situations in cities at the present time poor families have very little choice and probably need much more choice in order to be able to maximise their utility in other dimensions. Such problems cannot be explored using more elementary models.

The way in which such a model would be used depends, as in the case of the retail sales' model, on whose viewpoint is adopted. Ways of using the model in a predominantly public economy are outlined in another paper (Wilson, 1969f). It would also be possible to develop a measure of consumers' surplus from Equation (63), by analogy with the transport Equation (74), which took account of house price changes as well as transport price changes (cf.Wilson, 1969f).

6 Summary and conclusions

6.1 *Summary: the nature of the gravity model*
This paper has really been about the nature of the gravity model we all once thought we knew, and so the summary will be related to this central issue. In the next sub-section, a number of conclusions will be drawn in relation to the practical issue of how to build spatial interaction models.

In section 2.1, we saw that the gravity model is not a single model, but that there is a whole family of related spatial interaction models. In section 2.3, this family of models was derived using an entropy maximising methodology rather than the old Newtonian analogy. This helps considerably in our understanding of the nature of the gravity model: in particular, we saw that entropy maximising can be interpreted as a statistical averaging procedure, and thus that we can expect the model to work reasonably effectively when describing behaviour in large populations. For obvious reasons it begins to be much less effective if there are smaller numbers of decision making units in the system of interest.

Section 2.3 notes properties of the balancing factors in the production–attraction constrained model; section 2.4 is concerned with calibration methods; and section 2.5 with associated concepts such as accessibility, competition, potential and market areas. These sections all help with the practical development and with the interpretation of terms in the family of spatial interaction models.

Section 2.6 notes the existence of a number of other approaches to building spatial interaction models, but draws the important conclusion that many of the developments of this paper which have been applied to the old gravity model can be applied, *mutatis mutandis*, to other 'old' models such as the intervening opportunities model.

Section 3 reviews the basic concepts associated with the model family: a range of developments are possible, which will be summarised in the next sub-section as they relate to operational model building strategy. We can appropriately note two points at this stage, however. Firstly, section 3.6 shows how certain hybrid models can be developed—for example, partly attraction constrained, partly production–attraction constrained. This means that the family of models outlined in section 2 is capable of expansion to incorporate a considerable number of hybrid variants. Secondly, it is appropriate to note that the spatial interaction models we have discussed in sections 2 and 3 all apply to a single interaction sector, using a variable such as T_{ij}. These will be called *single-sector* models, in contrast to the *multi-sector* models to be discussed shortly.

In section 4, developments of a different order are described: in sections 2 and 3, we developed a family of single-sector spatial interaction models and discussed developments of the basic concepts associated with these models; in section 4, we discussed theoretical developments which considerably extend the family. In section 4.2, we discussed model disaggregation: this almost always means sector disaggregation to develop *multi-sector* spatial interaction models. Of course, a multi-sector model could be a set of single-sector models—one for each sector—but we are more interested in cases where different sectors interact. In an example, such sector interaction arose from travellers by different modes competing for the same job opportunities, from workers with different job locations and different wages competing for a stock of houses, and from flows of goods related by a set of input–output equations. In each case, we developed a sector-interacting set of spatial interaction models, thus further extending the concept of the 'family of models' introduced in section 2.

We went on to consider how to introduce *time* into spatial interaction models. In particular, it was argued that if the spatial interaction models are applied to *potential movers* instead of the whole population (this in the residential location model—similar notions could be applied in other contexts) then the models take on dynamic behavioural properties.

In section 4.4, we saw that certain requirements are imposed on the model—and interpretation of terms is made easier—if we recognise explicitly that the spatial interaction models we use are almost always sub-models in a more general framework.

In section 4.5, we saw how developments with these kinds of spatial interaction models can be related to other approaches to problems in urban modelling. Finally, section 5 outlines some applications of spatial interaction models in planning.

6.2 *Conclusions: a spatial interaction model building kit*

Another way of summarising the conclusions of this paper is to present a check list of steps for the spatial interaction model builder. This can be done as follows:

1 *Nature of spatial interaction variable.* The first task is to specify the spatial interaction variable, and in particular the level of sector aggregation. This will determine among other things whether the model will be single-sector or multi-sector. At the same time it is appropriate to decide how to treat time, and then to see whether the spatial interaction variables should relate to the total population or to potential movers (or the equivalent concept) only.

