MARKETING SPATIAL ANALYSIS: A REVIEW OF PROSPECTS AND TECHNOLOGIES RELEVANT TO MARKETING

Stan Openshaw

WORKING PAPER 93/25

SCHOOL OF GEOGRAPHY • UNIVERSITY OF LEEDS

WORKING PAPER 93/25

MARKETING SPATIAL ANALYSIS: A REVIEW OF PROSPECTS AND TECHNOLOGIES RELEVANT TO MARKETING

Stan Openshaw

School of Geography
University of Leeds
Leeds LS2 9JT

1. Introduction

Seemingly no one as yet has made any or much money out of applying spatial analysis in a marketing context. So far the major financial returns from the commodification of geography have mainly come from the sale of digital map products, and much smaller amounts from mathematical modelling of spatial interaction data in retail decision support, from the application of conventional aspatial statistical methods to census and client data, and from the re-organisation of space via redistricting and location-allocation modelling. This is perhaps surprising, see Openshaw and Goddard (1987); and suggests that there is still plenty of scope, given the right opportunities for the further commercial exploitation of the quantitative geography tool-box.

This chapter seeks to identify and develop spatial analysis technology that is, or could be, of interest in marketing contexts. First, though, Section 2 provides a review of expectations and perceptions from both a geographical and a marketing perspective. Section 3 then focuses on some of the key changes in IT infrastructure relevant to marketeers, focusing in particular on the new opportunities this creates for applied spatial analysis. Section 4 briefly reviews some of the more marketing relevant methods and Section 5 outlines some possible future developments.

2. Views, perceptions and expectations

2.1 Geographers and marketing

To many quantitative geographers, marketing geography is associated with retail geography, central business districts etc. Most geographers seem to have never really understood commercial needs or appreciated the potential applied opportunities for their technology in a marketing context, or indeed have displayed concern for the needs of applied users. As a result, in common with the rest of social science, there has never been any real stimulous into developing commercially relevant variants of it. Indeed even in those areas where the applied and commercial applications are now so glaringly obvious, it seems to take at least a decade for the linkages to be made and the technology transfer process to be started. It would appear, on reflection, that this is as much related to the life-cycle of the academics involved as to the commercial potential of the techniques. Indeed, 'nasty' profit making commercial activities have traditionally happened, if they happen at all, near the end of an academic's active research career. This is not surprising when academic performance and career prospects was and still is measured in terms of academic papers published and by research income rather, than by pounds sterling or dollars generated via commercialisation, product royalties, and licensing. All too often those who judge performance wrongly assume that making money is easy, facile, and not scholarship; reducing such activities to the "anyone could do it if they

wished" category. Reality is, of course, quite different. Very, very few academics have any of the necessary commercial skills to even get started, let alone be successful.

Geographers have been particularly slow to appreciate the commercial relevancy and monetary value of many basic geographical analysis and modelling skills. Indeed it is interesting that the GIS revolution continues this trend by not providing much support for real spatial analysis functionality, seemingly on the spurious grounds that there is no current market demand for it; whilst at the same time offering such basic spatial interaction and location-allocation facilities that are of extremely limited practical value.

It also needs to be said that from a marketing perspective, the principal attraction of spatial analysis is still probably a psychological one. Marketeers seem to feel that geography is important in that they know that there are major geographic variations in the demand for products. Maybe they feel that geographers should be able to help them perform better and there might be methods that geographers know about that could be beneficial to them. Of course those who believe this will probably already be displaying more confidence in the value of geography than many geographers!

At the same time the main problem that geographers face is knowing what might be needed in a marketing context. Seemingly few geographers know much or anything about the needs of marketeers and are thus unable to be of much assistance even if they wished to be. Getting started is another problem. Yet in principle any geographer reasonably proficient at stepwise regression, logistic regression, and discriminant and cluster analysis already knows much of the basic marketing relevant technology. This knowledge together with a pc, a statistical package, a London address, a persuasive demeanour, and a little luck in networking and in making initial contacts is probably sufficient to create a small marketing consultancy. A few unique selling points and the ability to hype-up, by mild exaggeration, mediocre results might also be regarded as useful. Particularly attractive at present is an ability to use neural net based methods, whilst waiting in the wings is the next wave of trendy technology based on genetic algorithms which is likely to create a major biomarketing industry sometime soon. However, it is arguable as to whether there is much or any need to be ahead of the field in terms of technical sophistication. Academic arrogance tends to be a factor here. Maybe the market does not need the highest quality and most advanced technology, but these factors are precisely what drive many academic geographers to aspire to excellence and simultaneous diminish their commercial prospects.

