UCL CENTRE FOR ADVANCED SPATIAL ANALYSIS

UCL

WORKING PAPERS SERIES

Paper 127 - Dec 07

Comparing Classifications: Some Preliminary Speculations on an Appropriate Scale for Neighbourhood Analysis with Reference to Geodemographic Information Systems

ISSN 1467-1298

casa

Centre for Advanced Spatial Analysis University College London 1 - 19 Torrington Place Gower St London WC1E 7HB
Tel: +44 (0)20 7679 1782 casa@ucl.ac.uk www.casa.ucl.ac.uk

Comparing Classifications: Some Preliminary Speculations on an Appropriate Scale for Neighbourhood Analysis with Reference to Geodemographic Information Systems.

Alex Singleton (a.singleton@ucl.ac.uk)

Department of Geography, Pearson Building, University College London, Gower Street, London, WC1E 6BT.

Centre for Advanced Spatial Analysis, University College London, 1-19 Torrington Place, London, WC1E 7HB.

Abstract

Geodemographic classifications represent the multidimensional socio-economic characteristics of people living within defined areas. They have been successfully applied in both private and public sector applications across a range of industries since the 1970s. In UK geodemographic classifications, neighbourhoods are predominantly defined using a full postcode or the 2001 Census Output Areas. Little research to date has been completed on: the potential loss of information one may experience using classifications derived at courser geographical aggregations; the most appropriate scale through which concepts of "neighbourhood" can be measured; or how the characteristics of private sector data sources affect classification structure. This paper presents the results of a comparative analysis between the National Statistics Output Area Classification (OAC) and the commercial postcode level classification Mosaic. It assesses the degree to which commercial classifications constructed at the scale of unit postcode leverage greater insight about the composition of neighbourhoods over those built at output area. It is argued here that the inclusion of additional data at finer spatial granularity in commercial classifications do

present added information about the composition of neighbourhood areas; however, the information loss experienced by switching to classification using output area is reasonably low. These losses are also shown to be socially and spatially heterogeneous. Finally, concerns are raised over the spatial and temporal reliability of those data which are used to inform commercial unit postcode classification, and solutions are offered as to how public sector users may maintain methodological robustness whilst retaining the advantages of finer scale neighbourhood classification.

Introduction

Abler *et al* (1971: 149) argue that "[i]f every object and event in the world were taken as distinct and unique – a thing itself unrelated to anything else – out perception of the world would disintegrate into complete meaninglessness". Humans therefore have an intrinsic desire to create classifications of reality and to seek an ordering framework through which information as perceived may be assembled and understood. Geodemographics presents an instance of such classification, specifically with regard to multidimensional socio-spatial differentiation between small neighbourhood areas. In more general terms, Sleight (1997: 16) describes the field of geodemographics as the "analysis of people by where they live", and as such, since the 1980s these classifications have predominantly been used by private enterprise to leverage competitive advantage through intelligent areal targeting of potential customers (Birkin *et al*, 2002). However, geodemographics per se is not a recent innovation (Batey and Brown, 1995), with lineage extending back to early 19th century urban models in an "ecological tradition" (Bassett and Short, 1980: 9) such as the models of Park and Burgess or Hoyt, through social area analysis (Shevky and Bell, 1955) and then later in more contemporary applications identifying urban deprivation clusters in Liverpool (Webber, 1975). The histories of geodemographics are covered extensively in the literature elsewhere and a full review is provided by Harris *et al* (2005). There are multiple geodemographics in the UK from both commercial and non commercial classification builders, and although they present alternate representations of neighbourhoods they share similar methods of construction. Modern geodemographic classifications are created using one of a small number of data reduction techniques (Ehrenberg, 1982), often generalised as cluster analysis (Everitt, 1974). In the creation of geodemographic classifications these algorithms search within large multidimensional datasets to find clusters of areas which possess similar aggregate socio-demographic characteristics. Once groupings of similar areas have been discovered through this software sorting of people and places (Burrows and Gane, 2006), typology descriptors are written by the classification builder which assign labels and descriptive supporting material to each of the clusters. As such, these "pen portraits" describe the average characteristics of those people living within a group of areas defined by the cluster.

Scale of Neighbourhood Classification

The scale at which geodemographic classifications may plausibly be created depends crucially on the resolution of the input data. For example, a classification created entirely using input data at the level of the Output Area (e.g. 2001 Census) could not purport to describe those socio-demographic patterns that may be exhibited between unit postcodes within an output area. The steady increase of classification resolution has formed a key innovation point in the creation of modern commercial geodemographic systems. Early classifications (e.g. Webber, 1975, Webber and Craig, 1978) grouped areas at the scale of Wards, Local Authorities and Parliamentary Constituencies. It was not until private sector organisations recognised that these techniques of area profiling offered insights into potential target customer behaviours that finer scale classifications were created (Harris *et al*., 2005). The first commercial classification were built at census enumeration district level, then later at unit postcode level and more recently classifications have been created at the level of individuals within households (Webber and Farr, 2001). Commercial classification builders argue that a classification created from data at a finer spatial granularity (where neighbourhood areas are defined as containing fewer numbers of people or households) provide a more precise description of those people living within these areas. Fundamentally this statement is true, however, the availability of data with comprehensive spatial coverage at a finer geographic and temporal resolution is less prevalent, and as such, informing a classification with these data may be problematic due to issues of sample size and bias (Fotheringham *et al*, 2000). Therefore, the definition of "neighbourhood" in a geodemographic sense is determined by the size of the geographic areas used to output the classification, e.g. Unit Postcode, Output Areas or Lower Super Output Areas.

In this paper two geodemographic classifications are compared to examine scale effects in the definition of homogenous neighbourhood areas. These classifications are the National Statistics 2001 Census Output Area Classification (OAC: Vickers, 2007) and Mosaic, a classification from the commercial vendor Experian. Neighbourhoods are defined using output areas in OAC (Martin, 1998) and unit postcodes in Mosaic. An output area contains on average around 39 unit postcodes, thus Mosaic purports to present a higher resolution description of neighbourhood composition. However, the added value that appending geodemographic classification at this increased resolution is debatable, and Harris *et al*.

(2005:219) recognise that “there is no optimal scale for classifying neighbourhoods. Consumer behaviour within some product categories is better predicted using demographic data for areas more geographically extensive than census output areas, while for others the appropriate granularity is as low as unit postcodes”. It is therefore important to choose a classification which is fit for purpose in the context of a specific application. However, these choices are based on assumed prior knowledge about the spatial heterogeneity of the target behaviours under investigation.

