## Title: Delineating retail conurbations: a rules-based algorithmic approach

Corresponding Author - Matthew D. Pratt

Matthew D. Pratt<sup>a, b</sup>, Jim A. Wright<sup>a</sup>, Samantha Cockings<sup>a</sup>, Iain Sterland<sup>c, d</sup>
<sup>a</sup>Geography and Environment, University of Southampton, University Road, Southampton, SO17 1BJ
J.A.Wright@soton.ac.uk; S.Cockings@soton.ac.uk
<sup>b</sup>Department of Geography, University College London, Gower Street, London, WC1E 6BT
matthew.pratt.12@ucl.ac.uk, phone: 07772 383059 fax: +44 (0)20 7679 0565
<sup>c</sup>Location Planning, Boots UK, 1 Thane Road West, Nottingham, NG2 3AA
<sup>d</sup>Location Planning, Sainsbury's Supermarkets Ltd, Unit 1, Draken Drive, Ansty Park, Coventry, CV7 9RD, iain.sterland@sainsburys.co.uk, phone: 02476 529034 m:07900 702411

## Abstract

Retail conurbations may be defined as market areas with high intra-market movement. A limited range of approaches has been used to delineate such retail conurbations. This paper evaluates a simplified version of an existing zone design method used to define labour market areas, the Travel-To-Work-Area algorithm (TTWA), for application in a retail context. Geocoded loyalty card spend data recorded by Boots UK Limited, a large health and beauty retailer, were used to develop retail conurbations (newly termed Travel-To-Store-Areas (TTSAs)) for several UK regions using this algorithm. The output TTSA boundaries displayed significantly greater intra-zone flows compared to existing retail conurbation delineation approaches. There is thus scope for researchers and analysts to broaden the zone design approaches used to develop retail conurbations.

*Keywords:* Retail Conurbations, Zone Design, Travel-To-Work-Areas (TTWAs), Self-Containment, Origin Destination Matrix

#### 1. Introduction

#### 1.1 Overview

The zone design problem refers to identification of the most appropriate method for subdividing a region into smaller areas for a given purpose. This paper is concerned with the particular challenge of defining zones for the spatial mapping and analysis of flow data i.e. interactions between zones. Approaches to the design of zones for flow data vary considerably, particularly between different sectors. Retail organisations face the problem when determining catchment areas and conurbations of stores based on customer shopping information. Retail conurbations differ from store catchments in that they often encompass many stores, representing regions in which customers travel internally. Thus, retail conurbations are shared, collective catchments for a group of stores drawing on a similar customer base. By delineating such conurbations, retailers can better understand customer movement to guide decisions such as store location, performance comparisons and investigation of competitor presence. In labour market analysis, an approach to delineating zones, known as the Travel-to-Work-Area (hereafter TTWA) algorithm, is now widely used (Coombes et al., 2010). The TTWA algorithm has been successfully employed internationally to delineate Local-Labour-Market-Areas (hereafter LLMAs) based on commuting flows (Coombes and Casado Díaz, 2007; Eurostat, 1992; Flórez-Revuelta et al., 2008). However, a similar rules-based approach has yet to be applied to retail conurbation delineation. The closest application of the TTWA algorithm in a retail context is to the design of bank catchments using synthetic data (Coombes, 2000). The aim of the research reported here is to demonstrate how the TTWA method (Coombes et al., 1986) can be adapted to create a new planning geography based on customer shopping patterns by defining nonoverlapping areas of retail conurbations while maximising the intra-zonal flow of spend between places of residence and stores. The method is generic in that it can be applied in any country or to any retailer with a sufficient number and coverage of stores and a flow matrix, ideally based on loyalty or bank card transactions.

In the rest of this paper, in addition to the terms TTWAs and retail conurbations, we use a newly introduced term 'Travel-To-Store-Area' (TTSA). Given that the computed areas are not Travel-To-**Work**-Areas, new terminology needs to be developed which reflects the application of this algorithmic approach to the delineation of retail conurbations. TTSAs are defined as 'retail conurbations delineated via the application of the Travel-To-Work-Area algorithm'. A distinction is therefore made between the TTWA algorithm (Coombes et al., 1986) and our newly developed TTSA algorithm. The algorithm itself is still, however, referred to as the TTWA algorithm.

