COMMENTARY: CURRENT ISSUES IN INTERACTION

Michael F. Goodchild
The University of Western Ontario

Introduction

The diversity of papers in this volume is a very clear indication of the breadth of the current interest in interaction models. It shows a wide spectrum of viewpoints and approaches, from the macroscopic paradigm of Warntz, with its physical analogies, to Webber's information minimization. At the same time there is a clear indication of the breadth of applications, which is appropriate in a geographical publication but often missing from the transportation or planning literature. Finally there is a diversity of purpose and methodology, with Warntz representing the most inductive and Webber the most deductive, and with the purely theoretical approach of Sheppard as well. In short, the collection is admirably representative of the current state of the interaction modelling literature, seen from the perspective of the discipline of geography.

It seems most appropriate for this paper to draw out and comment on a number of common threads as they are treated (or ignored) by the various authors. The sections which follow will deal in turn with the problems of goodness of fit and significance, potential measures, and the zoning or aggregation problem. The one issue which does not appear to have caused any great concern among the authors of this collection, unlike the literature as a whole, is that of calibration. The last section suggests a number of additional research issues.

Goodness of fit and significance

Interaction modelling presents a host of problems to those who have been used to dealing with simple bivariate regressions, and the literature as a whole is only now beginning to develop more appropriate methods of interpretation. One aspect of the problem is illustrated by Warntz's paper, and also by Pooler and de Abreu, when they regress various densities, including income density, against the income potential measure. One variable, income, is common to both income density and income potential, and so it is quite inappropriate to use the standard null bypothesis of linear regression, that x and y are independent, or R^2 for the sampled population is zero.

What, then, is the appropriate null hypothesis for the regression, and how can the observed correlation be interpreted? In essence, the relationship has to do with the spatial arrangement of income values. Income potential at a point is most strongly influenced by nearby income values, since the potential measure involves an inverse weighting by distance. The income potential surface is simply an averaged or smoothed version of the income density surface, and the calculation of potential can be regarded as the application of a smoothing operator or filter. So we may interpret the correlation between potential and density as an indication of the topography of the density surface. If the latter is smooth, the potential surface will be almost identical to it, and the correlation will be strong with a slope of 1.0. If the density surface is very peaked, applying the potential operator will tend to lower the peaks, giving a strong correlation but a slope of greater than 1.0. Finally if the surface is random with no spatial autocorrelation, it will be flattened by the operator to give a correlation near zero.

So the potential-density correlation is a test of the spatial configuration of the density surface, and should be interpreted in that context. An appropriate test of significance would be to compare an observed R^2 against the distribution obtained when income densities are repeatedly reassigned to cases, implying the null hypothesis of no spatial structure in the density map.

Regression of income potential with densities of parameters other than income also requires careful interpretation. Following the notion that income potential reflects the operation of a smoothing operator on income density, the regression is in fact between two densities, one straightforward and the other, income, averaged spatially over some neighbourhood. It implies that the density of, say, Who's Who students in Illinois is influenced not only by the income density of Illinois but also by that of Indiana and in fact by the income of the region as a whole. The smoothing operator is clearly crucial as it defines the concept of region. What is the result of regression when the operator is redefined to include a wider region by using $d^{-1/2}$, or a smaller region (d^{-2}), or even when income density alone is used without smoothing? The leads of course to a discussion of the basis of the potential index itself and the d^{-1} weighting, which will be taken up in the next section.

Other cases might be mentioned here as examples of the inappropriate null hypothesis problem. One is the so called rank size rule; a plot of a city's size against its rank in the national or regional city size distribution on double log paper invariably yields close to a straight line. Again there is really only one variable present, city size, and a null hypothesis of no relationship between x and y is inappropriate. A little inspection will reveal that the graph is a peculiar form of cumulative frequency distribution, and that the straight line represents a Pareto distribution. The appropriate significance test is a Kolmogorov-Smirnov against a Pareto null hypothesis.

