Quantitative Geography

Perspectives on Spatial Data Analysis

A. Stewart Fotheringham
Chris Brunsdon
Martin Charlton
9 Spatial Modelling and the Evolution of Spatial Theory

9.1 Introduction

A theme introduced in Chapter 1 is that quantitative geography has recently reached a stage of maturation whereby it is no longer a net importer of ideas and techniques from other disciplines but, rather, it is a net exporter, particularly in the areas of spatial statistics and geographical information science. The importation of ideas from other disciplines is not in itself to be discouraged and there is clearly a great deal to be gained by such cross-fertilization. However, since the beginnings of quantitative geography in the late 1950s and early 1960s and lasting until quite recently, there has been an unhealthy reliance on the use of statistical techniques and mathematical models developed in predominantly aspatial disciplines such as economics and physics. For instance, quantitative geographers have flirted with topics such as entropy (and appeals to the second law of thermodynamics!), portfolio analysis, Fourier analysis, neurocomputing, and new urban economics, all of which could be argued to have rather limited applicability to spatial processes concerning human decision making.

Concomitant with this overreliance on techniques developed in other disciplines, there has been a relative dearth in the development of statistical techniques and mathematical models aimed at understanding purely spatial processes. Consequently, a great deal of criticism, much of it reasonable, some of it not, has been levelled at modelling techniques which have been applied to human spatial behaviour but which make assumptions such as homogeneity of individuals, spatial non-stationarity, omniscience, rationality, equilibrium and optimal behaviour. Such criticism has at times been very powerful and influential (Sayer, 1976; 1992) with the result that there has been a decline of research into spatial modelling. Even the frameworks provided by classic spatial models such as central place theory (Christaller, 1933) and Weber's location theory (Weber, 1909) have been discarded by many geographers because of the unreasonableness of some of the assumptions embedded in them.

The criticism of certain areas of quantitative geography is not necessarily a bad thing. Some areas of research will undoubtedly, with hindsight, be seen as evolutionary dead-ends and the discarding of certain methodologies or topics is an inevitable part of the evolution and maturity of a discipline or subdiscipline. Who
knows what parts of geographical enquiry, currently popular, will be seen as the academic equivalent of the dinosaurs in 20, or even 10, years’ time? However, there are times when the criticism of spatial modelling has had an unwarranted negative effect. For instance, some frameworks, such as spatial interaction modelling, still get criticized for what they once were rather than for what they now are. Sometimes the initial criticisms, albeit reasonable, have led to calls for the complete termination of research into a particular modelling topic rather than for more research to correct what have turned out to be fairly easily correctable flaws.

Unfortunately, this course of action appears to have been particularly prevalent in human geography. The dominant philosophy appears to be that most types of spatial modelling efforts are fatally flawed because they fail to account for the complex attitudes, preferences and tastes of individuals. These latter attributes are influenced not only by personal circumstances and characteristics, but also by the cultural, social and political milieu in which individuals make spatial decisions. A classic target of such criticism is the mathematical modelling of movements over space, more formally known as spatial interaction modelling or spatial choice modelling (Fotheringham and O’Kelly, 1989; Fotheringham 1991a; Sen and Smith, 1995). This is despite the fact that spatial interaction modelling is one of the most applied geographical techniques (see, for instance, the range of applications and references in Fotheringham and O’Kelly, 1989). It is used in studies of retail location, in impact assessment, in forecasting the demand of housing, in regional population projections, in forecasting travel demand within urban areas, and in a host of other applied areas. Consequently in the remainder of this chapter, we describe the evolution of spatial interaction models and their theoretical basis to demonstrate three issues:

1. That spatial modellers are aware of the need to make models more realistic in terms of human behaviour and spatial interaction models provide a classic example of this type of evolution.

2. The research frontier in spatial interaction modelling has progressed well beyond that perceived by those who still view spatial interaction models as ‘gravity models’ using outdated ideas from social physics. Spatial interaction/spatial choice models provide a salutary lesson to those who would have us discard such approaches in human geography on how one type of spatial modelling can evolve in terms of its theoretical basis to provide fascinating insights into human spatial behaviour.

3. How the research frontier in this area has been dynamic with profound shifts in the theoretical bases used to justify the modelling approach. Spatial interaction modelling provides a good example of the evolution of quantitative geography from the stage of being fuelled primarily by the importation of ideas from other disciplines to a stage in which it is the seedbed for new spatial theories.

Four distinct phases of spatial interaction modelling can be identified, each of which has provided a quantum leap in our understanding of spatial movements over
the previous era. In chronological order, these phases are: (a) spatial interaction as social physics; (b) spatial interaction as statistical mechanics; (c) spatial interaction as aspatial information processing; and (d) spatial interaction as spatial information processing. The approximate periods in which each of these theoretical bases dominated our understanding of spatial interaction are shown in Figure 9.1.

The major criticisms levelled at spatial interaction modelling stem from the first three phases when geography borrowed heavily from concepts developed in other disciplines. Indeed, there was a lingering suspicion, based on empirical evidence on the spatial pattern of local parameter estimates, that the basic formulation for spatial interaction models was a gross misspecification of reality. It is only with recent developments in understanding the basis of spatial interaction models from the perspective of spatial information processing that we have been able to identify and correct this misspecification. The identification of the misspecification bias in traditional spatial interaction models thus provides a strong link to the types of research described in Chapter 5 on local analysis. It was only through the calibration of local forms of spatial interaction models that the symptoms of model misspecification became evident: they were completely hidden in the output of the global models. We describe the nature of the evidence from local spatial interaction models in Section 9.5, but for a fuller discussion, the reader is referred to Fotheringham (1981; 1983b; 1984; 1991b).