The rest of this section reviews zone design methods, current approaches to retail conurbation delineation and the rationale for employing the TTWA algorithm in this context. Section 2 then describes the proposed generic TTSA methodology, followed by an empirical example of its implementation. The results of the empirical example are then presented and evaluated. Finally, the paper concludes with a discussion of the implications for the delineation of retail conurbations, highlighting conceptual and practical issues which need to be addressed.

## 1.2 Zone design methods

Approaches to zone design for flow data aim to address the 'p-regions problem' which "...involves the aggregation of a finite set of *n* small areas into a set of *p* regions, where each region is geographically connected, while optimizing a predefined objective function" (Duque 2011, p1). In a retail context, one of the most common approaches to zone design is gravity modelling. Gravity modelling attempts to quantify the relationship between consumer movement and attractiveness of surrounding retail centres, while remaining confined by a distance-decay factor. The approach is tailored towards "what-if?" scenarios such as the opening of a competing store (Hernandez and Bennison, 2000). Gravity models are a form of spatial interaction model, whereby entities at different geographical points make contacts, location choices or supply/demand decisions (Roy and Thill, 2004). Simplistic gravity models allocate centres with largest flows to catchments or conurbations to give 'zones-of-dominance'. Many forms of clustering and gravity modelling require a pre-defined number of regions from the outset. Delineated regions through a gravity modelling methodology tend to create overlapping and non-tessellating regions.

Another set of approaches to zone design are hierarchical procedures. The automatic zoning procedure (AZP) of Openshaw (1977) and Openshaw and Rao (1995) has been described as a hierarchical procedure (Spence and Taylor, 1970) since a criterion for amalgamating areas is lowered so that the process iteratively groups input areas into larger zones until these output zones satisfy the criterion (e.g. a specified number of zones or zones of a uniform size). These methods emerged in the 1970s due to the increased availability of computing power and commuting flow matrices. Alvanides et al (2000) illustrate how the AZP algorithm can be employed to design zoning systems for census-based flow data. Hierarchical approaches typically form groups at early stages leading to restrictions at later stages on zones being broken-up and reallocated. A contiguity constraint is also required if contiguous zones are to be formed.

A further set of approaches are rules-based methods. These employ a set of rules in a multi-step procedure to determine zone design. Often these are contained within a theoretical model, guiding decisions as to how and when rules are applied (Flórez-Revuelta et al., 2008). The most widely used rules-based method is that for the delineation of TTWAs, used in the UK since the 1980s (Coombes et al., 1986). The TTWA algorithm is designed to identify LLMAs nationally, thereby aiding governmental bodies to understand commuting work flows and unemployment rates (Catton, 2002). The algorithm aims to maximise self-containment of flows within zones for a minimum flow size/population (Casado-Díaz and Coombes, 2005), such that as few flows as possible cross output zone boundaries; Goodman (1970) refers to this as 'external perfection'. Goodman (1970) first devised these concepts noting the inherent 'trade-off' between self-containment and flow size. It was Smart (1974) however who first defined the criteria and algorithm for analysing commuter flows for approximately 2,000 building blocks. Building blocks can be defined as small elemental units which represent the phenomena of interest and can be aggregated into larger output zones. Smart's (1974) algorithm was later computerised by Openshaw (Coombes and Openshaw, 1982). Around this time, numerous approaches were being developed with a similar objective to internalise larger commuting flows; a form of spatial analysis known as 'functional regionalisation' (Coombes, 2000). The first application of the TTWA algorithm was to the 1981 Census (Coombes et al., 1986) but in recent years, the algorithm has undergone several changes (Coombes and Bond, 2008). In 2001, advances in computational power allowed building block mergers to be broken-up and reallocated in a self-optimising fashion, while by 2007 the increased commuting in the UK (both length and frequency) necessitated a redefinition of the criteria for TTWAs. Zones became larger to internalise longer commuting flows: for example, 334 and 308 TTWAs for 1981 and 1991 respectively were delineated under identical criteria. As a result, the self-containment threshold is now lower and zones can cross international borders (Coombes et al., 2010).

