Entropy in Urban and Regional Modelling: Retrospect and Prospect

Alan Wilson
Centre for Advanced Spatial Analysis, University College, London, U.K.

Entropy in Urban and Regional Modelling introduced a new framework for constructing spatial interaction and associated location models. These ideas are reviewed briefly and then set in the wider context of the application of entropy in a range of disciplines. Related developments since 1970 are examined with particular reference to extensions of the core model, links to mathematical programming, the relationship to economics, and the introduction of a dynamic spatial structure hypothesis. The role of entropy maximizing in geography, and more widely in complexity science, is reviewed and two conclusions are drawn beyond the continued effective deployment of entropy models: geography could be a leading player as complexity science develops; and geography can use these ideas to update the presentation of its classic theories. The ongoing research agenda is extensive—particularly in relation to the modeling of imperfect markets in economics—in its links with network science and agent-based modeling. Recent developments in physics suggest a new role for entropy in understanding spatial structure. This article is organized into five main sections. The first summarizes the ideas of the book; the second sets this discussion in a wider context; the third surveys the developments that have taken place since its publication. The fourth summarizes the role of entropy maximizing in geography in the context of complexity science, and the fifth presents an ongoing research agenda.

Introduction

Entropy in Urban and Regional Modelling was published in 1970 as the culmination of 4 years of work in developing the idea and its applications (Wilson 1967). The original stimulus came from observing that the factors applied by American engineers to the gravity model had a resemblance to, but were not identical to, the partition functions used in statistical mechanics. This similarity made me think that a shift from a Newtonian analogy to a Boltzmann classical statistical mechanics one might be fruitful. This shift produced a reformulation of the transport model then in use. It provided some theoretical justification, but perhaps more important, it...
opened up a set of ideas that could be applied much more widely. It also connected
to a wider literature as entropy-maximizing models began to appear in diverse
fields. That the concept was not simply an application of the statistical physics
analogy became clear; rather, it was a high-level methodology that had applica-
tions in many disciplines for a certain kind of problem. Indeed, this contention was
argued in this journal before the publication of the book (Wilson 1969). The prob-
lem was what Weaver’s (1948, 1958) had categorized as one of “disorganized
complexity”: systems of large numbers of elements, interacting with each other
only weakly, and this fitted the populations of cities and regions very well. Weaver
argued strongly that the challenging problems of science were those of “organized
complexity”—systems in which the elements interact strongly. In urban and re-
gional geography, this takes us into economics and organizations that have strong
interactions. Weaver’s was a prescient argument and underpins what is now the
rapidly developing field of complexity science. Entropy maximizing turns out to
have a role to play here as well and one that is outlined in this article.

**Entropy in urban and regional modeling**

The history of the development of entropy in urban and regional modeling is pre-
sented in terms of its use, first, in transport modeling and, second, in retailing. The
role of generalized cost in relation to utilities and disutilities and various kinds of
disaggregation are explored. The functions and parameters of the models are in-
terpreted. There is a wide range of application of the models, including the pos-
sibility of estimating “missing data.”

**The transport model**

The initial challenge was to model the pattern of transport flows in a city or a region.
Assume a set of origin zones, labeled \( \{i\} \), and a set of destination zones, la-
beled \( \{j\} \), and let \( \{T_{ij}\} \) be the matrix of flows from \( i \) to \( j \). Assume that the numbers
beginning in each zone \( i \), \( \{O_i\} \), and similarly that the numbers ending in each zone
\( j \), \( \{D_j\} \), are known. Define \( \{c_{ij}\} \) to be a matrix of travel deterrence—\( c_{ij} \) is some
measure of the cost of travel from \( i \) to \( j \). The traditional (Newtonian) gravity model
may be written as

\[
T_{ij} = KO_iD_jc_{ij}^{-\beta}
\]

for a constant \( K \) of proportionality and a distance decay parameter \( \beta \). The problem
with this model is that the predicted \( \{T_{ij}\} \) do not satisfy

\[
\sum_j T_{ij} = O_i \quad (2)
\]

where \( \sum_j \) represent the sum over all destination zones and

\[
\sum_i T_{ij} = D_j \quad (3)
\]
Factors $A_i$ and $B_j$ may be added pragmatically to give

$$T_{ij} = A_i B_j O_i D_j c_{ij}^{-\beta}$$

(4)

with

$$A_i = \frac{1}{\left( \sum_k B_k D_k c_{ik}^{-\beta} \right)}$$

(5)

and

$$B_j = \frac{1}{\left( \sum_k A_k O_k c_{kj}^{-\beta} \right)}$$

(6)

to ensure that equations (2) and (3) are satisfied.

These $\{A_i\}$ and $\{B_j\}$ are the expressions that resemble partition functions in statistical mechanics. This led to the Boltzmann statistical mechanics analogy. The key idea is a very simple one. Think of a set of "boxes," one for each origin–destination ($i, j$) pair. A state of the system can then be an assignment of individual members of the population to these boxes. One extreme state would be one in which all of the population is assigned to one box; another is one in which there was an even distribution across all boxes. How many ways are there of assigning members of the population to these boxes? Because the people are assumed to be identical, the mathematics of permutations and combinations tells us that the number of possible combinations in this case is

$$W = \frac{T!}{(\Pi_{ij} T_{ij}!)}$$

(7)

where $T$ is the total population and $\Pi_{ij}$ is the product over all $i$ and $j$, and "!'" represents "factorial." The number of combinations associated with a particular matrix $\{T_{ij}\}$ can be taken as the probability that the given $\{T_{ij}\}$ occurs. What Boltzmann achieved in statistical mechanics is to show that—subject to any constraints that had to be satisfied—one of these states is overwhelmingly the most probable. We know that, in our case, equations (2) and (3) have to be satisfied, and we also add the transport equivalent of the physicists’ "energy" constraint,

$$\sum_{ij} T_{ij} c_{ij} = C$$

(8)

where $C$ is assumed to be the total expenditure on transport and the sum is over all $i$ and $j$. The most probable state can be found by maximizing $W$ in equation (7) subject to equations (2), (3), and (8). $W$ can be transformed by Stirling’s approximation (Wilson 1970a, p. 5) to the form

$$W = -\sum_{ij} T_{ij} \log(T_{ij})$$

(9)
which can be recognized as a form of entropy, yielding the notion of "entropy-maximizing models." This result produces

\[ T_{ij} = A_i B_j O_i D_j \exp(-\beta c_{ij}) \]  

(10)

with

\[ A_i = 1 / \left[ \sum_k B_k D_k \exp(-\beta c_{ik}) \right] \]  

(11)

and

\[ B_j = 1 / \sum_k A_k O_k \exp(-\beta c_{kj})^2 \]  

(12)

This result is more or less the same as equations (4)–(6) but with one difference: the power law for the cost function has been replaced by an exponential function. This is one of the instances where the entropy methodology provides new insight. The core model essentially assumes that people perceive their travel costs as increasing linearly. If this is changed to logarithmic—that is, the perception is \( \log(c_{ij}) \)—in effect giving less weight to longer trips, then replacing \( c_{ij} \) by \( \log(c_{ij}) \) in equation (8) reproduces the power function.

