Defining Retail Conurbations: Using a Rules-based Algorithm

Matthew D. Pratt[1], Jim A. Wright[2], Samantha Cockings[2], Iain Sterland[3]

[1]Department of Geography, University College London, Gower Street, London, WC1E 6BT
matthew.pratt.12@ucl.ac.uk
[2]Geography and Environment, University of Southampton, University Road, Southampton, SO17 1BJ
J.A.Wright@soton.ac.uk; S.Cockings@soton.ac.uk
[3]Location Planning, Boots UK, 1 Thane Road West, Nottingham, NG2 3AA
iain.sterland@boots.co.uk
KEYWORDS: Retail Conurbations, Self-Containment, TTWA Algorithm, Rules-based Method, Flow Matrix

1. Introduction

Retail conurbations are market areas which act as discrete entities with high-intra market movement. They provide valuable information for location decision-making processes, such as evaluating overall store performance, opening/closure of stores and competitor presence. However many retailers delineate them using simplistic gravitational modelling such as a 'zones-of-dominance' approach (Huff 1964; Nakanishi & Cooper 1974; Drezner 2009), often resulting in inaccurate conurbations and gaps or overlaps between conurbations. One issue has been that many retailers are limited by their data and rely on the marketing analyst CACI who use bank transaction data to produce conurbations using a relatively basic approach (CACI 2012). In recent years however, loyalty card use has been increasing, leading to a wealth of 'big data' (Donnelly et al 2012). Retailers are starting to employ more advanced methods using larger datasets (Reynolds and Woods 2010). This paper proposes the application of a rules-based zonation method known as the Travel-To-Work-Areas (hereafter TTWA) algorithm, previously constricted in use to mapping commuting flows for indicating job ratios and unemployment (Coombes and Bond 2007). Loyalty card data from Boots UK are used to form an origin-destination flow matrix aggregated to Super-Output-Areas (SOAs) to delineate Boots conurbations across Newcastle, Middlesbrough and the surrounding area. Aggregation to SOA level ensures that individual customer behaviour is anonymised, and the data is re-scaled to preserve commercial confidentiality . Flows represent the total transactions of customers in Boots stores over the course of a year. The aim is to produce non-overlapping conurbations with full scene coverage (i.e. no data gaps) and more accurate portrayal of conurbations acting as discrete shopping entities with high intra-boundary retail journeys.

2. The Travel-To-Work-Areas Algorithm

Rules-based methods use a sophisticated set of rules in a multi-step procedure to determine zone (or in this case conurbation) design. Often, these are contained within a theoretical model, guiding decisions as to how and when rules are applied (Florez-Revuelta, Casado-Diaz et al. 2008). The TTWA algorithm is one of the most widely accepted zone design rules-based algorithms. Its origins stem back to the early 1970's with Smart (1974) who defined the criteria and algorithm for analysing commuter flows for approximately 2,000 building-block areas. This was later computerised by Openshaw (Coombes and Openshaw 1982) and has been modified through various reviews such as in 2001 and 2007 to adjust for changing commuter dynamics and advances in computational capabilities (Coombes 2010). The algorithm produces defined, non-overlapping boundaries meeting a set minimum threshold for self-containment of flows ideally being internally active with high intra-market movement; Goodman refers to this as 'external perfection' (Goodman 1970). High self-containment would require low inward and outward flows from defined zones in relation to the total flows in that zone. The process of the algorithm is shown below:

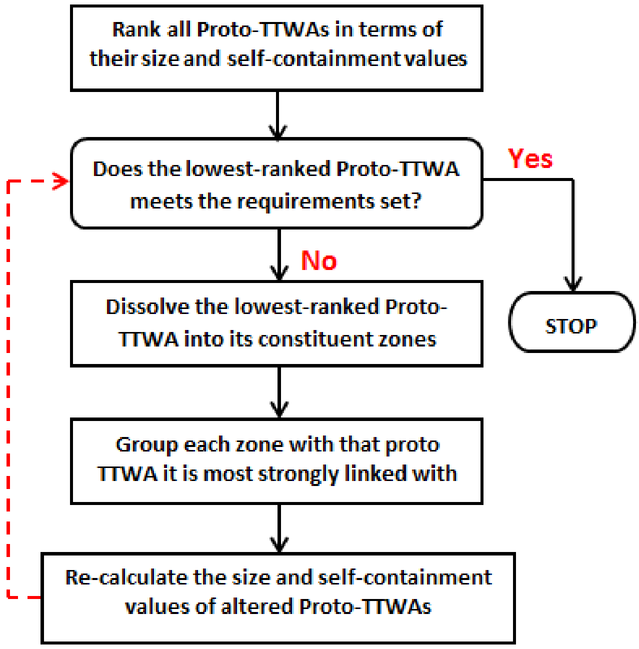


Figure 1: The Process for Delineating Travel-To-Work-Areas (Coombes 2010)

The TTWA algorithm evolved in response to key principles for defining local labour markets as outlined by Eurostat (Table 1). Thus, its objective is to define as many separate zones as possible while meeting the constraints outlined in Table 1. As a result, the TTWA algorithm was chosen as the accepted practice for delineating local labour markets in Europe. It has continued to develop and has been applied to 1981, 1991 and 2001 UK censuses, European countries such as Italy (Sforzi, Openshaw et al. 1997) and parts of Spain (Casado-Diaz 1996) and other Non-European countries such as New Zealand (Papps and Newell 2002). Frey and Speare (1995) considered it to be more advanced than any other alternative sub-regional statistical area definition method. To date, the TTWA algorithm has been applied to commuting flows, but no such rules-based approach has been used to define retail conurbations.

Principle	Practice
Objectives	
Purpose	To be statistically-defined areas appropriate for policy
Relevance	Each area to be an identifiable labour market
Constraints	
Partition	Every building block to be allocated to only one area
Contiguity	Each area to be a single contiguous territory
Criteria (*in descending priority*)	
Autonomy	Self-containment of flows to be maximised
Homogeneity	Areas' size range to be minimised (e.g. within fixed limits)
Coherence	Boundaries to be reasonably recognisable
Conformity	Alignment with administrative boundaries is preferable

Table 1. Principles for Local-Labour-Market-Area Definitions (EuroStat 1992)

3. Methodology

The research project aimed to implement the TTWA algorithm using VBA via an origin-destination flow matrix within MS Access, chosen for its ease of integration into Boots' planning department. This flow matrix represents customer transactions in Boots stores, whereby origins signify customer home addresses, destinations are Boots stores and flows are the amount spent for the year. This has no impact on the delineation of conurbations. The information is derived from loyalty card information, with home addresses and stores aggregated to Lower-Super-Output-Areas (LSOAs). Retail conurbations were delineated for Newcastle, Middlesbrough and the surrounding area, although eventually the approach will be extended from this pilot area to the entire UK. Boots recommended this pilot area because the rural nature of the surrounding area leads to fairly high self-containment of retail flows and due to accuracy concerns regarding current Boots conurbations in the region. The study area contains 150 Boots stores and 1748 LSOAs.