A recent application of a rules-based algorithm within a different sector is provided by Martinez et al. (2009), who employ a rules-based algorithm to define Transportation Analysis Zones (TAZs). Although not directly referencing Coombes' research or the TTWA algorithm, there are similarities in that regions are delineated via self-containment and flow size thresholds. A significant issue however is the prevalence of data gaps due to the relatively small survey-derived dataset. Full scene coverage is still achieved but only by sub-optimally merging unallocated building blocks based on their geographical proximity to the defined TAZs.

A more comprehensive review of zone design methods can be found in Duque's (2011) paper on the p-regions problem.

#### **1.3** Current approaches to defining retail conurbations

Approaches to zone design vary depending on the problem, input data and the purpose(s) for which the zones will be employed. The aim of this paper is to evaluate the TTWA method's suitability for defining retail conurbations. While retail conurbations are more appropriate for extensive store networks, many large retailers still currently favour single store catchments. In some instances, techniques used to delineate even single store catchmentsremain rudimentary. While 86% of UK retailers feel their methods for catchment area identification are sophisticated, only 65% were using techniques such as gravity modelling and as few as 29% were using expert systems (Reynolds and Wood, 2010). Single store catchments remain popular because of the added complexity involved in building retail conurbations and the relative ease of evaluating stores independently. In some cases, however, this potentially hinders planning because large store networks require conurbations to better reflect and understand customer geographies. Boots UK Limited, for example, map both single catchments and conurbations to guide planning decisions. Conurbations are also often wrongly considered as merely 'the sum of single store catchments' rather than denoting the complex relationships between stores.

Many researchers and retailers classify retail centres in terms of store data rather than customer data (Guy, 1998; Mackaness and Chaudhry, 2011); this presents difficulties in generating catchments for groups of stores. Furthermore, little is published on the use of customer data to examine retail geographies. Current retail sector practice instead uses a 'zones-of-dominance' approach, similar to gravity modelling based on Newtonian laws (Drezner, 1994; Huff, 1964; Nakanishi and Cooper, 1974), with many retailers relying on data and/or zones created by companies such as Experian or CACI, who currently employ this approach. CACI, for example, employ UK postcodes as building blocks, which are aggregated into output zones (retail conurbations) based on which areas have the strongest inter-zonal flows. If the flow between two areas does not reach a chosen threshold however, they remain unallocated to any output zones: while this avoids overlapping of conurbations, it instead results in an incomplete tessellation of the region being analysed, presenting difficulties for analysis and interpretation. Although the TTWA method was designed for the identification of LLMAs from commuter work flows, in principle its logic can be applied to most zone design problems, including the definition of retail conurbations. Yet, while there has been considerable discussion of the use of the TTWA algorithm in different sectors such as retail, health and crime (Coombes and Bond, 2008; Yusuf et al., 2011), there have been minimal practical applications. One reason for this is the large number of flows which need to be considered (Marble et al., 1997), although Alvandies et al. (2000, p116) note that "*Most flow data are of increasing commercial importance via data warehousing systems*". The primary reason for few practical applications however is the detached linkage between commercial organisations and academia. Retailers are often unaware of the existence or potential of other approaches, while academics are frequently limited by an inaccessibility of real-world data, remaining confined to synthetic datasets (ESRC, 2012). Coombes (2000) notes that there remain some fundamentally distinct opportunities discussed in the literature which have rarely, if ever, been implemented.

#### 1.4 Rationale for using the TTWA algorithm to define retail conurbations

In 1992, Eurostat outlined the key principles for defining LLMAs. Through time, the TTWA algorithm has become the accepted practice for delineating LLMAs in Europe, defining as many separate zones as possible while still meeting the Eurostat criteria. It has continued to develop and has been applied to 1981, 1991 and 2001 UK censuses, and in European countries such as Italy (Sforzi and Openshaw, 1997) and parts of Spain (Casado-Diaz, 1996) and Non-European countries such as New Zealand (Papps and Newell, 2002). In addition to its extensive deployment internationally, a review by Frey and Speare (1995) considered it to be more advanced than any other alternative sub-regional statistical area definition method.

