Papers Reg. Sci. 83, 339–361 (2004)
DOI: 10.1007/s10110-003-0189-4

Papers in
Regional
Science

Spatial interaction modelling

John R. Roy[1], **Jean-Claude Thill**[2]

[1] ETUDES, P.O. Box 96, Mallacoota, Vic. 3892, Australia (e-mail: johnmall@bigpond.com)
[2] Department of Geography, University at Buffalo, The State University of New York, Wilkeson Quad, Amherst, NY 14261, USA (e-mail: jcthill@buffalo.edu)

Abstract. Spatial interaction (SI) is the process whereby entities at different points in physical space make contacts, demand/supply decisions or locational choices. The entities can be individuals or firms and the choices can include housing, jobs, production quantities, exports, imports, face-to-face contacts, schools, retail centres and activity centres. The first SI models can be grouped under the generic heading gravity models. Their main characteristic is that they model the behaviour of demand or supply segments, rather than that of individuals and firms. This article traces the development of these models from their inception in the early part of the twentieth century to the present. The key advances include the replacement of the gravity analogy by the more general concepts of entropy or information theory, a statistical framework commonly used in physics. With the arrival of the regional science paradigm over 50 years ago, a key challenge has been to broaden these models compared to those arising in spatial economics, thus arriving at a more inclusive probabilistic framework. These efforts are discussed here, as well as inclusion of geographical advances, embracing activities as generators of travel, time-geography, recognition of spatial interdependencies, and use of neuro-computing principles.

JEL classification: C190, C450, C610, R00

Key words: Spatial interaction, gravity models, entropy, probabilistic, spatial structure, neural networks

1 Early developments

At the outset, let us state clearly that a single review article on the large number of contributions to spatial interaction (SI) modelling is inevitably selective. Readers can obtain further insights on the evolution of this field from the introductory portion of Sen and Smith (1995), the review chapter by Batten and Boyce (1986), the commentary in Roy (1990) and the first three chapters in Roy (2004).

Isaac Newton's Theory of Universal Gravitation reigned unchallenged in the latter part of the nineteenth century, before the appearance of Einstein's radical theories. If d_{ij} is defined as the distance between bodies i and j, m_i and m_j are the masses of bodies i and j respectively, and k is a constant, then Newton's theory yields the gravitational force F_{ij} between bodies i and j as

$$F_{ij} = k m_i m_j (d_{ij})^{-2} \tag{1}$$

Thus, it is hardly surprising that suggestions arose for using Newton's theory to explain the patterns in certain types of human activity between entities physically separated in space. In particular, during the mid-1850s the first applications of this theory to aggregate structures of movement and communication resulting from a behavioural decision process were undertaken, to either make contacts, demand/supply decisions, or locational choices; in short, spatial interaction occurred. In these pioneering applications, the gravitational force was replaced by the intensity of the interaction between the two areas, expressed as the number of trips or moves between the areas. The masses needed to be defined according to the type of activities being modelled. Not until the work of Stewart (1941) were the masses specified as the populations of the origin regions and destination regions of the movers, albeit whilst retaining Newton's squared distance effect. A more empirical determination of a "gravity parameter" awaited the efforts of retail modellers in the 1960s.

That one of the first texts recommending the use of gravity models in regional science was in the chapter 'Gravity, potential and spatial interaction models' of Isard (1960) is highly appropriate. Then, with the growth of large regional shopping centres in advanced economies in the 1960s and beyond, competition among the multiple urban regional centres, as well as between the regional centres and the Central Business District (CBD), needed to be accounted for. Huff (1963) rose to the challenge by developing a probabilistic retail model. His model evaluated the choices of alternative shopping centres by sets of shoppers, which in the end determine their viability and the (overlapping) trading areas associated with each centre. As with random utility theory, the probabilistic structure recognises that a typical shopper in a large urban area does not shop exclusively at one centre, but chooses among alternative centres. More importantly, Huff moved away from the rigidity of the Newtonian model by: (i) treating travel time rather than distance as the separation component, and (ii) allowing the gravity exponent to be calibrated from observations, rather than being set a priori as -2. Letting p_{ij} be the probability of a shopper in origin zone i shopping at centre j, t_{ij} the travel time between i and j and $(-\lambda)$ the gravity exponent, Huff's model is given as:

$$p_{ij} = W_j t_{ij}^{-\lambda} / \left[\sum_k W_k t_{ik}^{-\lambda} \right], \tag{2}$$

where $\sum_k$ denotes the sum over all the competing centres k. In general, W_j denotes a measure of the relative attractiveness of destination j, here taken as the floorspace.

Finally, if O_i denotes the number of customers (trip origins) in zone i, then the number of trips T_{ij} between i and j during a specified interval is simply given as:

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

From this pattern of shopping trips and assumptions on expenditure per trip, it is possible to evaluate expected turnover levels, parking requirements, and customer catchment contours for each centre. In practice, the households can be classified into different income groups and the retail goods into categories such as convenience goods and comparison goods, with a separate calibration being performed for each income group/good combination. Note that, by recognition of the gravity index $(-\lambda)$ as a parameter to be calibrated rather than an assumed constant, the Huff model must supplement Census data by surveys on the choice of the different centres by customers from the different origin zones. With the commercial use of Geographic Information Systems (GIS), it has established itself as one of the most widely used market forecasting approaches in business geographics (Birkin et al. 1996).

A further interesting analysis from the 1960s was Lakshmanan and Hansen (1966). In particular, they examine the long-term stability of the location and facility size distribution. For each class of goods, it assesses levels of turnover per square foot of floorspace as profitability thresholds, plotting these levels for the major centres in the Baltimore region. In projecting likely future patterns of facilities, it gives greatest credence to those having relatively small variations in the expected turnover per unit area. In fact, this work anticipates heuristically the more formal locational equilibrium framework introduced in the path-breaking paper of Harris and Wilson (1978).

An early alternative approach is the "intervening opportunities" model, first developed by Stouffer (1940) and refined for the journey-to-work by Schneider (1959) for the Chicago Area Transportation Study. Thus, for each origin zone i, we rank possible destination zones j as j_{im} in order $\{m\}$ of their proximity to zone i, where the 'closest' zone to i is denoted by $m = 1$. Let T_{ijm} be the number of trips from zone i stopping at the mth ordered destination j out from i. Also, let O_i be the number of commuters available at zone i. Then, approximating difference equations by a differential equation, T_{ijm} can be shown to be given as:

$$T_{ijm} = k_i O_i [\exp -(pW_{ijm-1}) - \exp -(pW_{ijm})], \tag{4}$$

in which W_{ijm} is the total number of jobs "passed" up to and including j_{im}. This model exceeds the basic gravity concept in attempting account for the spatial 'structure' of the opportunities, anticipating the more formal work on spatial structure spearheaded by Fotheringham (1983) and discussed later. A hybrid spatial interaction/intervening opportunities model was proposed by Long and Uris (1971). However, this and earlier incarnations of the concept of intervening opportunities have two key limitations: (i) their inability to satisfy the usual destination constraints:

$$\sum_{im} T_{ijm} = D_j, \tag{5}$$

where D_j is the number of jobs in zone j, and (ii) the fact that destinations can be distributed over 360 degrees surrounding the origin zone i, implying that the opportunities in successive destinations at increasing distance out from i may not be truly "intervening" (Guldmann 1999). This can be countered by introducing an extra route index along which opportunities directly "intervene" (Roy 1999).