9.2 Spatial interaction as social physics (1860–1970)

Attempts at understanding regularities in patterns of spatial flows began as early as the mid-nineteenth century with Carey’s (1858) observations, followed by Ravenstein’s (1885), that the movement of people between cities was analogous to the gravitational attraction between solid bodies. That is, greater numbers of migrants

![Figure 9.1 The changing theoretical basis of spatial interaction models](image-url)
were observed to move between larger cities than smaller ones, *ceteris paribus*, and between cities which were closer together than being farther apart, *ceteris paribus*. This led to the proposal of a simple mathematical model to predict migration flows between origins and destinations based on the Newtonian gravity model. This model has the following form:

$$ T_{ij} = k \frac{P_i P_j}{d_{ij}} \quad (9.1) $$

where $T_{ij}$ represents the number of trips between origin $i$ and destination $j$, $P_i$ and $P_j$ represent the sizes of origin $i$ and destination $j$, respectively (as measured, for example, by their populations), $d_{ij}$ represents the distance between $i$ and $j$, and $k$ is a scaling parameter relating the magnitude of $T_{ij}$ to the ratio $P_i P_j / d_{ij}$. The units of $k$ would be miles/person, for example, if distance were measured in miles and trips by the number of people going between $i$ and $j$.

Subsequently, it was recognized that the relationships embedded in the simple spatial interaction model shown in (9.1) might vary according to the type of movement being investigated as well as to the economic and social milieu in which the movement was taking place. For example, the deterrence of distance in trip-making behaviour is likely to be greater in less developed economies with relatively poor transportation facilities than in more advanced economies with relatively good transportation. Similarly, individuals are generally more deterred by distance when shopping for basic goods such as bread and potatoes than when shopping for luxury items such as jewellery or antiques. Hence the basic formulation in (9.1) was modified to allow for these variations in behaviour:

$$ T_{ij} = k \frac{P_i^\alpha P_j^\lambda}{d_{ij}^\beta} \quad (9.2) $$

where $\alpha$, $\lambda$, and $\beta$ are parameters to be estimated empirically and reflect the nature of the relationship between spatial flows and each of the explanatory variables.

A further refinement of the basic gravity model formula resulted from the recognition that many origin and destination attributes, rather than simply the two size variables, are likely to influence flow patterns so that the model could be written as

$$ T_{ij} = k \frac{V_i^{\alpha_1} V_j^{\alpha_2} \cdots V_i^{\alpha_f} V_j^{\lambda_1} V_j^{\lambda_2} \cdots V_j^{\lambda_g}}{d_{ij}^\beta} \quad (9.3) $$

where there are $f$ origin attributes, $V_i$, affecting the magnitude of the flows leaving $i$ and $g$ destination attributes, $V_j$, affecting the magnitude of flows entering $j$ (see
Haynes and Fotheringham (1984) for more details). The models in (9.2) and (9.3) can easily be made linear by taking logarithms of both sides of the equations which allows calibration by OLS regression with the caveats described by Fotheringham and O’Kelly (1989).

Whichever formula, (9.1), (9.2) or (9.3), is used to analyse a set of spatial flows, the underlying model framework is basically the same and has its origins as a social science analogy to a physical model of gravitational attraction, the so-called ‘social physics’ approach to model building. As such, it has been justifiably criticized for its lack of any theoretical grounding in the way individuals behave. However, even though this derivation of the model is theoretically empty, the model itself produces reasonably accurate estimates of spatial flows. Therefore, a great deal of effort has been expended on trying to develop an acceptable theoretical framework for the gravity model.¹

9.3 Spatial interaction as statistical mechanics (1970–80)

While various frameworks were espoused in attempts to justify the mathematical form of the gravity model (see for instance those of Dodd, 1950; Zipf, 1949; Huff, 1959; 1963; Nierderncorn and Bechdolt, 1969), the next major advance in providing a theoretical base came with the work of Wilson (1967; 1975). Wilson’s pioneering work, from his background in physics, produced what has become known as a ‘family of spatial interaction models’, each member of which can be derived from the principles of statistical mechanics (for further details, see Fotheringham and O’Kelly, 1989, Chapter 2).

Wilson considered a flow matrix, denoting the numbers of individuals moving between each origin–destination pair to be a ‘macrostate’ of the system. This macrostate results from the combination of many individual ‘microstates’ where a microstate in this context is defined as a description of the individuals moving between origins and destination. For instance, a macrostate description might be that five individuals moved from A to B and two moved from A to C; a microstate description would name the individuals. Clearly, there are many different microstates that could produce the same macrostate. If there are \(N\) individuals to be assigned to a set of categories and the number of individuals in category \(i\) is \(N_i\), then the number of microstates associated with any given macrostate is

\[
R = \frac{N!}{\prod_i N_i!}
\]  

(9.4)

For example, in the above situation where we assign five individuals from A to B and two individuals from A to C, there are \(7!/(5! \times 2!)=21\) possible ways of doing this. Different macrostates can have different numbers of microstates associated with them. For instance, if we assigned all seven individuals from A to
B and none from A to C, the number of microstates is only 1. If we assign six individuals from A to B and one from A to C, the number of microstates yielding this macrostate is 7, and if we assign four individuals from A to B and three from A to C, the number of microstates is 35. A full description of the possibilities is shown in Table 9.1.