Eurostat's criteria for defining LLMAs reflect well as criteria for defining retail conurbations (Eurostat 1992). For example, the definition of regions based on self-containment means that conurbations will have high intramarket movement; i.e. customers who reside in a conurbation will have a high probability of shopping in stores in the same conurbation. Analogous data to the commuter matrices also now exist in the form of loyalty card data reflecting customer flows between place of residence and retail centres. Various surveys (Byrom et al., 2001; Hernandez and Bennison, 2000; Reynolds and Wood, 2010) have underlined how data-rich many location planning departments are, although rarely has this been used to its full potential to date. The popularity of loyalty cards makes them a valuable source for geographic and non-geographic consumer data (Dennis et al., 2001).

Rather than make an a priori assumption about the most appropriate number of output zones for a given data set, the rules-based TTWA algorithm leaves the number of output zones to be determined by the patterns in a data set. This lends itself to the principles of exploratory data analysis (Goodchild et al., 1992) which enables analysis of complex spatio-temporal data sets without making prior assumptions about the nature of the data (Marble et al., 1997). Another benefit of rules-based methods are that they routinely produce contiguous regions without using an explicit contiguity constraint, instead relying upon the self-optimisation feature to maximise the inherent tendency of interaction data to link nearby areas (Coombes, 2000). Thus, the recognised theoretical benefits of using the rules-based TTWA algorithm for commuting work flows also apply to retail conurbation delineation.

## 2. Theory and Calculation

# 2.1 Proposed methodology for defining retail conurbations

This section describes the proposed generic TTSA algorithm. Figure 1 provides an overview of the algorithm, itself an adaptation of the TTWA algorithm.

## [Figure 1 about here]

At the outset, all building blocks are considered TTSAs (or 'proto' TTSAs), as none have yet been merged together. The first stage is to calculate the self-containment (i.e. the extent to which flows of customer spend are within zones) and size of all TTSAs. Retail-specific metrics for these variables are required. Self-containment can be measured as the proportion of spend originating from within a given TTSA, whilst TTSA size can be measured as the total spend in all stores within a TTSA. As a first step in designing conurbations, minimum target thresholds can be defined for both self-containment and size. When all TTSAs meet both these thresholds the algorithm will stop as it has converged. Self-containment has two components, supply-side and demand-side. These are calculated for each TTSA from the internal supply and demand flows. Supply flows are the total flows of spend <u>to</u> stores in other TTSAs, while demand flows are the total flows divided by the

supply/demand flows. The algorithm then takes whichever is the lowest value for supply-side and demand-side as the self-containment value. Figure 2 shows the self-containment calculations for an example TTSA (labelled Proto-TTSA 'A') in relation to two other TTSAs (B and C). TTSA 'A' currently has three internal flows from customer places of residence to stores, which do not cross its boundary. It also has three supply flows reflecting customers resident within TTSA 'A' who shop in stores in the two neighbouring zones, and three demand flows as customers from the two neighbouring zones shop in TTSA 'A'. The figure also illustrates how each TTSA can contain a variable number of stores and that TTSAs are formed by the aggregation of multiple LSOAs.

## [Figure 2 about here]

At this stage, the TTWA algorithm then calculates which proto-TTWA least meets the set thresholds. To determine this, a trade-off function known as 'The X Equation' produces a single value based on these two parameters. The X Equation can take many forms, from simple minimum or target thresholds through to the more complex equation implemented by Coombes (1986). In Coombes' (1986 p 951) approach, minimum and target thresholds are set for both self-containment and size. To meet the criteria a proto-TTWA must meet at least one minimum and one target threshold (figure 3). In principle, similar criteria and an equivalent to the 'X' equation can be employed for defining TTSAs, although the precise values and equation will need to be retail specific.