Some of the earliest work on modelling trade flows was performed by Heckscher (1919) and his student Ohlin (1933). For any given commodity flowing between any two regions or countries, let O_i be the production of the commodity in origin region i, D_j the consumption of the commodity in destination region j and $O(\equiv D)$ the total production/consumption) of the commodity over all regions. Then, the Heckscher-Ohlin (H-O) model gives the trade flows T_{ij} between any two regions i and j via the simple bi-proportional relation:

$$T_{ij} = O_i D_j / O. \tag{6}$$

If in the basic gravity equation (1) the masses are taken as O_i and D_j and the constant k as $(1/O)$, then the H-O model emerges just for the special case when the gravity exponent is zero, implying that trade flows are independent of the distances through which the goods are shipped. They are merely jointly proportional to the production of the commodity at the origins and its consumption at the destinations. As motivated by Deardorff (1995), a natural option is to turn to a gravity model in terms of the transaction costs d_{ij} between the regions. This would yield (6) in the form:

$$T_{ij} = O \left\{ O_i D_j d_{ij}^{-n} / \left[\sum_{ij} O_i D_j d_{ij}^{-n} \right] \right\}, \tag{7}$$

reducing to (6) when the gravity index n is zero. The denominator ensures satisfaction of the total flow conservation condition $\sum_{ij} T_{ij} = O$. Although this simple gravity model was used for some early studies following the lead of Tinbergen (1962) and Linnemann (1966), it was soon realised that the product of production potential, consumption potential and gravity influences did not explain some trade patterns well. For instance, many countries have historical trade patterns based on cultural affinities and perceptions of quality and reliability of delivery. This problem was addressed by Theil (1967), who multiplied T_{ij} in (6) by a bias factor Q_{ij} defined as:

$$Q_{rs} = (\overline{T}_{ij}\overline{O})/(\overline{T}_i\overline{D}_j), \tag{8}$$

where the bars refer to the corresponding quantities as observed at the base period. It is clear that such a model will perform well when mainly quantity changes occur between the base period and the projection period. However, when one is trying to determine the influence of expected changes in transport costs or import tariffs on future trade patterns, the use of (7) would be more appropriate.

As evidenced by its lineage, the distinctive contribution of SI models to flow and movement analysis is to separate explanatory factors into three multiplicative classes, site attributes of origins, site attributes of destinations and measures of relative distance/travel time separating origins and destinations, sometimes collected under the heading "impedance" effect.

2 Some important steps

2.1 The entropy approach

The entropy procedure (Wilson 1967) arises from the field of statistical mechanics, which reflects an enumeration approach to statistics in terms of combinatorial analysis. The theory can be regarded as a macro theory with micro foundations. The most probable macro distribution is that which can be replicated by the maximum number of micro level events. In other words, the most probable macro allocation of a given activity, such as work trips, between a set of spatial zones is a priori that which can be matched by the maximum number of potential trips between each individual worker in each zone of residence and each filled job in each zone of employment. The key assumption is that each micro level event associated with a given macro distribution is equi-probable. We note that this theory is applicable to any field where data availability restricts analysis to a macro level while identifiable discrete objects exist at the micro level, and is certainly not intrinsically a part of physics. In fact, the approach is used in many contexts where: (i) the number of individual objects is very large, (ii) data is only available at an aggregate level, (iii) one is only interested in having model results at this same aggregate level, and (iv) the within-aggregate variance of the individual behaviour being modelled is much less than the between-aggregate variance. In cases where condition (iv) is unlikely to hold, one should perform market segmentation and carry out a separate analysis for each segment.

A different concept of entropy was introduced by Shannon (1948) in terms of a representation of the uncertainty of a probability distribution, in which the most probable distribution is defined as that having the maximum entropy or uncertainty. Note that this specification allows for distributions to be defined either with respect to macro behaviour by *sets* of individuals (e.g., the probability of any randomly chosen worker commuting from a given residential zone and arriving at a given employment zone), or with respect to different discrete choices by specific individuals (e.g., his or her choice of travel mode for the work journey). For example, the entropy S associated with the probability p_{hm} that a sampled individual h will choose mode m for the work journey is:

$$S = -\sum_{hm} p_{hm} \log p_{hm}. \tag{9}$$

The possibility of using entropy for individual discrete choice was introduced into regional science by Anas (1983) and shown to yield identical results to multinomial logit models when estimated via maximum likelihood. This raises the question of the relevance of aggregate type models in contemporary regional science. The advent of information technologies has created an information-rich environment wherein travel behaviour databases encompassing movement and communication in space and time have become much more common (Cambridge Systematics 1996). An additional factor pertains to the dramatic advances in the econometrics of disaggregate choice models (Ben-Akiva and Lerman 1985; Cramer 1991) and in travel behaviour modelling (Ettema and Timmermans 1997; Hensher 2001) that expedite the analysis of disaggregate behavioural data. While some of the bad press directed to SI models may perhaps be justified (Ortuzar and Willumsen 2001), much of it seems to be 'ideological', omitting the fact that almost all recent applications of SI models cleverly segment the market to considerably reduce aggregation bias.

In addition, individual choice models are rarely estimated fully on individual data, with revealed choice travel models extracting travel times from an aggregate congestion assignment analysis and data on individual firm behaviour being restricted by confidentiality requirements. Also, well-tried market-segmented models, such as the Huff model, have proved surprisingly robust. Where multi-event decisions are being modelled, such as choice of different quantities of multiple goods and services at multiple stops along a trip chain, the required size of individual data samples can be awesome. Furthermore, flow modelling is used to forecast future demand, which typically has to be provided at the aggregate level due to the aggregate nature of future flow predictors. While deriving aggregate forecasts from a disaggregate model remains a challenge to this day, it is trivial with SI models. In particular, the mobility modules of large-scale integrated land-use/transportation models remain anchored in SI modelling (Mackett 1994; Wegener 1994; Martinez 1996). The main message for the modeller is to use judgement on which approach to adopt in a particular case, bearing in mind the above strengths and weaknesses, as well as the most current international studies.

Following Wilson (1967) and his successors, such as Fisk and Brown (1975), the statistical mechanics entropy formulation is now presented for journey-to-work travel. At a certain base period, let there be O_i workers observed in origin zone i and D_j filled jobs observed in destination zone j, yielding $T = \sum_i O_i = \sum_j D_j$ round trips in any one day. Let T_{ij} be the unknown number of trips between zones i and j. Also, let c_{ij} be the average generalised cost of travel (i.e., the money cost plus the time cost obtained with an assumed value of travel time) between zones i and j. From a base period survey, let the average generalised cost of travel over the entire urban area be c^0. Now, the objective is to find the 'macrostate' trip distribution T_{ij} at the base period, which maximises the number of associated micro level events (called microstates), consistent with satisfying all of the above base period observations, introduced via constraints that the model must replicate. These constraints are taken to add 'information' to the model.

The main principle for proceeding from macrostates to sets of associated microstates, as inferred from Roy and Lesse (1981) for regional science, is:

> Individual objects only define microstates upon descending from any macro (aggregate) level to the individual object level when they are non-homogeneous (that is, distinguishable) with respect to the macro level behaviour being modelled (e.g., when modelling interzonal journey-to-work travel, commuters travelling between zonal pairs have different commuting time vs. job benefit trade-offs).