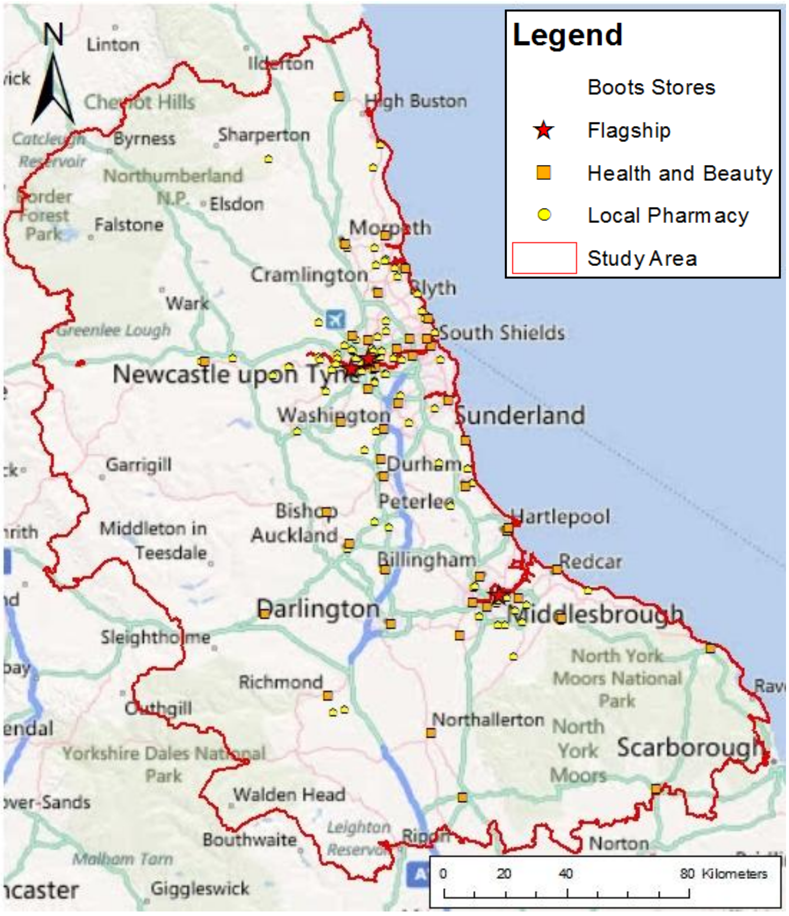


Figure 2. Greater Newcastle Study Area

For the purposes of the research presented here, two thresholds are set for defining conurbations: self-containment and internal flow size. The threshold on minimum internal flow size prevents the algorithm creating very small regions which could be sub-optimal. Self-containment however is the main driving force behind the algorithm. Self-containment is determined by calculating the internal, supply and demand flows. Supply flows are the total flows *to* other Proto-TTWAs (i.e. 'prototype conurbations'), while demand flows are the total flows *from other* Proto-TTWAs. Below is a simplified example for Proto-TTWAs A, B and C.

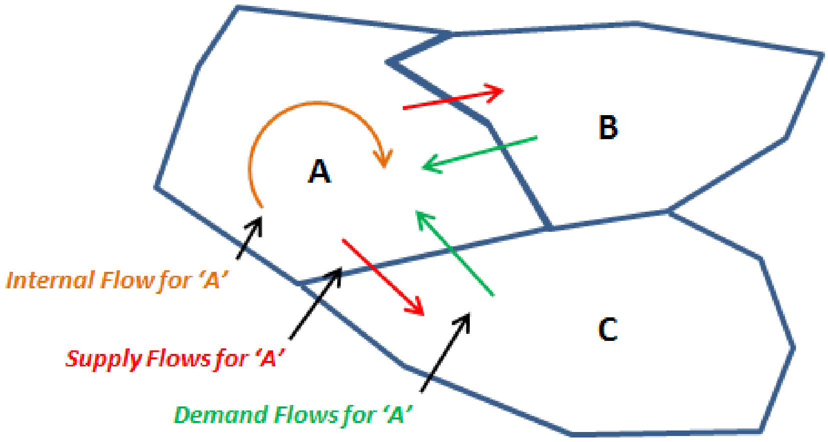


Figure 3. Internal, Supply and Demand Flows in Calculating Self-Containment Values

The self-containment value is determined by whichever is the lowest value for that Proto-TTWA for either the supply or demand containment. Supply and demand containment is calculated through the following equations; a value of 0.01 is added to avoid division by zero:

Supply Containment** = **Internal Flows** / (**Total Supply Flows** + **0.01)