#### [Figure 3 about here]

The algorithm then merges each of the proto-TTSAs with the TTSA with which it shares the greatest 'connectance flow'. The 'connectance flow' is determined by adding ratios of the supply and demand flows between different TTSAs (equation 1). This metric is designed to prevent TTSAs in major retail centres from engulfing all neighbouring regions and fewer TTSAs being formed than desired (Coombes et al., 2010). The algorithm then loops back and continues to iterate until the thresholds are met, reflected by the 'X' equation.

$$\left(\frac{T_{ij}}{\sum_i T_{ij}}\frac{T_{ij}}{\sum_j T_{ij}}\right) + \left(\frac{T_{ji}}{\sum_i T_{ji}}\frac{T_{ji}}{\sum_j T_{ji}}\right)$$

Whereby: 'Tij is the flow from TTSA i to TTSA j and  $T_{ji}$  is the flow in the reverse direction from TTSA j to TTSA i. The total flows from TTSA i ( $\sum_i T_{ij}$ ) include all flows from i to itself and the total flows from TTSA j includes flows from j to itself ( $\sum_j T_{ij}$ ). The same is true of the other terms in the denominator.

## Equation 1: Calculating the Strongest Connectance Flow, derived from (Coombes and ONS, 1998)

## 2.2 Data and Study Regions

The feasibility of implementing the generic TTSA methodology for the creation of retail conurbations is demonstrated using the empirical example of delineating conurbations for Boots UK Limited, a major UK health and beauty retailer. Boots UK Limited has invested heavily in data quality, resulting in large high-quality datasets which can be used for a range of purposes. For this example, Boots UK Limited provided customer origin-destination flows from their places of residence to stores, obtained from loyalty card information. Customers who sign up to the Boots UK Limited loyalty card scheme provide their home address. The amount they spend in each Boots UK Limited store is then recorded as an origin-destination spend flow from their home address to this store. Data were provided by Boots UK Limited at the aggregate level for Lower-Layer Super Output Areas (LSOAs), which are fairly small geographical areas with populations between 1,000 and 3,000, used to release Census and other data in England and Wales.

Stores located in airports were removed as customer spend flows would cover greater distances and therefore have an adverse effect on regional retail conurbation delineation. Flows from an origin LSOA to a store which had less than five customers per year, and flows that were less than half a per cent of the total flows leaving that origin LSOA per year, were removed from the database. The database, effectively an 'origin-destination' matrix, hence consisted of three columns: the origin LSOA where customer(s) live, the destination LSOA containing one or more stores, and the flow amount representative of yearly spend by customers in the origin LSOA in stores in the destination LSOA. Spend figures were multiplied by an undisclosed number by Boots UK Limited so as to avoid the release of market sensitive data. This does not affect the delineation of retail conurbations.

The pilot study area encompassed the settlements of Newcastle, Sunderland and Middlesbrough and consists of one hundred and fifty Boots UK Limited stores. The Newcastle study area was recommended by Boots UK Limited as there were few customer flows leaving and entering the study area, thereby minimising edge effects. Additionally, current retail conurbation coverage was poor leading to significant data gaps. In order to check that the methods are extensible to other areas, retail conurbations for Northern Ireland were also delineated as this is a relatively small study area with negligible flows in or out of the area. Results were compared to existing Boots UK Limited conurbations produced by the marketing analyst CACI using the 'zones-of-dominance' approach (see Section 1.3 above).

#### 2.3 Implementation of the Algorithm

As there is currently no publicly available software that implements the TTWA algorithm, a prototype version of the method was implemented in a relational database using Structured Query Language (SQL). A simplified version of the 'X' equation was employed here, since the definition of an appropriate level of trade-off between self-containment and size requires more detailed research. This simplified version sets a single target for self-containment and size (based on total spend), with the condition that both must be met. The equation works by dividing the minimum self-containment and size values by their desired target thresholds. If the ratio is equal to or greater than 1, then the value has met the criteria for the set thresholds. The results of both parts are then multiplied. If the outcome of this equation is also equal to 1 then both targets have been met and the algorithm has converged. If the target has not been met, the TTSA with this lowest score is broken up into its constituent zones.

Boots UK Limited requested a similar number of conurbations to the existing set produced through the 'zonesof-dominance' approach; this facilitates comparison between the two zone design methods. A self-containment threshold of 66.67% (similar to that used by Coombes and Bond, 2008) and a minimum total spend of 500,000 (in rescaled, dimensionless, units as described in Section 2.2) were chosen to meet these requirements. Through experimentation with a range of self-containment and size thresholds, this combination of values produced a suitable number of zones with known retail hierarchies self-contained. Feedback from Boots UK Limited was obtained to determine if the outputs looked reasonable for application in retail analysis.