In addition, the individual jobs themselves within each zone may usually be of variable intrinsic attractiveness. Thus, the microstates are defined here as each distinguishable worker within a set in a given origin zone travelling to a set of jobs in a given destination zone and occupying a distinguishable job there. Using an improvement introduced by Cesario (1973), the entropy Z associated with a certain macrostate distribution T_{ij} is the number of ways that O_i distinguishable workers may be allocated from residential zones i to job zones j in groups T_{ij} times the number of ways that the $D_j = \sum_i T_{ij}$ distinguishable workers arriving at j may be allocated among the D_j distinguishable jobs there. From permutation and combination theory, the number of microstates is given as:

$$Z = \{[\pi_i O_i!/\{\pi_j T_{ij}!\}][\pi_j D_j!]\}, \tag{10}$$

where π represents the product sign and ! the factorial. The first term denotes the number of ways (combinations) that O_i distinguishable workers in each origin zone i may be allocated into destination sets $\{T_{ij}\}$. With the second term, we have the number of ways (permutations) that the D_j distinguishable workers arriving in zone j may be allocated into the D_j distinguishable jobs there, with only one worker being allocated to each job. The objective Z to be maximised can be replaced by the monotonic transformation $S = \log Z$, where the log is the 'natural' log. Also, for cases with a large number of objects, the Stirling Approximation $x! = x(\log x - 1)$ can be applied. This enables (10) to be simplified to:

$$S = \log Z = \sum_i O_i[\log O_i - 1] - \sum_{ij} T_{ij}[\log T_{ij} - 1] + \sum_j D_j[\log D_j - 1]. \tag{11}$$

As the origin totals O_i and the destination totals D_j are known base period inputs, only the second term in (11) is involved in the differentiation. Now the maximisation of (11) is to be constrained by inducing the model flows to conform to certain aggregate base period quantities. Such constraints (except if they are strongly collinear) not only reduce the 'entropy' or 'uncertainty' of the final solution, but usually contribute to obtaining a $\{T_{ij}\}$ solution closer to that observed at the base period. Such a model then becomes potentially viable for making projections for a future period. The constraints introduced by Wilson are firstly the origin and destination constraints (12) and (13):

$$\sum_j T_{ij} = O_i \tag{12}$$

$$\sum_i T_{ij} = D_j. \tag{13}$$

In addition, as we are studying a particular class of behaviour, that is, of commuters, at least one behavioural constraint must be introduced. In Wilson (1967), this constraint reproduces the observed average base-period generalised cost of travel c^0 in terms of the interzonal travel costs c_{ij}, yielding:

$$\sum_{ij} T_{ij} c_{ij} = T c^0. \tag{14}$$

Now if expression (11) is maximised in terms of T_{ij}, applying constraints (12) with Lagrange multipliers λ_i, constraints (13) with multipliers η'_j and constraint (14) with multiplier β, the solution emerges in the form:

$$T_{ij} = A_i O_i B_j D_j \exp -\beta c_{ij}, \tag{15}$$

where the 'balancing factors' $A_i = \exp -\lambda_i$ and $B_j = \exp -\eta_j$ are expressed recursively in the form:

$$(A_i)^{-1} = \sum_j B_j D_j \exp -\beta c_{ij} \tag{16}$$

$$(B_j)^{-1} = \sum_i A_i O_i \exp -\beta c_{ij}. \tag{17}$$

This is the classical form of the doubly-constrained entropy model and has also become known as the trip distribution step in the 4-step transport planning model. The unknown balancing factors A_i and B_j are obtained by solving the recursive equations (17) iteratively. Parameter β is evaluated by solving (14) via linear extrapolation.

In making future projections where one has estimates of changes in the distribution of housing and jobs, as well as changed transport costs, the 'gravity impedance' Lagrange multiplier β above is usually treated as a parameter, with the average trip cost emerging as an output, overcoming the popular misconception that entropy models impose constant travel costs. A general formalism for transforming the model such that these relevant Lagrange multipliers at estimation become parameters for projection was introduced by Lesse (1982) and is discussed later.

Karlqvist and Marksjö (1971) demonstrated that if the model form (15) has its parameters A_i, B_j and β estimated by the method of maximum likelihood, not only are the likelihood equations identical to our entropy model constraints (12) to (14), but the parameters are identical to the Lagrange multipliers of the entropy model. This correspondence has not been fully appreciated.

Using the particular case of the doubly-constrained entropy model of Equations (14–17), Evans (1973) demonstrated that, as the gravity parameter β approaches infinity, the result approaches asymptotically that of minimising the total travel cost on the right-hand side of (14) under the origin and destination constraints (12) and (13), respectively. The latter is known as the transportation problem of linear programming (LP). Thus, there is a seamless transition, as β starts increasing towards infinity, between the solution of the probabilistic entropy model and the solution of the deterministic LP, the objective of which is the base period cost (behavioural) constraint of the entropy model. This result implies that the deterministic problem can always be obtained as a 'special case' of our entropy problem by making the behavioural parameter β sufficiently large. More importantly, we are free to construct probabilistic versions of classical deterministic models arising from different theories (Smith 1990), such as the profit maximisation supply model of identical perfectly competitive firms arising from microeconomic theory. In this case, the cost constraint (14) is merely replaced by a profit constraint, with the revenue term expressed via a production function (which gives output as a function of the quantities of all the material and human inputs) and the cost term given as the sum of the costs of all these inputs and their transport costs to the places of production. This property (Roy 2004) lies at the heart of efforts to link probabilistic SI modelling with some of the deterministic models of microeconomics, enhancing the latter to become probabilistic.

From the generic model (15), it is straightforward to derive the retail model of Wilson (1970) as:

$$S_{ij} = O_i W_j^{\alpha} \exp -\beta c_{ij} / \left(\sum_j W_j^{\alpha} \exp -\beta c_{ij} \right), \tag{18}$$

where S_{ij} is the number of retail trips between origin zone i and centre j, W_j the floorspace of centre j, c_{ij} the generalised cost of travel between zone i and centre j and α a scaling parameter estimated simultaneously via maximum likelihood. This model continues to be widely applied as a more elaborate cousin of the Huff model.

2.2 *Information theory models*

Information theory (Kullback 1959) is a statistical inference technique evaluating the change in a probability distribution due to the supply of certain new information. The information gain of a message predicting that an event will occur according to an a posteriori probability x_1 with respect to an a priori probability x_0 is given as:

$$I(x_1, x_0) = \log_e(x_1/x_0). \tag{19}$$

Expanding the concept to the n by n bilateral journey-to-work system T_{ij} of Wilson (1967), we can define $p_{ij} = (T_{ij}/T)$ as the probability that any trip chosen at random occurs between zones i and j. If q_{ij} is defined as the observed probability of such an event at the (usually) most recent time period at which data is available, then the information gain objective (19) is generalised to the bivariate distribution below:

$$I(p, q) = \sum_{ij} p_{ij} \log(p_{ij}/q_{ij}). \tag{20}$$

If both sides of the origin and destination constraint relations (12) and (13) are converted into probabilistic form by dividing by the total number of current trips T, then minimisation of (20) under these constraints yields:

$$p_{ij} = q_{ij} a_i o_i b_j d_j, \tag{21}$$

with $o_i = O_i/T, d_j = D_j/T$ and a_i and b_j expressed recursively as:

$$(a_i)^{-1} = \sum_j q_{ij} b_j d_j \tag{22}$$

$$(b_j)^{-1} = \sum_i q_{ij} a_i o_i. \tag{23}$$

As demonstrated by Snickars and Weibull (1977), the ‘Fratar’ method (21)–(23) usually performs better than the entropy gravity model (15)–(17). Note that, the average trip cost c represents a key property of the internal state of the transport system being modelled, being an output from rather than an input to our projection model. As the Fratar approach has no explicit representation of transport costs, it has no way of representing a major change in the transport network or costs between the base period (corresponding to our observed 'prior' distribution) and the projection period. Conversely, the gravity model (15)–(17) merely enters the new generalised transport costs c_{ij}^* in place of the original costs c_{ij} and re-balances (16) and (17) to compute the projected flows. However, if it did not do a very good job in fitting to the observed base period flows, we can only expect fair accuracy in projecting the *relative* changes in flows from the base period. This anomaly was addressed in Roy (1987), where an approach was presented to combine the greater projection reliability of the Fratar procedure with the sensitivity to transport network and cost changes of the conventional gravity model. If the ratios of the observed base period flows and the flows produced by the entropy model (15)–(17) are introduced as prior probabilities into an enhanced entropy model (15)–(17), the result corresponds identically to the observed base period flows. Thus, the enhanced model is not only sensitive to changes in transport costs via (14), but also likely to be a good predictor.