Equation 1. Calculating Supply Containment

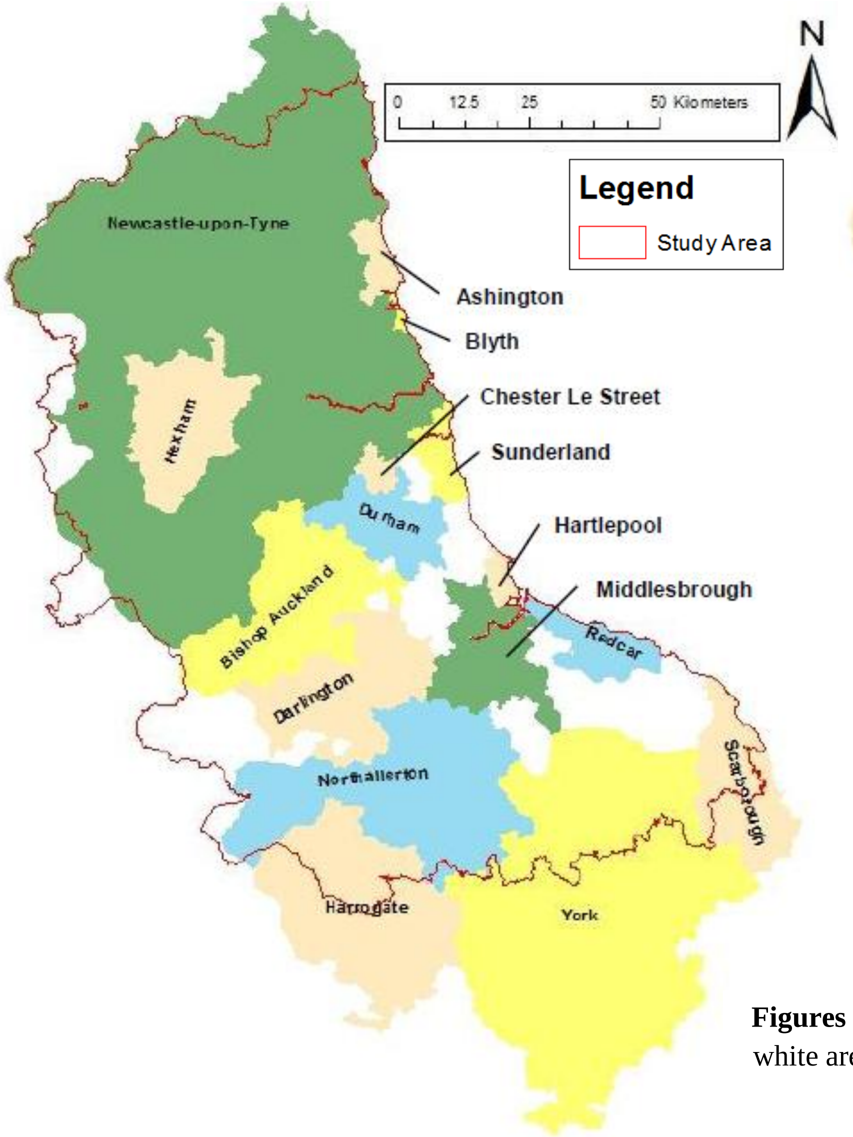
Demand Containment** = **Internal Flows** / (**Total Demand Flows** + **0.01)

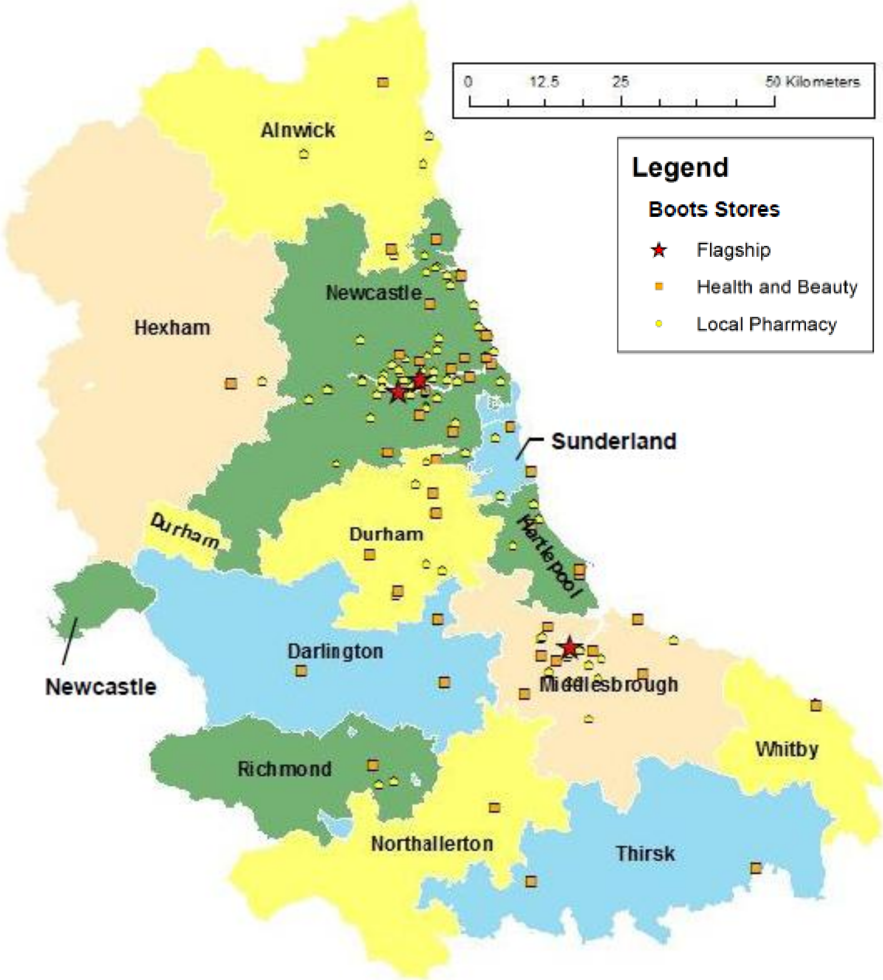
Equation 2. Calculating Demand Containment

