

Geodemographics and spatial interaction: an integrated model for higher education

A. D. Singleton · A. G. Wilson · O. O'Brien

Received: 15 February 2010 / Accepted: 27 September 2010 / Published online: 3 November 2010
© Springer-Verlag 2010

Abstract Spatial interaction modelling and geodemographic analysis have each developed as quite separate research traditions. In this paper, we present an integrated model that harnesses the power of spatial interaction modelling to behavioural insights derived from a geodemographic classification. This approach is applied to the modelling of participation in higher education (HE). A novel feature of the paper is the integration of national schools, colleges and HE data; a national model is then calibrated and tested against actual recorded flows of students into HE. The model is implemented within a Java framework and is presented as a first step towards providing a quantitative tool that can be used by HE stakeholders to explore policies relating to such topics as widening access to under-represented groups.

Keywords Spatial interaction · Geodemographics · Higher education · GIS

1 Introduction

In this paper, we present the results from a research project that integrated a geodemographic typology into a spatial interaction framework to model flow of students from English schools and colleges into the HE system. Our motivation for this work is that, despite a growth in the absolute numbers of students attending HE over recent years, differentials in the rates of participation between societal groups have remained persistently uneven (Comptroller and Auditor General 2008). Within

A. D. Singleton (✉)
Civic Design, University of Liverpool, The Gordon Stephenson Building,
74 Bedford Street South, Liverpool, Merseyside L69 7ZQ, UK
e-mail: alex.singleton@liverpool.ac.uk

A. G. Wilson · O. O'Brien
Centre for Advanced Spatial Analysis, University College London,
1-19 Torrington Place, London WC1E 7HB, UK

the geography and GIS literature, inequality in HE access has been examined separately using both spatial interaction (Wilson 2000) and geodemographic approaches (Batey et al. 1999; Singleton 2010). However, hitherto, there has not been a unified framework within which the relative benefits of both traditions can be maximised. In this paper, we define our integrated model mathematically and then demonstrate operational utility as a Java software application. Modelled results from this tool are encouraging, and we illustrate this in two case studies looking at the flows of students to the University of Manchester, and then to all institutions, from Norfolk. The overarching aim for creating this model is to provide a robust tool where various policy scenarios could be tested by key stakeholders to examine potential outcomes in terms of aggregate student participation, and how these rates may be disaggregated by institution and neighbourhood type.

Geodemographic modelling of neighbourhood conditions presents a successful applied approach to understanding socio-spatial differentiation at the neighbourhood scale and is widely used in the public as well as the private sectors to predict the consumption of products, services or resources. Often described as the ‘analysis of people by where they live’ (Sleight 1997), these classifications assign small geographical areas into a series of discrete types that are deemed to represent the average characteristics of the local population. Although the origin of such classification lies in public sector applications in urban deprivation targeting (Harris et al. 2005), geodemographics are today best known for their commercial successes in customer profiling and targeting (Birkin 1995; Birkin et al. 2002). There is currently considerable interest in the use of these classifications in public sector service delivery applications. This “renaissance of geodemographics” (Longley 2005) includes examples from policing (Ashby and Longley 2005), health (Aveyard et al. 2002) and education (Batey et al. 1999; Singleton 2010; Tonks and Farr 2003).

The underlying statistical models used in the creation of geodemographic classifications are founded upon social similarities and do not generally specify geographical proximity (as opposed to general distance to features such as the coast or town centres). The observed patterning of the resulting cluster types (see those maps produced by Vickers and Rees 2007 for example) arise from the old adage that “birds of a feather flock together” (Feng and Flowerdew 1998) and Tobler’s classic observation that “everything is related to everything else, but near things are more related than distant things” (Tobler 1970). As such, the clustering algorithms that are used to create the classifications are essentially aspatial, with similarity measured only in terms of the attribute space of the input dataset. Thus, geodemographic classifications posit that residents in any neighbourhood assigned to a specific type possess the exact same behavioural characteristics irrespective of geographical location. Geodemographics have been viewed as “methodologically unsatisfactory” (Twigg et al. 2000: 1111) because of their insensitivity to systematic geographical variations in behaviour patterns. Voas and Williamson (2001: 74), for example, found that despite broad-brush social similarities, “small areas are different in many different ways” and that “[a] few dimensions, however complex or carefully chosen, do not provide enough information to describe an area fully”. Yet despite this apparent methodological over-simplification, “[g]eodemographics has had a robust and enduring pedigree” (Birkin et al. 2002: 207) in commercial

applications and by extension are increasingly used by the public sector for social marketing and service delivery. The enduring popularity of these techniques lies in their usefulness in communicating salient patterns within complex multidimensional datasets to end-users who are not professional statisticians or geographers. However, the remit of such analysis is limited and there is a requirement to refocus the geodemographic research agenda on “theory, model building, and hypothesis testing” (Harris et al. 2007: 556).

Spatial interaction models represent and predict the size of spatial flows between origins and destinations in geographical regions of interest. They can incorporate a range of origin and destination constraints and take a number of forms according to this constraint structure. This has a long history and was developed in the USA during the 1950s and 1960s—see Bureau of Public Roads (1965). In this paper, we use the family of entropy maximisation models (Wilson 1970). There are alternatives such as discrete spatial choice or competing destinations models (Fotheringham 1983) though the use of these would not change the character of the argument presented here. Spatial interaction models remain “some of the most applied geographical techniques” (Fotheringham et al. 2000: 214) and have provenance in a range of application areas, including retailing (Birkin et al. 2003), transport (Wilson 1998; Erlander and Stewart 1990) and migration (Stillwell 1978). Early spatial interaction models were criticised for their “lack of behavioural theory” (Han and Timmermans 2006: 195). However, the integration of geodemographic neighbourhood classifications can provide a convenient and behaviourally plausible means of disaggregating the populations at origin zones of a spatial interaction model. This specification of geodemographic classes will better represent variability in the typical local aggregate behavioural characteristics of individuals and their interactions with the HE system (Han and Timmermans 2006). As such, a behaviourally informed spatial interaction model should not only provide greater predictive power, but also offer greater flexibility to support decision-making (Birkin and Clarke 1998).