2.3 Use of entropy as a constraint

In Erlander (1980) and Erlander and Stewart (1990), the entropy maximisation problem is turned on its head. The cost constraint, such as in (14), becomes the objective, with the entropy objective, such as in (11), being applied as a constraint. Note that this is not a dual formulation in the mathematical sense – it is rather an expression of the interchangeability of objective and constraint in Lagrangian analysis. In Erlander's framework, the minimisation of cost is restored, from being a constraint in the entropy model, to being the objective in its deterministic analogue. Then, to modify the solution to pick up the dispersion from the deterministic solution, the modelled entropy is induced to reproduce the entropy evaluated at the observed values of the flows. The derivation of Erlander's model is illustrated on the doubly-constrained journey-to-work model, given in (14) to (17). Let the objective C be the total transport cost to be minimised:

$$C = \sum_{ij} T_{ij} c_{ij}, \tag{24}$$

which occurs in constraint (14) in the usual solution. Then, if we know the observed base period flows T_{ij}^0, let the 'observed' entropy S^0 be computed from these flows and applied as the following constraint:

$$-\sum_{ij} T_{ij}[\log T_{ij} - 1] = -\sum_{ij} T_{ij}^0[\log T_{ij}^0 - 1] = S^0, \tag{25}$$

which is the objective (11) in the conventional case. Now, if the total transport cost C is minimised under the usual origin and destination constraints (12) and (13) plus the entropy constraint (25), we obtain the following:

$$T_{ij} = \exp -(\lambda_i^{''} + \eta_j^{''} + \beta^{''} c_{ij}), \tag{26}$$

where $(1/\beta'')$ is the Lagrange multiplier on the entropy constraint (25) and λ_i'' and η_j'' those on the origin and destination constraints (12) and (13), respectively. This relation can be compared with the solution of the conventional entropy model, where the functional form is identical, but the parameters are different. This difference in parameters is an empirical matter, depending on goodness-of-fit of the models.

The Erlander analysis above relates to a single event situation, such as destination choice. It remained for Boyce et al. (1983) to extend Erlander's ideas for distributions with at least two interdependent events (e.g., choice of job and choice of travel mode for the journey to work), adding at least two entropy constraints. Roy and Lesse (1985), inspired by Boyce's work, produced a primal entropy maximisation variation of the same models, recognising that the models possess a similar structure to nested logit models arising from random utility theory. The 'reverse model' was implemented by Abrahamsson and Lundqvist (1999). Boyce's own work is further developed in Boyce and Bar-Gera (2003). Although it is more a matter of taste whether the Erlander approach or the standard entropy procedure is to be preferred, there appear to be greater difficulties in computing the observed entropy than the average cost of travel, where robust sampling procedures are available.

2.4 Information, constraints and open models

In the first instance, we must regard model calibration merely as solution of the constrained entropy maximisation problem for both the unknown flows and the unknown Lagrange multipliers on constraints using data at a certain base period. In attempting to give probabilistic versions of the open models of economics, where quantities are endogenous and price-responsive, we are faced with an apparent contradiction. For our models to be information-rich and robust, information needs to be added to the model system during calibration in the form of constraints. On the other hand, such constraints would appear to close up the model when it is being used for projection. This conundrum is handled by using a formalism introduced by Lesse (1982), which was first used to open up SI models in (Roy 1983). Despite the elapsed time, the power of the approach seems not to have been fully appreciated. We illustrate it on the simplest possible case, the conversion of the Lagrange multiplier β on the base period cost constraint (14) to a parameter for projection.

The first step is to write out the Lagrangian function Z^0 associated with the entropy function (13) and its constraints (5), (7), and (14) as:

$$Z^0 = -\sum_{ij} T_{ij}\left(\log T_{ij} - 1\right) + \lambda_i'\left(O_i - \sum_j T_{ij}\right) + \eta_j'\left(D_j - \sum_i T_{ij}\right) + \beta\left(Tc^0 - \sum_{ij} T_{ij}c_{ij}\right). \tag{27}$$

Now, if we want β to be treated as a known parameter in (27) rather than as an unknown Lagrange multiplier, and its right-hand side, the total travel cost (Tc^0) as unknown output in (27) rather than as known input, we make the following invariant Legendre transform to the original Lagrangian Z^0, in the form:

$$Z' = Z^0 - (Tc^0)\partial Z^0/\partial(Tc^0), \tag{28}$$

in which the transformed Lagrangian Z' emerges as:

$$Z' = -\sum_{ij} T_{ij}\left(\log T_{ij} + \beta c_{ij} - 1\right) + \lambda_i'\left(O_i - \sum_j T_{ij}\right) + \eta_j'\left(D_j - \sum_i T_{ij}\right). \tag{29}$$

Now, if the transformed Lagrangian is maximised in terms of T_{ij} *given* the values of the parameter β, the theory of Legendre transforms, enunciated in Lesse (1982), indicates that not only are the flows T_{ij} from (29) identical to those (15) to (17) from the untransformed objective Z^0 in (27), but that when these T_{ij} values are substituted back into the original constraint on (Tc^0), it is perfectly satisfied. The transformed model emerging from maximisation of (29) is also identical in structure to the base period estimation model (15) to (17), except that now β is a known parameter rather than an unknown Lagrange multiplier.

At first sight, the above may appear to be a mere intellectual exercise. This is assuredly not the case. Having calibrated the parameter β on the base period data in the initial 'calibration' problem (11)–(17), the transformed problem (29) allows us to project the revised flows at a future period under potentially revised trip origins O_i^*, trip destinations D_j^* and travel times t_{ij}^*. We can also apply the Legendre transforms to the quantity constraints, thus truly opening up quantities in the projection time period to be endogenous and thus price-responsive (Roy 2004).

Although the Legendre transforms provide the mathematical formalism to convert Lagrange multipliers to parameters, we still should assure ourselves that their assumed constant value (at least over the short-term) is plausible. The case has already been well argued for the gravity impedance parameter β, which can be related to price elasticities of demand. Similarly, transforms with respect to quantities demanded or supplied can be related to quantity elasticities. Of course, we are still left with the same limitations surrounding use of estimated elasticities in economics – that is, no large global changes occurring, just large local changes or small global changes between the base period and the projection period.

2.5 Cost efficiency principle and most probable states

Some of the most rigorous and significant contributions to the area of SI modelling have been made by Tony E. Smith, formerly of the founding Regional Science Faculty, University of Pennsylvania. One of Smith's first major efforts was the development of a cost-efficiency principle for spatial interaction (Smith 1978). In this highly formal paper, Smith aimed to establish a more behavioural and intuitively plausible basis for SI modelling. The hypothesis is as follows: "A trip distribution P on X is said to satisfy the cost-efficiency principle if and only if for each possible (travel) cost configuration c on X and pair of trip patterns consistent with the same trip activity A on X, the pattern with the lowest travel cost is always the most probable one". Based on this definition of the principle, an Equivalence Theorem is proved: "A trip distribution P satisfies the cost-efficiency principle if and only if P is representable by an exponential gravity model". A corollary is that "All activity-equivalent trip patterns with the same total travel cost must be equally probable". Thus, a behavioural basis is provided for the exponential gravity model, independent of that implied by the conventional entropy maximisation technique, but consistent with its results.