2 *Nature of the mass terms.* It should then be possible to decide on the nature of the mass terms—whether they are interaction totals or attractiveness terms, and their sector aggregation. The answers to these two questions will determine the nature of the balancing factors. The nature of the mass terms and balancing factors can be checked further by writing down the constraint equations which would be used in the entropy maximising construction of the model. Occasionally, these will be of an unusual type—as with the input–output equations in the commodity flow model.

3 *Measurement of mass terms.* It will be clear how to do this for mass terms which are interaction totals, but for attractiveness terms there will be considerable degree of choice, the possibility of developing composite indices, and so on.

4 *Measurement of interaction cost.* This will be straightforward, in principle, if we use the notion of generalised cost, but may be difficult in practice.

5 *Choice of cost function.* There is a very wide choice. Several functions should be tried in the calibration process and the function used which fits best. It will then be possible to interpret this either in terms of the way in which cost is perceived, or as a 'hidden' disaggregation.

6 *Other terms, for example other probability factors.* In a multi-sector model there may be inter-sector interaction (not in any way spatial) which has to be incorporated. An example is the term in the residential location model which relates house price to ability to pay. More generally, is the theoretical behaviour of the model adequate?

7 *Design of spatial system.* It is necessary to define zones—particularly with reference to zone size and external zone structure.

8 *Calibration.* It is necessary to devise appropriate goodness-of-fit statistics for calibration, and to review the data which is available for this process.

9 *Relation to general model framework.* The model is almost certainly a sub-model of a more general model. This should be made explicit, and will help clarify the role of the model in planning and generally with what has been called the supply side.

We conclude this summary by showing, in Table 1, how these points can be related to the models presented as illustrations in section 4.2.

Table 1.

Check list	Transport model	Residential model	Commodity flow model
1 (1)	T_{ij}^{kn}	T_{ij}^{pw}	x_{ij}^{m}
	(time: all apply to whole population; residential location models should apply to potential movers only.)		
2	Masses: O_i^n, D_j Balancing factors: A_i^n, B_j	H_i^p, E_j^w A_i^p, B_j^w	- $\delta_i^m \epsilon_j^m$ (2)
3	trips	houses, jobs	-
4	generalised person trip cost	generalised journey to work cost	freight rates?
5 (3)	$\exp(-\beta c_{ij}^k)$	$\exp(-\beta^w c_{ij})$	$\exp(-\mu^m c_{ij}^m)$
6	-	$\exp\{-\mu^w[p-q^w(w-c_{ij}^1)]^2\}$	
7,8,9 (4)			

Notes

(1) These spatial interaction variables all apply to the whole population. In answering the question about time, we would conclude that the variable in the residential location model should ideally be redefined to apply to potential movers only.

(2) These balancing factors are not of the usual kind: they arise from the use of input-output equations as constraints.

(3) The models have all been recorded in the paper as using negative exponential functions as cost functions; a wider range of functions should be tested empirically.

(4) Only questions '1'-'6' are answered here as between them they determine the design of the model. The other questions relate mainly to the calibration and use of the model, and are not taken any further here. We should note, however, an obvious feedback between answers to these questions and the model design questions themselves: if, for example, certain data are not available, then this may force the model builder to a more coarse aggregation level.

References

Ambrose, P., (Ed.), 1969, *Analytical Human Geography* (Longmans Green, London).

Archer, E., 1970, "An investigation of a rider prediction technique for a demand-scheduled bus system", M.A.Sc. Dissertation, Department of Civil Engineering, University of Waterloo, Ontario.

Batty, M., 1970, "Some problems of calibrating the Lowry model", *Environment and Planning,* **2**, 95–114.

Bruton, M. J., 1970, *Introduction to Transportation Planning* (Hutchinson, London).

Bureau of Public Roads, 1965, *Calibrating and Testing a Gravity Model for any size Urban Area* (US Government Printing Office, Washington, D.C.).

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; also in Ambrose (1969), *op.cit.*

Chisholm, M., Frey, A., Haggett, P., (Eds.), 1971, *Proceedings, 22nd Colston Research Society Symposium* (Butterworths, London).