In this adaption of the TTWA-Algorithm, all conurbations must meet both self-containment and size (total transactions) thresholds for the algorithm to converge. Coombes refers to this stage of the algorithm as the 'X' equation (Coombes 2007). Coombes' method for allowing trade-off between these two criteria has been simplified in the pilot study presented here to using two thresholds as already stated.

4. Results

One of the main considerations of the project is the choosing of thresholds for the algorithm. A variety of different thresholds were tested from 50 per cent to 75 per cent self-containment, and 100,000 to 1,000,000 minimum internal flow size. Through consultation with Boots, it was decided that thresholds of 66.67 per cent and 500,000 generated an appropriate number of conurbations for retail planning and ensured high self-containment. Figures 5 to 7 present current and TTWA-derived conurbations, together with the self-containment value for each conurbation. As they are considerably outside of the study area, three current conurbations (Harrogate, York and Scarborough) are excluded from the self-containment calculations.





Figures 5 and 6. Current Boots Conurbations derived via gravity modelling; white areas represent LSOAs not assigned to any conurbation (left). TTWA-derived Conurbations (right)

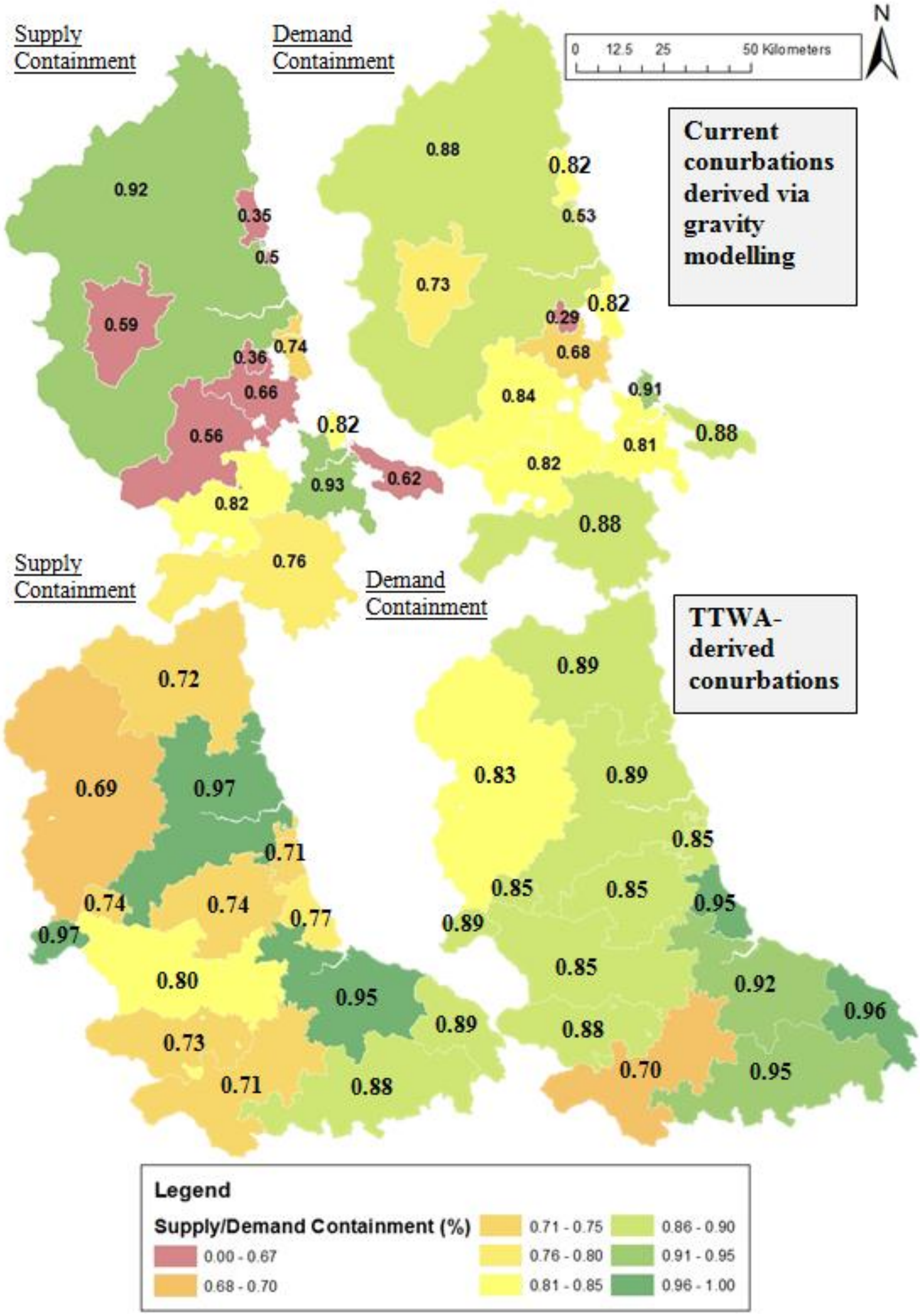


Figure 7. Self-Containment of Current Conurbations derived via gravity modelling and TTWA-derived Conurbations

For this study area there are 13 to 16 current (gravity model-derived) conurbations compared to 13 TTWA-derived conurbations. Discrepancies between the numbers of conurbations from the two methods are due to disparities in the study area boundary. In the TTWA-derived zones, all data gaps have been removed and full-scene coverage achieved whereas the gravity-derived models contained areas with no coverage. The self-containment of TTWA-derived conurbations is considerably higher, on average by 10 per cent for both supply and demand containment. Supply containment proves to be harder to achieve than demand. Many current conurbation supply containments are less