2 Data sources and geodemographic integration

The data and models presented in this paper relate to England only, although this could later be expanded if data were made available for other parts of the UK. The schools’ data were derived from the National Pupil Database (NPD) supplied by the Department for Education (DfE) that records a variety of attribute information for every state and privately funded pupil in England. Data were extracted from the NPD for those students who had qualified in the 2006–2007 academic year with A-Levels or equivalents (aged around 18) from schools and sixth form colleges. In addition, records were extracted from the Learning and Skills Council (LSC) 2005 Independent Learner Record (ILR) for those studying qualifications at a level enabling HE entry. The date mismatch of the FE records with the schools’ data is a result of a withdrawing of public access to ILR for projects outside of the DfE. The HE data were derived from the Higher Education Statistics Agency (HESA) and record attributes and home locations of all students studying courses of HE in

England during 2005–2006. The education data extracts were imported into a MySQL database at the student record level and then coded with the Office for National Statistics Output Area Classification (Vickers and Rees 2007: OAC). For FE, state school and HE records, these were coded using the home address location of each student. However, for independent schools, these records are missing geographical attribution, and as such, OAC codes were later appended on aggregate to these pupils based upon the OAC code of each school's location. OAC is a free geodemographic classification that disaggregates the UK population into a typology comprising 52 Sub Groups that hierarchically nest into 21 Groups and these into 7 Super Groups.

An estimated supply of students who could potentially attend HE was then created for each of the 7 OAC Sub Groups within 150 geographical zones based on local authorities (LA) in England. These boundaries are shown later in the results section (see Figs. 3, 4). The most aggregated level of 7 Super Groups was chosen over the other hierarchical levels of OAC to reduce the overall size of the spatial interaction matrix. The origin estimates for the spatial interaction model therefore comprised $150 * 7$ variables (1,050). This additional disaggregation of the supply population by OAC aims to capture heterogeneity in HE participation behaviours, which are known to stratify geodemographic clusters (see Batey et al. 1999; Singleton 2010). The demand capacity for each (university) destination in the model was estimated from the HESA data as the sum frequency of students with English domicile within each HE institution. Thus, because the analysis was limited to England, this comprised 88 demand universities or colleges.

For each of the OAC Super Groups, the cumulative frequency of students travelling to attend HE was calculated for a series of arbitrarily selected 20 km distance bins. Using these graphs, exponential decay functions were calculated to best fit these empirical observations (see Fig. 1). The exponential functions fit the empirically recorded distance decay reasonably well for the majority of OAC Super Groups, with the exceptions of “City Living” and “Countryside”. These observed travel behaviours may relate to the distribution of these clusters nationally, relative to the location of HE institutions. This hypothesis is explored in Fig. 2 where Output Areas coded as Super Group “Countryside” are shaded in darker grey and the location of the HE institutions plotted as black points. This map shows that the majority of HE institutions are located in populous areas and therefore any student attending an HE institution from the “Countryside” Super Group will first have to travel outside of their local area. Thus, exponential decay is only seen from this distance onwards and therefore creates the travel patterns exhibited in Fig. 1c. Similarly, the “City Living” Super Group are primarily concentrated in urban areas and there appears to be a considerable likelihood for students to travel to universities located in other major metropolitan centres over and above a natural exponential decay that might be expected with distance alone (see Fig. 1b). Because students from this Super Group are generally already living within larger cities (e.g. London, Manchester, Birmingham), it is not unreasonable to expect that they may desire a university place in a location of a comparable size, but, not necessarily their home city. We indicate below how to handle these issues in the model.