There was a valuable early exposition of the principle by Gould (1972) and another by Senior (1979). It even now appears in Wikipedia.

### A family of spatial interaction models

This application of the method reveals a general idea. If we want to model a system state for which we can define an entropy function and that has Weaver "disorganized complexity" characteristics, we then represent our knowledge of the system in a set of constraint equations and find the most probable state—which then becomes the model equations—by maximizing the entropy subject to these constraints. This argument led to the idea of a family of spatial interaction models (Wilson 1971).\(^3\) The transport model can be taken as production–attraction constrained.\(^4\) Dropping the attraction (destination) constraints generates a production-constrained model, and, similarly, attraction-constrained and unconstrained models can be defined.

Hybrid models also are possible; for example, where production constraints apply at the origin, but only some of the zones are constrained at the destination. This specification can apply, for instance, in a residential location model in which the origin is a workplace, and workers are being assigned to housing. Once the housing supply in a location is occupied, then it becomes constrained. This constraint is handled by defining a set, say \( Z_1 \), in which the destination zones are constrained, and a set, \( Z_2 \), in which they are not. This model takes the form

\[ T_{ij} = A_i B_j O_i D_j \exp(-\beta c_{ij}), \ j \in Z_1 \]  

(13)

\[ T_{ij} = A_i O_i W_j \exp(-\beta c_{ij}), \ j \in Z_2 \]  

(14)
In this case, an outer iteration is needed to determine which destination zones are in which sets.

**Entropy-maximizing models as location models: the retail model as an archetype**

The production-constrained model\(^5\) is particularly important for geography because it adds a locational dimension to the spatial interaction model. The retail model, which can be used as an archetypal location model as the argument develops, illustrates this as well as the general principle. The retail model has a long history—from Reilly (1931) to Lakshmanan and Hansen (1965). “Retail,” in what follows, can be interpreted very broadly: any system of interest where there is a flow from a population area to some kind of facility. It has been applied to the flow of students to universities, for example (Wilson 1995).

Let \( f_{Sij} \) be a matrix of money flows from residents in each zone \( i \) to retail centers in each zone \( j \).\(^6\) Let \( e_i \) be the per-capita expenditure in \( i \), and \( P_i \) the population, so that \( e_iP_i \) is the total retail expenditure leaving \( i \). Let \( c_{ij} \) be transport costs as before, with \( C \) being the total transport expenditure. We now introduce a new idea: suppose residents gain a benefit from using centers of a particular size that is proportional to \( W_j \), say, the size in floor space of a center. Let \( X \) be the total of such benefits. Then we can maximize an entropy function subject to the constraints that represent our knowledge of the system:

\[
\text{Max } S = -\sum_{ij} S_{ij} \log(S_{ij})
\]  

subject to

\[
\sum_j S_{ij} = e_iP_i
\]  

\[
\sum_{ij} S_{ij} \log(W_j) = X
\]  

\[
\sum_{ij} S_{ij}c_{ij} = C
\]

which, after the usual manipulations, leads to

\[
S_{ij} = A_i e_i P_i W_j^a \exp(-\beta c_{ij})
\]

with

\[
A_i = 1/\left[\sum_k W_k^a \exp(-\beta c_{ik})\right]
\]

Because the attraction “end” is unconstrained, we can calculate

\[
D_j = \sum_i S_{ij}
\]
which is the total amount of money attracted to (i.e., spent at) a center. We thus have a model that predicts a key locational variable, \( D_i \).

\( W_j \) is often referred to as an attractiveness factor and is modified by the deterrence term \( \exp(-\beta c_{ij}) \), so that \( W_j \exp(-\beta c_{ij}) \) is the pulling power of \( j \) for zone \( i \). 

\( 1/A_i \) can be viewed as a measure of the competition of other centers.

**Generalized cost: utilities and disutilities**

A number of discoveries were made en passant in the development of entropy-maximizing models. One of the most important was the use of the concept of generalized cost. In effect, this was a way of combining disutilities. For example, that travelers valued different kinds of time differently was well known; for instance, time to a public transport facility, travel time, or time to final destination. To these could be added fares, gasoline costs, or other costs, so that the generalized cost of travel could be represented by adding these components:

\[
c_{ij} = \sum_n a^n x_{ij}^n
\]  

We can write

\[
W_j a^\beta \exp(-\beta c_{ij}) = \exp\left[\alpha \log(W_j) - \beta c_{ij}\right]
\]  

so that a term proportional to \( \log(W_j) \) can be interpreted as the utility of a consumer using a facility with attractiveness, \( W_j \). As with generalized cost, this utility can comprise a number of components:

\[
W_j = \chi_j^{(1)} \chi_j^{(2)} \ldots \chi_j^{(n)}
\]

so that \( \alpha \log(W_j) - \beta c_{ij} \) can be interpreted as a utility function. Modifying the two functional forms in this expression is straightforward. This modification is very important in developing the economic interpretation of the model.

This result conveniently links to an earlier competing model that was offered in *Entropy in Urban and Regional Modelling*: Stouffer’s (1940) intervening opportunities model. His key idea is that the probability of going to a particular destination depends on the number of intervening opportunities: the more there are, the lower is the probability. This specification can be reconciled with the entropy-maximizing method by constructing a special cost function.