than 66.67 per cent (the selected threshold), in one instance as low as 35 per cent meaning that 65 per cent of customers come from outside the conurbation. While only two current conurbations fall below the 66.67 per cent threshold one has a demand containment of as low as 29 per cent. By comparison, the lowest TTWA-derived self-containment is approximately 68 per cent supply-side and 70 per cent demand side.

Although not presented here due to space constraints, the algorithm was applied to other similar sized regions across the UK to test the applicability of the algorithm and the accuracy of the observations. At the set thresholds slightly fewer conurbations were delineated than current conurbations but unsurprisingly with significantly higher self-containment and with the other benefits such as assignment of all LSOAs to a conurbation.

5. Discussion

Current methods to defining retail conurbations either produce overlapping zones or result in constituent areas not assigned to any conurbations. In contrast, the TTWA algorithm assigns all parts of a study area to a retail conurbation. Defining conurbations using self-containment also results in fewer flows (transactions) crossing output zone boundaries. There remain however some distinct issues drawn upon by this study and by Coombes (2010). Commuting flows are becoming larger and across greater distances. Since commuting and shopping flows are inter-related, the self-containment threshold used to define retail conurbations may need to be lowered to delineate future zones. Some regions are also highly dynamic meaning that the conurbation boundaries may need to be revised regularly. The most sizeable limitation of this approach however comes from the 'big data' required and the need for significant population coverage. The approach is therefore limited to large UK retailers, preferably with extensive loyalty card data. 'Big Data' however is becoming more prevalent enabling the use of these more sophisticated approaches. Finally, the current pilot implementation of the TTWA algorithm in MS-Access VBA requires optimisation in order to be applied to the entire UK.

