AN APPROACH TO THE STUDY OF NODAL GROWTH

Michael F. Goodchild
Nina Siu-ngan Lam
John D. Radke
University of Western Ontario

Introduction

In a recent review of the relevant literature, Griffith [6] has argu that the combination of theories of the size of spacing of cities wi models of spatial interaction and flow is a relatively unexplored b very promising and important field. Central place theory is primaril static, and its conclusions or predictions regarding spatial structur can only be accepted as an ideal to which real systems may tend witl variable consistency. If a dynamic element is to be added, as it mus if we are to understand real systems, it must comprise the growth anc decay of centres in response to changing patterns of spatial interaction, including migration, flows of goods, consumer behaviour, and so forth. Several studies have been made of the way in which centres respond to changing patterns of behaviour by dispersed consumers, represented through a spatial interaction model. White [9], for example, has simulated a system in which a gravity model is used to allocate dispersed consumers to centres. Each centre then responds by growing or declining in relation to the number of consumers it receives. Since the size of a centre appears in the interaction model, consumer behaviour must be re-evaluated in each cycle of the simulation. Goodchild [5] has described a similar problem as one of a large set of allocation problems in which consumer behaviour influences centre attraction (through growth, decay, crowding, congestion, and so on) which in turn influences consumer behaviour to complete the loop.

Stability in such systems occurs when the attraction level at each centre is precisely that level required to maintain the interaction which generated it. A major disadvantage as far as empirical research is concerned is that the state in which a system is found will depend very little upon the parameters of the processes involved; that is, the constants in the spatial interaction model and the attraction-interaction relationship. One state can be generated by wide ranges of parameters, and one set of parameters can lead to widely varying states depending on the initial conditions and history of perturbations in the system. Hence it is extremely difficult to infer process from form, or vice versa.

In the present paper we attempt to generalize some parts of this work by taking it out of the strictly tertiary context of central place theory. An increased market of dispersed consumers is treated as only one of the ways in which a centre can obtain an advantage and grow; comparable effects, of different duration, can be obtained by primary resource extraction or by increasing a base of secondary manufacturing. The approach is one of simulation, since such systems have a number of properties which tend to make analysis intractable.

The Simulation Procedure

At the beginning of each step in th

possess relative advantages or disadvantages then migration between pairs are symmetrical and no net population shifts occur. In reality a city's attraction for in-migrants is affected by a number of factors which perturb this symmetry. Job opportunities are attractive, and result from increased activity in all economic sectors, including the tertiary. Various environmental and residential qualities can either raise or lower attraction, leading to asymmetries in the migration flows between the city and others, and hence to net growth or decline of population.

Specifically, the flow between nodes i and j in period t is defined as

$$I_{ijt} = G\Delta_t^b P_{it} A_{jt}/D_{ij}^b$$

where G, b are constant, and

P_{it} is the population of i at the beginning of period t

A_{jt} is the attraction of j

D_{ij} is the distance from i to j.

Δ_t is a distance-like term introduced to improve the behaviour of the interaction model between time periods. Without it, the model would predict absurdly large variations in total flow between cities as the population is redistributed during the simulation; total migration between a few large cities would be far larger than if the same population were distributed in a number of smaller cities. Δ_t also improves the dimensional consistency of the model. It is set equal to the inverse of the square root of the density of cities, or

$$\Delta_t = 1/\sqrt{N_t/\alpha}$$

where N_t is the number of cities at the beginning of period t, and α is the area occupied by the simulation.

A_{jt} is taken to be equal to the population of city j perturbed by a temporary factor representing advantage. Each factor which contributes to a city's advantage has its own duration in time. An increase in basic employment, for example, may affect growth over a long period as associated non-basic opportunities develop and are filled. Duration is represented by ensuring positive autocorrelation of perturbations in the time domain as follows

$$A_{jt} = P_{jt}\, \varepsilon_t^{1-c}\, \varepsilon_{t-1}^{c} \qquad ,0 \leq c \leq 1$$

where ε_t is the perturbation at time t and c is a constant. With c = 0 there will be no persistence in advantage beyond one step. ε_t is found by generating a log-normally distributed random deviate. The parameter σ determines the average magnitude of the perturbation, and corresponds to the standard deviation of log ε_t.