**Disaggregation**

Different groups clearly have different characteristics of travel, consumer choice, and so on; and the characteristics of transport systems vary by mode. These features can be dealt with straightforwardly by disaggregation. If we define \( T_{ij}^{kn} \) to be the number of trips from \( i \) to \( j \) by people of type \( n \) by travel mode \( k \), then the constraint equations can be disaggregated appropriately, and a model can be specified in the following form:

\[
T_{ij}^{kn} = A_i^o B_j^o O_i^n D_j^o \exp(-\beta^n c_{ij}^k)
\]
Note that, in this case, the generalized cost varies by mode. And, of course, there could be further disaggregation by purpose of trip. A variety of such models have been widely used. For example, the decision-making sequence is sometimes considered to be sequential: “Where to go?” followed by “Which mode?” Accordingly, the model takes the form

\[ T_{ij}^n = A_i^n B_j^n O_{ij}^n D_{ij}^n \exp(-\beta^n C_{ij}^n) \]  

(26)

followed by a calculation of modal split, \( M_{ij}^{kn} \), defined by

\[ T_{ij}^{kn} = T_{ij}^n M_{ij}^{kn} = T_{ij}^n \exp(-\lambda^n C_{ij}^k) / \sum_k \exp(-\lambda^n C_{ij}^k) \]  

(27)

where \( C_{ij}^n \) is a composite generalized cost that can be shown to be determined from

\[ \exp(-\beta C_{ij}^n) = \sum_k \exp(-\beta c_{ij}^k) \]  

(28)

demonstrating that \( C_{ij}^n \) is some kind of exponential average over modal costs, which makes sense intuitively. A new parameter, \( \lambda^n \), now determines sensitivity to modal costs, with \( \beta^n \) being responsible for sensitivity to “distance.” This model was deployed in the SELNEC Transportation Study in the 1960s and anticipated the development of the nested logit model (cf. Wilson et al. 1969).

Similarly, a disaggregated retail model takes the form

\[ S_{ij}^{ng} = A_i^{ng} e_i^{ng} P_j^{ng} W_j^{gh} \exp(-\beta c_{ij}) \]  

(29)

where \( n \) is a person type, \( g \) is a type of good, and \( h \) is a type of retail center/store. The disaggregation of the component variables is obvious and can be changed and developed further if necessary. In this case, transport mode, for example, has been omitted but could be added if needed.

**Interpretations: relating functions and parameters to constraint equations**

One of the advantages of the entropy-maximizing formulation is the way it relates the model parameters, such as \( \alpha \) and \( \beta \) in the aggregate model, to the constraint equations. In the case of the retail model, for example, the nature of equations (17) and (18) allows the disutility of transport, \( c_{ij} \), to be perceived to grow linearly (which generates the exponential function in the model), while the increase in benefits can be seen to grow with the log of attractiveness (which produces the power function in the model). This argument can be turned around: when a model is fitted to data, if \( c_{ij} / C_0 \) fits better than \( \exp(-\beta c_{ij}) \), then transport disutility may be interpreted as being perceived logarithmically. This occurs, for example, for a model that involves many long trips—say, an interurban model. The argument can be taken further still: any function can be tested in a model and the constraints adjusted accordingly (cf. Haynes and Phillips 1987).
A wide range of application
We cite the modeling of transport flows and retail flows as prime examples, but the models are usually applicable—with appropriate adjustments for, say, attractiveness terms—to almost any situation that involves flows, and to locational structures, analogous to the retail case, that are based on flows. Often these models are components of other models. For example, models of migration are components of demographic models (cf. Rees and Wilson 1976; Willekens 1994; Rogers 2008). Straussfogel (1991) examines the dynamics of suburbanization. Models of trade flows are also important and more interesting, because in this case, there are additional constraint equations to those that determine the flows—the input–output relationships. The model-building principles still apply, and an integrated model can be built (Wilson 1970b and extensively developed, particularly by Hewings; see, e.g., Kim, Boyce, and Hewings 1983; Roy and Flood 1992; Oosterhaven 2005).

System representation and model calibration
The spatial representation we use is a system of discrete zones. Use of a continuous space representation is possible in principle, and a set of entropy-maximizing models was developed by Angel and Hyman (1976) on this basis. Their models work with trip densities and velocity fields. However, over the years, the discrete space representation has proved easier to handle, both in practice because of data sources and in mathematical terms.

Many other issues must be resolved when building models to represent real situations. So far, we have implicitly assumed that the system of interest is closed. In practice, with spatial interaction phenomena, this is not the case. A ring of zones external to the system of interest usually need to be added to close the system, including a “rest of the world” zone. The skill is to choose the system of interest and the external zones to achieve an effective and realistic representation.

The model then has to be estimated with real data. The simplest approach is to use maximum likelihood for this purpose. To estimate $\beta$, for example, simply requires solving the equivalent constraint equation in the entropy-maximizing formulation, for example, equation (8). However, a matrix of observations is necessary to do this. This usually can be achieved with good sample data that, for instance, determine the average cost of travel. Batty (1976a) provides a good account of the development of excellent calibration techniques. Issues of aggregation and spatial structure and their effect on parameters of spatial interaction models are explored by Fotheringham and Webber (1980) and Batty and Sikdar (1982). A mass of associated statistical analysis of spatial patterns that complement these core calibration methods also exist, not least in relation to spatial autocorrelation (Berry, Griffith, and Tiefelsdorf 2008; Griffith 2009; and the associated bibliographies). Relatively little work addresses errors or fluctuations (see Leung and Yan 1997). Indeed, econometric versions of spatial interaction have been developed by LeSage and Fischer (2010) and Fischer and Griffith (2008).
Missing data
Entropy-maximizing models are good for estimating missing data too. In one sense, this is what is done in a transport model that has been estimated using sample data. But the model can also be used to estimate, in nonspatial models, for example, the elements of a matrix when row and column sums are known or the cells of a three- (or higher) dimensional array when partial sums are known. Johnston provides the best example of this application through his work, with colleagues, about election data. He uses entropy-maximizing methods to estimate split-ticket voting, for instance, with results being tested with New Zealand data (see Johnston 1982; Johnston and Pattie 2003, 2009). He does make the point, however, that these techniques could be used much more widely in the social sciences, for example, to estimate linking data from the census, but this is rarely done. He cites Dempster, Laird, and Rubin (1977) as an early example, along with Chilton and Poet (1973) in relation to migration data. The entropy-maximizing method in these contexts provides an alternative to King’s methods relating to the ecological inference issue (cf. King 1997).

The wider context: statistical mechanics, statistics, and information theory
As noted in the introduction, my own motivation in developing the method that was published in my 1967 article was to pursue an analogy with statistical mechanics to find a way of reproducing the partition function-like terms in the then-standard “gravity” model. The simplest way this could be done was through the method originally used by Boltzmann in his theory of a classical gas (cf. Wilson 2008). This approach deployed what the physicists would describe as a microcanonical ensemble. I also explored what the canonical and grand canonical ensembles could offer. Because most of the texts dealt with statistical mechanics in terms of quantum mechanics, I reverted to the magnificent text by Tolman (1938), which gives a full classical treatment, including a method attributed to Darwin and Fowler based on contour integration. All of these approaches led to the same result. However, much more recently, the methods of statistical mechanics have become more widely recognized as applying to a larger class of problems involving more than analogy (Ruelle 1984), and there are new contributions that will be noted later in this essay.