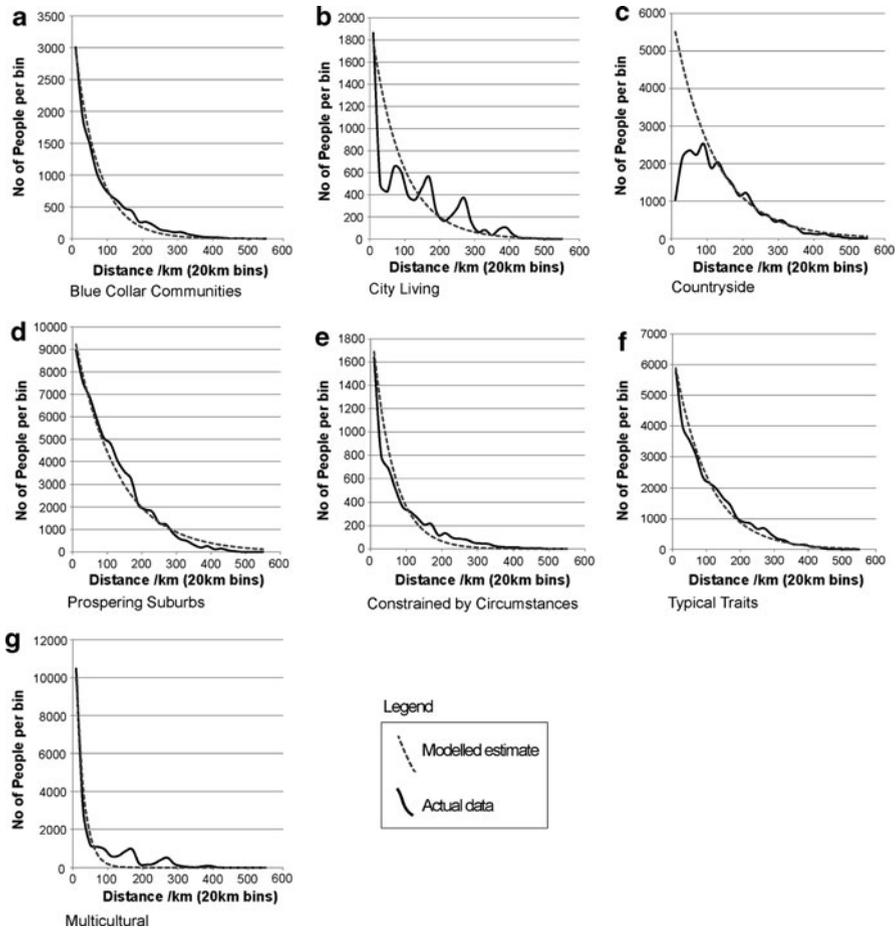


Fig. 1 Distance decay of English HE participation by OAC

3 A model of student flows from school to university

As we have seen, the study area consists of 150 origin zones, generically labelled *i* (local authorities) and 88 destination zones, labelled *j* (universities and colleges). The origin zones each represent the final-year school and further-education college population in that local authority and are divided into separate subpopulations based on geodemographics and attainment characteristics. The set of *i* are locations defined from the centroid of each local authority in England. The set of *j* are locations that are the centroids of the main campus of each non-specialised university in England. A partially constrained spatial interaction model, calculating flows between the origins and destinations, consists of singly and doubly constrained parts, depending on whether the destination is at capacity or not. For universities which are not at capacity, the singly constrained model is used. The choice of model is related to the observation that only a subset of universities at the end of the main application cycle

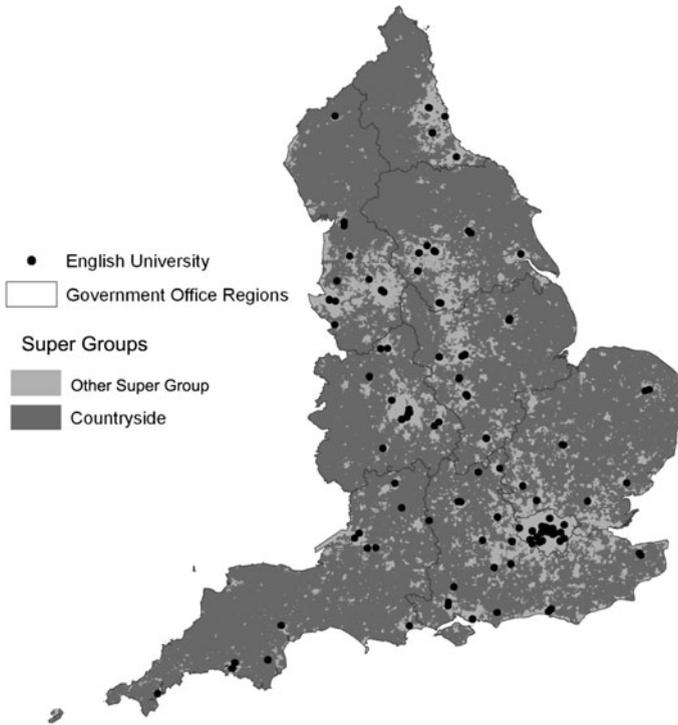


Fig. 2 Countryside Output Area Classification Super Group and university locations

are left with surplus places. These course vacancies are typically entered into a clearing process where potential students are matched to open places; however, these activities vary considerably between ‘recruiting’ and ‘selecting’ institutions (Singleton 2010). This research is being conducted during a period of very rapid HE policy change. The model presented below is a mixed model, partially constrained, with a facility to change the degree of constraint according to circumstances.

The OAC groups and the levels of attainment are labelled with superscripts f and g , respectively. The model can thus be written as

$$S_{ij}^{fg} = A_i^{fg} O_i^{fg} (R_j)^{\alpha} K_j^g \exp(-\beta_{1i}^f \beta_{2ij} c_{ij}) \quad \sum_i S_{ij} \leq K_j^g \quad (1)$$

$$S_{ij}^{fg} = A_i^{fg} B_j^g O_i^{fg} K_j^g \exp(-\beta_{1i}^f \beta_{2ij} c_{ij}) \quad \sum_i S_{ij} > K_j^g \quad (2)$$

$$A_i^{fg} = \sum_j (R_j)^{\alpha} K_j^g \exp(-\beta_{1i}^f \beta_{2ij} c_{ij}) \quad \sum_i S_{ij} \leq K_j^g \quad (3)$$

$$A_i^{fg} = \sum_j B_j^g O_i^{fg} K_j^g \exp(-\beta_{1i}^f \beta_{2ij} c_{ij}) \quad \sum_i S_{ij} > K_j^g \quad (4)$$

$$B_j^g = \sum_i A_i^{fg} O_i^{fg} \exp(-\beta_{1i}^f \beta_{2ij} c_{ij}) \quad \sum_i S_{ij} > K_j^g \quad (5)$$

where S_{ij}^{fg} = the number of students from local authority i in OAC group f , with attainment level g going to university j . O_i^{fg} = the number of students from local authority i in OAC group f with attainment level g going to university. R_j = the attractiveness of university j . K_j^g = the capacity of university K for admitting students of attainment level g . β_{1i}^{fg} = an element of the distance parameter relative to distance decay. β_{2ij} = an element of the distance parameter—a device for relative attainment to enhance requirement. c_{ij} = a measure of the ‘distance’ between i and j as a surrogate for impedance.

In the singly constrained model, there is no B_j coefficient. In the doubly constrained model, both A_i and B_j coefficients are present and depend upon each other, so the equations are solved iteratively. The parameter β_{1i}^f is calibrated with the actual flows observed on the decay functions of Fig. 1 with a binary attainment band g split at the national average A-Level points score of 730.

Each university also has a composite attractiveness factor $R_j^\alpha K_j$. R_j^α is an approximate measure based upon the 2010 Times Good University Guide rankings score. It is raised to the power of α , which is set to 0.5 of this experiment, and then multiplied by the university capacity K_j to make larger institutions more ‘attractive’ than smaller ones. This attractiveness factor is only used in the singly constrained part of the model. The capacity itself is used in the doubly constrained model. We are aware that concerns have been raised about the methodological robustness of the metrics created by university rankings (Turner 2005); however, they are also a resource commonly used by students when making institutional choices (Bowden 2000). And, of course, this metric could be changed in future models.