The resolution of the data available to the classification builder is a key to deciding at which scale a geodemographic will be created. OAC is a National Statistic and has been created entirely from the 2001 Census data which are disseminated at Output Area level. Those data included in the classification were limited to variables which occurred on the 2001 censuses of all the UK devolved countries. Mosaic contains around 54% 2001 census data and 46% derived from the Experian consumer segmentation database (Experian, 2004). The latter of these data sources contains records for 46 million adult residents and 23 million households across a range of domains including lifestyle survey information, credit activity, shareholders register, house prices, and council tax information (Experian, 2004).. This additional information is used by Experian to differentiate between unit postcodes within an output area. Unlike OAC, which was created using an open and public domain methodology (Vickers, 2007), the precise nature of the data, methods and variable weightings used to create Mosaic and other private sector classification are closed, raising concern over their use in public sector applications (Singleton and Longley, 2008; Longley and Goodchild, 2008) where the implications of misclassification can deprive stakeholders in public services of real life chances.

The output of clustering algorithms used to construct geodemographic classification can often be organised in multiple ways depending on the data used, variable weightings or stability in the optimisation routines (Singleton and Longley, 2008), and as such, geodemographics produced from cluster analysis cannot present a definitive or uncontested representation of reality (Gordon, 1999). The purpose of this paper is not to assess how closely OAC or Mosaic depict reality, i.e. ground truth the classification, but instead consider both classification as alternate representations or interpretations of neighbourhood organisation, limited by their data input, clustering methodology and classification builder biases. The central aim of this paper is thus to evaluate and explore how the ingredients of commercial data input which purport to represent spatial differentiation at finer scales than the Output Area, represent and characterise neighbourhood areas, and whether the addition of these data to a classification increases postcode level differentiation within Output Areas.

Assessing the Social Similarity of Neighbourhood Areas

The first part of this paper considers social similarity between the most aggregate level of Mosaic, a unit postcode level classification, and OAC, a classification at output area level. The aim of this analysis is to identify those neighbourhood characteristics that tend to be shared between the two classifications. OAC Super Groups and Mosaic Groups were appended to every UK unit postcode and then cross tabulated to calculate a series of index scores and percentages. The index scores are calculated by comparing the distribution of all postcodes by OAC against a subset taken for each Mosaic Group and highlight which of the target Mosaic groups are proportionally overrepresented. An index score of 100 indicates the distribution of the target variable is represented in the same proportion as that in the total population. A score of 50 indicates the variable is half as frequent, and 200 twice as frequent. This scoring system is widely used in the geodemographics literature as a method of measuring the concentration of people from particular neighbourhoods within a target population.

Table 1 shows that while a number of Mosaic Groups have significant overrepresentation within particular OAC Supergroups, others show less extreme overrepresentation with more even distributions. A shared characteristic of those neighbourhoods which are more evenly distributed is that they are predominantly identified by Mosaic as comprising people living in more affluent areas. If reality is assumed to be accurately represented by the Mosaic Classification, it could be interpreted that these more affluent neighbourhoods are less well defined in the OAC classification, i.e. by the variables available from the 2001 Census (which, of course, exclude income). However, what is encouraging for those using OAC in applications targeting less deprived areas (as interpreted from the descriptive “pen portraits” accompanying Mosaic), is that this classification shows a greater similarity to Mosaic across these neighbourhood areas.

Table 1: Index Scores and Percentages (Within Brackets), Comparing Mosaic Groups with OAC Supergroups

Mosaic \ OAC	Blue Collar Communities	City Living	Countryside	Prospering Suburbs	Constrained by Circumstances	Typical Traits	Multicultural
Symbols of Success	1	127	108	277	1	48	20
	(0.2)	(8.7)	(23.4)	(57)	(0.1)	(8.9)	(1.7)
Happy Families	62	27	43	203	16	186	31
	(8.2)	(1.8)	(9.3)	(41.7)	(1.7)	(34.5)	(2.6)
Suburban Comfort	23	7	93	225	4	126	72
	(3)	(0.5)	(20.1)	(46.4)	(0.4)	(23.4)	(6.2)
Ties of Community	111	78	19	39	56	238	209
	(14.6)	(5.3)	(4.1)	(7.9)	(5.9)	(44.2)	(18)
Urban Intelligence	0	747	1	3	8	101	330
	(0.1)	(51.1)	(0.3)	(0.5)	(0.8)	(18.8)	(28.4)
Welfare Borderline	16	196	0	0	469	0	404
	(2.1)	(13.4)	(0)	(0)	(49.7)	(0)	(34.7)
Municipal Dependency	381	0	0	0	397	0	89
	(50.2)	(0)	(0)	(0)	(42.1)	(0)	(7.7)
Blue Collar Enterprise	511	1	8	2	186	26	71
	(67.2)	(0)	(1.7)	(0.4)	(19.7)	(4.9)	(6.1)
Twilight Subsistence	91	46	7	2	699	21	57
	(12)	(3.1)	(1.5)	(0.4)	(74.2)	(3.9)	(4.9)
Grey Perspectives	28	180	117	104	49	166	14
	(3.7)	(12.3)	(25.3)	(21.5)	(5.2)	(30.7)	(1.2)
Rural Isolation	1	0	452	7	0	2	0
	(0.1)	(0)	(98)	(1.4)	(0)	(0.4)	(0)

Under the assumption that Mosaic presents an accurate representation of reality, it is possible to further consider the effects of using a larger areal aggregation (e.g. output areas) to assign a typology description to a neighbourhood. This is a useful analysis which explores the degree of information loss that OAC may incur because it is created and disseminated only from data aggregated at larger geographical units. This analysis also examines the geodemographic profile of those neighbourhoods where Mosaic clusters uniformly within an Output Area. An algorithm was created which counts the frequency of different Mosaic Groups at unit postcode level within each output area. Those postcodes without Mosaic information (e.g. business addresses) were excluded from this analysis, creating a dataset of around 1.4 million postcodes. For each output area the frequency of different Mosaic Groups was recorded and the total population within these areas appended. The population was then aggregated by the different frequency of Mosaic Groups scores present across all output areas. The results are shown in Table 2. If we again assume that the Mosaic classification is 'correct', this table provides a guide to the numbers of people who may be misclassified through failure to detect variability at coarser levels of granularity than the unit postcode.