That entropy maximizing sat in a wider context than statistical mechanics, with ideas rooted in physics but having much wider implications within statistics, became clear to me between 1967 and the publication of Entropy in Urban and Regional Modelling in 1970. Two key figures were Jaynes and Tribus. Jaynes (1957a, b, 1979) published two early, hugely influential papers, and his account of the history of the concept in a book edited by Levine and Tribus (1979) remains fascinating reading. He describes the work of Bayes that led to Bayes’ Theorem, as developed and articulated by Laplace before statistics was taken over by a “frequency” perspective. Only from about 1960 onward was Bayesian statistics fully
developed (cf. Good 1950) and not fully implemented until around 1990. The point in this context is the core of the Bayesian methodology is essentially entropy maximizing, and the maximum likelihood equations to be solved in parameter estimations are the constraint equations in an entropy formulation. This equivalence is noted in my 1970 book, though I had a far-from-full understanding of the historical context. Later (Berzins and Wilson 2003), I also pursued it in what I now see to be a frequency context using the concept of Fisher information. This pursuit stemmed from a book by Frieden (1998) titled Physics from Fisher Information. Again, this approach produced an equivalent result. (Also see Smith (1976) for alternative statistical approaches; Smith and Hsieh (1997) for a Markov equivalent.) Batty (1974, 1976b) added spatial weightings to the entropy function, an idea that has been developed by others but not extensively applied (though see Paulov 1991).

A different approach that brought the idea of entropy into a wider arena was developed by Shannon and presented by Shannon and Weaver (1949). Shannon wanted a measure of the amount of information that was transmitted in signals and used $-\sum p_i \log(p_i)$, where $\{p_i\}$ can be taken as the probabilities in a distribution. He also realized that the inverse of “information” was “uncertainty”: the higher the value of the expression, the more uncertain was the knowledge. Tribus (1979) reports a 1961 interview with Shannon in which Shannon told of his explorations of what to call his function. “Information” was already overused. He consulted von Neumann, whose “response was direct”: “You should call it ‘entropy’ and for two reasons: first, the function already has a use in thermodynamics of that name; and second, and more importantly, most people don’t know what entropy really is, and if you use the word ‘entropy’ in an argument you will win every time!”

Another different kind of wider context also existed: the introduction of “balancing factors,” such as $A_i$ and $B_j$, turn up in a number of areas (cf. Fratar 1954; Bacharach 1970). In economics, Richard Stone (1967) used them in his “RAS method” to adjust input–output tables.7 As we note, the equations involving two sets of balancing factors have to be solved iteratively. In practice, convergence is very rapid (when a constant of proportionality is included). It became clear that there was some deep underlying mathematics that was at the root of this: Brouwer’s (1910) fixed point theorem, which appears in many contexts.

**Developments since 1970**

By 1970, much had been revealed through the introduction of the entropy-maximizing model. The breadth of application was in place, particularly the assembly of internally consistent comprehensive models, building on Lowry’s (1964) original masterpiece. Much polishing and many developments, however, were to follow. Here, we consider, in turn: extensions to the core model; the recognition that the formulation was a nonlinear mathematical program so that all the tools of that field could be applied; its role in economics; and the major leap to incorporate spatial interaction into a dynamic location model.
Extensions of the core model

Snickars and Weibull (1977) show how to incorporate prior information into a model by using an entropy term in the form $-\sum_{ij} T_{ij} \log(T_{ij}/T_{ij}^0)$, where $T_{ij}^0$ is a known estimate of the matrix. Fotheringham (1983) modifies the model to give emphasis to the destinations that most compete with a particular zone. Pooley (1994) extends the family of models. A major task was to integrate the spatial interaction model with network assignment models. This arises because the transport model usually comprises four components: generation, distribution, modal split, and assignment. The trip generation component is straightforward in principle: submodels to estimate $O_i$ and $D_j$. The entropy-maximizing model is the distribution model, and disaggregation copes with the modal split element. Assignment is more challenging. The transport flows are, of course, over a real network, which can be represented formally as sets of nodes and links. Some of the nodes are the centroids of zones and represent origins and destinations. A route from an origin to a destination is a set of intermediate nodes and links. A number of issues then have to be resolved: do all travelers use the least-cost route, or are they spread across several routes? The generalized cost itself can be thought of as a sum of link costs. Typically, each link carries a number of origin–destination flows, and the cost on any given link is a function of the volume of the total flow on that link; this formulation is necessary to represent congestion. This conceptualization suggests that a grand iterative scheme is needed: calculate the trip matrix, given $\{c_{ij}\}$, load these trips onto the network, and then recalculate each $c_{ij}$ as a sum of link costs, taking traffic volume into account. The idea of an integrated model lies in Beckmann, Maguire, and Winsten (1956). It was the achievement first of Evans (1973, 1976) and further developed by, for example, Boyce (1978) to offer methods of solution. These models, though, are big and expensive to develop and run, resulting in less work with them than might be desirable. When this integration can be achieved, it fully incorporates congestion into spatial interaction models. However, the resulting model is obviously complex to implement in full, which is rarely done. The methods used by Boyce rely on mathematical programming; we pick this point up in the following subsection.

One of the problems of adding more and more detail, through disaggregation or combination with other submodels, is that the arrays become bigger and bigger and an increasing number of their cells become zero. From a computational point of view, the storage of the arrays is very inefficient. Griffith (2009) provides an example of intercounty commuting in Texas in which the core array has 65,000 cells, of which 55,000 contain zeros. This led to the idea of what has become microsimulation: generating a hypothetical population where each member has a set of characteristics that is consistent with the aggregate arrays. This formulation also builds on the work of Orcutt (1957) and, in the context of entropy-maximizing models, Wilson and Pownall (1976), and is now in common use. See Clarke (1996) for an exposition and Ballas et al. (2007) for some applications.
Links to mathematical programming
The core entropy-maximizing model is a nonlinear mathematical programming problem in which an objective function is maximized subject to a number of constraints. This means that the mathematical apparatus associated with this general class of problems becomes available. One important example is the use of dual formulations in which the Lagrangian multipliers become the main variables, which then can be interpreted as comparative advantage—and hence as rents or shadow prices—an important notion for economic interpretations. The $A_i$s and the $B_j$s in the core model (equations [5] and [6]) are proportional to the logarithms of the Lagrangian multipliers. The full argument is presented in Wilson and Senior (1974)8 (also see Coelho and Wilson 1977).