The c_{ij} values were taken as Euclidean distances; again as an approximation for demonstration purposes. Two adjustments were made to deal with the issues raised at the end of section 2. For origin subpopulations within the OAC Super Group (f) equal to ‘Countryside’, the c_{ij} was reduced by 50 km while ensuring a minimum distance of zero. This accounts for such origin subpopulations needing to travel at least this distance on average to get to even their nearest university (see Fig. 2), causing the initial increase in the participation with distance seen for this geodemographic in Fig. 1. The second issue, relating to the other poorly fitting Super Group in Fig. 1, ‘City Living’, was discovered to arise in major urban areas in England where this OAC Super Group is typically found. Adjustments were made as part of model calibration to c_{ij} terms where the i and j are within the same metropolitan area, particularly for newer universities. These calibrated distance costs are shown in Table 1. Unusually, these are negative, indicating a positive attraction of ‘nearness’.

A further ‘not at university’ destination capacity was created to account for those students who do not attend HE. For this destination, all c_{ij} were set to be 0, R_j set to a very low, but non-zero value, and K_j was set as a very large number. From an operational perspective, those students who are assigned to the ‘not at university’ destination capacity by the end of the model run can be profiled to examine origin characteristics more prevalent in non-participating students. Using this information, the ‘not at university’ group forms a benchmark upon which potential widening participation scenarios could be modelled and tested.

Table 1 Intra-zonal flow adjustments

Local authority <i>i</i>	University <i>j</i>	Adjusted distance c_{ij}/km
Bath and North East Somerset	Bath Spa University	-110
Birmingham	Birmingham City University	-60
City of Bristol	University of the West of England	-30
Leeds	Leeds Metropolitan University	-40
Liverpool	Liverpool Hope University	-120
Liverpool	Liverpool John Moores University	-100
Manchester	Manchester Metropolitan University	-20
Norfolk	Anglia Ruskin University	-100
Nottingham	Nottingham Trent University	-20
Sheffield	Sheffield Hallam University	-30
Southampton	Southampton Solent University	-180

Before the model was run, the balancing factors were initialised (β_j values to 1, A_i to 1). A second beta factor, β_{2ij} , acts as a step function and is used with an origin subpopulation's attainment band and a destination segmented population's minimum attainment requirement. If the former is lower than the latter, then β_{2ij} is set to a very large number for very rapid exponential decay in effect to set the value to zero, otherwise it is 1. Thus, $\beta_{2ij} = \begin{cases} \infty, & g < G \\ 1, & g \geq G \end{cases}$ and the overall beta value takes the form $\beta_{ij}^f = \beta_{1i}^f \beta_{2ij}$. This has the effect that for some institutions, capacity is not available for potential students with lower attainment (see Table 2) and is required to simulate approximate variable entry requirements within the HE sector.

4 Evaluating student flows from school to university

The set of results generated by the model are very large and it is impossible to describe them all. To illustrate, flows to a single university from all local authorities are considered, followed by the examination of flows from a local authority to all universities. These analyses aim to demonstrate the utility of the model and indeed highlight areas where further refinement is needed. The performance can be assessed by comparing the origin–destination flows that are predicted against those which were observed in reality. The aggregate (combined geodemographics and attainment) top 40 flows from all local authorities to Manchester University are shown in Table 3. Graphical representations of these aggregate flows are shown in Figs. 3 and 4.

These actual and predicted results show strong regional flows of students from Local Authorities in the North West region to study at the University of Manchester. The majority of these local flows are predicted reasonably well by the model with the exception of Trafford, which is broken down by OAC Super Group in the first row of Table 4. In particular, more students than expected travel from the “Multicultural” Super Group, which is assigned one of the sharpest beta decay

Table 2 Institutions specified within the model to have “higher” entry requirements

	Institution
	Imperial College London
	King’s College London
	Lancaster University
	Loughborough University
	Royal Holloway, University of London
	University College London
	University of Bath
	University of Birmingham
	University of Bristol
	University of Cambridge
	University of Durham
	University of East Anglia
	University of Exeter
	University of Leeds
	University of Leicester
	University of Liverpool
	University of Manchester
	University of Newcastle upon Tyne
	University of Nottingham
	University of Oxford
	University of Reading
The London School of Economics and Political Science would also be included in the table above, but is excluded from the model to remove an anticipated unusual student flow due to its restricted range of subjects on offer	University of Sheffield
	University of Southampton
	University of Sussex
	University of Warwick
	University of York

values in the model (see Fig. 1). This under prediction could relate to specific local influences such as the provision of a good transportation (e.g. Manchester Metrolink) system that links Trafford to the City of Manchester or, perhaps, specific widening participation activities that may have occurred within the local area. If the actual flows from within the “Multicultural” Super Group in Trafford are broken down by attainment, it is found that 18 are in the high band and 2 are in the low band. This is unsurprising, given that the University of Manchester has high entry requirements; however, it also suggests that there may be justification in segmenting the beta decay values further by attainment band in addition to OAC Super Group.