Wilson formulated the problem of deriving a mathematical model of spatial flows in terms of choosing the particular macrostate which can be constructed from the largest number of microstates. This would be most likely to occur in the absence of any other information. In the above problem, this would be either four individuals going from A to B and three going from A to C or three going from A to B and four going from A to C. In a spatial flow context, the problem is then to select the set of $T_{ij}$ values which maximizes $R$ in Equation (9.4). Equivalently, one can maximize the natural logarithm of $R$ divided by $T$, the total number of trips in the system (this will not alter the optimization result but is mathematically more tractable). This can be written as

$$H \equiv (1/T) \ln R = (1/T) \left( \ln T! - \sum_{ij} \ln T_{ij}! \right) \quad (9.5)$$

where $\ln$ represents a natural logarithm. If all the $T_{ij}$ values are large, use can be of Stirling’s approximation that

$$\ln T! = T \ln T - T \quad (9.6)$$

and

<table>
<thead>
<tr>
<th>Table 9.1</th>
<th>An example of microstates and macrostates: allocate seven individuals between two flows (A → B; A → C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macrostate</td>
<td>Number of microstates</td>
</tr>
<tr>
<td>A → B</td>
<td>A → C</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>0</td>
<td>7</td>
</tr>
</tbody>
</table>
\[
\ln T_{ij}^* = T_{ij} \ln T_{ij} - T_{ij} 
\]

(9.7)

to obtain

\[
H = \left( \frac{1}{T} \right) \left( T \ln T - T - \sum_j T_{ij} \ln T_{ij} + T \right) 
\]

(9.8)

On rearranging, this can be written as

\[
H = - \sum_j (T_{ij}/T) \ln(T_{ij}/T) 
\]

(9.9)

or, equivalently, defining \( p_{ij} \), the proportion of all trips that originate at \( i \) and terminate at \( j \), as \( T_{ij}/T \),

\[
H = - \sum p_{ij} \ln p_{ij} 
\]

(9.10)

which is the formula for the entropy of a distribution (Shannon, 1948; Jaynes, 1957; Georgescu-Roegen, 1971) and which can be interpreted as a measure of the uncertainty about which microstate actually produces the observed macrostate.

However, it is perhaps obvious from the results in Table 9.1 that finding the set of trips which maximizes the entropy measure in Equation (9.9) or (9.10) is actually a trivial matter – \( H \) will be at a maximum when the \( T_{ij} \) values are all equal (or as near to equal as is possible). Wilson’s further contribution to the theoretical derivation of spatial interaction models was to add constraints to this maximization procedure. The constraints that have been imposed on the system are

\[
\sum_j T_{ij}^* \ln P_i = P_i 
\]

(9.11)

where \( T_{ij}^* \) represents the model prediction of \( T_{ij} \) and \( P_i \) is the population of origin \( i \);

\[
\sum_j T_{ij}^* \ln P_j = P_2 
\]

(9.12)

\[
\sum_j T_{ij}^* \ln d_{ij} = D 
\]

(9.13)
where $D$ is the total distance travelled in the system;

$$
\sum_{j} T_{ij}^* = K \tag{9.14}
$$

where $K$ is the known total interaction;

$$
\sum_{j} T_{ij}^* = O_i \quad \text{for all } i \tag{9.15}
$$

where $O_i$ is the known total flow from each origin; and

$$
\sum_{i} T_{ij}^* = D_j \quad \text{for all } j \tag{9.16}
$$

where $D_j$ is the known total inflow into each destination.

Imposing different combinations of these constraints on the maximization of Equation (9.9) produces different types of spatial interaction models – the so-called ‘family’ of spatial interaction models (Wilson, 1974; Fotheringham and O’Kelly, 1989). For instance, maximizing (9.9) subject to the constraints in (9.11) to (9.14) produces the gravity model formulation in Equation (9.2). Maximizing (9.9) subject to (9.12), (9.13) and (9.15) produces what is known as a production-constrained spatial interaction model:

$$
T_{ij} = O_i P_j d_{ij}^\beta \div \left( \sum_{j} P_j^\alpha d_{ij}^\beta \right) \tag{9.17}
$$

Maximizing (9.9) subject to (9.11), (9.13) and (9.16) produces what is known as an attraction-constrained model:

$$
T_{ij} = D_j P_i^\alpha d_{ij}^\beta \div \left( \sum_{i} P_i^\alpha d_{ij}^\beta \right) \tag{9.18}
$$

and maximizing (9.9) subject to (9.13), (9.15) and (9.16) produces a production–attraction-constrained or doubly constrained model:

$$
T_{ij} = A_i O_i B_j d_{ij}^\beta \tag{9.19}
$$
where

\[ A_i = \sum_j (B_j D_j d_{ij}^{\beta})^{-1} \]  \hspace{1cm} (9.20)

and

\[ B_j = \sum_i (A_i O_i d_{ij}^{\beta})^{-1} \]  \hspace{1cm} (9.21)

Examples of the application of these models and details on how to calibrate them are given in Fotheringham and O'Kelly (1989). Using a framework first recognized by Alonso (1978), Fotheringham and Dignan (1984) show how each of these four spatial interaction models can be considered as the extremal points on a continuum of models defined in terms of how strictly the various constraint equations are enforced.