Table 2: Total Population within OA by Frequency of Different Mosaic Groups

Different Mosaic Group Frequency	Total Population	Percentage Population
0*	785 009	1.3
1	28 525 957	48.6
2	21 838 386	37.2
3	6 553 279	11.2
4	912 243	1.6
5	68 253	0.1
6	2 413	0.0
7	408	0.0

* = The population within OA that have 0 postcodes are OA in which all postcodes are classified as unknown in the Mosaic classification.

Around 48.6% of the total population live within output areas that are considered socially homogenous at the unit postcode level by Mosaic Groups. A further 37.2% live within output areas which contain 2 different Mosaic Groups. Only a very small proportion of the total population live within output areas which are very diverse at the level of the unit postcode. These data are further illustrated by two maps which are presented in Figure 1 and Figure 2 showing an output area which is homogeneous at unit postcode level by Mosaic Groups, and a second output area which is very heterogeneous.

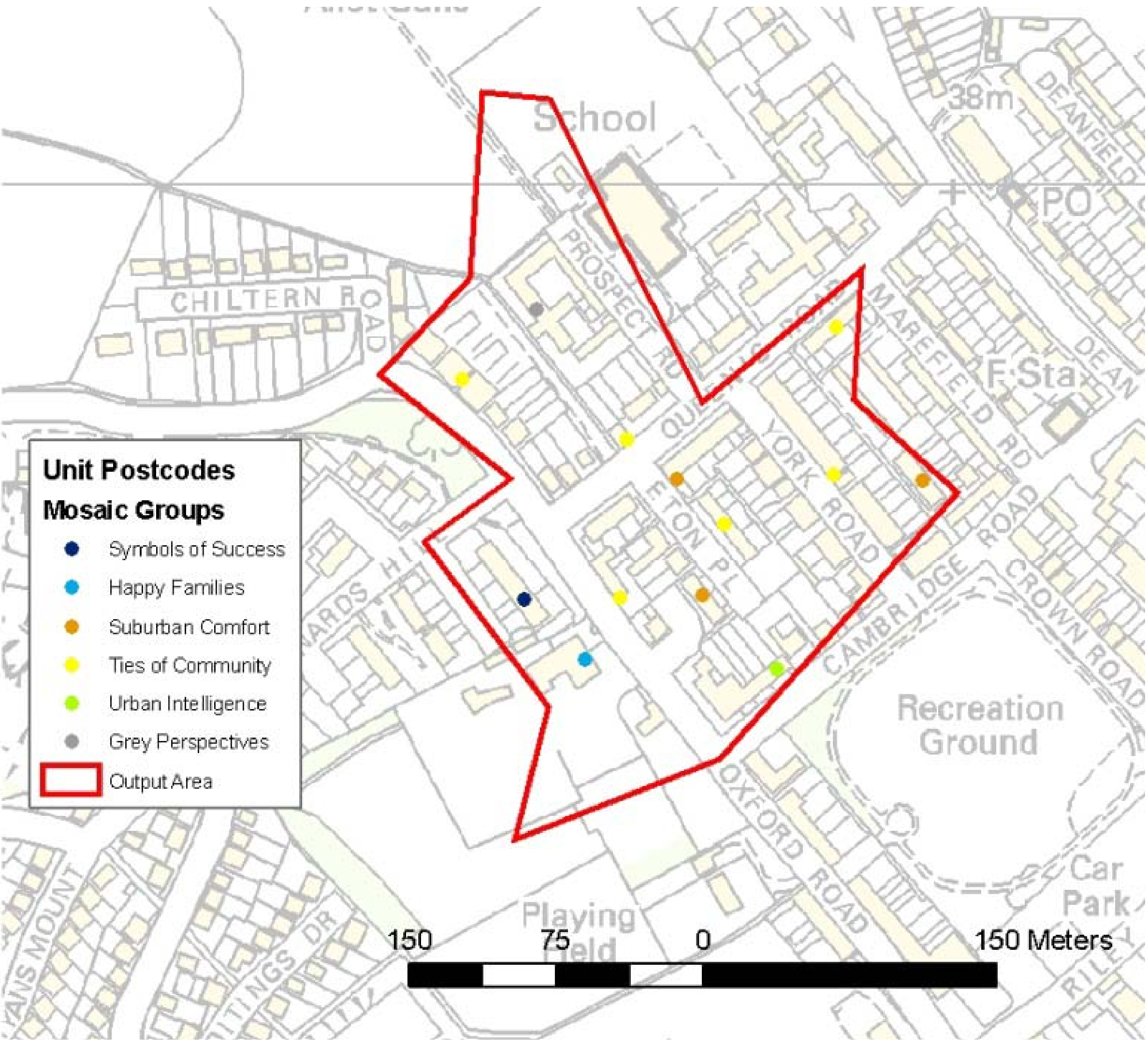

Figure 1: A Heterogeneous Output Area "11UFHC0012" (Boundaries © Crown copyright 2007)