A particular consequence of this representation relates to the transportation problem of linear programming. In this case, the cost constraint (8) in the formulation of the usual model becomes an objective function to be minimized subject to the trip end constraints, that is, to find a way of generating a distribution matrix that minimizes overall costs. Because in the usual model the parameter $\beta$ determines average trip length, and the higher the value of $\beta$, the shorter this trip length is, then intuitively we would expect that as $\beta \to \infty$, the solution of the entropy model would converge on the solution of the linear-programming model. This conjecture is proved in Evans (1973). A major insight follows from this result that is not yet fully appreciated: entropy generates a most probable suboptimal—in the economic sense—model but still incorporates rents or shadow prices through the dual. Much of economics is concerned with optimization of a form that is analogous to the transportation problem of linear programming. Adding entropy makes this kind of model realistic.

A further alternative perspective is provided by a form of mathematical programming known as geometric programming (cf. Scott and Jefferson 1977; Samanter and Majumber 2007).

Economics
The entropy-maximizing model was developed in the course of a project on cost–benefit analysis in transport planning, building on Foster and Beesley (1963). Therefore, beginning a review of links with economics by noting its contribution to the foundations of cost–benefit analysis is appropriate. This was fully worked out by Williams (1977; and see Wilson et al. 1981; Williams, Kim, and Martin 1990), taking advantage of the mathematical programming basis of the model (also see Miller 1999a, b).

More traditionally, in economics, the demand for travel, interpreted through probabilistic expressions—the probability that a commuter will select a given mode, or a resident will choose to live in a particular area—has been widely studied through the microeconomics of discrete choices. The usual treatment involves an assumption of a rational economic agent selecting an alternative that results in maximum satisfaction or utility. Dispersion is introduced either through assuming
that the parameters of the utility function are random and/or that an observer (the modeler) has limited knowledge of the factors influencing choice. By attributing random elements to a utility function and, in particular, by assuming that the utility function can be expressed as a “representative value” plus a random component, a set of random utility maximizing models can be formulated. More specifically, if the random component is expressed as a Gumbel distributed variable, the probability of choice can be expressed as a multinomial logit model. Various researchers draw attention to the similarities and differences between the utility-maximizing and entropy-maximizing approaches, and, in particular, to their respective derivations of the multinomial logit model (for more details, see Wilson 1981a; Anas 1984; Train 2003).

Spatial structure: the slow dynamics
A significant breakthrough appears in an article by Harris and Wilson (1978) that extends entropy maximizing in a fundamental way by embedding the spatial interaction models within a framework that represents slow dynamics. We have seen that the entropy-maximizing model functions as a location model, using the retail system as an example. If we continue to use the retail model as an archetypal example, we can show how the argument can be taken a step further. That model functions as a location model in that revenues accruing to a retail center can be totaled. However, the measure of attractiveness of each center, $W_j$, in a model, is taken as given. The deeper location question is can we model the structural vector $\{W_j\}$? And following implicit assumption in the case of the spatial interaction model generates the array $\{S_{ij}\}$ because this is an equilibrium solution, if a system is disturbed, the return to equilibrium will be rapid. This assumption seems reasonable intuitively and can be thought of as “fast dynamics.” In the case of the $\{W_j\}$, intuition also tells us that change is relatively slow and to model this, we move into the territory of “slow dynamics,” which is both interesting and more difficult.

The Harris and Wilson hypothesis is

$$\Delta W_j(t, t+1) = \varepsilon[D_j(t) - KW_j(t)]W_j(t)$$

(30)

where $\Delta W_j(t, t+1)$ is the change in a time period. This equation simply says that if the center at $j$ is profitable, it grows, and vice versa. The parameter $\varepsilon$ measures the strength of the response. The factor $W_j$ is added for convenience. It changes behavior near the origin but not the equilibrium position. The equations (30) then look like a multispecies Lotka (1925) and Volterra (1938) model, where $\{W_j\}$ are populations of species competing for resources. In our case, the “species” are retailers competing for consumers. These equations resemble those by May (1973), who shows the possibilities of chaos in such systems. The equilibrium position is given by

$$D_j = KW_j$$

(31)
If substitutions are made from equations (19)–(21), then equation (31) can be rewritten as

$$X_i/C_2 e^{PiW} a_j \exp(-C_0 b c_{ij}) = X_k W_k a_k \exp(-C_0 b c_{ik}) = KW_j$$

(32)

showing a quite complicated set of simultaneous nonlinear equations in \( \{W_j\} \). This result is a rich seam for geographers as it creates the possibility of modeling existing geographical structures (insofar as they are at or near equilibrium) and their evolution over time. This broader theme is developed in the penultimate section of this article. Because the system embraces a Boltzmann (B) methodology for the interaction terms and a Lotka–Volterra (LV) methodology for spatial structure, equation (32) can be usefully thought of as a BLV model. Although this argument has been cast in terms of the retail model, BLV models embrace a wider class in a number of disciplines (Wilson 2008).

A significant number of authors apply this model in different situations. Wilson and Oulton (1983) show how it offers an explanation of the corner shop to supermarket shift in food retailing—which was very rapid and could be considered to be a phase change. Wilkinson (1990) applies the model in a marketing context. Other explorations include Phiri (1980), Clarke (1981), Rijk and Vorst (1983a, b), Borgers and Timmermans (1986), Clarke, Clarke, and Wilson (1986), Fotheringham and Knudsen (1986), Lombardo (1986), Oppenheim (1986), Pumain, Saint-Julien, and Sanders (1987), Baker (1994), Clarke, Langley, and Cardwell (1998), and Lombardo, Petri, and Zotta (2004). The possibilities of abrupt change with these models can be explored in terms of catastrophe theory (Wilson 1981b), and Batty (2009b) notes the possibility of sequences of such changes, “catastrophic cascades.” Alternative explorations of structural dynamics are, for example, furnished by Allen and Sanglier (1979) and Nijkamp and Reggiani (1987, 1988a, b, 1989). More recent explorations are much more fully developed with vastly improved computer visualization capabilities, and this topic is further discussed in “A mathematical puzzle” as part of the research agenda.

**Entropy maximizing in geography and complexity science**

Weaver’s (1948, 1958) notions of disorganized and organized complexity, introduced in the opening section of this article, provide a framework for recognizing entropy as a superconcept (Wilson 2010) that crosses disciplines. In this section, I review the significance of entropy’s use within geography, noting that this also reveals the role of geography as a discipline that permeates—or should permeate—many others and that acts as a lead agent in the development of complexity science.