Other interesting results from this example include flows from local authorities in the South East of England including Surrey and Buckinghamshire, and the London Boroughs of Barnet and Harrow. All of these local authorities exhibit actual flows that are greater in magnitude than those predicted by the model. Again, these flows are disaggregated in Table 4. The under predictions exhibited in Surrey and

Table 3 Actual and predicted flows from local authorities to Manchester University

LEA	Predicted population	Actual population	Ratio
Lancashire	312	284	0.9
Cheshire	282	179	0.6
Manchester	194	175	0.9
Stockport	124	155	1.3
Trafford	79	128	1.6
Derbyshire	111	110	1.0
Leeds	87	110	1.3
Hertfordshire	77	106	1.4
Surrey	63	104	1.7
Cumbria	79	96	1.2
Staffordshire	123	92	0.7
Wigan	126	88	0.7
Tameside	52	87	1.7
Bolton	73	83	1.1
Kent	46	82	1.8
Oldham	106	81	0.8
Wirral	79	81	1.0
Buckinghamshire	44	78	1.8
Barnet	9	76	8.4
Kirklees	122	73	0.6
North Yorkshire	121	72	0.6
Rochdale	18	72	4.0
Birmingham	74	72	1.0
Bradford	62	69	1.1
Sheffield	70	69	1.0
Bury	188	67	0.4
Leicestershire	69	67	1.0
Nottinghamshire	71	67	0.9
Hampshire	50	59	1.2
Warwickshire	65	57	0.9
Oxfordshire	53	57	1.1
Northamptonshire	53	56	1.1
Essex	51	56	1.1
Worcestershire	64	54	0.8
Warrington	54	52	1.0
Harrow	8	49	6.1
Lincolnshire	83	48	0.6

Results in bold are expanded by OAC group in Table 4

Buckinghamshire mainly originate in the OAC Super Group “Prospering Suburbs”. Although geodemographic clusters are constructed with the aim of maximising similarity within and differences between geographical areas, there may be some degree of regional difference within a classification, and specifically so if some

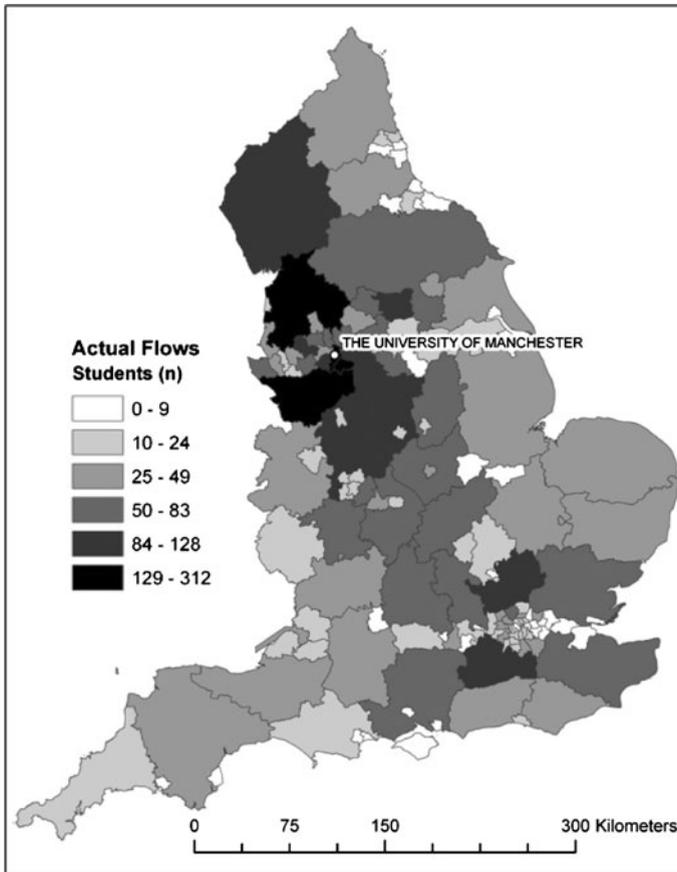


Fig. 3 Actual flows to Manchester University

dimension is not effectively accounted for within the input data used in the build process. Because OAC is created entirely from the 2001 Census, there is no direct measure of household income and as such, one might hypothesise that the under predictions in the model could be induced by regional differences in income. The “Prospering Suburbs” areas of Buckinghamshire and Surrey are perhaps more affluent than in other parts of the country and therefore may influence greater propensity for students to travel to institutions at further distances from their domicile. This is illustrated in Fig. 5 by examining the average distance that students from “Prospering Suburbs” areas travel to university by local authorities. It can be seen that these results partially relate to the isolation of the area; however, both Surrey and Buckinghamshire are areas where student travel is more prevalent. Also of note is the attractiveness of the North West Region for students from these affluent areas to attend more local institutions.

The model also under predicts the flow of students from “Prospering Suburbs” neighbourhoods in the London Boroughs of Harrow and Barnet. However, in these

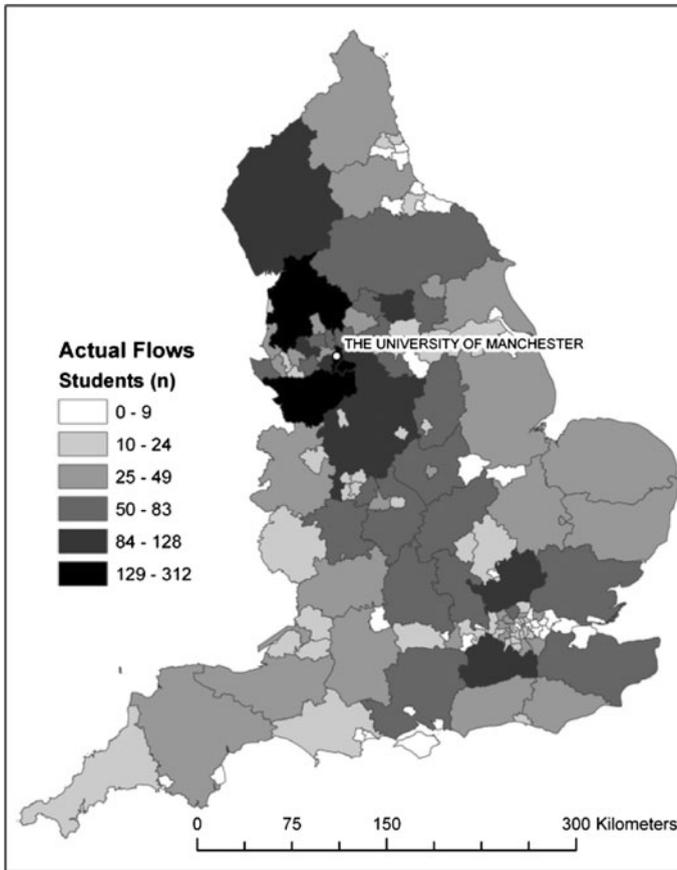


Fig. 4 Predicted flows to Manchester University

two local authorities, there is also an additional under prediction of the flow of students from “Multicultural” neighbourhoods, which typically have a very low propensity to travel. There may be specific draws within the local area of Manchester which account for these flows. For example, like Barnet and Harrow, Manchester has a large Asian population and there may be family ties that enhance links between the two sets of places and, combined with good transport links, increase the flow of students from the “Multicultural” Super Group. A possible enhancement to the model could be achieved by implementing a different geodemographic classification that better accounts for patterns of affluence and additionally has a more disaggregated clustering of multicultural neighbourhoods that identifies those students who are more willing to travel.