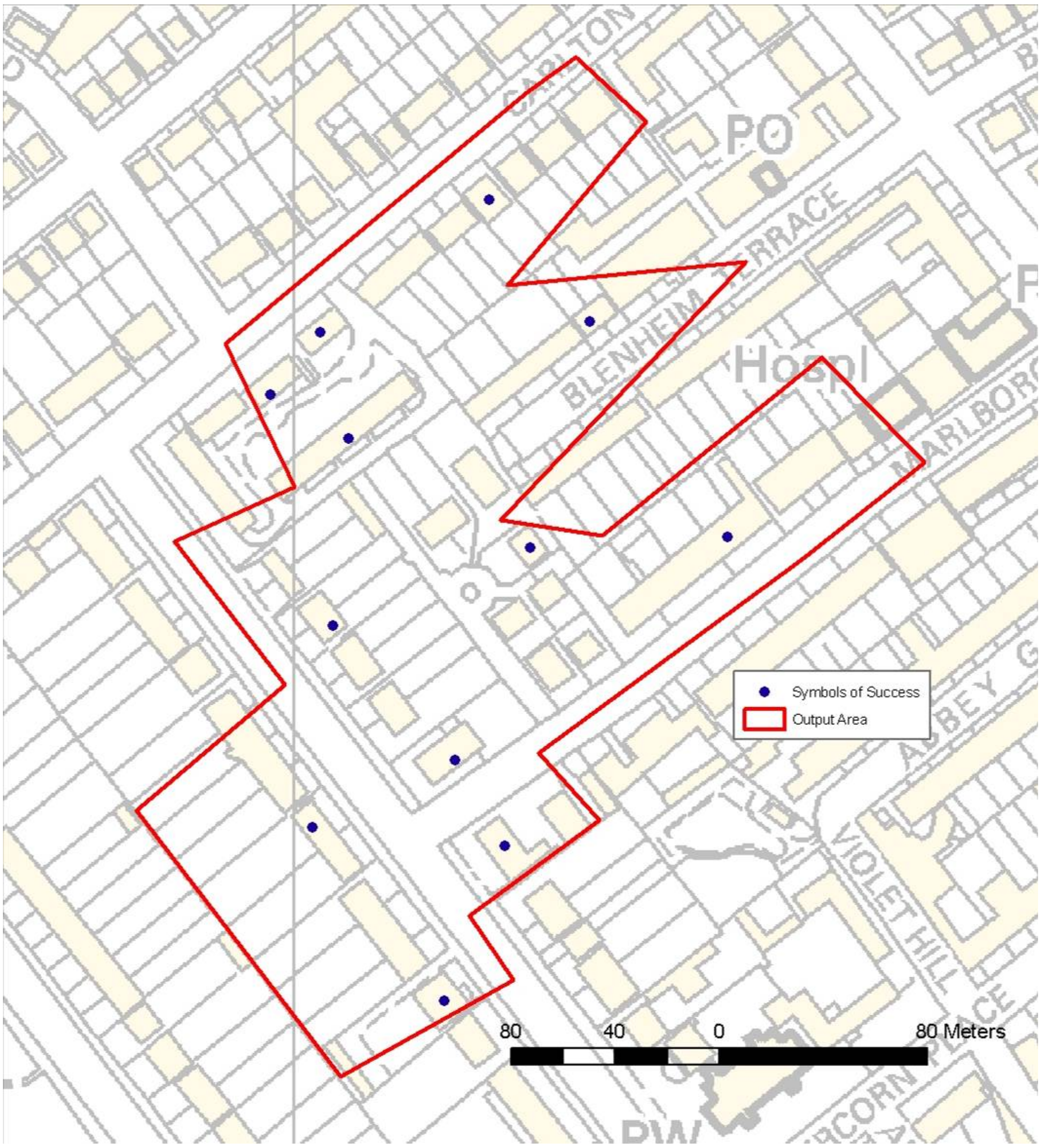

Figure 2: A Homogeneous Output Area "00BKGA002" (Boundaries © Crown copyright 2007.)

Table 2 shows that 48.6% of the total population live within apparently homogenous output areas, i.e. those that contained unit postcodes categorised by a single Mosaic Group. The postcodes within these output areas can be profiled to assess the types of OAC neighbourhoods where Mosaic coded postcodes are more typically found to cluster at output area level (see Figure 3). An explanation for the differentiation could be that these homogenous areas (by Mosaic) are identified from within the attribute space predominantly by census variables (disseminated at output area level), or, these areas are defined by neighbourhoods which typically may show greater spatial homogeneity at output area level.

The latter interpretation is supported by a large overrepresentation of the Type "Just Moving In" from within the Group "Happy Families" (See Figure 4). These areas are typically newly built houses within homogeneous estates where one may expect a greater degree of social similarity across an area typically measured by the extent of an output area.

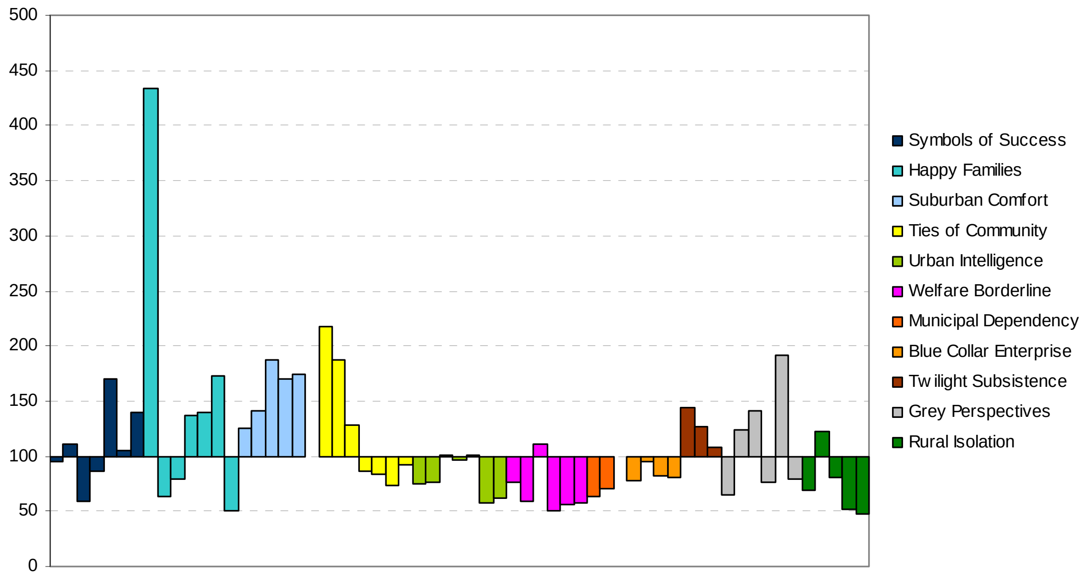

Figure 3: Homogeneous Output Areas by Mosaic Types (Source: Mosaic – Experian Ltd)

Figure 4: A Photograph of the Mosaic Type "Just Moving In": Babbage Way, Bracknell, Large new houses on small plots, RG12 7GN (Source: Mosaic – Experian Ltd)

Heterogeneous output areas with two Mosaic Groups were shown in Table 2 to account for a significant proportion of the total population (37.2%). The following analysis considers the magnitude of the distribution between the two groups. If an output area contained a large proportion of postcodes from one Mosaic Group and a single postcode of another group, this could perhaps be considered less significant than a output area in which there were an even split in unit postcodes between two neighbourhood groups. Alternatively, it could be argued that these micro level differences between unit postcodes, and no matter how small in magnitude, are a key discriminator that makes classification disseminated at the unit postcode level more powerful.