## 2.4 Evaluation of Output Retail Conurbation Zones

Two methods are employed for comparing the retail conurbations produced using the existing zones-ofdominance approach and the new results from the TTSA algorithm. Firstly, the self-containment of flows for each conurbation is calculated. Self-containment is an appropriate measure as conurbations should maximise internal flows to best convey customer behaviour. Secondly, retail hierarchies are produced, which show whether two stores with a strong hierarchal relationship are contained within the same conurbation. Strong hierarchal relationships are determined by the geographic overlap of two store catchment areas, with each catchment defined using customer records. A smaller store proximal to a larger store will typically have a strong hierarchical link, as the smaller catchment will sit within the larger one, whereas neighbouring stores of similar size may show only a low overlap as customers mostly shop in one or the other. If these hierarchal relationship lines cross retail conurbation boundaries then it casts doubt on their validity.

## 3. Results

## **3.1 Algorithm Execution**

Figure 4 (after Coombes and Bond, 2008)) shows the minimum self-containment and total spend values from any TTSA after each iteration of the algorithm. Supply-side containment is generally lower than that of demand-side throughout the iterative process. As the algorithm runs, the minimum TTSA self-containment gradually increases to meet the threshold. Sharp declines are evident throughout as TTSA are broken up to search for a better merger. Total spend within a TTSA also increases, but more gradually than self-containment.

## [Figure 4 about here]

Supply containment is very high for conurbations containing the largest settlements of Newcastle and Middlesbrough. Customers close to these major settlements are much less likely to shop in other regions and settlements for convenience reasons. Demand-side, however, is more difficult to contain for these regions.

## **3.2 Output TTSA Conurbations**

Figure 5 shows the zones-of-dominance (5A) and TTSA (5B) conurbations for the Newcastle study region. Since the spatial extents of the study region and the zones-of-dominance conurbations are not precisely coterminous, there are slight discrepancies in the number of conurbations delineated by each approach. In Figure 5B, the locations of Boots UK Limited stores, categorised by store type, have been plotted on the TTSA conurbations. Flagship stores are the largest stores in terms of floor size and typical range of stock, while local pharmacy stores are the smallest. Typically, TTSA conurbations in Figure 5B contain at least one health and beauty store in order to facilitate the needs of the customers within the conurbation.

## [Figures 5A and 5B about here]

The outputs from the two zone design methods differ a great deal. The zones-of-dominance approach produces more variation in the geographical size of conurbations than the TTSA method. Perfect tessellation is achieved with the TTSA algorithm, however the lack of a contiguity constraint means that the Newcastle, Durham and Darlington conurbations are non-contiguous.

Figure 6 shows the conurbations derived for Northern Ireland. Unlike for the Newcastle study region, all TTSAs for Northern Ireland are contiguous. The data gaps between zones-of-dominance conurbations have been filled and the TTSA algorithm has delineated 10 conurbations compared to the 11 produced by the zones-of-dominance approach. For both study regions, slightly fewer conurbations were delineated by the TTSA algorithm compared to those defined via the zones-of-dominance approach.

#### [Figures 6A and 6B about here]

## **3.3 Evaluation of Outputs**

Figure 7 presents box plots of the supply-side, demand-side and overall containment values, together with size values, for the zones-of-dominance and the TTSA-derived conurbations for the Newcastle study region It shows that supply-side, demand-side and overall containment values are all higher using the TTSA algorithm than the zones-of-dominance approach and exhibit a much smaller range in values. Some of the zones-of-dominance conurbations perform poorly, with less than 40% self-containment. The size of conurbations delineated using the TTSA algorithm is more uniform with lower variance.

#### [Figure 7 about here]

Table 1 shows the number of strong retail hierarchical links which are contained within (intra-zonal), and which cross (inter-zonal), boundaries of retail conurbations for the Newcastle study area. The TTSA conurbations contained a third of the inter-zone links of the zones-of-dominance conurbations. This means that TTSA conurbations more closely reflect the retail hierarchies between stores by containing these strong store relationships within the same conurbation. 6 out of the 10 inter-zone links for the zones-of-dominance conurbations were due to links being drawn to stores located in data gaps.