After each step, the new population of each place is computed from the net in-migration

$$P_{j,\, t+1} = P_{jt} + \sum_i I_{ijt}$$

The total population of the system is thus conserved. If the total out-migration exceeds the population of a city, that node is assumed to die and is removed from simulations in subsequent time periods.

Initial Conditions

At time 1 the system consists of a number (N_1) of nodes of equal population distributed over the study area. It can be interpreted as a representation of villages acting as centres for an economy based on subsistence agriculture. Provided σ is finite, the effect of simulation on this system must be to concentrate population, since the initial state of equal populations can only reappear given a very unlikely combination of advantages. In other words, the system behaves analogously to statistical thermodynamics; the initial state is one of zero entropy, and entropy almost surely increases through time.

Fifty initial locations are used in each simulation, distributed over an area of 100 x 100 units. The literature suggests that the most appropriate distribution pattern would be intermediate between random (x and y uniformly and independently distributed) and regular (hexagonal lattice), in other words "more regular than random". (For a review see Haggett [7, pp. 414-47].) This can be interpreted to mean either that the variance in quadrat counts should be less than the mean, or that the mean distance between nth nearest neighbours should be greater than the value expected in a random pattern, for small n. Several stochastic models exist for determining appropriate quadrat counts, such as the binomial and two models proposed by Dacey [3; 4]. But there have been few studies of stochastic models of point location. The method used here allocates points sequentially using uniform distributions in x and y, but rejects any point which would lie within a critical distance λ of a previously located point. λ was set to 4.0.

Parameters

Four parameters affect the system described thus far; G,b, σ and c. Of these, G, σ and c are all effective in the time domain. G affects the numbers of people moved in each step. σ determines the amount of asymmetry in interactions by controlling the magnitude of the average advantage perturbation, and thus again affects the total movement in each step. Finally, c determines the persistence of a perturbation in time. Intuitively, a small G and a large c would give small movements and persistent perturbations, while a large G and small c, giving large, non-persistent movements, might be expected to produce the same effects in a smaller number of steps. We have therefore regarded G, σ and c as crudely interchangeable in their effects on the system, and have varied only σ while holding G and c constant in each simulation. c was set to 0.3, and G to 1.0.

The constant b, on the other hand, is uniquely effective in the spatial domain, and can be expected to affect a number of spatial properties of the simulation. Thus values of b and δ were both permuted during simulations of the system.

Simulations

Simulations were carried to five time periods in all cases. In some instances only a small fraction of nodes remained at the end of the fifth step, but in all cases the system was still undergoing changes. The only possible stable end point for the system is the complete concentration of population into one place, so there seems little to be gained by continuing simulations until this state is reached.

The state of the system at any step is determined by the initial conditions (in other words the initial distribution of points), by the particular sequence of advantages generated, and by the parameters of the processes, in particular by b and σ. Each aspect of the form or state of the system reflects these in varying degrees. For example, the population of one place at one step is very much affected by the particular advantage value generated for that place. On the other hand, average or aggregate measures of the system should reflect b and σ if they are to be in any way indicative of the processes occurring. The analysis of the simulations was therefore designed to separate the relative importance of specific conditions (initial locations and specific advantages) and general process parameters on a selection of measures of the system's state.

Analysis

Six structural measures were compared in the analysis, as follows; each is based on the N_t nodes existing at each step t.

i) Interaction, $\sum_i \sum_j I_{ijt}$. As noted above, the term Δ_t allows a more effective comparison of interaction from period to period. The changes in total interaction observed reflect the tendency for a concentrated population to interact more than a dispersed one, as simulations proceed.

ii) Mean length of interaction, $\sum_i \sum_j I_{ijt} D_{ij} / \sum_i \sum_j I_{ijt}$. This parameter should reflect the relative scale of concentration of population. If concentration occurs at a number of regional centres, mean interaction will remain high; on the other hand it will drop rapidly if a single dominant centre emerges.

iii) Information, $-\sum_i \left[P_{it}/\pi\right] \log \left[P_{it}/\pi\right]$ where $\pi = \sum_i P_{it}$. This is a measure of the equality of distribution of population among places. If all N_t existing places have equal population, as they do at the outset, the measure reduces to $\log N_t$, which is the maximum possible. On the other hand, when all population is concentrated in one centre at time infinity, the measure is zero. We can therefore expect it to decline almost monotonically as simulation proceeds.