To summarise, it would seem that many spatial analysts, because of their knowledge of multivariate statistical methods and latterly in GIS, possess most of the basic and essential analysis skills needed to be successful in marketing. Indeed when such people join marketing consultancies they often do well, or at least appear to. However, if they remain as academic geographers and deprived of both the contact networks and key marketing databases they are not likely to make much, if any progress, except as a limited source of consultancy advice and ideas for others to exploit. Yet at the same time it is evident that there are mutual benefits to both the marketing industry and geography from closer collaboration. The geographer might gain access to data not in the public domain, new publishing opportunities arise and there is at least some prospect for technology transfer and commercialisation. The marketing industry gains access to a largely untapped skill base. The question is, however, what methods, what applications, and what new products might be created via these means. It is also pertinent to ask whither the explicit spatial component and how can explicitly geographical analysis rather than applied statistical analysis be involved in marketing. These are hard but fundamental questions to ask a geographer, especially when as a profession they have in general been extremely slow even reluctant to develop applied spatial analysis skills. If lack of awareness is the main problem then maybe the comments in this chapter may help.

2.2 Whither applied spatial analysis?

If little applied use is made of spatial analysis anywhere, why should marketeers be interested in being amongst the first to use it? There are a number of possible responses that can be reduced to five key factors:

(1) a perceived opportunity to gain an advantage of some kind by commercial exploitation of potentially useful methods of analysis and modelling of spatial data;

(2) better processing of spatial data constitutes an information industry that can generate wealth by adding value;

(3) concern that others might be gaining a commercial advantage by being first to appreciate the spatial analysis opportunities;

(4) concern about the lack of basic academic research in many of the core areas of marketing and a feeling that geographers could or should be interested; and

(5) there is some evidence of usefulness, for example, the retail modelling activities of GMAP.

Another related factor is the rapid rate of change in many of the important aspects of the marketing world, although seemingly not all these developments are visible to insiders. In general though it is clear that the world in which marketeers operate is becoming ever more complex: there are more competitors, profit margins are under pressure, but at the same time there is a much greater data availability, much faster and cheaper computers, as well as new and more complex processing technologies. It is both doubtful whether these technological developments can either be safely ignored or, indeed, safely exploited! It is in these difficult conditions that the geographer has to try and find a niche.

Perhaps most important is a growing need for an awareness of what is going on and where the technology is leading to. All this takes place at a time when competitive tensions are increasing with a growing concern about whether or not suddenly new, better, and much improved methods and technology might not have emerged. Maybe, if you understood what was happening and knew something about the nature of the new technologies you might be able to do better in the applications areas relevant to your business concerns. As many concern geographical things, then maybe geographers may have something to offer here.

2.3 A geographical opportunity

It is argued then that there is an emerging opportunity to use geographic information and to apply spatial analysis tools in marketing and that for human geographers, at least, there is the prospect of an interesting new area of academic study. In seeking to build this partnership it is important to be aware of the potential constraints on both parties and to appreciate the institutional barriers that may exist and which can so easily get in the way of the research. However, it is also useful to consider the changes that are currently underway in IT which may well be regarded as a major driving force in making spatial analysis a viable technology in a marketing context.

3. Significant changes in marketing IT

There are three key processes of change that are relevant here:

(1) computing hardware and software;

(2) databases; and

(3) analysis technology

3.1 Computing hardware

The computing hardware changes are perhaps most notable. Everyone knows that computers are becoming faster, smaller, and cheaper; but the rate of change is now quite spectacular. It is perhaps of only passing historic interest to note that in 1970 a large IBM mainframe (the largest model 370/165) came with 3 megabytes of RAM and each massive disk drive had a 100 megabytes capacity. Twenty years later most pc's would outperform this monster for about 10,000 times less money in real terms. The improvement in cost performance has been amazing. For example, a large Amdahl mainframe (model 5860) would have cost at least £2 million in 1985 and was rated at about 3.1 LINPACK megaflops (million floating point operations per second); at least 5 times faster than the IBM 370/165 could manage in 1970. Yet by 1990, a Sun Sparc 1 workstation costing about 200 times less in real-terms was rated about half the Amdahl's speed. By 1992, the latest Sun workstations are now running about seven times faster, have a bigger RAM, and could handle equivalent or more disks. This workstation was advertised as the equivalent of a Cray

I supercomputer which in 1975 would have cost 20 million dollars.

