Investigating catchment area anomalies for a north England store

David Lloyd, Jason Dykes

City University London, giCentre, School of Informatics,
Department of Information Science, Northampton Square, London EC1V OHB.
Tel. +44 20 7040 0212, Fax: +44 20 7040 8584
d.lloyd@soi.city.ac.uk, www.soi.city.ac.uk/~at775

KEYWORDS: retail catchment; gravity model; demographics; newspaper circulation; census

1. Introduction

Planning regulations constrain the development of new retail sites in the UK, and a consequent area of focus is evaluating expected performance against actual sales for existing individual stores (Birkin et al, 2002). A UK national retailer noticed significant differences between the predicted and recorded customer catchment areas of a large out-of-town store sited near Pudsey, West Yorkshire. Unable to account satisfactorily for this spatial anomaly the retailer provided a map of the store's catchment area (CA). An anomalous area (AA) was constructed for locations outside the CA but within a travel time of 45 minutes, corresponding to the longest travel times to the store from within the CA (Figure 1). Possible causes for the anomaly were investigated including the impact of changing administrative boundaries, the route of the M62 motorway, the practicalities of access to the site by road and public transport, road signage issues and the impact of the nearest competitor store. Here, two approaches are applied to evaluate the impact of self-contained communities and travel to work flows. In addition, sales data for the retailer are modelled quantitatively to see whether spatial effects can be identified that would indicate the source of the anomalous catchment area of the Pudsey store.

Figure 1. Location of Pudsey store in relation to the Catchment (CA) and Anomalous Area (AA).
Source: Borders: Crown copyright.

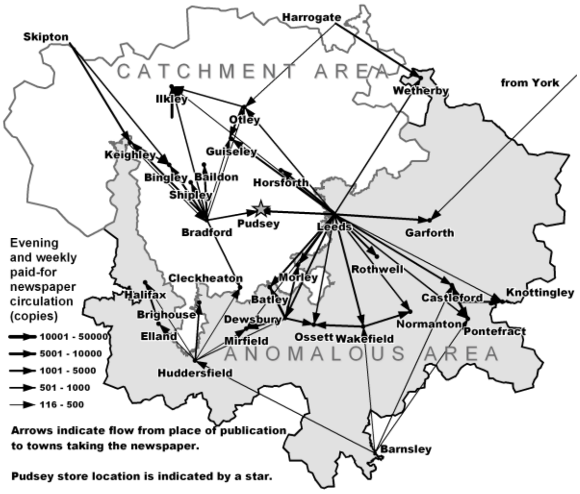

Figure 2. Paid-for evening and weekly newspaper circulation flows around the CA and AA border.
Sources: JICREG; Borders: Crown copyright.

2. Newspaper circulations and communities

The geography of local newspaper circulation has been used to study community boundaries (Park and Newcombe, 1933). JICREG (2006) gives circulation figures for regional/local newspapers titles by town. Of 21 "paid for" local and regional newspapers in West Yorkshire, fourteen have circulations of more than 300 copies that cover portions of the AA. A plethora of titles cover the areas south of Leeds and Bradford, consistent with the AA having a large number of separate communities. Data relating to these publications were mapped to reveal the magnitude and direction of newspaper flows and explore possible community boundaries. These town-based circulation patterns might be indicative of self-contained retailing areas.

Figure 2 shows that all around the borders of the CA, paid-for newspaper circulation suggests a community focus away from the CA, linking towns in the AA. In particular, towns to the south of Bradford and Pudsey do not take newspapers from towns inside the CA, but are generally focused on neighbouring communities to the east and west. Batley and Cleckheaton, only five kilometres apart either side of the CA/AA border, have very different newspaper consumption patterns. If retail patterns follow "community of interest" boundaries then there is a case for hypothesising that this contributes part of the explanation for the existence and line of the CA/AA boundary.