iv) Spatial autocorrelation,

$$N_t \sum_i \sum_{j \neq 1} w_{ij} (P_{it}-\bar{P}_t)(P_{jt}-\bar{P}_t) \Big/ \left[\sum_i \sum_{j \neq 1} w_{ij} \sum_i (P_{it}-\bar{P}_t)^2\right]$$

where $\bar{P}_t$ denotes the mean node population at time t [2]. w_{ij} is a weight assigned to city j at city i, set equal to $2^{-Dij/\beta}$. The constant β determines how rapidly weight declines with distance, being interpreted as the distance over which a weight halves. The autocorrelation measure is an index of the smoothness of the city size distribution in space. The more similar neighbouring cities in size, the closer the index to 1. On the other hand, a value of zero indicates no spatial ordering of node populations. β controls the interpretation of "neighbouring", and was set equal to 18 in the study, so that for example a city 10 units away is twice as much a neighbour as one 28 units away.

v) Nearest neighbour statistic, $\sum_i \operatorname{Min}_{j \neq i} D_{ij}/2 \; N_t \sqrt{\Delta_t}$. This is an index of the spatial pattern of nodes at each step, without regard to population. A value of 1 is expected for a random pattern, and 2.1 for a hexagonal lattice. In view of the process used to generate nodes, we expect the index at time 1 to be greater than 1.

vi) Number of nodes, N_t. This statistic is expected to decline monotonically from $N_t = 50$, eventually to 1.

Results

All simulations were carried out to six time periods. While this only reduced the number of nodes to 1 in a few cases, it was sufficient to identify significant trends. The series of specific conditions, namely the initial point pattern and the sequence of perturbations, were different in each case. In addition the process parameters b and σ were varied, b in the range 1 to 4 in steps of 1, representing increasing resistance to travel, and σ in the range 0.15 to 0.35 in steps of 0.05. Twenty replications were made for each of the twenty combinations of b and σ.

Most of the trends in the measures can be anticipated from their definitions. The effect of increasing σ, the strength of perturbation, is to accelerate changes in the system, so that for example N_t decreases most rapidly for high σ systems. Increasing b tends to restrict the spatial range of interaction, leading to shorter mean lengths. The nearest neighbour statistic increases with b, since for high b places with close nearest neighbours tend to be extinguished before those whose nearest neighbours are more distant. Spatial autocorrelation is lower for high b values for similar reasons.

The standard deviation in each measure over the 20 replications was calculated for each set of process conditions in each time period. In general, standard deviations increase with time and with σ, since perturbations tend to have a cumulative effect on the variability of the system.

The central point of the study concerns the ability to distinguish variation due to process from that due to specific conditions. As an example, Table 1 shows the means and standard deviations for the interaction measure at time 3, for the twenty combinations of b and σ. Let us suppose that a value of 215 was observed for a system at time 3, with b and σ unknown. 215 is the expected value for a range of combinations of b and σ, whose locus is indicated in Figure 1 as the line "equal expected value". The point of minimum standard deviation will be the point of maximum likelihood at (1.4, 0.15), which is therefore the most likely process for an observed interaction of 215. In addition it is possible to compute the probability that any other (b,σ) process would yield an observed value of 215 from the appropriate mean and standard deviation and the t statistic, to create a surface of probability. The 95 per cent limits shown in Figure 1 enclose all (b,σ) combinations which cannot be rejected with at least 95 per cent confidence.

Table 1

MEANS AND STANDARD DEVIATIONS FOR INTERACTION, TIME 3

σ	b = 1		b = 2		b = 3		b = 4	
	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s
0.35	363.7	51.2	206.5	41.8	351.1	130.9	167.5	164.5
0.30	332.0	32.1	197.6	33.3	276.2	107.9	819.9	679.5
0.25	288.9	26.7	181.3	25.1	248.7	86.1	656.8	368.0
0.20	261.5	19.3	168.6	16.0	211.4	55.0	535.0	308.5
0.15	245.2	10.0	165.0	15.2	206.3	37.2	430.2	177.4