The importance of Wilson's entropy-maximizing derivation of a family of spatial interaction models was that it provided a theoretical justification for what had been until that time only an empirical observation. However, there are several criticisms which can be levelled at the derivation. First, it simply replaces one physical analogy — that of gravitational attraction — with another — that of statistical mechanics. The derivation is still sterile in terms of the processes by which individuals make spatial decisions. Secondly, while some of the constraint equations have a behavioural interpretation, others, such as those on the populations, are more difficult to justify except that they lead to a particular model form. Thirdly, the use of Stirling's approximation to derive the entropy formulation is highly suspect in most situations because of its assumption that all the \( T_{ij} \) values are large. When the \( T_{ij} \) values are small, the approximation is rather poor.

The latter criticism has been addressed within this framework by Webber (1975; 1977), Tribus (1969) and Dowson and Wragg (1973), amongst others, who eliminate the need to derive the entropy formulation from a discussion of microstates and macrostates. They argue that the entropy formulation satisfies certain reasonable requirements of a measure of statistical uncertainty and therefore they simply state this as a starting point for the derivation of mathematical models, the argument being that we should maximize our uncertainty about any outcome subject to constraints. To do otherwise would be adding bias into the model-building procedure. Despite this argument, however, there are still considerable difficulties with entropy maximization as a framework for the development of models of human spatial behaviour. Consequently, despite stimulating large amounts of research into spatial interaction modelling, the Wilson framework has been largely discarded in favour of more behavioural approaches.
9.4 Spatial interaction as aspatial information processing (1980–90)

The major advance in giving spatial interaction models a theoretical foundation that was based on human behaviour and information processing was the recognition that models such as those in (9.17) and (9.18) are ‘share’ models and have a logit formulation. The model in (9.17), for example, which is the most widespread form of spatial interaction model in use today, allocates shares of the number of people leaving an origin, $O_i$, amongst the destinations according to the attributes of these destinations. The derivation of this model could then be based on the derivation of the discrete choice model by McFadden (1974; 1978; 1980).

Consider an individual at origin $i$ about to make a choice of a single destination $j$ from a set of possible destinations. Suppose the individual evaluates every alternative and then selects the one yielding him or her maximum benefit or utility. The utility accruing from a person at $i$ selecting alternative $j$, $U_{ij}$, can be thought of as being composed of two parts, a measurable component, $V_{ij}$, and an unmeasurable component, $\mu_{ij}$. That is,

$$U_{ij} = V_{ij} + \mu_{ij}$$  \hspace{1cm} (9.22)

If we knew the values of $U_{ij}$ for every alternative we could say with certainty which destination an individual would choose. That is,

$$P_{ik} = \begin{cases} 
1 & \text{if } U_{ik} > U_{ij} \text{ (for all } j \in N, j \neq k) \\
0 & \text{otherwise} 
\end{cases}$$  \hspace{1cm} (9.23)

where $k$ is a specific alternative from the set of $N$ alternatives. However, given that each alternative has an unknown component to its utility, we cannot say for certain which alternative an individual will choose. We can only evaluate each alternative based on the measurable component and then state a probability that a destination will be selected based on a comparison of the observable components of the utility functions. That is,

$$p_{ik} = \text{prob}[U_{ik} > U_{ij} \text{ (for all } j \in N, j \neq k)]$$  \hspace{1cm} (9.24)

which, on substituting (9.22) and rearranging, can be written as

$$p_{ik} = \text{prob}[\mu_{ij} < V_{ik} - V_{ij} + \mu_{ik} \text{ (for all } j \in N, j \neq k)]$$  \hspace{1cm} (9.25)

Recognizing that the $\mu_{ik}$ and the $\mu_{ij}$ terms are random elements drawn from a
continuous distribution ranging from \(-\infty\) to \(+\infty\), Equation (9.25) can be written as

\[
p_{ik} = \int_{x=-\infty}^{+\infty} g(\mu_{ik} = x) \prod_{j \neq k} \int_{y=-\infty}^{V_{ij} - V_{ij} + x} g(\mu_{ij} = y) \, dy \, dx \tag{9.26}
\]

where \(g()\) represents some probability density function. McFadden (1974) demonstrated that if the \(\mu\) terms are distributed according to a Type I extreme value distribution (Fisher and Tippett, 1928), the model which results is

\[
p_{ik} = \exp(V_{ik}) / \sum_j \exp(V_{ij}) \tag{9.27}
\]

and which has a behaviourally intuitive logistic response rate between the probability of choosing a particular alternative and its observed utility (see Fotheringham and O’Kelly (1989, p. 74) for a demonstration of this).

The expression in (9.27) has the same basic form as the spatial interaction models in (9.17) and (9.18). To see this more clearly, the relationship can be made more explicit by defining the observable utility associated with each destination as a function of its attributes. For example, consider the case of retail choice in which an individual selects a supermarket from a set of supermarkets in an urban area. The individual is likely to gain more benefit from selecting a large store as opposed to a small store in terms of having a greater variety of produce to choose from and also possibly cheaper prices. Similarly, the individual will gain more benefit, in terms of cost and time savings, from selecting a store in close proximity rather than far away. Obviously, there are many other attributes that can be added to an individual’s utility function but these two will serve our purpose in demonstrating the behavioural relationships embedded in the model derivation at this stage.