These changes in price performance are set to continue, indeed at the top end even more dramatic changes are anticipated. Many computer scientists now talk about an impending new era in computation due to the emergence of massively parallel computing hardware; see Daedalus (1992) for a review. It now seems likely that during the 1990's the economics of information processing will undergo a major revolution. In 1993, the UK's fastest supercomputer, the Cray ymp8 at the Rutherford Atlas Laboratories, is rated at a peak speed of 2,667 megaflops. It has 8 processors. Most users will probably be lucky to gain much more than a factor of 10 speed-up compared with a Sun Supersparc workstation. The current world fastest supercomputer is rated at about 131,000 megaflops, it has 1,024 processors (Dongarra et al, 1993). Users would be lucky to get half or less of the peak speed but nevertheless a speed up of several thousand times compared with a Sun workstation. By the late 1990's it is confidently expected that peak speeds in excess of a teraflop will be commonplace (1 teraflop = 1,000,000 megaflops). The implications for marketing is that fairly soon, indeed if not now, affordable hardware will exist which is able to handle any conceivable present or future marketing relevant database and analysis problems. As the leading hardware developments enter teraflop speeds, so the

rest of us will be left with cheap gigaflop (1,000 megaflop) hardware; if indeed it is possible to create marketing applications big enough to need it!

However, there are at least three areas where low gigaflop speeds might be useful now: retail modelling, credit scoring, and database marketing. As the hardware costs fall during the next few years and gigaflop workstations appear, so the number and range of marketing applications will dramatically increase. Targeting of clients and modelling predictions of their needs for consumer products will reach unparalleled degrees of sophistication and accuracy. This claim rests on the view that for many commercial applications it is possible to trade memory and compute speed for lack of knowledge engineering, and even software effort. Waltz (1990) writes "Today's massively parallel machines present AI with a golden opportunity to make an impact, especially in the world of commercial applications. The most striking near-term opportunity is in the marriage of research on very large databases with case based and memory based AI" (p1). Currently, there are few, if any, marketing applications of this technology. Faster computers do not merely speed-up current applications but they create major new opportunities for new products and new approaches. This is as true in the academic world as it is in the commercial world.

As compute speeds increase and costs diminish, so the economics and capabilities of information processing relevant to marketing will be revolutionised. New forms of analysis and modelling will become practicable commercial propositions and, perhaps more importantly, spatial analysis technologies might well become a mainstay of a new computational marketing era. As marketing databases become more sophisticated and customer information systems become operational and mature, so important new opportunities and needs for spatial analysis and modelling will emerge.

It is interesting to note that in some areas of marketing how the economics of geodemographic processing have already changed dramatically. The 1981 Census ACORN system was allegedly based on a minicomputer. The development of the 1981 Superprofile system consumed the equivalent of £2m of computer time on a large mainframe (Charlton et al, 1985); and a single run took about 2 days of CPU time. Today a 1981 SuperProfile system could be built on a pc in about 2 days, for about 1/20,000th of the actual cost. No wonder Openshaw (1993) can claim that most large users should today have their own customised geodemographic systems, indeed, there is no good reason why they cannot have several, each tuned to different market sectors. Today the principal cost is the census data, not the computer processing. On the other hand, who wants to create in 1993 a geodemographic system no better than was possible in 1983? Maybe the extra computing capability needs to be

invested in building significantly better geodemographic systems. A much better strategy here as in other areas of marketing involves finding better ways of using unlimited and cheap computing power to improve what you do, not just make it cheaper. This is discussed further in Section 4; indeed the development of better targeting tools is probably the single most important present day application area for spatial analysis in marketing.

3.2 Databases

The last few years have seen the increased use of large marketing databases. Database vendors or brokers such as NDL have databases containing details of over 14 million people, CMT of over 7 million; ICD, Equifax, CCN, GUS, and Littlewoods have some details about nearly everyone aged 18 and over; based on computer copies of telephone directories; electoral registers; and other public information such as shares registers and county court judgements for bad debt. Also very large scale sample surveys provide details of the lifestyle and behaviour patterns of millions of people. This has been supplemented by other survey forms distributed with guarantee cards, sponsored shopper surveys, and by estimates of key variables such as income, age, and credit worthiness. In addition, many large organisations hold comprehensive customer information databases and have often spent millions building usable systems.