A subset of the cross tabulated data was created which consisted of those output areas containing unit postcodes split across two Mosaic Groups. Across the various possible combinations of these splits, the proportion of the largest of the two groups was calculated as a percentage of the total. These scores are influenced by the number of postcodes within an output area. For example, an output area which contained only two postcodes could only ever be assigned a single score of 50%. The range of scores that can be assigned to an output area will therefore increase as the frequency of postcodes increases (see Table 3). The frequency of output areas with the number of different postcodes within them is shown in Figure 5.

Table 3: Possible Homogeneity Scores

Postcodes Frequency within OA	Frequency of maximum possible scores	Those scores possible
2	1	50.0%
3	1	60.7%
4	3	50.0%, 75.0%
5	2	80.0%, 60.0%
6	2	83.3%, 66.7%, 50.0%
7	3	85.7%, 71.4%, 57.1%
8	4	87.5%, 75.0%, 62.5%, 50.0%
9	4	88. 9%, 77.8%, 66.7%, 55.6%
10	5	90.0%, 80.0%, 70.0%, 60.0%, 50.0%

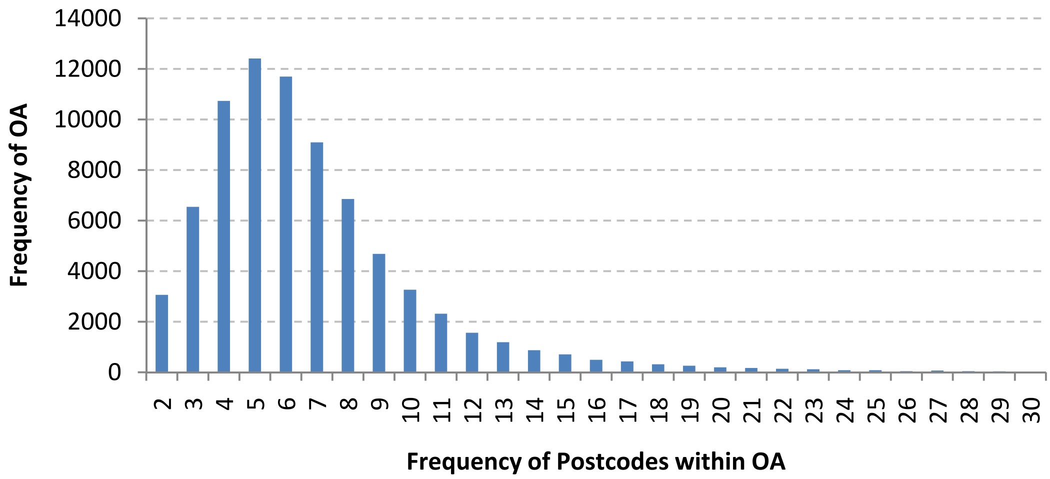

Figure 5: Frequency of Postcodes within Output Areas

One could therefore expect certain percentage scores to be more overrepresented based on their probability of appearing within the dataset, e.g. 50% could be possible for every output area with even number of postcodes. Despite this caveat the analysis still provides a useful measure of output area level neighbourhood heterogeneity. The distribution of the maximum percentage scores created by this analysis is shown ranked by order of frequency in Figure 6. With the exception of 50%, the top five output area frequencies all indicate over three quarters allocated to a single Mosaic Group. In the top five percentage groups ranked by frequency of output areas, there are scores of 80% and 83% of postcodes allocated to a single Mosaic Group. These findings are encouraging for users of the OAC classification as it appears that the loss of postcode level information may be minimal within the majority of output areas split into two or fewer Mosaic Groups.

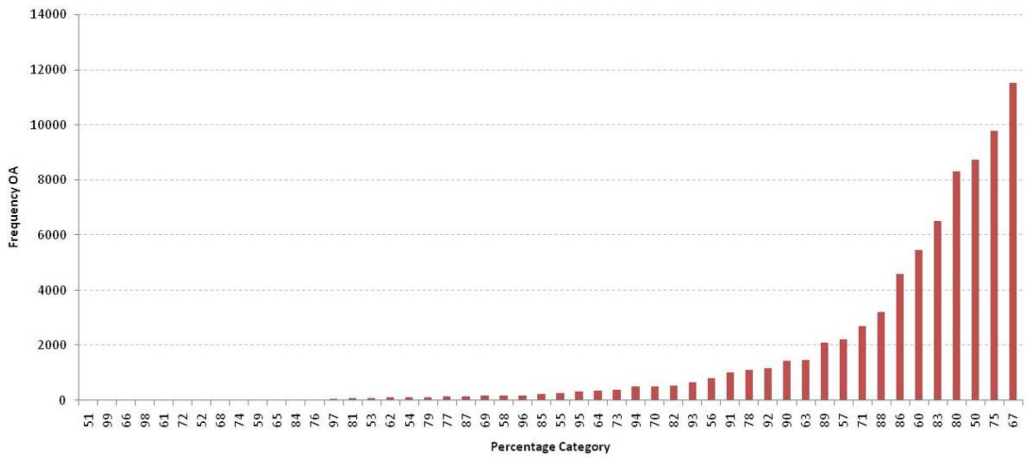

Figure 6: Distribution of Maximum Percentage Scores for OAs containing two Mosaic Groups

The analysis was repeated for those output areas which contained unit postcodes that were split into 3 Mosaic Groups (approximately 11.2% population). The results are shown in Figure 7 and it can be seen that the ranges of Mosaic Groups within these output areas are more heterogeneous than those output area with dichotomous Mosaic Groups; however, there are still a large number of output area with 80% and 70% of their postcodes assigned to a single Mosaic Group.