3. Travel to work flows

Commuting patterns between parts of the CA and AA past the store were analysed and compared with patterns relating all AA commuters to see whether the CA was being avoided in some way. The flows of road users past the Pudsey store were derived from the "travel to work and study" data of the 2001 Census. The CA and AA were segmented into areas north, south, east and west of the Pudsey such that road journeys from west to east (and vice versa) and north to south (and vice versa) would take plausible major road routes passing within 1 km of the Pudsey store. The four-way segmentation is appropriate in the context of the particular geography of major roads near the Pudsey location - the Bradford-Leeds A647, and two arms of the Leeds ring road, A6110 and A6120, are the only nearby major roads, and converge together with 1km of the store (Figure 3).

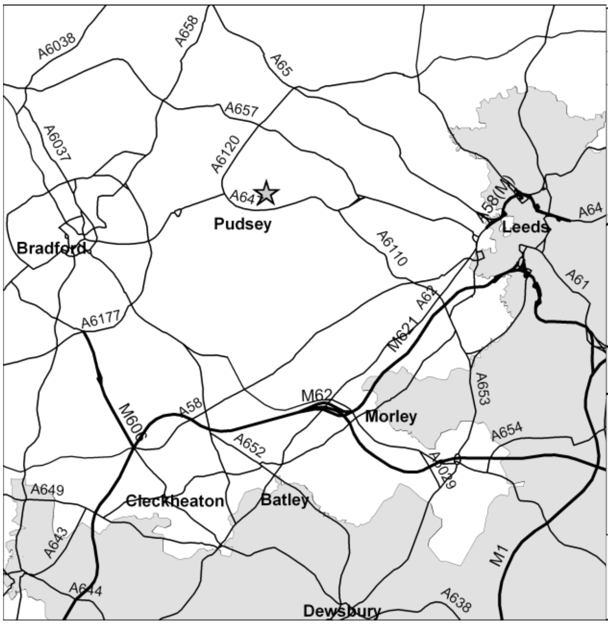

Figure 3. Access routes to the Pudsey store (shown as a grey star) showing CA (white background) and AA (grey background). *Source: Strategi®: Crown copyright.*

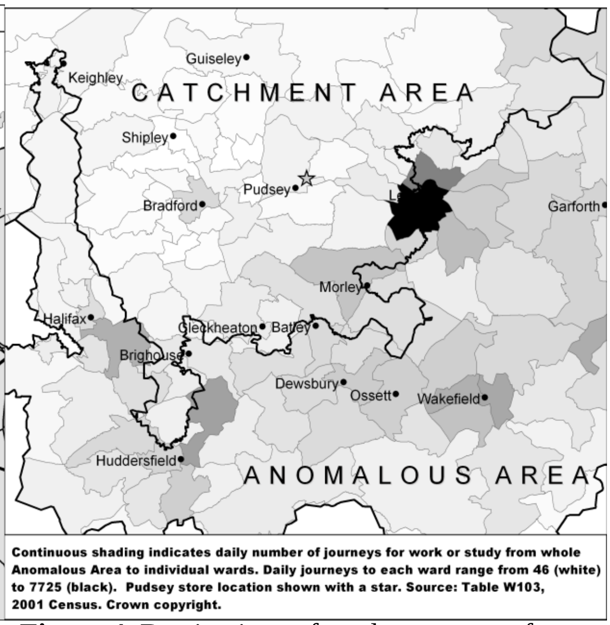

Figure 4. Destinations of road commuters from the whole Anomalous Area, showing the relative pattern of commuting to AA and CA destinations around the CA/AA boundary.
Sources: W103, 2001 Census and borders: Crown copyright.

Travel to work and study direction	% of people	% of economically active people
West to East	0.9	1.5
East to West	0.5	0.8
North to South	10.5	15.3
South to North	3.1	4.6

Table 1. Analysis of travel to work and study data from 2001 Census. West to East, for example, indicates commuters travelling from wards west of the Pudsey store to wards east of the store that would pass within 1km of the store. *Source: Table W103, 2001 Census, Crown copyright*.

The numbers of people travelling to work from north to south are an order of magnitude higher than the east-west and west-east flows (Table 1). 15% of economically active people travel by road from the defined area in the north, pass close to the Pudsey store and on to destinations in the south. This represents a large “passing trade” market, domiciled in the CA. In contrast, the corresponding south to north figure is less than 5% - a clearly spatially differentiated result.