Models of spatial interaction and location lie at the heart of one branch of geographical theory, with entropy-maximizing models representing a substantial subset. That there will be further development, there is no doubt. Applications are widespread, and as new tools become available (currently through complexity
science), major research issues can be resolved, ones that would facilitate building effective models of the functioning and evolution of cities. A wide range of successful and useful applications of the models exist in many industrial and service sectors. However, much of this current development and application is taking place outside geography. “Quantitative geography” is out of fashion (cf. Berry, Griffith, and Tiefelsdorf 2008; Johnston 2008) precisely at a time when demand is at its greatest. The insights of the quantitative and qualitative approaches are complementary but are not always regarded as such. They should not be seen as competing approaches within the discipline. If the ongoing challenges are not resolved within geography, they will be taken on by others—probably in newly developing fields such as complexity science—that are already well funded by research councils all over the world. In the rest of this section, I explore some of these challenges.

Spatial interaction systems—systems of flows—are nearly always systems of disorganized complexity, and in these cases we have seen that statistical averaging produces good models, the essence of entropy maximizing. Spatial structures are components of systems of organized complexity, but we have shown that, by incorporating interaction models within a dynamic framework, such as variants of Lotka–Volterra equations, these can be modeled as well. This structural evolution model is at the heart of complexity science. If we combine both of these arguments (i.e., spatial averaging and spatial structure), we can see that many systems of interest have both interaction and locational elements, and, as such, the understanding of these systems achieved through modeling should offer a platform for geography to have a bigger role in the interdisciplinary research of the future. A good, straightforward example is provided by history: many examples exist of interaction and location being critical to historical analysis. Two examples can be cited. First, in the work of Rihll and Wilson (1987a,b, 1991) about Greece in the ninth century BC, point data for settlement structures were available from archaeological explorations and were used within a “retail” model to make estimates of population sizes at different locations. This approach is further developed by Graham and Steiner (2006). More recently, Wilson and Dearden (forthcoming) have used the BLV framework to model the impacts of railways on urban systems’ development in the United States between 1790 and 1870. Earlier approaches also essentially use Lotka–Volterra dynamics but without the spatial interaction submodels; see, for example, Dendrinos and Mullally (1981), Orishimo (1987), and Cappello and Faggian (2002).

In what sense is geography a lead agent in complexity science? If we restrict ourselves to urban geography, the methods of BLV modeling are essentially those that complexity theorists are working with in other disciplines, though the connection is not widely recognized. Cities are more than complex enough to be interesting; through BLV and other modeling, more progress can be made than with, say, complex systems in biology. Therefore, geographers must make this case within the broader complexity science area; much mutual learning and research could materialize that would be important to geography.
What about entropy-maximizing models—in their extended BLV form—in geography itself? An extensive range of applications exists. A search on Google Scholar reveals almost a thousand citations of key references to *Entropy in Urban and Regional Modelling* (the magnitude of this outcome takes a comprehensive review of this literature beyond the scope of this essay). Many more citations of related papers also exist, and much current work appears in which the methods are used but the core references are not cited! This is, perhaps, the price of longevity. The fruits of the commercial exploitation of these methods, particularly in the retail industry, are not represented in the literature to an extent that reflects the number of applications carried out. For example, national and international companies that need retail outlets—supermarkets, motor manufacturers, and banks, for example—routinely use these methods for optimizing branch networks and for testing potential new sites. An advantage of these applications is that the models have been very thoroughly extended (through disaggregation) and tested. Much of this work was carried out by GMAP Ltd. from the 1980s onward (Birkin et al. 1996; Birkin and Culf 2001; Clarke and Clarke 2001; Birkin, Clarke, and Clarke 2002). More broadly, the range of application in applied geography is reviewed in Bennett and Wilson (2003); for a contemporary review, see Batty (2009a).

However, perhaps the most important impact of entropy-maximizing models—in the BLV framework—in geography is their potential for rewriting the “classical” models of the discipline, with implications for teaching as well as research. This effort is largely completed, in principle, first in a series of papers (Wilson 1978; Birkin and Wilson 1986a, b; Wilson and Birkin 1987), summarized in Wilson (2000). The “classical” models can be taken as

- von Thünen (1826): agricultural land use (cf. Stevens 1968);
- Weber (1909): industrial location;
- Hotelling (1929), Palander (1935), and Hoover (1937, 1967): market areas;
- Christaller (1933) and Lösch (1940): central place theory;
- Reilly (1931) and Zipf (1946): retail and gravity models;
- Burgess (1927), Hoyt (1939), and Harris and Ullman (1945): urban structure.

These models are well known in their core forms (e.g., von Thünen’s rings, Weber’s triangle) though in many cases the authors in their writings were well aware of further real-world complexities. These models are restrictive in one or both of two respects: monocentricity and/or nonoverlapping market areas. Von Thünen, Hoyt, Burgess, and Harris and Ullman all had monocentric systems of interest; all had nonoverlapping areas of various kinds. These restrictions are easily overcome in the entropy-maximizing BLV formulation. The models are articulated explicitly in Wilson (2000), with the classical models being reproduced as special cases. However, the insights into the archetypal problems that can be gained from this position have not been taken advantage of. Given the relative lack of mathematical expertise in the discipline and in most university geography courses, what may be needed is an equivalent of Gould’s (1972) exposition on this wider canvass.

379
The ongoing research agenda

There is a development agenda to be implemented, building on existing knowledge, and a number of research challenges to be met. These challenges include the integration with other approaches, the full exploitation of the properties of nonlinear systems, including path dependence (which leads to the notion of urban “DNA”), and the full articulation of the underlying mathematics.

Implementing the development agenda

The preceding section indicates an ongoing richness of potential for applications of the entropy-maximizing concept. The essence of the idea is to predict the most probable state of a system of interest subject to what is known. “What is known” is what is captured in a set of constraint equations, both in the equations themselves and in the forms of functions within them. An initial part of the research agenda, therefore, is to collect ideas for the further development of what we already understand. We can then move on to new challenges.

In “Links to mathematical programming,” I outline some extensions to the core model. Some of these ideas can be implemented straightforwardly, such as the use of prior information or competing destinations. In these cases, the hypotheses of what is known about individual decision making and how to incorporate that into either the objective function or the constraints are the issue. The integration with network assignments is different: this would mean progressing to a level of detailed system representation that is often considered too costly to take on in an academic context, but it is widely used in practice.