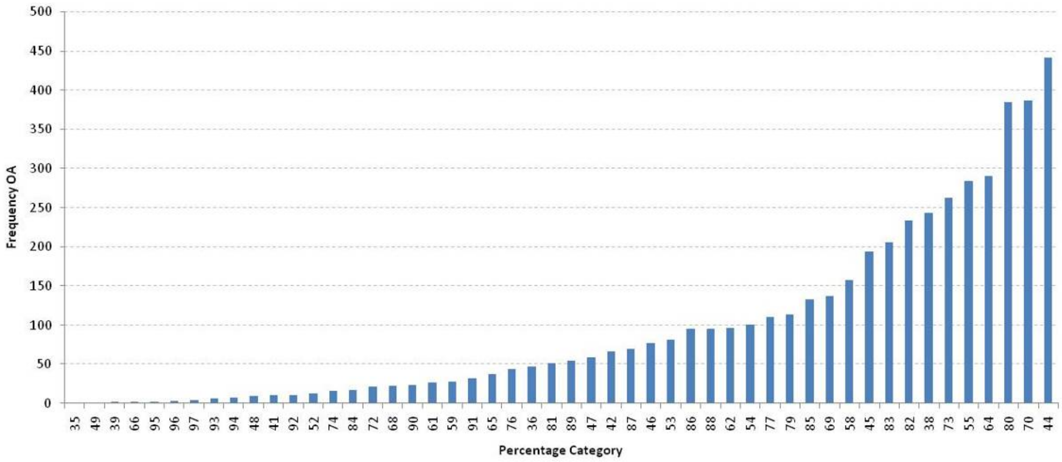

Figure 7: Distribution of Maximum Percentage Scores for OAs containing three Mosaic Groups

Indices of Dissimilarity

A further method to measure the postcode level heterogeneity within output areas can be gauged is through the use of diversity indices. These are established techniques predominantly used in ecology to measure the heterogeneity of species within biological areas such as transects. These techniques can be adapted to provide a further method of measuring the homogeneity of Mosaic Groups within output areas. The Simpson Diversity Index (Simpson, 1949) measures the balance of neighbourhood groups within an output area and the scores range between 1 and 0. A more heterogeneous output area would contain a number of different Mosaic Groups and have a score closer to 0, whereas a homogeneous output area would have fewer Mosaic Groups and a score closer to 1. The Simpson Diversity Index (*D*) in output area *e* is calculated by summing the square roots of *m* Mosaic Groups ranging from *i*=1 to =11 which correspond to each of the 11 Mosaic Groups (see Equation 2).

$$D_e = 1 - \sum_{i=1}^{11} \left(m_{ei}{}^{2} \right)$$

Equation 2

Once the diversity scores have been created for each output area they may be mapped to examine the spatial distribution of output area level homogeneity (see Figure 8). It is shown that in urban areas there is a tendency for output areas to have a higher diversity with rural areas such as North Devon, Mid to North Wales, North Lancashire and the majority of Scotland all having lower diversity. An exception to this broad pattern is Northern Ireland which appears quite diverse, even in areas which would be considered predominantly rural. Furthermore, London appears to have lower diversity in many of the central areas outside of the main city (see Figure 9), possibly due to the density of people living within these areas and the consequential creation of smaller output areas.