By comparison, the 1991 Census and the SAR looks much less attractive. In a marketing context, accurate census data for April 1991 based on 100% of the population is probably less valuable than a large, biased, sample of mail responsive persons for 1993. The statistical bias in the marketing database which may render the data of reduced academic value in certain applications (ie for estimating 5-10 year olds) is now a positive advantage; (5 to 10 year olds do not order by mail). In many marketing applications the population of interest is that fraction of the total census population who are likely to be mail order responsive. If the database is biased against the poor, the homeless, and the extreme rich then this might actually make it more valuable not less.

Another key aspect is that all the data is postcoded. There is an almost universal use of postcodes with address data; the principal exception being the OPCS! Postcodes make it very easy to introduce a geographical dimension to address data and thus add value to it (Raper et al, 1992). For example, a list of names and addresses input from the electoral roll has no value. Once postcoded, other information can be added to it using geographic file linkage (ie a geodemographic code) and it becomes a valuable commodity. However, what is perhaps a little surprising is that whilst postcodes can be converted into geographical coordinates most users do very little with the map information. Location is clearly important but apart

from drawing maps of such things as the distribution of clients, it is not really used. Maybe this reflects the GIS revolution of the late 1980's and the popular appeal of mapping systems. Indeed most business applications for GIS seem to be largely mapping orientated. Dangermond (1993) lists what he terms several major areas of business: real estate, energy, electrical utilities, forest products, tourism and recreation, transportation, communications, publishing, and insurance. He writes "In just the last few years the application of GIS to business has begun to grow very rapidly. This seems to be because the technology is getting significantly easier to use, much less expensive, and because the business community now accepts the value of GIS in performing many kinds of necessary work" (p9)

There is no doubt also that there is a growing number of pc based marketing systems that allow various GIS operations to be performed on marketing data. It is fairly easy to plot maps of clients or market penetration rates against various map backcloths; for example, an OS map base or streetlines or travel times etc. Various site evaluations can be done; for example, population or potential sales profiles within x km of a retail site, indeed this can be a highly profitable and popular activity. Dangermund (1993) notes that this can be useful as a means of locating new business. However, it is also a very "noddy" spatial analysis approach to what is a complex activity and provides no real basis for multimillion dollar investments.

Superficial site evaluations and displays of customer and outlet locations may look picturesque and spatial query may be helpful in explaining what is going on to senior company executives, but it often has no other value. Spatial decision support using GIS often reduces to no more than just eyeballing map displays, or querying spatial databases by pointing a mouse. At the very least you need a 'what if' capability and ideally much more as well. A desktop mapping-cum-GIS does not provide a cheap route to spatial decision support although it can aid understanding and prepare the way for better systems to come.

The problem is that the rapid development of marketing databases and information systems has outstripped our ability to make good use of them. If a marketeer knows exactly what profile clients should have then there really is no problem. For example, you might want 100,000 names and addresses of golfers aged 40 to 55 with £30,000 plus income as being the ideal profile for a given product. However, performance now depends on (a) you being right about this ideal profile, (b) a good performance of a very simple buying model that assumes people of this profile will respond at a much greater rate than other types of people, and (c) accuracy of the database variables used in the selection. In practice, much must depend on the identification of the target profile, but can you be that precise or that confident. What if you are wrong, or don't know? If does not help to know that you could also have

made selections based on at least 57 other variables as well! Finally, why are you not using the geographic coordinates as an additional targeting aid?

3.3 Emerging new marketing technologies

Clearly then data is being generated faster than it can be digested. Very large (over 50 GByte) marketing databases exist now and provide a plethora of information; some undoubtedly useful, some totally useless but in a context dependent way. There are details of personal and household characteristics, of consumer behaviour, of lifestyle variables, of wealth and possessions, and of behaviour; and much much more. Yet what do we do with it? Most databases exist on hosts that can perform simple queries and print address labels, probably in a mainframe and batch environment. Yet the speed-up in computing and the revolution in computing economics makes it feasible to apply dramatically more complex and intelligent analysis procedures than is currently practised. There are emerging new ways of mining databases for far more effective selections and performance. The technological basis for marketing has started to change rapidly, although seemingly most marketeers are still using old fashion and obsolete 1960's technology, aimed at solving problems from the 1970's, clothed in 1980's GUIs, and run on 1990's hardware. There is a danger that many are using simple, sloppy, sixties systems because they yield understandable results.