Consider first the relationship between \(V_{ik}\) and \(S_k\), the size of the store. Two possibilities are shown in Figure 9.2: a linear relationship and a logarithmic relationship. The linear relationship might at first seem the more reasonable but on reflection the behaviour it represents is rather implausible. It implies that an additional 1000 square metres of retail space to an existing store would generate the same added utility regardless of the original size of the store. That is, the additional utility from the extension is predicated to be the same whether the original store is 2000 square metres or 10000 square metres. An alternative and more plausible behavioural relationship is that provided by the logarithmic function which implies that the addition to the size of the store generates less benefit to the consumer when the store is already very large than when the store is small. Consumers are unlikely to reap much benefit from an addition when the store is already very large. Hence, the relationship between \(V_{ik}\) and \(S_k\) can be expressed as
Figure 9.2  The relationship between observed utility and size

\[ V_{ik} \propto \alpha \ln S_k \]  \hspace{1cm} (9.28)

where \( \alpha \) represents the particular shape of the logarithmic relationship.

For a similar behavioural reason, the relationship between \( V_{ik} \) and \( d_{ik} \), the distance to alternative \( k \) for an individual at \( i \), is also likely to be logarithmic as depicted in Figure 9.3. The addition of 10 kilometres to a journey which is already 500 kilometres is likely to affect one's assessment of the journey less than the addition of 10 kilometres to a journey which is only 5 kilometres. Hence, the appropriate relationship is given by

\[ V_{ik} \propto \beta \ln d_{ik} \]  \hspace{1cm} (9.29)

Figure 9.3  The relationship between observed utility and distance
where $\beta$ is a distance-decay parameter to be estimated. Substituting the relationships described by Equations (9.28) and (9.29) into (9.27) yields

$$p_{ik} = \frac{S_k a d_{ik}^\beta}{\sum_j S_j a d_{ij}^\beta} \quad (9.30)$$

which is the form of production-constrained spatial interaction model given in (9.17).

It should be noted that this form of model has two well-known properties which, in the context of spatial choice, are undesirable. The first is termed the independence from irrelevant alternatives (IIA) property and stated simply is that the ratio of the probabilities of an individual selecting two alternatives is unaffected by the addition of a third alternative. This is easy to see: in Equation (9.30) consider two alternatives, $j$ and $k$. As the denominator in (9.30) is a constant, the ratio of the probabilities of choosing $k$ over $j$ is

$$\frac{p_{ik}}{p_{ij}} = \frac{S_k a d_{ik}^\beta}{S_j a d_{ij}^\beta} \quad (9.31)$$

which is indeed independent of any other alternative. In a spatial choice context this property implies that the model's predicted flow to any destination is unaffected by the location of the other alternatives. That is, the ratio of the probabilities of choosing two stores is unaffected by the fact that one store might be surrounded by competitors and the other store is in relative isolation. This seems unrealistic in terms of most spatial choice contexts where the location of an alternative with regard to its competitors would seem to be important (Fotheringham, 1989).

The other undesirable property of the model in (9.30) is what Huber et al. (1982) term 'regularity'. That is, it is impossible for the model to predict an increase in the flow to any existing alternative due to the addition of a new alternative. Again, in many spatial choice situations, this is unnecessarily restrictive and unrealistic. Consider, for example, clothing stores where agglomeration effects are present (one of the reasons why shopping malls exist). The location of a new store can increase the sales at nearby stores.

More recent developments in the theoretical derivation of spatial interaction models have used spatial choice principles, which avoid both of the above problems. These developments and the resulting models are now discussed.

### 9.5 Spatial interaction as spatial information processing (1990 onwards)

The derivation of spatial interaction models described above in the framework of discrete choice represents an advance over the previous physical analogies to gravitational attraction and entropy. However, it is a framework which has also
been 'borrowed' from economics, another predominantly aspatial discipline. The
discrete choice framework was developed for aspatial contexts such as brand choice
and choice of transportation mode. As such, it contains an assumption that, while
tenable in the context of most aspatial choice, is untenable when applied to most
spatial choice problems. In Equation (9.24), and subsequently, the assumption is
made that an individual is able to evaluate all alternatives. That is, alternative $k$
is compared with all the alternatives in the full choice set $N$. In essence, the
individual is assumed to be omniscient and able to process vast amounts of
information. In simple choice situations where $N$, the number of alternatives, is
small, such as in most brand choice and mode choice situations, it is reasonable to
assume that individuals can process information on all the alternatives. However, it
is generally recognized that individuals have a limited capacity for processing
information (inter alia, Simon, 1969; Lindsay and Norman, 1972; Newell and
Simon, 1972; Norman and Bobrow, 1975) and indeed this limit might be reached
very quickly. Bettman (1979), for example, argues that our limit to information
processing might be reached with as few as six or seven alternatives. Consequently,
in most spatial choice situations the number of alternatives is far too large to
assume individuals can evaluate all possible alternatives. Consider, for example,
buying a place to live in a large metropolitan area where there might be thousands
of properties for sale, or a migrant choosing between cities within a country, or
even a consumer selecting a store within an urban area. In such situations the
number of alternatives is generally very large, far too large in fact for us to be able
to retain the assumption that individuals evaluate every alternative.

The manifestation of this untenable assumption and the subsequent misspecification
of the traditional spatial interaction model formulation only comes to light
when local forms of spatial interaction model are calibrated (see Chapter 5).
Typically, local spatial interaction models are calibrated separately for each origin
in the system (although they can also be calibrated separately for each destination).
In so doing, a set of parameters, one for each origin, describes the relationship
between spatial movement and an attribute of the destinations. These sets of origin-
specific parameter estimates replace the single parameter estimates obtained in the
calibration of a global model. An indication of a problem with the classical spatial
interaction model formulation was uncovered when origin-specific estimates of the
distance-decay parameter were mapped. The maps showed a consistent and puzzling
spatial pattern in which peripheral origins generally had more negative
distance-decay parameter estimates than more central ones (see Fotheringham
(1981) for examples of this pattern and a summary of the discussions, sometimes
intense, that such patterns evoked).