In Smith (1989), the general interaction model of Alonso (1978) was examined. An important result was that the form of Alonso's model where the total number of interactions is exogenous, can be simplified to the following:

$$T_{ij} = gA_i^{\alpha-1}B_j^{\beta-1}f_{ij}, \tag{30}$$

where T_{ij} is the number of trips between i and j, g is a normalisation parameter, A_i and B_j are sets of systemic variables, akin to balancing factors, defined at the origins and destinations, respectively, α and β are positive scalars lying between zero and one and the f_{ij} are exogenous spatial impedance factors between i and j. The Alonso model is shown to be interpretable within the decision theory developed in the same paper. Like balancing factors, systemic variables A_i and B_j can be interpreted as accessibilities (Hua 1980). Variable A_i is large for origins with a short mean trip length and vice-versa. Variable B_j takes on a similar interpretation in relation to destinations. Given this property, Fotheringham and Dignan (1984) can reinterpret parameters α and β as the elasticities of total outflows and inflows with respect to the inverse of the mean trip length. They also point out that α and β are strongly influenced by the accuracy of constraints on marginal flow totals, so that it may be optimal to use a form of quasi-constrained interaction model in some forecasting circumstances where there is uncertainty about marginal flow totals.

Arguably, Smith's most recent major personal contribution of relevance to this chapter was the formalisation in 1990 of entropy models within the Most Probable State (MPS) approach. "The fundamental hypothesis of MPS-analysis is that all micro population behaviour is describable probabilistically in terms of certain of its macro properties". In Smith's terminology, a *probabilistic theory of micro behaviour* (that is, the 'theory') is defined as a function of the macro properties (e.g., total travel cost), themselves given as functions of the trip variables. The chosen macro properties are not selected at random, but *are inextricably bound up with the theory itself.* The important results, proved rigorously by Smith, are stated qualitatively here:

- If theory T is true, then we can compute an exact form for the conditional probabilities of macro states given any observed values of the macro properties.
- The conditionally most-probable macro states are precisely those which can occur in the largest number of ways (as in the entropy microstate approach).
- If theory T is true, then for all sufficiently large population realisations, the maximum entropy state is the overwhelmingly most-probable macro state which can occur.
- Each choice of macro variables constitutes a different theory of micro behaviour. For instance, total kinetic energy is defined as a macro property in Boltzmann's

probabilistic equilibrium theory of an ideal gas, lying at the very heart of the theory.

The second to last statement represents a major generalisation of the restricted proof in Wilson (1967), which was proved for an entropy model with just one constraint, that is, on the total population in the system. Note that, testing of any MPS-based probabilistic theory does *not* require that the samples be statistically independent.

To continue, Smith's work is of paramount importance in its attempts to generalise SI models to include microeconomics concepts. For instance, in developing a probabilistic model for the supply behaviour of profit-maximising competitive firms, profit is the key macro property (Roy 2004). The main results from Evans (1973) are adapted to indicate that the results of any classical deterministic theory are obtainable as limiting cases of the probabilistic model in which the relations of the deterministic theory are introduced as a macro constraint reflecting observations. As the entropy (uncertainty) approaches zero, the solution of the probabilistic model converges asymptotically towards that of its deterministic counterpart. Thus, the work of Smith and Evans is pivotal in introducing microeconomics into the field of SI modelling.

2.6 Spatial structure and spatial interdependencies

A pervasive interpretation of the 'gravity impedance' multiplier β in model (15)–(17) and its derivatives is that it accounts for the distance decay effect on spatial interaction, so that differences in β have been ascribed to different preference structures brought about by differences in social and economic conditions. Not until the early work by Porter (1956), followed by Linneman (1966), Curry (1972), Greenwood and Sweetland (1972), Johnston (1973), and others, was it realised that such differences might also reflect the locational pattern of flow origins and destinations, as well as the spatial distribution of origin and destination attributes. The detailed econometric analysis of Canadian journey-to-work data conducted by Griffith and Jones (1980) shows that the rate of distance decay in SI models and the spatial structure associated with origins and destinations are interdependent. It vindicates the view of Sheppard (1976) that the entropy model (14)–(17) is an accurate estimator only if all the pertinent characteristics of the process under study are used as prior information. For a spatial structure effect, a spatial autocorrelation function can conveniently summarise spatial interdependencies embedded in the spatial interaction process. If the one-lag spatial autocorrelation is known a priori, a maximum entropy model can be estimated by including autocorrelation as an extra constraint.

Fotheringham and Webber (1980) tackle the underlying spatial structure of SI systems from a different angle. With prior information on the spatial autocorrelation of flow attractions, an unconstrained SI model can be formulated as a simultaneous equation system:

$$T_{ij} = m_i^{\alpha} M_j^{\gamma} \exp -\beta c_{ij} \tag{31}$$

$$m_i = g(T_{ij} : j \in J) \tag{32}$$

$$M_j = h(T_{ij} : i \in I), \tag{33}$$

where m_i refers to the 'mass' of origin i, M_j is 'mass' of destination j, c_{ij} is the generalised cost of travel between zone i and zone j, α, β and γ are estimable parameters, $J(I)$ is the set of destination (origin) zones, and $g(h)$ is a certain function. This theory assumes that the system dynamics reach an equilibrium state where interactions have adjusted to changes in the masses and that the masses have adjusted to changes in interaction. Equations (32) and (33) are feedback mechanisms that create spatial structure out of interactions in the spatial system. Fotheringham and Webber (1980) posit that the specification of functions g and h should be based on prior information that is part of the corpus of regional science theory. For instance, the theory of growth centres implies that centres spawn export-oriented growth through a process that affects other centres in direct relation to their accessibility. Thus, in a migration system, the logic of this theory suggests that Equation (32) can be specified as follows:

$$m_i = \exp \rho \textstyle\prod_{j \in J} \exp \phi T_{ij}, \tag{34}$$

where ρ and ϕ are estimable parameters. A far-reaching implication of the theory advocated by Fotheringham and Webber (1990) is that spatial interaction and spatial structure are inextricably and mutually linked through generative processes. As Gould (1991) put it, "no connections, no geography" (p. 4). Interestingly, Getis (1991) independently discovered that SI models such as (14)–(17) are formally equivalent to a broad range of statistics of spatial autocorrelation, which are commonly used to capture the structure embedded in map patterns. Hence SI models are clearly positioned to play a crucial role in understanding spatial processes and spatial structures. In addition, the unified framework uncovered by Getis (1991) paves the way to spatial inference in SI modelling.

Another approach to incorporating spatial structure in SI models is to "spatially contextualise" their parameter(s). According to Casetti's (1972) expansion method, parameters of a model designed to capture first-order (or 'global') relationships are expressed as a linear function of other attributes, including location or accessibility, so that trends in parameter estimates can be revealed. Thus, complex spatial patterns in flow origins and/or destinations can be explicitly incorporated in a SI model. Surprisingly, little use has so far been made of this practical approach to SI modelling, except in trade flow analysis (Zhang and Kristensen 1995) and interregional migration (Roy 2004).

As discussed in Fotheringham (1983), when a system with, say, two identical retail centres spaced well apart is enhanced by the addition of an identical third retail centre, will the *relative* competitive position of the two existing retailers be influenced by alternative location options of the third centre? The conventional model would say no! However, what if the third retail centre were located quite close to say centre number 1? If each centre were equally accessible to customers, then the conventional model would indicate that addition of the third centre would change the market shares from $\{1/2, 1/2\}$ to $\{1/3, 1/3, 1/3\}$, with no effect of the proximity of centres 1 and 3 on the relative competitive position of the original two centres.