Openshaw (1992) terms this the Catherine Cookson approach to marketing. At one time there was no other choice; simple methods, understandable by non-technically minded users reigned supreme. However, the new technologies are inherently more complex and are increasingly incapable of yielding simple results. The problem is how to use them safely and painlessly despite their complexity. Indeed, this can be done if effective use is made of graphical interfaces. Most car drivers have no good idea as to how their car works but it does not stop them either driving cars or trusting them. The equivalent systems in marketing also need to be user proofed and rendered intrinsically safe. Once this required extensive statistical and mathematical skills, now it merely needs an automated process with built in checking and validation mechanisms. The extra computing power can be used to power a fail safe, almost fool proof technology. Openshaw (1992) discusses this further.

One aspect of this complexity is the greater emphasis being placed on statistical methods. However, this amounts to little more than a catching-up process; whereby the less sophisticated increase their levels of statistical competence. It is also a filtering down process; as once advanced methods become sufficiently affordable to be more widely applied. None of this is particularly relevant to marketing. There are problems with the statistical tools, often due to the non-ideal nature of the data. Also

building stable models of what are always low probability events (ie success in a marketing context typically has a response probability of less than .05, and often less than .001). There is also the increasing problem of making best or even good use of the available data.

The current enthusiasm is for neural nets and perhaps soon biomarketing. The problems tend to be severe when a largely nontechnical user community start combining very large and massively multivariate databases with advanced analysis and modelling tools that offer good performance at the expense of brittleness. The growing list of available methods (scoring methods, regression, discriminant analysis, logit versions, cluster analysis, AID, CHAID, and neural nets) does nothing to help either. Moreover, even after all this apparent sophistication, there is still no obvious spatial analysis component.

In a modelling and response prediction context Openshaw (1992) argues that developing black-box automated analysis tools might be the only safe solution. If there is uncertainty as to which methods to use, then try them all. The Marketing Machine (MM/1) concept offers a conceptual framework for this approach. It recognises that the most important factor is "safety" and this puts considerable emphasis on evaluating the results, checking for robust good performance. This approach may not be very intellectually satisfying but would seem to be a pragmatic

and extensible solution to an emerging problem. However, it is still not spatial in anyway even if it does provide a framework within which spatial analysis tools might one day be applied.

4. Spatial analysis in marketing

4.1 Where are the opportunities?

Merely because marketing data has implicit geographical referencing does not automatically mean that spatial analysis is needed. Indeed discovering how to use the spatial element is not easy. Location may well be considered by geographers to be an a priori important variable in everything they do. The problem is that you cannot simply use x y coordinates as merely another pair of predictor variables in a regression model. This has been attempted, but it will seldom be sensible in a marketing context. Instead, one obvious approach is to try and utilise the great strength of spatial data by focusing on methods of analysis that looking for geographical patterns at regional and local scales. A common and recurrent aspect of geographical information of all kinds is that there are strong often localised patterns. These spatial heterogeneities or non stationarities or localised extra-Poisson variation (or whatever it might be called) can present horrendous problems to conventional statistical models (which expect global relationships without any or many spatial singularities). Yet other methods, often quite simple ones, can be used to define these patterns.

It is argued, therefore, that the most useful spatial analysis techniques for marketeers will be the more spatial pattern descriptive procedures. Table 1 lists some of these methods. Note that mapping raw data is not analysis and not sufficient by itself. However, once the data are processed to separate the 'rough' from the 'smooth' components then mapping becomes much more useful. It is with spatial pattern recognition and modelling methods where the main opportunities for spatial analysis lie in a marketing context.

4.2 Spatial classification

The existing geodemographic systems are one example of spatial pattern seeking. Clusters of geographically contiguous census areas provide useful descriptions of multivariate census data. Space is being used as a convenient metric for simplifying complex data. The 'old' geodemographic argument is applied so that it is assumed that areas with similar characteristics might well behave similarly in response to marketing stimulus or share similar needs. However, to make the most of the benefits of geodemographics the classifications need to handle rather than ignore both the varying spatial reliability of the data (due to small number problems) and the varying precision of the variables themselves; see Openshaw (1992). The smaller the spatial units the greater these problems become.