There are many similar instances in which it is possible to test data against a model which is itself the outcome of a stochastic process and can therefore be used as the null hypothesis. Suppose that one wishes to model an interaction consisting of an integer count, such as the number of migrants between two states i and j in a five year period. Let the observed count be I^*_{ij}. Suppose further that a model has been devised which gives the predicted count I_{ij}. Since the observed count has been taken over some limited period and is subject to sampling error, differences between I^*_{ij} and I_{ij} are to be expected. However they can be compared statistically with a chi-square test, using the model as the null hypothesis. This gives a very powerful, positive test of an interaction model, compared to the more conventional procedure of rejecting a null hypothesis of no relationship between I^*_{ij} and I_{ij}, a double negative. It tends to avoid the problems of R^2 interpretation which Fesenmaier comments on. Finally, it is significant that early interaction modelling, using R^2, tended to claim substantial success, whereas more recent work using the stochastic process approach is more critical (Openshaw, 1976, Goodchild and Kwan, 1978).

As a final note consider the expected or model values, I_{ij}. In a simple regression the cases are all of one class, whereas in interaction modelling additional structure is implied by the double subscripts i and j. Information thus exists at more than one level. Observations and model can be compared by testing I^*_{ij}

against I_{ij} for all i and j, but also by comparing marginal totals for all i or all j. Equally the mean travel costs can be compared for the observed and expected tables, giving three potential levels of test.

Now suppose one wished to test some observed I^*_{ij} against an expected set representing purely random interactions, a null hypothesis of no order. Clearly one would not want the observed and expected to differ in mean trip cost, because then the null hypothesis could be rejected out of hand. So the observed mean trip cost should be a constraint on the expected table. Equally the marginal totals should be constraints, so that the differences between I^*_{ij} and I_{ij} occur at one level only, that of the table entries themselves. As we know very well from the analysis originated by Wilson (1967), of all possible sets of I_{ij} under these constraints the most likely is itself the spatial interaction model. So in another, more powerful sense the model is the appropriate null hypothesis. Furthermore, this argument seems to avoid many of the conceptual difficulties of other interpretations of the entropy paradigm.

Potential measures

In the previous section it was argued that potential measures can be interpreted as the result of the application of a somewhat arbitrary smoothing operator, and that a density-potential correlation is simply a comparison of a surface with a smoother version of itself. A great many more interpretations of potential measures can be found in the literature, including the one which is generally accepted as the original rationale, which compares the relationship between potential and interaction to that between scalar potentials and vector forces in such areas as magnetism, electrostatics and gravitation. Although he rejects any need to depend on physical analogies, Sheppard defines potential from interaction in the same way.

Another interpretation, which is perhaps the commonest in the geographical literature, is also based on interaction but yields a different result. In most interaction models there is an origin term which is a measure of size or population, or gross ability of the origin to interact. Take the example of population. When the model is divided through by this term we have per capita interaction. Potential is then interpreted as the total per capita interaction with all destinations, or the total interaction to be expected for an individual at the origin location.

Unfortunately this interpretation is not consistent with the previous one, since it would require the power of distance to be the same in both potential and interaction equations, rather than differing by one. It is easy to see why it has become popular, however, since the word 'potential' is much closer to its common meaning. In the physical analogy 'potential' is the potential to do work not in any sense potential interaction. Since there is no obvious social analogy to the concept of physical work, it is in fact difficult to see how the idea of a physical analogy ever arose.

Zones and aggregation

Another recurrent issue in the collection is that of the choice of zones for interaction modelling. Fesenmaier has explicitly considered the effect of the number of zones on the goodness of fit of a misspecified model. Within each level of aggregation or spatial resolution there is also the problem of how the actual configuration of zones affects a model, and Openshaw (1977) has made a very innovative contribution to this topic.

The spatial resolution problem is evident in the Webber paper, where the city of Hamilton is divided into fourteen zones, defining the scale of the data and also of the eventual result. Since the number of zones determines the number of possible solutions, which rises very rapidly with increasing numbers of zones, the problem can only be solved in a reasonable amount of computer time if very large zones are used. It is interesting that each of the solutions has five shopping centre locations, although this does not appear to be a requirement. As we noted in a previous section, the solution to an information-minimizing or entropy-maximizing strategy can be regarded as a null hypothesis, the most likely state of the system given the constraints. Thus it is the deviations from the maximum entropy solution which are interesting, since they represent the effects of additional constraints and unequal probabilities which the system imposes, in other words information not previously known to the researcher. A good fit on the other hand indicates that there is nothing in the system unknown to the researcher. How good then is the fit of Webber's model? This is clearly the most important topic for any continued research.