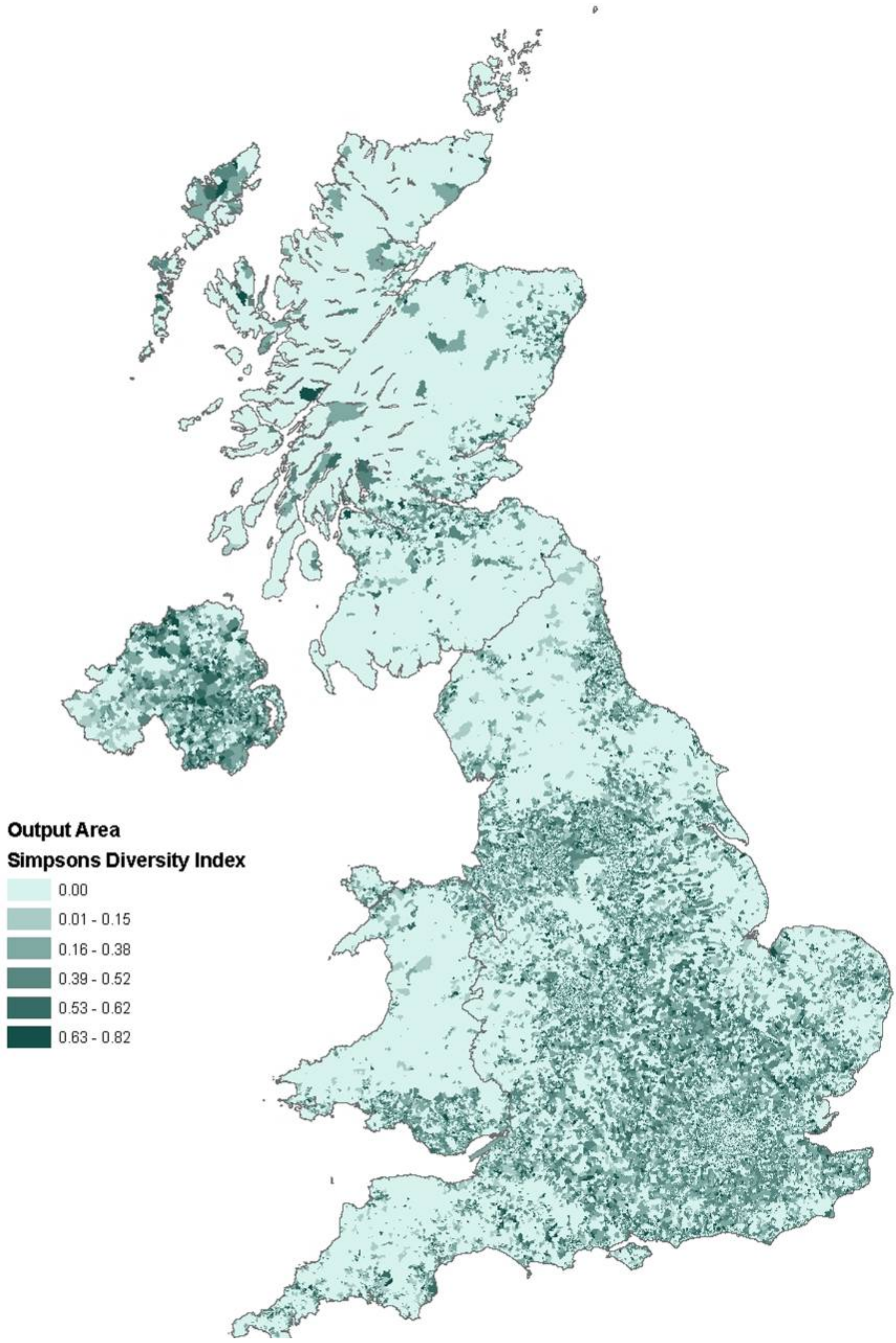

Figure 8: UK Spatial Distribution of Simpson’s Diversity Index

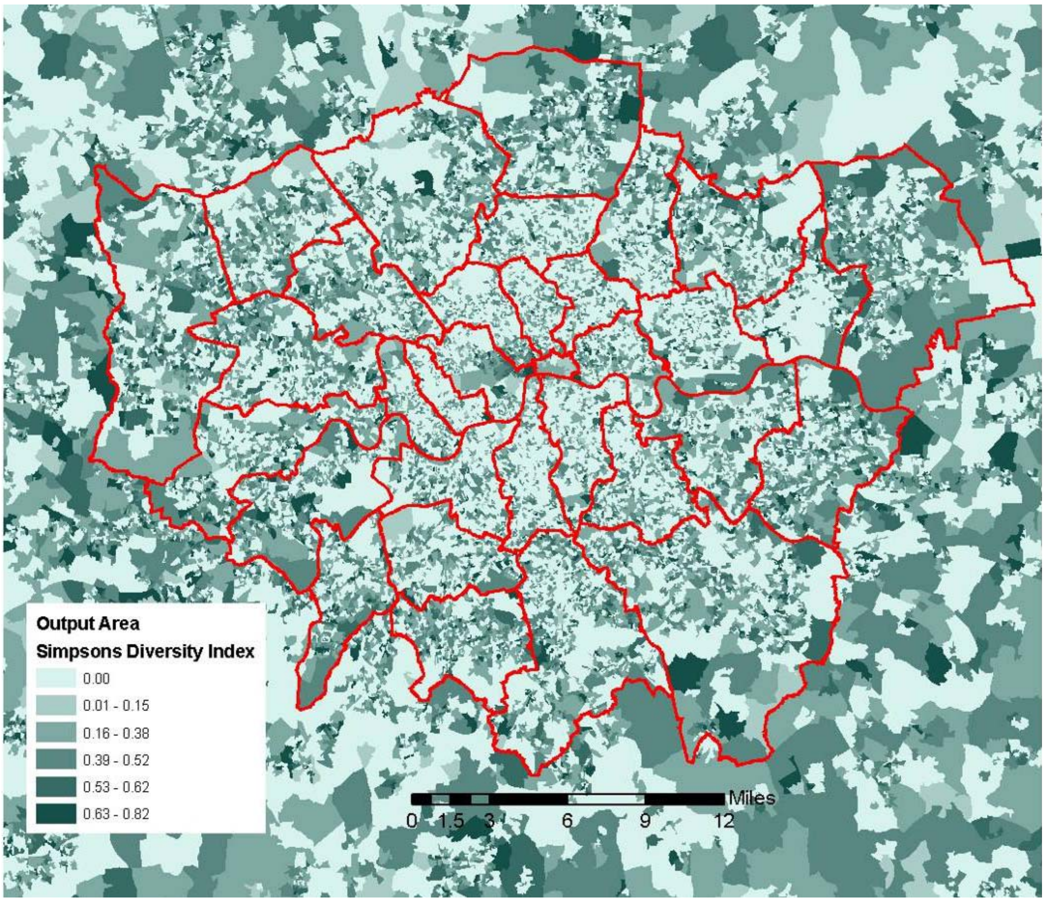

Figure 9: London Spatial Distribution of Simpson's Diversity Index

By aggregating the output area diversity scores into the OAC typology it is possible to assess whether particular neighbourhood types demonstrate heterogeneity within their assigned output areas. Because OAC clusters vary in size, these aggregated scores were transformed into index scores using the complete OAC Group distributions as a base. Thus, Figure 10 represents the Mosaic postcode level diversity within each of the OAC neighbourhood Groups. A high index score indicates an area which is heterogeneous, and a low index score more homogeneous. There are differences shown both between and within the Super Groups. The OAC Super Group "Constrained by Circumstances" contains three Groups. The first ("Senior Communities") and last ("Public Housing") of these Groups are both more homogeneous, while the middle Group ("Older Workers") contains increased diversity between unit postcodes. From within the Super Group "Blue Collar Communities", the last

Group “Older Blue Collar” are more heterogeneous. Within the Super Group “Countryside”, the Group “Agricultural” is shown to be more homogeneous. The other two groups within this Super Group are areas which are more urban such as small villages or towns.

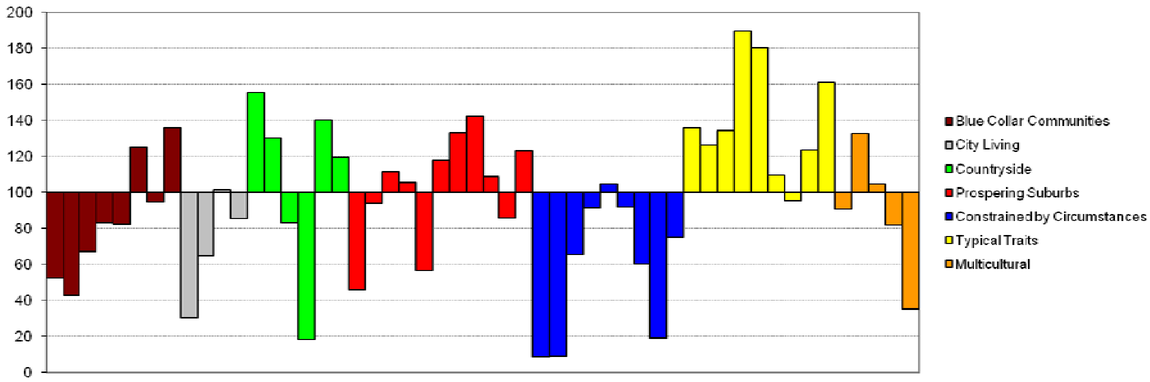

Figure 10: Incidence of High Simpson Diversity Index scores by OAC Subgroups

Harris *et al* (2005) imply that that in the production of the Mosaic classification Experian may re-cluster non urban areas in a separate analysis in order to create clusters which represent a finer level of detail. This could possibly explain why the OAC Groups “Village Life” and “Accessible Countryside” show greater heterogeneity at the level of unit postcode. The Super Group “Typical Traits” represents some of the most heterogeneous areas with the exception of one Group, “Young Families in Terraced Homes” which presumably because of the high spatial concentration of the characteristic housing stock, appear to have greater social similarity at the level of output area. In the Super Group “Multicultural” the “Afro Caribbean Communities” Group are more homogeneous than the “Asian Communities” Group. Through examination of the Mosaic typology and pen portraits it appears that the Mosaic classification achieves a high level of differentiation between wealthy and deprived Asian neighbourhoods. Thus, these neighbourhood Types appear across two Mosaic Groups and as such may be associated with greater recorded diversity within output areas.