Intuitively, for this case of identical centres of equal accessibility, one may expect the market share of centre 1 to be equally split with that of the new centre, leading to shares of $\{1/4, 1/4, 1/2\}$! The appreciation of this anomaly and of the misspecification of the conventional model gave Fotheringham the incentive to develop a new form of SI model called the competing destinations (CD) model (Fotheringham 1983, 1986; Fotheringham and Knudsen 1986), which is in some sense hierarchical, as given explicitly in Fotheringham (1986). In other words, customers are not only attracted to a particular centre per se, but to the milieu of that centre with respect to adjacent centres. In this way, both spatial competition and spatial agglomeration influences can be identified. It should also be noted at this point, that inspired by Fotheringham's 1983 work, Roy (1985) developed via entropy a hierarchical destination choice or 'cluster' model, which shares the structural dependence properties of the nested logit random utility model and overcomes the Independence from Irrelevant Alternatives (IIA) weakness. In this approach, there is a pre-determined nested hierarchical choice between a cluster and then a particular destination within that cluster. Fotheringham's formulation, on the other hand, influences the choice of a primary destination by its accessibility to a set of alternative secondary destinations.

In illustrating Fotheringham's ideas, we turn to his new competing destinations retail model presented, for instance, in Fotheringham and Knudsen (1986). In contrast to the conventional retail model in (18), upon defining flows S_{ij} in terms of trip origins O_i, floorspace W_j and home-based travel costs c_{ij}, the CD model is expressed as:

$$S_{ij} = O_i W_j^{\alpha} A_j^{\gamma} \exp -\beta c_{ij} / \left[\sum_j W_j^{\alpha} A_j^{\gamma} \exp -\beta c_{ij} \right], \tag{35}$$

with the potential measure A_j defined for competing destinations as:

$$A_j = \sum_{l \neq j} W_l c_{jl}^{-\delta} d_{jl}, \tag{36}$$

where c_{jl} represents the unit travel cost between the primary destination j and a nearby destination l. The binary variables d_{jl} are unity when destination l lies within a pre-specified range of destination j and zero otherwise. The introduction of the 'competing destinations' potential terms A_j^{γ} distinguishes the new model from the conventional model in (18). A positive value of the scaling index γ demonstrates the presence of consumer agglomeration economies, with a negative value indicating the presence of consumer competition or congestion forces.

Today, the CD model is grounded in behavioural and cognitive theories backed by extensive empirical work. It derives from principles of utility maximisation under the constraint that individuals resort to simplifying information processing strategies due to their limited ability to process the large amounts of information that reach them in typical spatial choice situations. Fotheringham's (1988) contention is that a hierarchical processing strategy is the most likely, whereby choice alternatives are perceived in clusters according to their similarity. Thus, the probability that an alternative is in the individual's choice set is a function of the similarity of this alternative to all other alternatives under consideration. It should be pointed out that similarity among alternatives is not only to be conceived in geographical space, but also in the space of non-locational attributes of destinations. Various formulations have been devised, ranging from the gravity accessibility measure mentioned earlier, to a Manhattan distance measure (Borgers and Timmermans 1987, 1988), to a Pearson correlation coefficient (Meyer and Eagle 1982). All these approaches purport to account for substitution effects among alternatives, thus avoiding the IIA fallacy so pervasive in discrete choice modelling.

The spatial choice foundation of the CD model implies that SI behaviour may involve constraints and restrictions on the size and composition of the choice sets. In essence, constraints of various sorts define a range of feasible options within the universe of possible destinations. As discussed by Sheppard (1980) and Thill (1992), a constraint-oriented approach to SI modelling is desirable on two counts at least. On the one hand, an erroneous interpretation of interaction behaviour may ensue from the failure to account for constraints on the formation of choice sets. The model explains interaction flows in terms of deliberate decisions, whether or not they are the product of the structure of the choice sets. Furthermore, misspecified choice sets lead to biased model estimates (Pellegrini et al. 1997; Thill and Horowitz 1997a). Most of the research to alleviate choice-set definition problems has focused on the implicit definition of choice sets, as in the CD model or the Approximate Nested Choice-Set Destination Choice (ANCS-DC) model proposed by Thill and Horowitz (1997b).

The competing destinations modelling structure clearly represents a great advance in recognition of interdependencies in spatial choice. It is also supported by a great deal of empirical work. Interestingly, the approach advocated by Fotheringham to handle substitution and complementarity relationships among flow destinations can also account for interdependencies created by multi-purpose and multi-stop travel. As indicated by Thill (1995), the potential measure A_j in (36) can be recast into:

$$A_j = \sum_{l \neq j} M_l c_{jl}^{-\delta} d_{jl} \tag{37}$$

where M_l measures the attractiveness of destination l for purposes other that the one being analysed, while other notation is as defined above. Although the analysis of multi-purpose and multi-stop activities within the CD modelling framework remains to be fully formalised, a related attempt is made in Chapter 1 of Roy (2004).

2.7 Neural networks and complex flow systems

Over the past two decades, a variety of techniques of evolutionary computation and artificial life have permeated regional science research to grasp more fully the inherent complexity of many spatial and regional systems. Openshaw (1993) argued that the mapping function of SI models between propulsiveness, attractiveness, and spatial impedance, on the one hand, and flows, on the other hand, can be built into supervised artificial neural networks such as the feed-forward back-propagation network. Empirical results reported by Openshaw (1993), Fischer and Gopal (1994), Fischer et al. (1999), Reggiani and Tritapepe (2002), and others leave no doubt that neural network models may out-perform conventional SI models in many cases. However, the limits of entropy-type models have still not been reached.

Artificial neural networks are based on an analogy with the workings of the brain. As such, they are composed of a number of elements whose function is to process and pass along information to other elements. Multiple parameters associated with these elements are estimated iteratively in parallel, following rather well established optimisation routines.

Because neural networks make no assumptions on the form or distributional properties of interaction data and predictors, they can be viewed as non-parametric methods. This has a significant advantage over conventional SI models in that spatial interaction can be modelled even when the only data available are explicitly noisy or statistically ill-conditioned. Neural SI models also offer greater representational flexibility than many existing entropy-based models and, very much in the spirit of exploratory data analysis, relax many constraints on possible model designs. As in Sen and Smith (1995), they are consistent with the modelling philosophy of fitting an appropriate model to data, rather than forcing data into an assumed model structure. Thus, the potential exists for incorporating spatial interdependencies explicitly into the network representation, although this remains to be implemented.

Neural SI models come in a large variety of architectures (number of layers, number of elements, degree of connectivity, etc.), and involve multiple mathematical specifications of transfer functions and estimation procedures. Suffice it to say here that the generic unconstrained model is:

$$T_{ij} = \Psi \left[\sum_{h=0}^{H} \gamma_h \; \varphi_h \; \left(\beta_{h1} p_i + \beta_{h2} a_j + \beta_{h3} c_{ij} \right) \right] \tag{38}$$

where T_{ij} is the output (flows from zone i to zone j), p_i (propulsiveness of i), a_j (attractiveness of j), and c_{ij} (travel cost from i to j) are the inputs, H is the number of hidden elements, γs and βs are parameters, and Ψ and ϕ are transfer functions often taken to be sigmoid. While a common criticism of model (38) is that it is no more than a black box, it is clear that a better understanding of the inner workings of the model makes for better and richer SI models. In this perspective, it can be noted that singly constrained models can be derived from (38) by imposing accounting flow constraints *ex post* (Mozolin et al. 2000; Thill and Mozolin 2000). Alternatively, constraints can be built into the model to produce a one-stage model (Fischer et al. 2003). The notion of preset model architecture can be questioned and the architecture best suited to a particular SI situation can be inferred by genetic algorithms.