Clough, D. J., 1964, "Application of the principle of maximising entropy in the formulation of hypotheses", *Canadian Operational Research Society Journal,* **2**, 53–70.

Cordey Hayes, M., Wilson, A. G., 1970, "Spatial interaction", *Socio Economic Planning Sciences,* to be published; CES-WP57 (Centre for Environmental Studies, London).

Cox, K. R., Golledge, R. G., (Eds.), 1969, *Behavioural Problems in Geography: A Symposium* (Northwestern University Press, Evanston, Illinois).

Cripps, E. L., Foot, D. H. S., 1969, "The empirical development of an elementary residential location model for use in sub-regional planning", *Environment and Planning,* **1**, 81–90.

Dickey, J. W., Hunter, S. P., 1970, "Grouping of travel time distributions", *Transportation Research,* **4**, 93–102.

Edens, H. J., 1970, "Analysis of a modified gravity model", *Transportation Research,* **4**, 51–62.

Eilon, S., Tilley, R. P. R., Fowkes, T. R., 1969, "Analysis of a gravity demand model", *Regional Studies,* **3**, 115–122.

Evans, A. W., 1970a, "Some properties of trip distribution methods", *Transportation Research,* **4**, 19–36.

Evans, A. W., 1970b, "The calibration of trip distribution models with exponential or similar cost functions", *Transportation Research,* to be published.

Ferragu, C., Sakarovitch, M., 1970, "A class of 'structural' models for trip distribution", *Transportation Research,* **4**, 87–92.

Golledge, R. G., 1969, "The geographical relevance of some learning theories", in Cox and Golledge (1969), *op.cit.*, 101–145.

Halder, A. K., 1970, "An alternative approach to trip distribution", *Transportation Research,* **4**, 63–69.

Hansen, W. G., 1959, "How accessibility shapes land use", *Journal of the American Institute of Planners,* **25**, 73–76.

Herbert, J., Stevens, B. H., 1960, "A model for the distribution of residential activity in urban areas", *Journal of Regional Science,* **2**, 21–36.

Hotelling, H., 1929, "Stability in competition", *Economic Journal,* **39**, 41–57.

Huff, D. L., 1964, "Defining and estimating a trading area", *Journal of Marketing,* **28**, 34–38; also in Ambrose (1969), *op.cit.*

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

Hyman, G. M., 1970, "Trip distribution and modal split by categories of households", *Transportation Research,* **4**, 71–76.

Hyman, G. M., Wilson, A. G., 1969, "The effects of changes in travel costs on trip distribution and modal split", *Journal of High Speed Ground Transportation,* **3**, 79–85.

Isard, W., 1960, *Methods of Regional Analysis* (MIT Press, Cambridge, Mass.).

Kirby, H. R., 1970, "Normalizing factors of the gravity model—an interpretation", *Transportation Research,* **4**, 37–50.

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

Lambe, T. A., 1969, "The opportunity, gravity and Thurstone models of individual choice", *Socio Economic Planning Sciences,* **3**, 411–419.

Leontief, W., Strout, A., 1963, "Multi-regional input–output analysis", in *Structural Interdependence and Economic Development,* Ed. T. Barna (Macmillan, London).

Loubal, P., 1968, "A mathematical model for traffic forecasting", Ph.D. Thesis, University of California, Berkeley.