TABLE 1 Spatial analysis tools relevant to marketing

spatial classification
spatial cluster detectors
fuzzy spatial targeting
spatial relationship seekers and modellers
smart modelling
spatial pattern recognition
spatial data value adding

4.3 Spatial cluster detectors

An obvious development of crude spatial classification technology is to focus more on clustering inherent in client data or in responses to direct mail campaigns. It is interesting that spatial epidemiological methods have seemingly not been used much, or at all, in marketing, despite a strongly implicit assumption that response is a highly contagious process (ie neighbourhood effects). If responders cluster at some spatial scale, then location can be used as a filter on the selection process. Also, the locations of the clustering may themselves be of marketing relevancy; if recurrent meta-patterns can be identified, then they have an obvious use in response prediction.

The underlying spatial analysis assumption is that socioeconomic characteristics are not by themselves sufficient to describe or predict response, and that it is interaction of these variables with relative geographic location that is most critically important. These interactions are of two kinds: neighbourhood effects and location effects. The former is dealt with to some extent by spatial classification but might also be made more explicit. The assumption here is that responders cluster, more or less regardless of socioeconomics. The relative location effects are probably either evidence of recurrent patterns or spatial proxies for missing variables; for example, place, culture, accessibility, locality, etc. Examples would be: M25 corridor, rural not urban, town

centre, peri-urban, seaside location, market town rather than industrial town, small town rather than large town, New Town, high/low unemployment area, areas recent population influx, etc. A systematic search might well be worthwhile especially if it could cover a large number of marketing response data sets.

The cluster detecting tools of relevant here are those developed to detect localised anomalies in rare cancer databases; see Openshaw and Craft (1991) Besay, and Newell (1991), and Cuzick and Edwards (1990), Draper (1991). These technologies are of considerable potential value in marketing especially once the Ordnance Survey's Address Point product provides accurate locational referencing of all addresses in the UK. However, even without it, considerable progress could be made and new dimensions added to direct marketing targeting systems. This spatial epidemiological technology is particularly appealing because even the lowest marketing response rates will exceed rare cancer incidences by at least one order of magnitude. Also, it is apparent that the marketing industry could do themselves a great favour by mining their data sets for clustering in response to their mailouts, with a particular concern for persistent distributions, especially clusters that are in predictable location or clusters which are common to different types of products.

4.4. Fuzzy spatial targeting

The concept of a fuzzy geodemographic system is not new (Openshaw, 1989, 1991) but none exist as yet. The basic idea is to exploit the spatial fuzziness in the standard geodemographic systems. There are two areas of fuzziness: first in the assignment of a census enumeration district to a cluster and second in the assignment of a postal address to the census enumeration district.

In all multivariate classifications of spatial data it is unavoidable that some areas will be mis-assigned, and others may well be 'near' in a taxonometric sense to more than one cluster. Yet the current systems are based on all-or-nothing assignments, so no fuzziness exists. A fuzzy geodemographic system would allow you to ask the question which addresses are not in the target groups but 'near' to them in the similarity space of the classification. This is important in a spatial context because the internal heterogeneity of census eds varies so much.

The second source of fuzziness concerns the assignment of addresses to census eds. Currently, there is no 100% accurate directory but even when there is, it might still be beneficial to consider fuzzy queries such as, identify addresses not in the target areas but within 250 metres of them. Obviously both sources of fuzziness can be combined to yield queries such as find addresses within 500m of

target areas within clusters which are different but near the target ones.

The problem now is that this requires much larger databases and much more complex user interfaces. The ideal system would automatically calibrate itself by identifying optimal levels of fuzziness for any particular marketing application. The benefit offered here is a rejuvenation of geodemographics by targeting areas previously excluded by all or nothing approaches.

4.5 Spatial relationships seekers and modellers

The other value of location is as a proxy for other missing variables. For example, a high response from addresses near to a railway may well reflect either accessibility difficulties due to the barrier effect of the track or be a proxy for some socioeconomic or community or cultural effect. It is important to pick-up local spatial associations, that seem to matter. Whilst the focus on individuals rather than areas in targeting may have been very beneficial, the opportunities for further segmenting individuals based on their spatial relationships with surrogate map features might also be useful.