The daily commuting destinations of people living in the AA are shown in Figure 4. AA commuting destinations generally fall off in the area of the CA/AA boundary, and here its line can be delineated clearly, suggesting the existence of the boundary in terms of an external variable independent of retailing data. This may be a further indication of an AA community focus away from the CA as suggested by the results of the newspaper circulation analysis.

4. Gravity models

To supplement these approaches to the investigation, a series of models of retail sales were developed using, progressively, data on store attractiveness, time of travel, and demographics. The models considered both the competition from other company outlets, as well as the effect of competition from other retailers.

The Huff gravity model (Huff, 1963) has been widely used to model shopping patterns of customers given multiple destination choices (Birkin et al, 2002). It derives the probability (P_{ij}) of a consumer from a given area, i, travelling to a given shopping facility, j from:

$$P_{ij} = S_j / T_{ij}^{\lambda} / \sum_{j=1}^{n} S_j / T_{ij}^{\lambda} \qquad (1)$$

where: S_j is size of shopping facility; T_{ij} the travel time from area i to shopping facility j; λ is an empirically derived parameter.

This model was fitted using company-provided square footage of the 31 company stores in the north of England (S_j) for a range of λ.

In order to differentiate between the effect of competition from other outlets of the company and that from other retailers, an alternative formulation of the model used the aggregated competitive space around the Pudsey store with retail floorspace data (ODPM, 2003) as the model denominator.

A range of gravity models fitted aggregated company store card sales data with limited success. When demographic data were incorporated in the Huff model as a scaling factor in an attempt to improve the model, a marked deterioration was noted. Both Mosaic (Experian, 2005) and 2001 Census groups (ONS, 2005) were used. The fact that the additional demographic information not only failed to improve the gravity model, but worsened the fit, implies that the demographic information is duplicating factors already included within the model, requiring another approach.

5. Demographic models

The effects of using the gravity model with demographic data suggested modelling the sales data in a multiple regression with the demographic data alone. These models gave extremely good fits (r^2 around 0.8) for both Mosaic and Census group versions. To extract any spatial component in the residuals, geographically weighted regression (Fotheringham et al., 2002) was employed. A small increase in fit (to r^2 around 0.85) was obtained, but the small spatial residuals resulting were not visually associated with any spatial feature(s) in the CA or AA. Subsequently a t-test was used to compare the mean incidences of the different demographics in AA and CA.

The Mosaic t-test yielded significant differences between the CA and AA in the mean incidences of demographic types A01, A04, B07, B08, G30, K45 and K47 (positive for sales), and C12 and D13 (negative, but unimportant, for sales). A01 types are particularly important for sales and a map of modelled A01 sales delineates the CA/AA boundary similarly to the map of AA commuters in,Figure 2.

There is a band stretching from the CA/AA boundary into the AA where the numbers of the sales-positive Mosaic types identified by the t-test are very low. This "banded" area corresponds to households with higher deprivation as indicated by ODPM Index of Deprivation data and are of low sales potential. Further out, towards the outer AA boundary, the numbers of A01 and other important demographic types increase, but here the time of travel to Pudsey approaches 40 minutes.

Similarly, the Census group t-test yielded significant differences between the CA and AA for a number of demographics: "rooms/household", "higher education", "2+ cars/household", "% people aged 16-74 in employment working in financial intermediation" (positive for sales and numbers significantly higher in CA than AA), "% people aged 16-74 in employment working in mining, quarrying or construction" and "non-pensioner one-person households".

Correlation coefficients indicated the major difference in sales between the CA and AA were due to a cluster of demographic types sharing Mosaic and Census variables in common - A01, A04, B07, B08, K45, and K47 Mosaic types owning 2+ cars, well-educated, and with a large number of rooms per household.