Lowry, I. S., 1964, "A model of metropolis", RM-4035-RC, Rand Corporation, Santa Monica.
Ministry of Transport, 1966, *Portbury* (HMSO, London).
Morrill, R. L., Kelley, M. B., 1970, "The simulation of hospital use and the estimation of location efficiency", *Geographical Analysis,* **2**, 283–300.
Neidercorn, J. A., Bechdolt, B. V., 1969, "An economic derivation of the 'gravity law' of spatial interaction", *Journal of Regional Science,* **9**, 273–282.
Olsson, G., 1965, *Distance and Human Interaction* (Regional Science Research Institute, Philadelphia).
Osofsky, S., 1958, "A multiple regression approach to forecasting urban area traffic volumes", Proceedings of the American Association of State Highway Officials, Washington.
Quandt, R. E., Baumol, W. J., 1966, "The demand for abstract transport modes: theory and measurement", *Journal of Regional Science,* **6**, 13–26.
Quarmby, D. A., 1967, "Choice of travel mode for the journey to work: some findings", *Journal of Transport Economics and Policy,* **3**, 273–314.
Scott, A. J., (Ed.), 1969, *London Papers in Regional Science,* volume 1, *Studies in Regional Science,* Proceedings of the Second Annual Conference of the British Section of the Regional Science Association (Pion, London).
Stewart, J. Q., 1947, "Empirical mathematical rules concerning the distribution and equilibrium of population", *Geographical Review,* **37**, 461–485.
Swerdloff, C. N., Stowers, J. R., 1966, "A test of some first generation residential land use models", *Highway Research Record,* **126** (Highway Research Board, Washington, D.C.), 38–59.
Theil, H., 1965, *Economics and Information Theory* (North-Holland, Amsterdam).
Tomazinis, A. R., 1962, "A new method of trip distribution in an urban area", Highway Research Board Bulletin 347 (Highway Research Board, Washington, D.C.).
Tomlin, S. G., 1969, "A kinetic theory of traffic distribution and similar problems", *Environment and Planning,* **1**, 221–227.
Tomlin, S. G., 1970, "Time dependent traffic distributions", *Transportation Research,* **4**, 77–86.
Tressider, J. O., Meyers, D. A., Burrell, J. E., Powell, T. J., 1968, "The London transportation study: methods and techniques", *Proceedings of the Institution of Civil Engineers,* **39**, 433–464.
Wilson, A. G., 1967, "A statistical theory of spatial distribution models", *Transportation Research,* **1**, 253–269.
Wilson, A. G., 1968, "Models in urban planning: a synoptic review of recent literature", *Urban Studies,* **5**, 249–276.
Wilson, A. G., 1969a, "Notes on some concepts in social physics", *Papers, Regional Science Association,* **22**, 159–193.
Wilson, A. G., 1969b, "Disaggregating elementary residential location models", CES-WP37 (Centre for Environmental Studies, London); to be published, *Papers, Regional Science Association.*
Wilson, A. G., 1969c, "The integration of accounting and location theory in urban modelling", in A.J.Scott (Ed.) (1969), *op-cit,* 89–104.
Wilson, A. G., 1969d, "Developments of some elementary residential location models", *Journal of Regional Science,* **9**, 377–385.
Wilson, A. G., 1969e, "The use of entropy maximising models in the theory of trip distribution, mode split and route split", *Journal of Transport Economics and Policy,* **3**, 108–126.
Wilson, A. G., 1969f, "Metropolitan growth models", CES-WP55 (Centre for Environmental Studies, London); to be published, *Conference Proceedings,* The mastery of urban growth, Mens en Ruimte, Brussels.
Wilson, A. G., 1969g, "Generalising the Lowry model", CES-WP-56, Centre for Environmental Studies, London; to be published in A.G.Wilson (Ed.) (1971).
Wilson, A. G., 1970a, "Advances and problems in distribution modelling", *Transportation Research,* **4**, 1–18.
Wilson, A. G., 1970b, "Interregional commodity flows: entropy maximising approaches", *Geographical Analysis,* **2**, 255–282.
Wilson, A. G., 1970c, *Entropy in Urban and Regional Modelling* (Pion, London).
Wilson, A. G., 1970d, "The use of the concept of entropy in system modelling", *Operational Research Quarterly,* **21**, 247–265.
Wilson, A. G., 1970e, "On some problems in urban and regional modelling", in M.Chisholm, A.Frey, and P.Haggett (Eds.) (1971), *op.cit.*
Wilson, A. G., (Ed.), 1971, *London Papers in Regional Science,* volume 2, *Urban and Regional Planning,* Proceedings of the Third Annual Conference of the British Section of the Regional Science Association (Pion, London), in the press.

Wilson, A. G., Hawkins, A. F., Hill, G. J., Wagon, D. J., 1969, "Calibrating and testing the SELNEC transport model", *Regional Studies,* 3, 337–350.
Wilson, A. G., Kirwan, R., 1969, "Measures of benefits in the evaluation of urban transport improvements", CES-WP-43, Centre for Environmental Studies, London.