An example of the trend in origin-specific distance-decay parameter estimates
commonly found is shown in Figure 9.4. This describes the relationship between
origin-specific distance-decay parameter estimates and a measure of the centrality
of each origin. The parameters are derived from a production-constrained model
(such as that shown in Equation (9.17)) applied to 1970–80 migration flows
between the 48 contiguous states of the USA. The destination attributes in the
model are: population size; distance from each origin; mean personal income; average temperature; unemployment rate; and median house price. The pattern is quite striking in that it appears to indicate that distance-decay rates are less negative for centrally located states than for peripherally located states. Behavioural explanations for the types of pattern exhibited in Figure 9.4, such as people in more central origins being more spatially mobile, have not been convincing (see Fotheringham, 1981) and the search turned to a theoretical explanation to this puzzle. The answer came in understanding the type of information processing which is likely to take place in spatial choice, which we now describe.

The generic spatial choice problem can be stated as: 'How does an individual at location \(i\) make a selection from a set of \(N\) spatial alternatives?' This type of spatial choice must precede most types of spatial interaction, whether it be shopping, migration, house selection which determines commuting patterns, vacationing, or any other of a myriad types of movement over space. The spatial choice process as shown in Figure 9.5 has three characteristics (Haynes and Fotheringham, 1990):

1. It is a discrete, rather than a continuous, process. That is, either a destination is selected or it is not and there is a finite number of alternatives.
2. The number of alternatives is generally large, and in some cases, very large.
3. The alternatives have fixed spatial locations, which limits the degree to which alternatives are substitutes for one another. It also means that, unless the spatial distribution of alternatives is perfectly regular, each alternative faces a unique spatial distribution of competing alternatives.
The first property suggests that the discrete choice framework introduced from aspatial choice, and described above in Section 9.4, provides a basis for understanding spatial choice. The second and third properties suggest the framework will need some modification as these two properties are not usually found in aspatial choice.

With the spatial choice framework in mind, and the realization that there are limits to individuals’ abilities to process information, Fotheringham (1983a; 1983b; 1984; 1986; 1991b) develops a new form of spatial interaction model, termed a competing destinations model. This model results from the starting point of how individuals process spatial information and its derivation does not depend on the assumption that individuals are able to evaluate every alternative.

The derivation of the competing destinations model closely follows that of the logit discrete choice model described above with the added flexibility of allowing individuals to make choices from restricted choice sets rather than from the complete set of alternatives. This is achieved by replacing Equation (9.24) with

\[ p_{ik} = \text{prob}[U_{ik} > U_{ij} + \ln p_i(j \in M) \text{ (for all } j \in N, j \neq k)]p_i(k \in M) \quad (9.32) \]

where \( M \) is the restricted choice set in which an individual at \( i \) evaluates all the alternatives and \( N \) is the full set of alternatives. Alternatives that are not within \( M \) are not evaluated and therefore cannot be chosen by the individual. For example, suppose an individual looking for a house within a city only looks at housing in the south-east quadrant of the city. Houses in other parts of the city are never considered and so are not evaluated and therefore will not be chosen, even if they have all the attributes the individual seeks. To see how Equation (9.32) represents this type of behaviour, consider two situations. First, if the alternative \( k \) is not in the restricted choice set \( M \), \( p_i(k \in M) \) will be zero and hence \( p_{ik} = 0 \), no matter how
large $U_k$ is. Secondly, if $j$ is not in the restricted choice set, $p_i(j \in M) = 0$ and $\ln p_i(j \in M) = -\infty$ so that $U_j + \ln p_i(j \in M) = -\infty$ and hence $k$ will be selected in preference to $j$ despite $j$'s attributes. Thus, Equation (9.31) allows sub-optimal choices to be made: individuals sometimes select alternatives which do not yield maximum utility because they cannot evaluate every alternative.

Substituting Equation (9.32) for (9.24) into the framework described in Section 9.4 yields the following general spatial choice model:

$$p_{ik} = p_i(k \in M) \int_{x=-\infty}^{+\infty} g(\mu_{ik} = x) \prod_{j,j \neq k}^{n} \int_{y=-\infty}^{V_k - V_0 + x - \ln p_i(j \in M)} g(\mu_{ij} = y) \, dy \, dx \quad (9.33)$$

(Fotheringham, 1988). Under the assumption that the $\mu$ terms are independently and identically distributed with a Type I extreme value distribution, the resulting spatial choice model is

$$p_{ik} = \exp(V_{ik}) \times p_i(k \in M) \sum_j \exp(V_{ij}) \times p_i(j \in M) \quad (9.34)$$

Essentially this is a logit model where each alternative’s observable utility is weighted by the probability of the alternative being evaluated. Three specific models can be derived from this general form depending on the definition of $p_i(j \in M)$:

1. If $p_i(j \in M) = 1$ for all $j$, then Equation (9.34) is equivalent to (9.27). That is, if all the alternatives are evaluated by individuals, then the model in (9.34) degenerates to the classic logit model.

2. If $p_i(j \in M) = 1$ for all $j \in M$ and 0 otherwise, the model is equivalent to a logit model applied to only those alternatives in the restricted choice set. Obviously to operationalize such a model it would be necessary to know in advance what the restricted choice set was for each individual. Generally such information is not known a priori. Also, the situation is complicated by the likelihood that individuals in different locations have different mental images of space and different cognitive constructions of clusters of spatial alternatives (Gould and White, 1974).