6. Conclusions and Significance

The investigation indicates that the anomalous catchment problem has a number of components:

- Communities in the AA may be more distinct from communities in the CA than their geographic locations might suggest - as evidenced by the newspaper circulation data (Figure 2)
- Residents of the AA do not appear to commute to the CA in large numbers on a daily basis - as evidenced by the 2001 Census travel to work data (Figure 4)
- Those living within the AA do not contribute to the passing trade of the Pudsey store in significant numbers (Table 1).
- The CA has an identifiable element of high sales corresponding to particular demographic types.
- Far fewer numbers of these types live in the AA and those that do are near the outer AA boundary with unattractive journey times to the store.
- The predominant demographic types living in the AA just outside the CA are not significant spenders at the retailer's stores.

Furthermore, the geodemographic datasets used here (both Mosaic and Census groups) contain intrinsic spatial information that yielded better fits to sales data in regression models than were achieved using traditional gravity models.

Methodologically, the use of newspaper circulation data to explore the boundaries of communities of interest appears to be underexploited and a fruitful avenue for further exploration. Additionally, the Census groups dataset is a freely available demographic source which has been used here successfully in a comparable way to a commercial alternative (Mosaic), indeed the Census groups have a more transparent meaning (e.g. "2+ cars") than Mosaic descriptors (e.g. "A01 – global connections").

This study has presented an investigation whose findings suggest differences in the AA and CA populations that might account for the Pudsey store's anomalous catchment area. The analysis has used a diverse range of datasets in innovative ways with potential for application elsewhere, and draws attention to the degree of spatiality inherent in UK demographic data.

References

Birkin, M., Clarke, G., Clarke, M. *Retail Geography and Intelligent Network Planning* (2002) (Chichester: Wiley).

Experian Mosaic (2006) http://www.business-strategies.co.uk/Content.asp?ArticleID=566

Fotheringham, A.S., Brunsdon, C., Charlton, M. (2002) *Geographically Weighted Regression*, (Chichester: Wiley).

Huff, D.L. (1963) A Programmed Solution for Approximating an Optimum Retail Location, *Land Economics*, 42, pp.293-303.

Joint Industry Committee for REGional press research Readership data (2006) http://jiab.jicreg.co.uk

Office of National Statistics (2004) Origin-Destination Statistics: Output Areas, Table W103

Office of National Statistics (2005) UK Census based classification of output areas http://www.statistics.gov.uk/census2001/cn_139.asp

Office of the Deputy Prime Minister (2003) Commercial and industrial floorspace and rateable value statistics 2003 http://www.odpm.gov.uk/stellent/groups/odpm_planning/documents/page/odpm_plan_025922.hcsp

Office of the Deputy Prime Minister (2004) Indices of Deprivation, http://www.odpm.gov.uk/index.asp?id=1128445#P200_14232

Park, R.E. and Newcomb, C. (1933) Newspaper Circulation and Metropolitan Regions *The Metropolitan Community*. Chicago, Russell & Russell, quoted in Rosenberg, M. Metropolitan Newspaper Circulation, *About Geography*, http://geography.about.com/library/weekly/aa082498.htm

Biography

David Lloyd completed his Masters in Geographic Information (MGI) at City University London. In 2005 he began PhD studies at City on the usability of geovisualization applications for Leicestershire County Council supported by an EPSRC iCASE award. This paper represents part of his M.Sc. dissertation.

Jason Dykes is a Senior Lecturer in Geographic Information at City University London with interests in geovisualization. He has developed a number of software applications and published research papers on dynamic cartography. A member of the ICA Commission on Visualization and Virtual Environments, Jason is co-editor of 'Exploring GeoVisualization' (Dykes, MacEachren, Kraak, 2005).

Acknowledgments

2001 Census data and Output Area Boundaries. Crown copyright. Crown copyright material is reproduced with the permission of the Controller of HMSO. HMSO click-use licence C02W0003601.

Strategi®. Crown Copyright/database right 2006. An Ordnance Survey/EDINA supplied service.

Retail sales, store size and other company data provided by the Pudsey store owners under confidentiality restrictions that include anonymity.

Thanks to the GISRUK reviewers of this paper for their helpful comments and suggestions.