The second example illustrates how the model can produce predictions of flows to all universities from a single local authority. Norfolk was chosen as this offered large aggregate flows and represented a good mix of both rural and urban areas. The results for this local authority are shown in Table 5 for the highest 30 actual flows

Table 4 Actual and predicted flows from local authorities to Manchester University by OAC Super Group

Local authorities	Predicted							Actual						
	1 (blue collar)	2 (city living)	3 (countryside)	4 (prospering suburbs)	5 (constrained by circumstances)	6 (typical traits)	7 (multicultural)	1 (blue collar)	2 (city living)	3 (countryside)	4 (prospering suburbs)	5 (constrained by circumstances)	6 (typical traits)	7 (multicultural)
Trafford	4	5	2	33	6	24	6	6	7	0	59	5	31	20
Surrey	0	6	6	41	0	10	0	0	1	14	73	1	15	0
Buckinghamshire	1	3	8	26	0	6	0	1	3	11	45	0	13	5
Barnet	0	3	0	4	0	1	0	0	17	0	34	0	9	16
Harrow	0	5	0	2	0	1	0	0	4	0	20	0	8	17

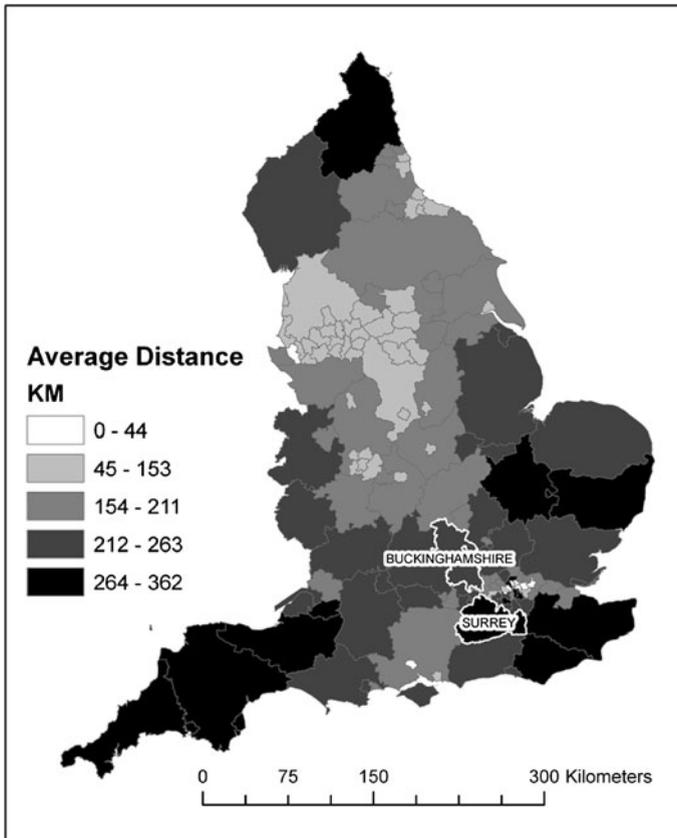


Fig. 5 Average distance students from Prospering Suburbs travel to university

and for all flows in Figs. 6 and 7. Overall, the model performs reasonably well; however, an interesting result was for the University of Kent, which received fewer students from Norfolk than one might expect, given close geographical proximity. A possible explanation could be the difficulty in travelling to this destination on public transport (e.g. via London) which might reduce the attractiveness of the destination to students. A similar hypothesis could be applied to the over prediction of student flows for the more disparately located University of Manchester and the University of Leeds. Both of these locations may be less attractive to students in reality, given long public transport travel times requiring multiple changes during the route. These effects could be explored in future iterations of the model by further refinement of the impedance term based on travel times or connectivity.

A further interesting result is that the model over predicts the number of students attending the University of Cambridge, which is in close proximity to Norfolk and has high entry grade requirements. In this case, the supply of students with the levels of attainment necessary to obtain places at Cambridge might be limited, and as such account for the over prediction. To some extent, the University of Cambridge and

Table 5 Flows of students from Norfolk to university

University	Predicted flows	Actual flows	Ratio
Anglia Ruskin University	358	376	1.1
University of East Anglia	266	354	1.3
University of Lincoln	76	115	1.5
Sheffield Hallam University	60	84	1.4
University of Sheffield	58	79	1.4
Nottingham Trent University	77	78	1.0
University of Hertfordshire	63	70	1.1
University of Nottingham	74	70	0.9
University of Leeds	70	61	0.9
De Montfort University	82	53	0.6
University of Kent	63	44	0.7
Loughborough University	49	41	0.8
University of Hull	70	39	0.6
University of Essex	57	38	0.7
University of Reading	18	37	2.1
University of York	29	37	1.3
University of Warwick	25	37	1.5
University of Cambridge	62	36	0.6
University of Durham	29	35	1.2
Oxford Brookes University	28	34	1.2
University of Birmingham	41	31	0.8
University of Newcastle upon Tyne	25	30	1.2
Leeds Metropolitan University	60	30	0.5
University of Brighton	30	29	1.0
University of Manchester	45	29	0.6
University of Portsmouth	27	25	0.9
University College London	21	24	1.1

other institutions of similar stature are so dissimilar to other universities, both in terms of their admission procedures and their entrance requirements, that they might be better represented as a disaggregated destination group within the model.