A final comment on neural SI models is in order. Recent work by Mozolin et al. (2000) and Thill and Mozolin (2000) indicates that, in spite of their good prediction of base-year data, commonly used feed-forward back-propagation models may be inferior to conventional constrained models estimated by maximum likelihood when it comes to transferability through time. If corroborated, this research invalidates this approach as a long-term planning tool. More research is needed to single out the causes of, and remedies for, this deficiency.

3 The future

After the path-breaking formalisation of SI models by Wilson (1967, 1970), Evans (1973) demonstrated the fundamental asymptotic character of the approach, providing a link to the deterministic theory implied by the behavioural constraints. Fisk and Brown (1975) and Roy and Lesse (1981) identified alternative forms of entropy for alternative decision contexts, freeing the models from the limited set of functional forms introduced by Wilson. Most importantly, the entropy formulation was generalised to the Most Probable State (MPS) approach by Smith (1990), not only demonstrating the most-probable result for all constrained models in cases of large populations as *overwhelmingly* the most probable, but also classifying the behavioural constraint of the model as the fundamental relation of the theory being tested. Lesse (1982) showed how the Lagrange multipliers on the model constraints can be transformed into parameters when the models are to be used in projection. At about the same time, Fotheringham (1983, 1986) was able to free the simpler SI models from the naïve IIA restriction by incorporating spatial interdependencies in a quasi-hierarchical framework. Progress was meanwhile being made on formulation and implementation of discrete choice models of individual behaviour. There should be no "black and white" comparison of such models with SI models – each has its own appropriate application context. Finally, Roy (2004) has attempted to draw some of these threads together to provide a more coherent whole. At the same time, as mentioned recently by Smith (private communication), "It does little good to know what the overwhelmingly most probable response to a given policy will be unless this can be translated into actual confidence bounds on key response variables". We can be heartened in this endeavour by the intimate links between entropy models and maximum likelihood models, established very early (Karlqvist and Marksjö 1971). It should be possible to exploit the recent large amount of work done on error analysis for maximum likelihood to establish the confidence bounds, which Smith has mentioned. At the same time, the links with spatial autocorrelation models (Getis 1991) should be further explored.

Some of the latest SI research has evidenced that there is room for further enhancing our modelling of spatial interaction. Several approaches are now available to incorporate the effects of spatial structure and spatial interdependencies either in the form of modifications to existing models or of entirely new models. Certainly our ability to develop these new avenues is limited by our knowledge of spatial patterns embedded in SI data. In this respect, techniques of exploratory spatial data analysis, as in Tobler (1975) and Marble et al. (1997), should prove useful in inducing hypotheses on local patterns in spatial interaction, which in turn will enrich SI theory.

Another very promising research direction is that entailing computational systems. The past decade of research along this line has established it as a viable framework for SI analysis. It has involved a lot of trial and error, as well as learning the extreme flexibility of geo-computational methods. Research currently underway will lead to a clearer methodology for their valid application to SI analysis. Their inherent flexibility supports the emergence of hybrid models incorporating many elements from modelling traditions rooted in behavioural theory and the entropy principle.

References

Abrahamsson T, Lundqvist L (1999) Formulation and estimation of combined network equilibrium models with applications to Stockholm. *Transportation Science* 33: 80–100

Alonso W (1978) A theory of movement. In: Hansen NM (ed) *Human settlements*. Ballinger, Cambridge, MA

Anas A (1983) Discrete choice theory, information theory and the multinomial logit and gravity models. *Transportation Research* 17B: 13–23

Batten DF, Boyce DE (1986) Spatial interaction: Transportation and interregional commodity flow models. In: Nijkamp P (ed) *Handbook of regional and urban economics,* vol. I. *Regional economics*. North-Holland, Amsterdam

Ben-Akiva ME, Lerman SR (1985) *Discrete choice analysis: Theory and application to travel demand*. MIT Press, Cambridge, MA

Birkin M, Clarke G, Clarke M, Wilson AG (1996) *Intelligent GIS. Location decisions and strategic planning*. GeoInformation International, Cambridge

Borgers AWJ, Timmermans HJP (1987) Choice model specification, substitution and spatial structure effects: A simulation experiment. *Regional Science and Urban Economics* 17: 29–47

Borgers AWJ, Timmermans HJP (1988) A context-sensitive model of spatial choice behaviour. In: Golledge RG, Timmermans HJP (1988) (eds) *Behavioural modelling in geography and planning*. Croom Helm, London

Boyce DE, Bar-Gera H (2003) Validation of multiclass urban travel forecasting models combining origin-destination, mode and route choices. *Journal of Regional Science* 43: 517–540

Cambridge Systematics (1996) *Scan of recent travel surveys*. US Department of Transportation, Washington, DC

Boyce DE, Le Blanc LJ, Chon KS, Lee YJ, Lin KT (1983) Implementation and computational issues for combined models of location, destination, mode and route choice. *Environment and Planning A* 15: 1219–1230

Casetti E (1972) Generating models by the expansion method: Application to geographical research. *Geographical Analysis* 4: 81–91

Cesario FJ (1973) A note on the entropy model of trip distribution. *Transportation Research* 27: 331–333

Cramer JS (1991) *The logit model: An introduction for economists*. Edward Arnold, London

Deardorff AV (1998) Determinants of bilateral trade: Does gravity work in a neoclassical world? In: Frankel JA (ed) *The regionalization of the world economy*. University of Chicago Press, Chicago

Erlander S (1980) *Optimal spatial interaction and the gravity model* (Lecture Notes in Economics and Mathematical Systems 173). Springer, Berlin Heidelberg New York

Erlander S, Stewart NF (1990) *The gravity model in transportation analysis*. VSP, Utrecht

Ettema DF, Timmermans HJP (1997) *Activity-based approaches to travel analysis*. Pergamon, Oxford

Evans SP (1973) A relationship between the gravity model for trip distribution and the transportation problem in linear programming. *Transportation Research* 7: 39–61