The problems here are purely methodological. If you wish to use spatial relationships as surrogates then you need an exploratory analysis technology to help you do it. It is interesting to note, therefore, that the Geographical

Correlates Exploration Machine (GCEM) of Openshaw et al (1990) started out as a market research tool. Maybe it is time to re-apply it.

A different type of spatial relationship might be found in consumer responses to regionalised variables. Spatial trend surface regression methods have been used in research for at least a decade. It is possible that these methods may also be useful at more macro-geographic scales as a means of modelling spatial response. A good review of relevant methods is provided in Haining (1990). Spatial regression methods should also be used in preference to conventional regression methods whenever zonal data (eg postcode sectors) are being analysed, especially if there is concern with significance testing or interpreting regression parameters by reference to their standard errors.

4.6 Smart modelling

The real leading edge today lies in the development of smart modelling and analysis systems. Market research is an industry that quite often cannot see the forest for the trees. Users need to find or discover insightful, interesting, and commercially valuable patterns or relationships without necessarily being able to define in advance what they are looking for. They know what they want, it is just they don't know how to get it in anything like an optimal manner. The required paradigm is quite

straightforward. You input all the available data, you provide a sample of successes and failures, and you hope the system will do the rest. Amongst the input data will be geographical variables, the system not you determines whether it matters and how best to use it. You could use a neural net but quite often better results will be obtained by using a much simpler memory based reasoning paradigm, a form of case based machine learning. Basically, you either seek to classify cases against a library of good and bad profiles noting k^{th} nearest matches, or you build-up a library of good performers. There are various ways of achieving both objectives. The point to note here is that the results can be extremely good in situations where no other model would work at all. You are using the knowledge that is acquired as a result of the marketing to improve the performance of subsequent marketing efforts.

An interesting variant of this paradigm is to allow a genetic algorithm (GA) to create good customer profiles from the library database. In essence the GA is breeding query operators which would result in clusters of customers with the highest response rates. This might well be considered to be an artificial life approach. How you explain the results to marketeering end users is left to the readers imagination.

Another important aspect concerns the need for feedback. The ideal systems are adaptive and dynamic. They are not based on any detailed process understanding, that is just too hard, also we know too little to do this at present. Ideal systems sense their environment and they respond to it without being told how to do it. The Alife clients are 'alive'; they are living symbolic creatures. They will adapt to change. They are ideally placed for real-time systems or for marketing operations that generate feedback. In a phased mailshot you start from a position of little or no information, you gain response information and the targeting should improve if it can learn as it goes along and is sensitive to success and failure, and also exogenous changes. The technology exists, it is just not used much at present.

4.7 Spatial pattern recognition

A final area of potential spatial analysis relevancy concerns what might be termed spatial pattern recognition. There is a risk in marketing that major spatial and aspatial regularities are being overlooked. Maybe the available analysis tools are too sophisticated and are attempting to be too clever and too smart. By focusing on the minutiae maybe we are failing to see the whole. Instead of trying to predict response for item x or campaign k, maybe we should be considering the overall spatial pattern of response. Are we simply seeing different versions of the same generic pattern or different

spatial patterns; and if the latter is there a library of different spatial pattern types that might exist? Maybe there is a macroscale geographic pattern to most responses, and if there is and if it can be parameterised, then there is a basis here for an entirely new approach; a macro spatial pattern framework within which to fit the detail. It would certainly be interesting to investigate further.

4.8 Spatial data value adding

The final use of spatial analysis in marketing is as a means of enhancing databases by geoprocessing and linking in other spatial data; see Table 2. Geodemographics is one example. More importantly there is now a tremendous wealth of digital map data from the Ordnance Survey, the AA, and Bartholomews. Much of the information is not directly relevant but some is. In particular, the various feature overlays in the digital map databases can be used to tag postcode and census data with contextual information such as distance to various linear features (ie coastline, certain land uses, rivers, woods by type, golf courses, urban boundary), or location within polygons of interest (ie built-up areas) or distance to polygon of interest (ie parks).