5 Concluding comments

The tests of this model represent a useful beginning. However, there is obviously much scope for future work. It would be possible to disaggregate the model further and, for example, to incorporate information on the subjects that are on offer at institutions. This might be especially important when a course at a specific institution is perceived as having a strong 'reputation'. For example, if a course in

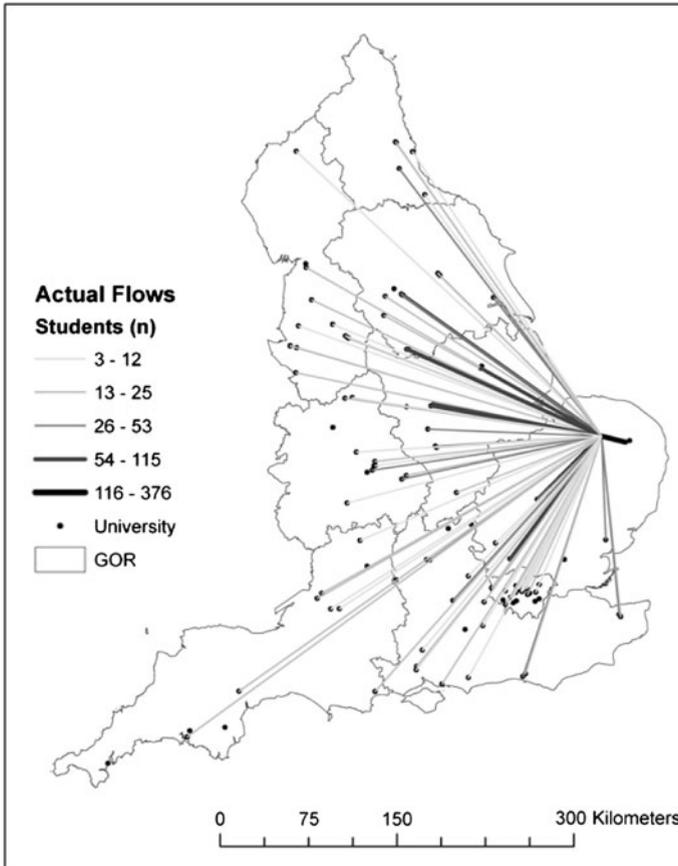


Fig. 6 Actual flows of students from Norfolk to universities

Sociology at the University of York was especially well regarded within the applicant population, this might be represented in a model by stronger attractiveness parameters than the same course more generally. By having a model disaggregated at subject level, this might better account for geographical variations in course availability. For example, courses with a marine focus are usually found in institutions that are accessible to the coast.

Some institutions with specialist focus were excluded from the model as they typically offered a very limited subject range and had skewed entrance profiles. For example, in an Arts university, the entrance criteria are typically based on a portfolio of creative achievement as opposed to attainment entry criteria. In future work where HE subject differences are better accounted for, these institutions could be reintroduced.

Those data made available for this research related to England only and in future research, it would be useful to integrate Welsh and Scottish data should these be made available. For local authorities and institutions in border regions (e.g.

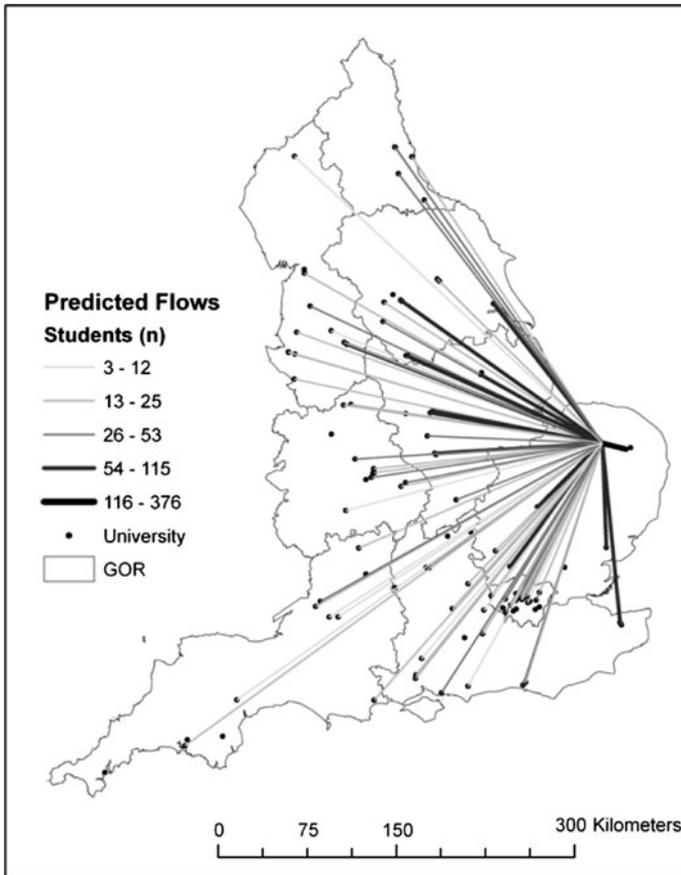


Fig. 7 Predicted flows of students from Norfolk to universities

Northumbria), these data may be particularly useful. However, these data would also add an extra layer of complexity to the model due to a variety of administrative differences between the countries of Great Britain. For example, Scottish students are given benefits of studying HE within Scotland through fee waiver, and Welsh students studying within Wales are required to pay a lower fee.

The geodemographic classification chosen for this research was the Office for National Statistics Output Area Classification; however, any one of a number of different classifications could have been used. In future work, it would be useful to examine whether the model is more effective when utilised with a commercial classification (e.g. Mosaic from Experian) or a bespoke model designed specifically for HE (see Singleton and Longley 2009).