Fischer MM, Gopal S (1994) Artificial neural networks: A new approach to modelling interregional telecommunication flows. *Journal of Regional Science* 34: 503–527
Fischer MM, Hlavackova-Schindler K, Reismann M (1999) A global search procedure for parameter estimation in neural spatial interaction modelling. *Papers in Regional Science* 78: 119–134
Fischer MM, Leung Y (1998) A genetic-algorithms based evolutionary computational neural network for modelling spatial interaction data. *Annals of Regional Science* 32: 437–458
Fischer MM, Reismann M, Hlavackova-Schindler K (2003) Neural network modelling of constrained spatial interaction flows: Design, estimation and performance issues. *Journal of Regional Science* 43: 35–61
Fisk C, Brown GR (1975) A note on the entropy formulation of distribution models. *Operational Research Quarterly* 26: 755–758
Fotheringham AS (1983) Some theoretical aspects of destination choice and their relevance to production-constrained gravity models. *Environment and Planning A* 15: 1121 – 1132
Fotheringham AS (1986) Modelling hierarchical destination choice. *Environment and Planning A* 18: 401–418
Fotheringham AS (1988) Consumer store choice and choice set definition. *Marketing Science* 7: 299–310
Fotheringham AS, Dignan T (1984) Further contributions to a general theory of movement. *Annals of the Association of American Geographers* 74: 620–633
Fotheringham AS, Knudsen DC (1986) Modeling discontinuous change in retailing systems: Extensions of the Harris-Wilson framework with results from a simulated urban retailing system. *Geographical Analysis* 18: 295–312
Fratar T (1954) Vehicular trip distribution by successive approximation. *Traffic Quarterly* 8: 53–65
Getis A (1991) Spatial interaction and spatial autocorrelation: A cross-product approach. *Environment and Planning A*: 1269–1277
Gould P (1991) Dynamic structures of geographic space. In: Brunn SD, Leinbach TR (eds) *Collapsing space and time: Geographic aspects of communications and information*. HarperCollins Academic, NY
Greenwood MJ, Sweetland D (1972) The determinants of migration between standard metropolitan areas. *Demography* 9: 665–681
Guldmann JM (1999) Competing destinations and intervening opportunities interaction models of inter-city telecommunications flows. *Papers in Regional Science* 78: 179–194
Harris B, Wilson AG (1978) Equilibrium values and dynamics of attractiveness terms in production-constrained spatial interaction models. *Environment and Planning A* 10: 371–388
Heckscher E (1919) The effect of foreign trade on the distribution of income. *Ekonomisk Tidskrift* 21: 497–512
Hensher DA (2001) *Travel behaviour research: The leading edge*. Pergamon, Oxford
Hua C (1980) An exploration of the nature and rationale of a systemic model. *Environment and Planning A* 12: 1121–1132
Huff DL (1963) A probabilistic analysis of shopping center trade areas. *Land Economics* 39: 81–90
Johnston RJ (1973) On frictions of distance and regression coefficients. *Area* 5: 187–191
Isard W (1960) *Methods of regional analysis*. MIT Press, Cambridge
Karlqvist A, Marksjö B (1971) Statistical urban models. *Environment and Planning* 3: 83–98
Kullback S (1959) Information theory and statistics. Wiley, New York
Lakshmanan TR, Hansen WG (1966) A retail market potential model. *Journal of the American Institute of Planners* 31: 134–143
Linnemann H (1966) *An econometric study of international trade flows*. North-Holland, Amsterdam
Lesse PF (1982) A phenomenological theory of socio-economic systems with spatial interactions. *Environment and Planning A* 14: 869–888
Long W, Uris R (1971) Distance, intervening opportunities, city hierarchy, and air travel. *Annals of Regional Science* 5: 152–161
Mackett R (1994) Land use transportation models for policy analysis. *Transportation Research Record* 1466: 71–78
Marble DF, Gou Z, Liu L (1997) Recent advances in the exploratory analysis of interregional flows in space and time. In: Kemp Z (ed) *Innovations in GIS 4*. Taylor & Francis, London
Martinez F (1996) MUSSA: Land use model for Santiago City. *Transportation Research Record* 1552: 126–134

Meyer RJ, Eagle TC (1982) Context-induced parameter instability in a disaggregate-stochastic model of store choice. *Journal of Marketing Research* 19: 62–71

Mozolin M, Thill, JC, Usery EL (2000) Trip distribution forecasting with multilayer perceptron neural networks: A critical evaluation. *Transportation Research B* 34: 53–73

Ohlin BG (1933) *Interregional and international trade*. Harvard University Press, Cambridge, MA

Openshaw S (1993) Modelling spatial interaction using a neural net. In: Fischer MM, Nijkamp P (eds) *Geographic information systems, spatial modelling and policy evaluation*. Springer, Berlin Heidelberg New York

Ortuzar J, Willumsen LG (2001) *Modelling transport*. Wiley, Chichester New York

Pellegrini PA, Fotheringham AS, Lin G (1997) An empirical evaluation of parameter sensitivity to choice set definition in shopping destination choice models. *Papers in Regional Science* 76: 257–284

Reggiani A, Tritapepe T (2002) Neural networks and logit models applied to commuters' mobility in the metropolitan area of Milan. In: Himanen V, Nijkamp P, Reggiani A (eds) *Neural networks in transport applications*. Ashgate, Aldershot

Roy JR, Lesse PF (1981) On appropriate microstate descriptions in entropy modelling. *Transportation Research B* 15: 85–96

Roy JR (1983) Estimation of singly-constrained spatial-interaction models. *Environment and Planning B* 10: 269–274

Roy JR (1985) On forecasting choice among dependent spatial alternatives. *Environment and Planning B* 12: 479–492

Roy JR, Lesse PF (1985) Entropy models with entropy constraints on aggregated events. *Environment and Planning A* 17: 1669–1674

Roy JR (1987) An alternative information theory approach for modelling spatial interaction. *Environment and Planning A* 19: 385–394

Roy JR (1990) Spatial interaction modelling: Some interpretations and challenges. Commentary, *Environment and Planning A* 22: 712–716

Roy JR (1999) Areas, nodes and networks. *Papers in Regional Science* 78: 135–155

Roy JR (2004) *Spatial interaction modelling: A regional science context.* Springer, Berlin Heidelberg New York

Schneider M (1959) Gravity models and trip distribution theory. *Papers and Proceedings of the Regional Science Association* 5: 51–56

Sen A, Smith TE (1995) *Gravity models of spatial interaction behaviour*. Springer, Berlin Heidelberg New York

Shannon CE (1948) A mathematical theory of communication. *Bell System Technical Journal* 27: 379–423; 623–656

Sheppard ES (1976) Entropy, theory construction and spatial analysis. *Environment and Planning A* 8: 741–752

Sheppard ES (1980) The ideology of spatial choice. *Papers of the Regional Science Association* 45: 197–213

Smith TE (1978) A cost-efficiency principle of spatial interaction behaviour. *Regional Science and Urban Economics* 8: 313–337

Smith TE (1989) *A simple decision theory of spatial interaction: The Alonso model re-visited*. Working Paper 135, Department of Regional Science, University of Pennsylvania

Smith TE (1990) Most-probable-state analysis: A method for testing probabilistic theories of population behaviour. In: Chatterji M, Kuenne RE (eds) *New frontiers in regional science*. MacMillan, London

Snickars F, Weibull JW (1977) A minimum information principle: Theory and practice. *Regional Science and Urban Economics* 7: 137–168

Stewart JQ (1941) An inverse distance variation for certain social distances. *Science* 93: 89–90

Stouffer SA (1940) Intervening opportunities: A theory relating mobility and distance. *American Sociological Review* 5: 845–867

Theil H (1967) *Economics and information theory*. North Holland, Amsterdam

Thill JC (1992) Choice set formation for destination choice modelling. *Progress in Human Geography* 16: 361–382

Thill JC (1995) Modeling store choices with cross-sectional and pooled cross-sectional data: A comparison. *Environment and Planning A* 27: 1303–1315

Thill JC, Horowitz JL (1997a) Travel-time constraints on destination choice sets. *Geographical Analysis* 29: 108–123

Thill JC, Horowitz JL (1997b) Modelling non-work destination choices with choice sets defined by travel-time constraints. In: Getis A, Fischer MM (eds) *Recent developments in spatial analysis, spatial statistics, behavioural modelling and neurocomputing*. Springer, Berlin Heidelberg New York

Thill JC, Mozolin M (2000) Feed-forward neural networks for spatial interaction: Are they trustworthy forecasting tools? In: Reggiani A (ed) *Spatial economic science: New frontiers in theory and methodology*. Springer, Berlin Heidelberg New York

Tinbergen J (1962) *Shaping the world economy: Suggestions for an international economic policy*. The Twentieth Century Fund, New York

Tobler W (1975) Spatial interaction: Extracting meaning from masses of data. In: Swain E (ed) *The IISA project on urban and regional systems*. IISA, Laxenburg

Wegener M (1994) Operational urban models: State of the art. *Journal of the American Planning Association* 60: 17–29

Wilson AG (1967) A statistical theory of spatial distribution models. *Transportation Research* 1: 253–269

Wilson AG (1970) *Entropy in urban and regional modelling*. Pion, London

Zhang J, Kristensen G (1995) A gravity model with variable coefficients: The EEC trade with third countries. *Geographical Analysis* 27: 307–320