Exploiting the theorems and concepts of mathematical programming as indicated in “Economics” is also clearly possible. Too little has been made of dual formulations and shadow prices, for example. This connects in particular to the economic interpretations of the model (cf. Herbert and Stevens 1960; Alonso 1964; Senior and Wilson 1974; Wilson and Senior 1974). Much of the theory has been carefully worked out, for example, for consumers’ surplus, but this result is not being exploited systematically. Similarly, the relationship to logit models of various kinds has been fully worked out by Williams (1977), but the spirit of entropy maximizing—essentially in this instance, a means of modeling consumers’ (or other sources of) imperfect knowledge and hence representing forms of imperfect markets—has not been absorbed into economic theory. Entropy maximizing provides a tool that facilitates the development of internally consistent models that can incorporate whatever economic hypotheses are required. The modeling of trade flows within multiregional input–output models is a good example of this, but there is almost certainly a wider range of potential applications.

An appropriate conclusion to this section is a comment that applies across the development agenda. I return to the general idea of representing “what is known” in constraint equations. A particular aspect of this known information is the distinction between exogenous and endogenous variables, and much of the ongoing research agenda can be focused on transferring variables from the exogenous to the
endogenous category. For example, consider the “total cost” term, $C$, in a transport model, as in equation (8). This quantity is normally considered exogenous, meaning that the associated Lagrangian multiplier, the parameter $\beta$, has to be estimated statistically, in effect from sample data that estimate $C$ and then $\beta$ through the use of equation (8) with a maximum-likelihood technique. If we could model $C$ in a separate submodel as a function of other variables, then both $C$ and $\beta$ would become endogenous to the model (cf. Wilson 1970a; Haynes and Phillips 1987).

A second example concerns the cost term, $K$, in the spatial structure model, in equation (31). If, for the purposes of a simple illustration, we have been taking $W_j$ as measured by size, then $K$ is a cost-per-unit size. More generally, however, we could see $KW_j$ as the production function for zone $j$, which could be modeled separately and then incorporated endogenously into the spatial structure model. And, of course, $K$ should have spatial variation.

**Research challenges**

Although much of the theory of entropy-maximizing BLV models is either nearly fully worked out, as in the spatial interaction case, or established in principle, as in the spatial structure case, much work remains to be done. In this section, I consider some of the more challenging areas for future research. In the next subsection, six examples are presented for which we might seek a fuller integration with other approaches. Next, the idea of urban “DNA” is introduced. Then a specific mathematical challenge is posed. Finally, the need for a deeper understanding of the underlying mathematics is established.

**Integration with other approaches**

First, I already noted connections between entropy modeling and economics in a number of respects, but a full integration with economic modeling could be very fruitful: in particular, using the notion of an entropy term in an objective function to simulate imperfect markets. Urban economic modeling is all too often trapped by a continuous space representation, which forces an analyst into an unrealistic monocentric city assumption. A major associated challenge is to be able to model the development of technology. A potential fruitful area for development is provided by the work of Medda, Nijkamp, and Reitveld (2009) in applying Turing’s (1952) idea of morphogenesis to the problem.

Second, applying the BLV methodology to ecology, and in particular to spatial analysis in ecology, has implications for that aspect of physical geography. Most ecological models are aspatial and neglect the movement of animals or, for example, wind-blown seeds, across space. An explicit discrete spatial system can be incorporated within a Lotka–Volterra framework, and then entropy-maximizing spatial interaction theory can be used to model the flows (Wilson 2006; also see Phillips 2004). Extending the argument further, the BLV-modeling framework can be applied to data from remote sensing; such techniques may provide the equivalent of X-ray spectroscopy for the identification of geographical “DNA” (Wilson forthcoming-a, forthcoming-b).
Third, a significant challenge arises from what has become named as network science. This is now an enormous enterprise with many researchers from a range of disciplines and with (probably) many thousands of papers in a diverse literature. This research focuses on the topology of a wide range of networks, often in the counting of links from each node and the distribution of these counts. Insofar as these counts can be said to relate to the importance of a node, this network analysis connects to our treatment of spatial structure through a vector such as \( \{W_j\} \), leading to an emphasis on scale-free networks and scale-free distributions of node sizes (cf. Wilson 2008). However, the bulk of this research is not connected to spatial interaction modeling, which can load flows onto networks and create a much richer picture. Above all, networks should be recognized as representing one element of a much more complex system of interest; embedding them within a BLV framework acknowledges this point. An application that does combine these ideas is Tomlin’s (2003) work about the World Wide Web. A major research problem exists in the study of networks: the evolution of the network system itself. The evolution of arrays like \( \{W_j\} \) relates to one aspect of urban structure. We may now be in a position to tackle harder problems (Wilson 1983; Wilson and Dearden forthcoming).

Fourth, the issue of integration with agent-based modeling (ABM) can be addressed. In this case, the system of interest is populated by individual agents who are given (probabilistic) rules of behavior. Interesting explorations link ABM with spatial interaction; compare Bura et al. (1996) and Heppenstall, Evans, and Birkin (2005). A research challenge is to investigate whether a set of rules can be specified guaranteeing that the model outcome is equivalent to the entropy-maximizing one when the ABM results are aggregated (cf. Dearden and Wilson forthcoming). This research may involve making assumptions about the possibilities of interaction beyond what are often taken as nearest-neighbor assumptions in ABM.

ABM, when the agents are “people,” can be considered to be a people-centered approach. This feature foreshadows the fifth major area: the extent to which the modeling paradigm that has been the main subject of this article is too place based. The point is eloquently argued by Miller (2005) in a paper that echoes Hagerstrand’s (1970) plea for an activity-based approach. In essence, this perspective requires a combined space–time representation within spatial interaction models in which the time dimension suggests that new constraints representing the combination of activities should be added; hence, this is also referred to as an activity-based approach. A long history of exploration of this idea exists, stimulated by Hagerstrand’s paper, but it has never been fully resolved and remains a research challenge. The predominant use of entropy maximizing in spatial interaction has been in the modeling of single trips. The activity approach is the endpoint of attempts to model “tours” and “circuits.” I made an early attempt to formulate the problem in economic terms (Wilson 1972), building on the work of Lancaster (1965), but it was neither fully developed nor integrated with entropy maximizing.
Sixth, much can still be learned from new developments in physics, particularly in relation to modeling spatial structure. Physicists have long had an interest in modeling the structure of molecules on lattices; for example, through Ising (1925) models. These models tend to be rooted in nearest-neighbor interactions, but this formulation was extended by Potts (1952) to incorporate longer-range interactions. This extension opened up, in contemporary times, a huge field of research in physics (see, e.g., le Bellac, Mortessagne, and Batrouni 2004). This extension reveals the possibility of a different kind of entropy-maximizing model applied to the structural variables, that is, applied to slow dynamics. The solution involves including an entropy term like $-\sum W_j \log(W_j)$ into a suitable maximizing framework. This formulation has been attempted by setting up a Hamiltonian framework (Wilson 2008); the results are identical to those of the Harris–Wilson model. Working out an explicit spatial “thermodynamics”—a “thermodynamics of the city” (Wilson 2009)—may also be worthwhile.