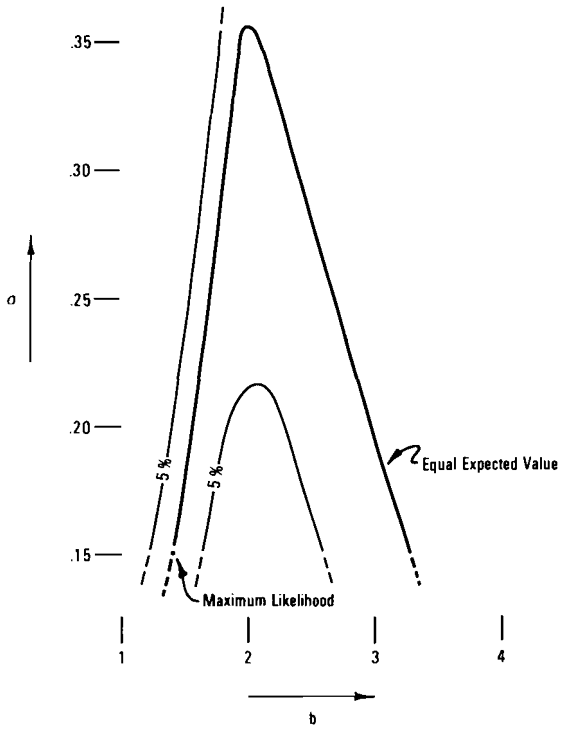

Figure 1

PROCESS DISCRIMINATION FOR INTERACTION, TIME 3

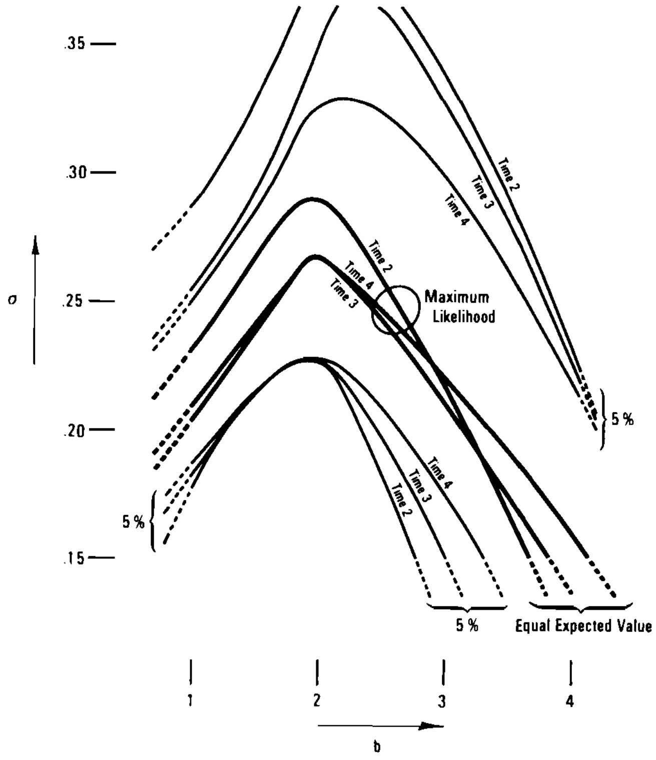

Figure 2

PROCESS DISCRIMINATION TO TIME 4: NUMBER OF NODES

At time 2 the corresponding figure for interaction shows narrow vertical bands, indicating a strong ability to distinguish b values but no discrimination of σ. In later time periods the accumulation of random effects makes it less easy to discriminate between high b values, and also tends to limit the range of possible σ. Figure 2 shows the pattern up to time 4 for the number of nodes measured, based on the observed value equal to that expected at (2.8, 0.23). Here there is a tendency for the confidence limits to narrow through time, indicating greater discrimination.

The discrimination limits for all six measures are summarized in Table 2, based on observed values equal to the expected value for (2.0, 0.25). Neither spatial autocorrelation nor the nearest neighbour statistic are capable of any degree of discrimination at any time period, within the range of (b,σ) studied. Interaction and mean length are useful discriminators of b in early time periods, and of σ later. But the most effective discriminator by far is the number of nodes, with 95 per cent confidence limits which continue to narrow through the observed time periods. (The range must widen again at later times since $N_\infty = 1$ for all processes.)