3. If $p_i(j \in M)$ is a function of the location of $j$ with respect to the other alternatives, then the model in (9.34) holds and the task is to determine how to define $p_i(j \in M)$. We now pursue this line of enquiry in order to derive the competing destinations spatial interaction model (Fotheringham, 1983b; 1986).

By defining a likelihood as a probability divided by a constant, the model in (9.34) can be rewritten as
\[ p_{ik} = \exp(V_{ik}) \times l_i(k \in M) \bigg/ \sum_j \exp(V_{ij}) \times l_i(j \in M) \]  

(9.35)

where \( l_i(j \in M) \) represents the likelihood that alternative \( j \) is in the restricted choice of alternatives which are evaluated by an individual at location \( i \). A great deal of research has been undertaken into how to define this likelihood function, or its equivalent (inter alia, Meyer, 1979; Batsell, 1981; Meyer and Eagle, 1982; Fotheringham, 1983b; 1991b; Fik and Mulligan, 1990; Lo, 1991; Fik et al., 1992; Thill, 1992).

It can be argued that there are two distinct processes represented in Equation (9.35): one is how individuals evaluate an alternative and the other is whether it is evaluated (see Figure 9.6). Associated with these two processes are three types of attributes:

1. The attributes of an alternative that affect how individuals evaluate it. These are already incorporated in the model through the expression \( \exp(V_{ij}) \), the exponential of the observable utility component.

2. The attributes of an alternative that affect both how individuals evaluate it and whether they evaluate it. Again, such attributes are already included in the spatial choice framework through the expression \( \exp(V_{ij}) \).

3. The attributes of an alternative that affect only whether an alternative is evaluated. It is these attributes which need to be incorporated into the

![Figure 9.6 The spatial choice process](image-url)
modelling framework through the definition of the likelihood of an alternative being evaluated, \( l_j (j \in M) \). The definition of these attributes will lead to new and improved forms of spatial choice/spatial interaction models. We now describe the rationale behind the definition of one of these attributes which describes the location of an alternative vis-à-vis all other alternatives and which leads to a model known as the competing destinations model (Fotheringham, 1983b; 1986).

Assume that, because of limits on our ability to process information, individuals do not evaluate all the alternatives available in a typical spatial choice context. Instead, they process spatial information and make spatial choices in such a way that they first evaluate clusters of alternatives and only evaluate particular alternatives from within a selected cluster. For instance, a migrant might first select a region in which he or she would like to live and then evaluate alternatives only within that region. Equivalently, an individual looking at housing in a city might have strong opinions about which parts of the city he or she would like to live in and which parts are to be avoided. Support for the hypothesis of hierarchical information processing is given by, *inter alia*, McNamara (1986; 1992), Walker and Calzonetti (1989), Hirtle and Jonides (1985), Stevens and Coupe (1978), Curtis and Fotheringham (1995) and Fotheringham and Curtis (1992; 1999).

Unfortunately for the modeller it is very difficult, if not impossible, to know how people mentally cluster alternatives in space. Individuals themselves might not be able to tell you how they process spatial information and the mental clusters they form might well be fuzzy (Zadeh, 1965; Pipkin, 1978) rather than discrete. Certainly, there is evidence to suggest that such clusters will vary according to one's location in space (Gould and White, 1974; Downs and Stea, 1977; Curtis and Fotheringham, 1995; Fotheringham and Curtis, 1999). However, Fotheringham (1991b) and Pellegrini and Fotheringham (1999) argue that it is not necessary to define the clusters of alternatives that individuals evaluate. It is only necessary to assume that some form of hierarchical spatial information processing takes place whereby clusters of alternatives are first evaluated and then only alternatives within a selected cluster are evaluated. The task then becomes one of identifying any destination attributes associated with the hierarchical processing of spatial information that affect the likelihood of an alternative being evaluated.

One such variable, identified by Fotheringham (1983b; 1986), is the proximity of an alternative to all other possible alternatives. There are many ways such a variable could be measured, but a common measurement is the accessibility of an alternative \( k \) to all other alternatives, \( A_k \), as defined by

\[
A_k = \sum_{j \neq k} \frac{P_j}{d_{jk}}^\alpha / d_{jk}^\beta \tag{9.36}
\]

where \( P_j \) represents the population of alternative \( j \) and \( d_{jk} \) is the distance between
alternatives \( j \) and \( k \). This formulation measures the ‘competition’ faced by alternative \( k \) from all the other alternatives: when an alternative is centrally located near many other alternatives, \( A_k \) will be large; when it is relatively isolated, \( A_k \) will be small. In practice, the values of \( \alpha \) and \( \beta \) are often set to 1 and \(-1\), respectively, although they could be estimated either in the model calibration procedure or externally.