Nonlinear systems and urban “DNA”

A particular property of nonlinear systems possibly underplayed in urban and regional research is that, for the structural variables, in progressing, say, from time $t$ to time $t+1$, the equilibrium state at $t+1$ will be highly dependent on the initial conditions at time $t$. This property arises because of the possibility of multiple equilibrium solutions. It is this property that, through time (in effect a sequence of initial conditions) gives rise to Arthur’s (1988) notion of path dependence. It is a striking idea that the “future” of the structure of a system of interest may be so sensitively dependent on this sequence that the initial conditions, which are either exogenous and/or slowly changing, can play the same role as DNA in biological systems (Wilson forthcoming-a, forthcoming-b; cf. Silva 2004). A particularly interesting consequence is that, following a sequence of planned, exogenous changes, or simply by adding some noise, a possibility cone (in a large number of dimensions) of achievable future states (Wilson and Dearden 2010) can be generated. This in turn produces the idea of “genetic planning” by analogy with genetic medicine: a planning goal can be achieved by adjusting one of the initial conditions in such a way that a more substantial change is effected. This outcome can be seen intuitively by the way a landmark investment, such as Canary Wharf in London, can be the trigger for major subsequent development.

A topical example might be transport planning for a sustainable future. Consider the aggregate transport model for simplicity. Suppose we aim to reduce total expenditure on travel, $C$ in equation (8), or, equivalently, to increase the $\beta$ parameter. A research question would be to explore the following conjecture for some kinds of city: a switch to shorter trips and to public transport demands a higher-density residential and workplace structure. Such a structural change would be a shift in the “DNA,” and the model could be used to explore consequences in terms of both fast dynamics (e.g., mode switching) and slow dynamics (perhaps further private-sector high-density development).
The representation and understanding of these features of nonlinear systems—particularly path dependence and phase changes—is hugely facilitated by contemporary computer visualization techniques (see Batty, Steadman, and Xie 2006; Dearden and Wilson 2009).

A mathematical puzzle
Geographical analysts have long recognized that the distribution of settlement sizes in a system of interest is likely, empirically, to follow a power law—and hence to be scale free. A BLV model of spatial structure predicts such size distributions, through the $\{W_j\}$ vector, and when they are plotted (cf. Dearden and Wilson 2009), they have the same characteristic. Is it possible to derive from the BLV model a formula for the resulting size distribution of settlements? A solution to this problem might explain the property of scale freeness and power distributions in spatial structure—a prize worth having!

The underlying mathematics
The set of variables and parameters that describe a geographical system of interest constitute a state space: a set of values of those variables and parameters, a point in that space. One research challenge is to connect the models based on this description to some deeper underlying mathematics. For example, the constraint equations that limit the feasible solution space form a manifold in a very high-dimensional space; exploring the properties of such manifolds in a general way may furnish insights to spatial analysts.

This description leads to another area of research that has been partially, but far from fully, worked out: the extent to which alternative mathematical representations of entropy-maximizing BLV models are equivalent. I present examples of alternatives in this article, most of which are equivalent though with some differences in the case of random utility theory. As noted previously, these equivalences are rooted in Brouwer’s (1910) fixed-point theorem (and see Scarf 1973 for a review of fixed points in economics), and the role of this theorem in game theory and the uniqueness of Nash (1950) equilibria suggest that a game theory interpretation is worth exploring (cf. Williams, Kim, and Martin 1990). Similarly, early explorations of a formulation using neural networks (e.g., Openshaw 1993; Leung 1997; Reggiani et al. 1998; Fischer 1998) have now been significantly developed (Fischer 2002; Fischer and Reismann 2002; Fischer, Reismann, and Hlaváčková-Schindler 2003). These offer an interesting calibration procedure. A third example is the Markov approach of Smith and Hsieh (1997), and a fourth is the master equations approach of Haag (1990), essentially an alternative way of presenting the statistical mechanics. Each of these cases involves the probability of a person making a trip; they are more or less connected to entropy maximizing and through modeling this probability.

Expanding the earlier analysis of equivalences by Macgill and Wilson (1979) should be possible. Perhaps, if equivalences are established, the particular
approach adopted in any one case is a matter of taste. And we should also bear in mind that Occam’s razor applies.

Concluding comments

The entropy-maximizing approach delivers models that represent the most probable state of a system of interest subject to constraints representing a modeler’s knowledge of that state. The approach is particularly powerful when it is linked to constraints that are accounts. In the core applications to spatial interaction modeling, these account constraints connect trips to structural variables, such as number of workers at a residential location and number of jobs at employment locations. The models are essentially statistical-averaging models and so work effectively only when Weaver’s disorganized complexity conditions apply: large numbers of elements weakly interacting. These are the circumstances that govern most spatial interaction situations, and this is why the model and its variants are so widely applicable. The models contribute to multiregional demographic and input–output models, to transport, and to location modeling, such as retail. They are applied in historical and contemporary contexts. Substantial progress has been made by combining spatial interaction models with structural models that mimic Lotka–Volterra equations, which leads to the current major research challenge: to model the evolution of urban and regional structures. An outstanding “entropy” challenge in this field is to determine the extent to which the methods now being used by physicists to model solid-state and molecular structures can be translated into geography. There may yet be a generation of new models!

Acknowledgements

I am grateful to the editor, the three referees, and Dr. Martyn Senior for helpful comments that have improved this article. I am especially grateful to Professor Huw Williams, who sent me very detailed comments on an early draft. Any errors and omissions are my responsibility.

Notes

1 Throughout this paper, the braces { } denote a set of items.
2 Interestingly, the \{A_i\} and \{B_j\} are not precisely partition functions (see Wilson 2009).
3 The term “spatial interaction” was introduced to attempt to shift the analogy from “gravity model.”
4 “Production” = origin; “attraction” = destination.
5 Or its mirror image, the attraction-constrained model.
6 The \{i\} and \{j\} zone systems can be different.
7 The “R” and the “S” were, of course, his initials.
8 Martyn Senior (private communication) pointed out to me that the dual variables in this article should be corrected for zone size effects.
The book *Knowledge Power* (Wilson 2010) introduces the notion of superconcepts more broadly with “entropy” as a significant example.

**References**