Discussion and Conclusion

Voas and Williamson (2001) argue that general purpose classification built at geographies similar to output areas (enumeration districts) can conceal hidden diversity at a finer scale. Harris (2001) also argues that finer scale information such as those available in lifestyle databases at individual or household level may produce more satisfactory classification to discriminate between areas. These provide persuasive arguments for the commercial classification builders, as the imputation of higher resolution data where access is prohibitively expensive to general or public sector users, allows their classification to be created at finer resolutions, and therefore leverage market advantage from those without access to these data. However, the analyses presented in this paper have started to address these issues systematically and to re-assess at what scale it is appropriate to create and disseminate geodemographic classifications. The evidence appears to suggest relatively little heterogeneity between postcode level Mosaic Groups within output areas and this supports the argument that an output area may be an appropriate scale at which to categorise neighbourhoods. This is also supported by Vickers (2006), who contends that finer scale geographies are more prone to error induced through aging of the data. The finer the scale at which input data area collected, the increased temporal refreshing is required to maintain that they are accurate and safe to use. At a coarser level of granularity, such as the OA, small changes in the population will not alter how an area should be most appropriately classified, as the average characteristics of those residents still within the area will predominantly remain aligned to the classification representing these aggregate characteristics. As discussed earlier, the micro-level data which are used by commercial classification builders often originate from sample surveys which are not representative of the total population, and as such, the implied precision of a classification at unit postcode level accuracy may be superfluous. Indeed, the use of these data may explain why in the analysis presented in this paper a minority of output areas were found to be very heterogeneous, with one output area in Northern Ireland containing 7 different Mosaic Groups, perhaps caused by erroneous imputation of sample data, or the unavailability of household level databases. Despite concerns over the appropriate scale at which to create geodemographics, Voas and Williamson (2001) also identified that geodemographics can be improved through creation of task specific classification, which is augmentation of census data with domain specific data or re-weighting of variables to reflect a given purpose. The development of bespoke classifications offer significant potential for public sector users wishing to obtain geodemographic classifications at scales finer than Census output areas while maintaining confidence that areas are not miss-specified through use of geographically disparate sample data. There appears to be significant opportunity in developing these domain specific classifications at unit postcode level using person level attributes from public sector administrative databases. These classifications could offer the benefits of increased methodological robustness while retaining the performance advantages attributable to finer scale neighbourhood classification, and at a lower cost than current commercial offerings.

References

ABLER, R., ADAMS, J. S. & GOULD, P. (1971) *Spatial Organisation: The Geographer's View of the World,* London, Prentice-Hall International.

BASSETT, K. & SHORT, J. (1980) *Housing and Residential Structure: Alternative Approaches,* London, Routledge.

BATEY, P. W. J. & BROWN, P. J. B. (1995) From Human Ecology to Customer Targeting: the Evolution of Geodemographics. IN LONGLEY, P. & CLARKE, G. (Eds.) *GIS for Business and Service Planning.* Cambridge, GeoInformation International.

BIRKIN, M., CLARKE, G. & CLARKE, M. (2002) *Retail Geography and Intelligent Network Planning,* Chichester, Wiley.

BURROWS, R. & GANE, N. (2006) Geodemographics, Software and Class. *Sociology,* 40**,** 793-812.

EHRENBERG, A. S. C. (1982) *A Primer in Data Reduction,* Chichester, John Wiley and Sons.

EVERITT, B. (1974) *Cluster Analysis,* London, Heinemann Educational Books Ltd.

EXPERIAN (2004) *Mosaic United Kingdom: The Consumer Classification for the UK,* Nottingham, Experian.

FOTHERINGHAM, S., CHARLTON, M. & BRUNSDON, C. (2000) *Quantitative Geography: Perspectives on Spatial Data Analysis,* London, Sage.

GORDON, A. D. (1999) *Classification, 2nd ed,* Boca Raton FL, Chapman and Hall/CRC.

HARRIS, R. (2001) On the diversity of diversity; is there still a poace for small area classification? *Area,* 33**,** 329-336.

HARRIS, R., SLEIGHT, P. & WEBBER, R. (2005) *Geodemographics, GIS and Neighbourhood Targeting,* London, Wiey.

LONGLEY, P. A. & GOODCHILD, M. F. (2008) The use of Geodemographics to Improve Public Service Delivery. IN J, H., C, D., C, S. & M, W. (Eds.) *Managing to Improve Public Services.* Cambridge Cambridge University Press.

MARTIN, D. D. (1998) Optimizing census geography: the separation of collection and output geographies. *International Journal of Geographical Information Science,* 12**,** 673-85.

SHEVKY, E. & BELL, W. (1955) *Social Area Analysis: Theory, Illustrative Application, and Computational Procedures,* Stanford, Stanford University Press.

SIMPSON, E. H. (1949) Measurement of Diversity. *Nature,* 163**,** 688.

SINGLETON, A. & LONGLEY, P. (2008) Creating Open Source Geodemographics - Refining a National Classification of Census Output Areas for Applications in Higher Education. At *Papers in Regional Science,*

SLEIGHT, P. (1997) *Targeting Customers: How to Use Geodemographic and Lifestyle Data in Your Business,* Henley-on-Thames, NTC Publications.

VICKERS, D. (2006) Multi-level Integrated Classifications Based on the 2001 Census (Unpublished thesis). *Department of Geography.* Leeds, University of Leeds.

VICKERS, D. (2007) Creating the UK National Statistics 2001 output area classification. *Journal of the Royal Statistical Society. Series A. Statistics in society,* 170**,** 379.

VOAS, D. & WILLIAMSON, P. (2001) The diversity of diversity: a critique of geodemographic classification. *Area,* 33**,** 63-76.

WEBBER, R. (1975) Liverpool Social Area Study, 1971 Data: PRAG Technical Paper 14. IN STUDIES, C. F. E. (Ed. London.

WEBBER, R. & CRAIG, J. (1978) A Socio-Economic Classification of Local Authorities in Great Britain. IN HMSO (Ed. London.

WEBBER, R. & FARR, M. (2001) MOSAIC: From an area classification system to individual classification. *Journal of Targeting, Measurement and Analysis for Marketing,* 10.