Potential measures are very dependent on the number and size of data zones. They are often defined as integrals, but then computed by summing over discrete zones, as in the Warntz paper. The sum is clearly biassed with respect to the integral, and yet there has never been to the author's knowledge any research on the strength of the bias. The contribution to self, which is often the dominant term in the sum, is highly dependent on zone size. In interaction modelling there has been a tendency in recent years to replace the negative power distance function by a negative exponential, particularly following Wilson's work. Because the negative exponential is finite at a distance of zero there is no problem of contribution to self, and one would suspect also that the sum is less biassed with respect to the integral. These seem therefore to be good reasons for using the negative exponential as the basis for potential as well.

Tobler's paper illustrates another geographical influence on interaction models, the somewhat arbitrary closure introduced by the boundary. One wonders also what effect boundary definition has on the Webber model. Such aspects of interaction modelling are strictly spatial, and tend to be areas where geographers can make a valuable contribution, both through awareness of the practical problem, and possession of the necessary tools.

Conclusions

In its thirty or so years as a field of interest, interaction modelling has developed very much along the lines indicated by the sequence of papers in this volume. The early literature relied on physical analogy, coupled with an inductive approach that searched for empirical regularity with little theoretical guidance or structure. A lull in the early sixties was followed by a spate of theoretical activity which created a deductive basis for the models and developed much more acceptable methods of calibration.

The old physical analogy basis has been replaced by not one but two systems of theoretical propositions, however, with corresponding schools of thought. On the one hand is the statistical school, still incidentally retaining a form of physical analogy in its strong links to statistical mechanics. It sees the gravity model as the most likely state of a system behaving randomly, within essential constraints. Any behavioural preferences take the form of additional constraints, and unequal microstate probabilities, and cause deviations from the model, which is expected to perform best at the most aggregate level, where local differences in behaviour can be regarded as a form of randomness. The success of the model is seen by this school as due to the averaging of large amounts of variable microbehaviour. Although the microbehavioural processes may be perfectly systematic at the individual level, their parameters vary from group to group and from place to place, and the observed form of behaviour is also affected by spatial opportunities. Thus microbehaviour and macrobehaviour are essentially different fields, with their own models and techniques, and little progress will be made in trying to derive one from the other. Again, the analogy to statistical mechanics is valid, that it is futile to try to derive the macroscopic properties of a gas from the study of its individual molecules and vice versa.

Opposed to this is the view that the gravity model is behaviourally based, and that there is some appropriate level of aggregation or disaggregation at which its fit is optimized. Above or below this level fit will deteriorate through the mechanisms illustrated by Fesenmaier's paper. The task of the researcher is therefore to find that level. The viewpoint is essentially inductive in contrast to the statistical view: although the behavioural gravity model theories of Neidercorn and Bechdolt (1969) and others fall into this category, they must be regarded as rationalizations of an observed empirical regularity.

We return, then, to the very current topic of aggregation and zoning as holding the key to resolve the difference between the behavioural and statistical schools. To reiterate, this seems to be the area to which geographers can be expected to make a unique and exciting contribution in the next few years.

References cited

Goodchild, M.F., and M.Y.C. Kwan (1978). "Models of hierarchically dominated spatial interaction", *Environment and Planning A*, 10, 11, pp. 1307-1318.

Neidercorn, J.H., and Bechdolt, B.V. (1969). "An economic derivation of the 'gravity law' of spatial interaction", *Journal of Regional Science*, 9, pp, 273-282.

Openshaw, S. (1976). "An empirical study of some interaction models", *Environment and Planning A*, 8, pp, 23-41.

Openshaw, S. (1977). "A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling", *Institute of British Geographers*, Transactions New Series 2, 4, pp. 459-472.

Wilson, A.G. (1967). "A statistical theory of spatial distribution models", *Transportation Research* 1, pp. 253-269.