Table 2

DISCRIMINATION LIMITS FOR THE MEASURES
BASED ON 95 PER CENT CONFIDENCE

		Time 2		Time 3		Time 4		Time 5		Time 6	
		lower	upper	lower	upper	lower	upper	lower	upper	lower	upper
Interaction	b	1	4	1		1		1			
	σ			.15		.15		.20		.20	
Mean length	b	1	3	1	3	1	3	1			
	σ								.35	.15	
Information	b		4		4					1	
	σ	.20		.20		.20		.20	.35	.20	.35
Spatial auto-correlation	b										
	σ										
Nearest Neighbour statistic	b										
	σ										
Number of nodes	b		4	1	4	1	4	1	4	1	4
	σ	.15		.20		.20	.35	.20	.35	.20	.30

Discussion

The results show that at any particular stage of the system aggregate measures have little power to discriminate between process because of the degree of influence by specific conditions. This section discusses the implications of these results regarding the testability of a theory of nodal growth.

Classical, static central place theory leads to a number of predictions about the form of the system which are directly testable. A great deal of work in the late 50s and early 60s was devoted to searches for the predicted hexagonal arrangements of places, and discontinuous hierarchies of city sizes predicted by the theory [1]. A series of logical arguments in the theory enable one to predict that if certain processes occur, certain forms will appear on the landscape. The relationship between process and form is not one to one, since presumably other processes might be devised which would lead to the same form, but it is sufficiently direct to render the theory testable. Furthermore, alternative processes can be rejected as non-parsimonious.

As we now know, the regularities of form predicted by the theory were found not to exist except as weak statistical tendencies, the implication being that the assumptions about processes which form the basis of the theory were oversimplified or incorrect. Attention was directed from studies of form to the development of new models of process, as seen in the intensive work on consumer spatial behaviour models (see for example [8]). This, it was hoped, would eventually lead to better process models and a new theory that would finally satisfy the goal of explaining the sizes and spacings of settlements.

To return briefly to the model of this paper, in which b and σ represent process and the six aggregate measures, form, the results indicate very little connection between particular process and unique form. Given a form, but no knowledge of the time of the system or its initial conditions, it is difficult to establish process with any satisfactory degree of discrimination or certainty, or to test a specific hypothesized process. In general, process must be regarded as essentially untestable from the form of the system.

The results of the model defined for this paper must of course be generalized with extreme caution, but certain implications can be identified. Initial conditions, external influences and non-tertiary activities, represented in this model by random disturbances, are undoubtedly present in any central place system, and must be regarded as largely unknown. The stage of the system, in this case time, is also virtually unknowable in any real system. Under these circumstances the observed form of the system provides no power to discriminate processes or to test process hypotheses. Processes must be studied and verified directly, and their implications for the form of the system must be unknown.

When emphasis in central place studies shifted in the late sixties from form to process, there was a clear implication that the eventual goal remained the same, and that new processes would eventually be integrated into better predictions of spatial form. The broadest (and of course most tenuous) implication of this paper is that the shift is permanent, and that the goal of explaining the sizes and spacing of cities is virtually unattainable.

References

1. Berry, B.J.L. Geography of Market Centres and Retail Distribution. Englewood Cliffs: Prentice Hall, 1967.

2. Cliff, A.D. and J.K. Ord. Spatial Autocorrelation. London: Pion, 1973.

3. Dacey, M.F. "A Compound Probability Law for a Pattern More Dispersed than Random with Areal Inhomogeneity", Economic Geography, 42 (1966), 172-179.

4. Dacey, M.F. "A County Seat Model for the Areal Pattern of an Urban System", Geographical Review, 56 (1966), 527-542.

5. Goodchild, M.F. "Spatial Choice in Location-allocation Problems: The Role of Endogenous Attraction", Geographical Analysis, 10 (1978), 65-72.

6. Griffith, D.A. "Spatial Structure and Spatial Interaction: A Review", Environment and Planning A 8, (1976), 731-740.

7. Haggett, P., A.D. Cliff and A. Frey. Locational Analysis in Human Geography. 2nd edition. London: Arnold, 1977.

8. Rushton, G. "Analysis of Spatial Behaviour by Revealed Space Preference", Annals, Association of American Geographers, 59, 2 (1969), 391-400.

9. White, Roger. "Dynamic Central Place Theory: Results of a Simulation Approach", Geographical Analysis, 9 (1977), 226-243.