6. Conclusion

To date, application of the TTWA algorithm outside of commuting flows and the use of sophisticated rules-based approaches to delineate retail conurbation design has been limited or non-existent. This paper shows that such applications can drastically improve on current methods for delineating retail zones. As retail conurbations present a base map for customer planning geography in many location decision-making processes, their accurate and reliable delineation is extremely important. Future studies should continue to add to these findings by improving the efficiency and application of the TTWA algorithm to large geographical areas and datasets.

Acknowledgements

The authors would like to thank ESRC, the Demographic Users Group (DUG) and the 'Retail Data Navigators' Keith Dugmore and Paul Longley, for creating a medium whereby academia and industry could work together on co-operative projects such as this. Thanks are also to other colleagues at Boots for their role in data pre-processing.

References

CACI (2012). "Catchment Modelling." Retrieved 10/07, 2012, from http://www.caci.co.uk/CatchmentModels.aspx

Casado-Diaz, J. (1996). "Local labour markets: preliminary analysis of the Valencian case." Revista de Estudios Regionales **45**: 129 - 155.

Coombes, M. (2010). Defining Labour Market Areas by Analysing Commuting Data: Innovative Methods in the 2007 Review of Travel-To-Work-Areas. Technologies for Migration and Commuting Analysis: Spatial Interaction Data Applications. J. Stillwell, O. Duke-Williams and A. Dennett. Hershey, New York, Information Science Reference.

Coombes, M. and S. Bond (2007). "Travel-to-Work Areas: the 2007 review." Centre for Urban & Regional Development Studies in Newcastle University) and the (Office for National Statistics.

Coombes, M. and S. Openshaw (1982). "The use and definition of Travel-to-Work Areas in Great Britain: some comments." Regional Studies **16**: 141 - 149.

Donnelly, C., et al. (2012). "Marketing planning and digital customer loyalty data in small business." Marketing Intelligence & Planning **30**(5): 515 - 534.

Drezner, T. (2009). "Location of retail facilities under conditions of uncertainty." Annals of Operations Research **167**(1): 107 - 120.

Eurostat (1992). "Study on Employment Zones." Eurostat (E/LOC/20), Luxembourg.

Florez-Revuelta, F., M. Casado-Diaz, et al. (2008). "An Evolutionary Approach to the Delineation of Functional Areas Based on Travel-to-work Flows." *International Journal of Automation and Computing* **05**(1).

Frey, W. and A. Speare (1995). Metropolitan areas as functional communities. Metropolitan and Nonmetropolitan Areas: New Approaches to Geographical Definition. D. Dahmann and J. Fitzsimmons. Washington, DC: US Bureau of the Census**:** 139 - 190.

Goodman, J. (1970). "The Definition and Analysis of Local Labour Markets: Some Empirical Problems." *British Journal of Industrial Relations* **8**(2): 179 - 190.

Huff, D. l. (1964). "Defining and estimating a trading area." Journal of Marketing **28**: 34 - 38.

Nakanishi, M. and L. G. Cooper (1974). "Parameter estimate for multiplicative interactive choice model: least squares approach." Journal of Marketing Research **11**: 303 - 311.

Papps, K. and J. Newell (2002). "Identifying Functional Labour Market Areas in New Zealand: A Reconnaissance Study Using Travel-to-Work Data " Discussion Paper No. 443, Februrary 2002.

Sforzi, F., S. Openshaw, et al. (1997). The procedure to identify local labour market areas. I Sistemi Locali del Lavoro. F. Sforzi. Rome, ISTAT**:** 235 **-** 242.

Smart, M. (1974). "Labour market areas: uses and definitions." Progress in Planning **2**: 239 **-** 353.

Biography

Matthew Pratt completed his Masters in Applied GIS and Remote Sensing from the University of Southampton in 2012. He is now undertaking a PhD relating to geodemographics at UCL under the supervision of Professor Paul Longley.

Jim Wright is currently a lecturer in Geographical Information Systems, with a particular interest in environmental management and health applications of GIS. Jim also has research interests in water and health in developing countries.

Samantha Cockings is currently Lecturer in Geography. Her research interests are focused around automated zone design, space-time modelling of populations and links between environment and health.

Iain Sterland leads the Portfolio Analytics team at Boots UK. His focus is to understand how location dynamics impact site performance, reduce risk around investment decisions and leverage geography to drive sales and profit growth across the wider business.