The justification for the inclusion of this competition variable is based on ‘the psychophysical law’ that individuals tend mentally to underestimate the size of large objects (Stevens, 1957; 1975). Consider an individual evaluating spatial clusters of alternatives, such as cities within a country. One of the factors in the evaluation of each cluster is likely to be its size, reflecting the number of opportunities for employment etc. However, the psychophysical law suggests that the cognized size of a large cluster is likely to be smaller than its objective size and that the magnitude of this mental underrepresentation will increase with the size of cluster, resulting in the logarithmic relationship depicted in Figure 9.7. The linear relationship between the cognized and objective size of clusters would only occur if individuals did not underrepresent large clusters mentally. The probability of an individual selecting a large cluster of alternatives, and consequently evaluating the individual alternatives within that cluster, will obviously be less if the relationship is logarithmic than if it were linear. Hence, the likelihood of a spatial alternative being evaluated is a function of the location of an alternative with respect to all the other alternatives; alternatives in large clusters are less likely to be evaluated than those in small clusters, \textit{ceteris paribus}. Rather than having to identify the exact nature of the clusters cognized by individuals, all we therefore need to define \( l_j( j \in M) \) is the measure of destination competition given in Equation (9.36). That is,

\[
l_j( j \in M) = A_j^\beta \tag{9.37}
\]

![Figure 9.7](image_url) Relationships between cognized and objective cluster sizes
where \( \delta \) is a parameter to be estimated. If \( \delta = 0 \), all destinations are equally likely to be evaluated (the aspatial logit assumption). Given the above theoretical reasoning, we would expect that \( \delta \) will be negative in most spatial choice situations; the more centrally located an alternative, the more likely it is to be in a large cluster cognized by an individual and the less likely it is to be evaluated, \textit{ceteris paribus}. However, for certain types of spatial choice, such as non-grocery shopping, there could well be substantive agglomeration effects so that individuals are attracted to large clusters of alternatives in order to minimize the costs of comparison shopping. In such instances, \( \delta \) will be positive.

Substituting (9.37) into (9.35) yields the spatial choice model,

\[
p_{ik} = \frac{\exp(V_{ik}) \times A_k^\delta}{\sum_j \exp(V_{ij}) \times A_j^\delta}
\]

which Fotheringham (1983; 1986) calls the ‘competing destinations model’. A large number of empirical studies have demonstrated the superiority of this model over the aspatial logit formulation in modelling spatial choice (\textit{inter alia}, Fotheringham 1984; 1986; Fotheringham and O’Kelly, 1989; Boyle et al., 1998; Atkins and Fotheringham, 1999; Pellegrini and Fotheringham, 1999). Typically, it produces demonstrably less biased parameter estimates and significant improvements in goodness of fit.

As an example, consider the pattern of origin-specific distance-decay parameters described above in Figure 9.4. This pattern was described as evidence of a misspecification problem with the model because there appears to be a regular change over space in the effect of distance on spatial interaction which is difficult to account for in behavioural terms. When a model having the form of (9.38) is calibrated with exactly the same destination attributes as those used to generate the results in Figure 9.4, the spatial pattern of the origin-specific distance-decay parameters disappears as shown in Figure 9.8, drawn to the same scale. This suggests strongly that the distance-decay parameters generated by standard logit forms of spatial interaction models contain a severe misspecification bias and that this bias is eliminated in the calibration of the competing destinations model. Fotheringham (1986) demonstrates theoretically the nature of the misspecification bias in the parameter estimates obtained from the aspatial logit model when the relationship in (9.38) holds.

By incorporating the appropriate spatial weight on each observable utility function, both of the undesirable properties identified with the aspatial logit model in the previous section, the IIA property and regularity, are also removed in the competing destinations model. The removal of the former can be seen by taking the ratio of the probabilities of selecting two alternatives, \( j \) and \( k \), from the model in (9.38), which is

\[
p_{ik}/p_{ij} = \frac{\exp(V_{ik}) \times A_k^\delta}{\exp(V_{ij}) \times A_j^\delta}
\]
and which is unlikely to be constant under the addition of a new alternative to the system. The addition of a new spatial alternative will have differential effects on the measurement of $A_k$ and $A_j$ depending on the location of the new alternative vis-à-vis the locations of $k$ and $j$. Similarly, the regularity property is removed because within the competing destinations framework, it is possible that the addition of a new spatial alternative will result in an increase in the probability of selecting an existing alternative. This would happen, for instance, in a shopping context in which $\delta$ were strongly positive; then, the addition of a new store in close proximity to an existing one could result in an increase in the predicted patronage of the existing store.

### 9.6 Summary

The development of spatial interaction models highlights the progress made over the last few decades in this area of quantitative geography. We have witnessed the progression from models whose only justification was an empirical regularity and an analogy to gravitational attraction; to models derived from either non-behavioural or aspatial theories imported from other disciplines; and finally to models based on principles of spatial information processing, sub-optimality, hierarchical decision making and spatial cognition. Current research lies in consolidating and improving this latest framework.

However, despite the progress in making spatial interaction models more behaviourally based, it is probably the case that many geographers still associate spatial interaction modelling with its early social physics background (as demon-
strated by the lingering usage of the term 'gravity model'). This is unfortunate for two reasons. The first is that, despite its very widespread application to many facets of the real world, these geographers ignore or even dismiss spatial interaction modelling, not because of what it is but because of what it was 20 or 30 years ago. The second, and more important, is that spatial interaction modelling provides a very fertile area for understanding spatial behaviour and for developing theories which are explicitly spatial. It is an area that is quintessentially geographical; it is an area where geographers should be leading the way by exporting their ideas to other disciplines.

**Notes**

1. Additionally, efforts in obtaining data for trip models have led to other spinoffs. For example, in the UK the first file linking postcodes to grid references arose out of government-sponsored trip modelling efforts (Raper et al., 1992, p. 43).
2. Centrality here is measured with respect to overall population so that in this US example, the states with the highest centrality index are those in the north-east and the states with the lowest centrality index are those in the west.