TABLE 2 Adding value to marketing data by geoprocessing

distance to selected digital map features
locations within selected digital map polygons
non-standard aggregation
place name processing
house name processing

In addition, there are other ways of adding value by geoprocessing, principally by the analysis of addresses for place name and house name content. There is almost certainly an extremely high association between social status, wealth, and income and postal address and, when present, house names. The associations are not straightforward nor are they linear; but they are undoubtedly strong. It works at both extremes; in that there must be many streets in the UK where the response rates are always low or high. Modern GIS and computer technologies provide the tools for investigating further these spatial associations; if only to add new layers in the spatial query process and new predictor variables in the response prediction models. Indeed, as the latter become cleverer and less dependent on human beings to provide all the intelligence in the form of specification and recodes of variables so it becomes possible to make use of the wider range of contextual spatial information that is available or could easily be generated; see Openshaw (1993).

5. Conclusions

This chapter has attempted to stimulate interest in applied spatial analysis relevant to marketing amongst both geographers and marketeers. Maybe similar comments could also be written about other areas of potential application. However, marketing has the distinction of being the best commercial prospect for exploiting spatial analysis skills

and tools. A number of generic spatial analysis approaches have been defined as being of potential and real marketing relevancy. Now is the time to try them out. Do they work? Can they be made to work? What further developments are needed? What further research is needed? How can we best respond to the needs of the end users? These are all challenges for quantitative geographers.

It should be appreciated that geographers have no proprietorial rights in this area; they are not that unique in what analysis skills they offer. Whilst it might take 5 years or even a life-time to convert a geographer into a 'good' computer scientist, it may only take a few months for a computer scientist to learn or pick-up the bits of geography that matter. To this extent, time is running out and as spatial analysis tools are eventually packaged for popular application, so the opportunity could slip by. This may, of course, be no bad thing but equally it might also be a significant waste of new research opportunities, a turning away from new areas of scientific investigation and scholarship, and a neglect of the possible commercialisation gains from technology transfer.

REFERENCES

Besag, J., Newell, J., 1991, 'The detection of clusters in rare diseases', J. of the Royal Statistical Soc Ser A, 154, 143-155

Charton, M., Openshaw, S., Wymer, C, 1985, 'Some new classifications of census enumeration districts in Britain: a poor man's ACORN', J. of Economic and Social Measurement 13, 69-98

Cuzick, J., Edwards, R., 1990, 'Tests for spatial clustering in heterogenous populations', J. of Royal Statistical Society Series B, 52, 73-104

Daedalus, 1992, 'A new era in computing', Journal of the American Academy of Arts and Sciences Winter, 1992

Dangermund, J., 1993, 'Overview of business applications for GIS', GIS in Business '93: Conference Proceedings, GIS World, Colorado p7-10

Dongarra, J.J., Meuer, H.W., Strohmaier, E., 1993, 'TOP500 Supercomputers', RUM 33/93, Computing Centre, University of Mannheim, Mannheim, Germany

Haining, R., 1990, Spatial data analysis in the social and environmental sciences Cambridge University Pion, Cambridge

Openshaw, S., 1989, 'Learning to live with errors in spatial databases', in M Goodchild, S Gopal (eds) The accuracy of spatial databases Taylor & Frances, London p264-276

Openshaw, S., 1989, 'Making geodemographics more sophisticated', J. of the Market Research Society 31, 111-131

Openshaw, S., 1992, 'A review of the opportunities and problems in applying neurocomputing methods to marketing applications', J. of Targeting, Measurement and Analysis for Marketing, 1, 170-186

Openshaw, S., 1993, 'Special classification', in B Leventhal, C Moy, J Griffin (eds) An introductory guide to the 1991 Census NTC Publications, Henley, p69-82

Openshaw, S., Craft, A., 1991, 'Using Geographical Analysis Machines to search for evidence of clusters and clustering in childhood leukaemia and non-Hodgkin lymphomas in Britain', in G Draper (ed) The geographical epidemiology of childhood leukaemia and non-Hodgkin lymphomas in Great Britain, 1966-83 Studies in Medical and Population Subjects No. 53,, OPCS, HMSO, London P109-126

Openshaw, S., Goddard, J.B., 1987, 'Some implications of the commodification of information and the emerging information economy for applied geographical analysis in the UK', Environment and Planning A, 19, 1423-1440

Openshaw, S., Cross, A., Charlton, M., 1990, 'Building a prototype geographical correlates exploration machine', Int J. of GIS 3, 297-312

Raper, J.F., Rhind, D.W., Shepherd, J.W., 1992 Postcodes: the new geography Longmans, London

Waltz, D.L., 1990, 'Massively parallel AI', Proceedings of National Conference on AI, Boston, August 1990