The zonal geography used for this model was chosen with the specific aim of reducing the frequency of zero values in the origin–destination matrix, given that these can introduce problems when running the spatial interaction model. In future iterations, we would aim to examine some alternate disaggregated zonal geography

that may mitigate potential misspecification issues related to the centroid of large zones not being a good fit for the actual local distribution of the student population.

Finally, more experimentation with the precise form of model—for example in relation to policy changes affecting the financial capacity of institutions to accept students—and more systematic calibration of parameters is also needed. The attractiveness exponent was set at 0.5, for example, and the OAC beta values are visually rather than programmatically calibrated. Other impedance functions could be evaluated and tested through more detailed statistical evaluation (see Black and Salter 1975).

However, the research presented here represents some steps towards a model that can be used by HE decision makers to test potential recruitment, selection and widening participation scenarios. The model demonstrates how a geodemographic framework can be used to simplify complex socio-demographic characteristics of origin populations within a spatial interaction model. Through integration of administrative data with a geodemographic classification, aggregate differences in HE participation behaviour between geodemographic clusters were used to inform the calibration of parameter estimates within the model. This approach has demonstrated two methodological innovations; first, it has infused behavioural insights into a spatial interaction model and secondly, it has demonstrated how geodemographic classifications can be incorporated into a traditional modelling framework that is more sensitive to local spatial contexts.

References

- Ashby DI, Longley PA (2005) Geocomputation, geodemographics and resource allocation for local policing. *Trans GIS* 9(1):53–72
- Aveyard P, Manaseki S, Chambers J (2002) The relationship between mean birth weight and poverty using the Townsend deprivation score and the super profile classification system. *Public Health* 116(6):308–314
- Batey P, Brown PJB, Corver M (1999) Participation in higher education: a geodemographic perspective on the potential for further expansion in student numbers. *J Geogr Syst* 1(3):277–303
- Birkin M (1995) Customer targeting, geodemographic and lifestyles approaches. In: Longley PA, Clarke GP (eds) *GIS for business and service planning*. Geoinformation, Cambridge, pp 104–149
- Birkin M, Clarke G (1998) GIS, geodemographics, and spatial modeling in the UK financial service industry. *J Hous Res* 9(1):87–111
- Birkin M, Clarke G, Clarke M (2002) *Retail geography and intelligent network planning*. Wiley, Chichester
- Birkin M, Clarke G, Clarke M, Culf R (2003) Using spatial models to solve difficult retail location problems. In: Stillwell J, Clarke G (eds) *Applied GIS and spatial analysis*. Wiley, Chichester, pp 35–54
- Black JA, Salter RJ (1975) A statistical evaluation of the accuracy of a family of gravity models. *Proc Inst Civ Eng* 59:1–26
- Bowden R (2000) Fantasy higher education: university and college league tables. *Qual High Educ* 6(1):41–60
- Bureau of Public Roads (1965) *Calibrating and testing a gravity model for any size urban area*. US Dept of Commerce—Bureau of Public Roads, Washington DC
- Comptroller Auditor General (2008) *Widening participation in higher education*. National Audit Office, London

- Erlander S, Stewart NF (1990) The gravity model in transportation analysis: theory and extensions. VSP BV, The Netherlands
- Feng Z, Flowerdew R (1998) Fuzzy geodemographics: a contribution from fuzzy clustering methods. In: Carver S (ed) *Innovations in GIS 5*. Taylor & Francis, Oxford, pp 119–127
- Fotheringham AS (1983) A new set of spatial interaction models: the theory of competing destinations. *Environ Plan A* 15(1):15–36
- Fotheringham AS, Brunsdon C, Charlton M (2000) *Quantitative geography: perspectives on spatial data analysis*. Sage, London
- Han Q, Timmermans H (2006) Towards models of strategic spatial choice behaviour: theory and application issues. *GeoJournal* 67(3):195–206
- Harris R, Sleight P, Webber R (2005) *Geodemographics, GIS and neighbourhood targeting*. Wiley, Chichester
- Harris R, Johnston R, Burgess S (2007) Neighborhoods, ethnicity and school choice: developing a statistical framework for geodemographic analysis. *Popul Res Policy Rev* 26(5–6):553–579
- Longley PA (2005) Geographical information systems: a renaissance of geodemographics for public service delivery. *Prog Hum Geogr* 29(1):57–63
- Singleton AD (2010) *Educational opportunity: the geography of access to higher education*. Ashgate, Farnham
- Singleton AD, Longley PA (2009) Creating open source geodemographics—refining a national classification of census output areas for applications in higher education. *Prog Hum Geogr* 88(3):643–666
- Sleight P (1997) *Targeting customers: how to use geodemographic and lifestyle data in your business*. NTC Publications, Henley-on-Thames
- Stillwell JCH (1978) Interzonal migration: some historical tests of spatial interaction models. *Environ Plan A* 10:1187–1200
- Tobler WR (1970) A computer movie simulating urban growth in the Detroit region. *Econ Geogr* 46(2):234–240
- Tonks D, Farr M (2003) Widening access and participation in UK higher education. *Int J Educ Manage* 17(1):26–36
- Turner D (2005) Benchmarking in universities: league tables revisited. *Oxford Rev Educ* 31(3):353–371
- Twigg L, Moon G, Jones K (2000) Predicting small-area health-related behaviour: a comparison of smoking and drinking indicators. *Soc Sci Med* 50(7–8):1109–1120
- Vickers D, Rees P (2007) Creating the UK national statistics 2001 output area classification. *J R Stat Soc Ser A Stat Soc* 170(2):379–403
- Voas D, Williamson P (2001) The diversity of diversity: a critique of geodemographic classification. *Area* 33(1):63–76
- Wilson AG (1970) *Entropy in urban and regional modelling*. Pion, London
- Wilson AG (1998) Land-use/transport interaction models: past and future. *J Transp Econ Policy* 32(1):3–26
- Wilson AG (2000) The widening access debate: student flows to universities and associated performance indicators. *Environ Plan A* 32(11):2019–2031