#### [Table 1 about here]

## 4. Discussion

## 4.1 Comparison of Local Labour Market Areas and Retail Conurbations

The use of the same algorithm to generate both LLMAs and retail conurbations enables comparison of these two sets of zones. Many of the characteristics observed in the TTSAs are similar to the TTWA-derived LLMAs reported by Coombes and Bond. As per Coombes and Bond (2008), supply-side containment is lower than demand-side containment. This is because a labour-force from one region may work in a variety of areas, whereas demand flows are likely to be easier to contain to a region. This population movement is similar for customers to stores. The perfect tessellation is also a common feature between LLMAs and retail conurbations. Non-contiguous regions do feature in Coombes' LLMAs, but they appear to be less frequent than in the initial trials reported here with retail data.

The selection of the self-containment and total spend (flow size) thresholds should be carefully considered in a retail context. Occasionally, the thresholds for delineating LLMAs have been revised, particularly the self-containment constraint, as people commute increasingly longer distances (Coombes and Bond, 2008). Further implementations therefore should experiment with these thresholds in determining an optimum calibration with consideration of the desired retail zonal geography.

## 4.2 Comparison of Zones-of-Dominance and TTSA-derived Retail Conurbations

Compared to the zones-of-domination approach, the TTSA algorithm generated retail conurbation boundaries that were crossed by a lower proportion of spend flows between customer residences and stores, as shown by their higher self-containment values in Figure 7. This finding is unsurprising, given that the algorithm explicitly optimises self-containment. However, an analysis of stores sharing significant numbers of loyalty card holders in common suggested these were more frequently located in the same TTSA conurbation, compared to the zones-of-dominance zones (Table 1). This metric was not an explicit criterion optimised by the TTWA algorithm, and therefore forms a more independent means of evaluating the coherence of the two sets of zones.

The TTSA algorithm also generated retail conurbations that were more uniform in size (Figure 7). From a retail management point of view, it is more desirable to have more uniform areas for a sales team to oversee. Small conurbations were avoided by the TTSA algorithm through the use of the customer spend (size) threshold,

which allows retailers to create conurbations with a desired minimum customer spend. Typically, the zones-ofdominance approach generated larger conurbations for city centres than the TTSA algorithm. For example, the Newcastle conurbation derived by the zones-of-dominance approach is roughly twice the geographical size of the TTSA-derived conurbation; however it has a self-containment 5 per cent lower. The complete tessellation of the entire study area is also an obvious advantage to the TTSA approach, though it does generate noncontiguous retail conurbations (Figures 5 and 6). These non-contiguous regions have flows travelling over other conurbations but are not considered as being external customer spend flows. While some may see these as misleading and harder to visualise, they arguably allow for more realistic conurbations.

## 4.3 Applications of TTSA Retail Conurbations

Retail conurbations are most useful when they are highly self-contained in terms of customer spend, when they perfectly tessellate producing no data gaps and when they are neither geographically very small nor large. The TTSA outputs are therefore highly applicable to the design of retail conurbations. Retail conurbations have many uses in understanding customer behaviour. They provide excellent zonal boundaries for evaluating stores as a network rather than a 'sum of its parts'; stores within a conurbation will have strong hierarchical links with each other, and without the identification of TTSAs it is very difficult for a retailer to assess how well a local network of stores are performing as a whole due to the complexity of shoppers behaviour. The TTSA approach allows the analyst to abstract from individual stores and benchmark performance across TTSA areas. This can help indicate competitor presence, gaps in the market and help to guide decisions as to the opening, closure, expansion and relocation of stores.

#### 4.4 TTSA algorithm performance and implementation issues

For this research, we implemented the TTSA algorithm using SQL in a relational database. Our prototype implementation remains computationally intensive and would currently be difficult to run for the whole of the UK. Since this version of the implementation has been run only for sub-regions rather than nationally, 'edge' effects are introduced at the margins of sub-regions such as the northeast and Northern Ireland. Coombes encountered performance problems in earlier pre-2000 forms of the TTWA algorithm when available computational power was lower (Coombes, 1986), but the long-term revision and development of the local labour market implementation of the algorithm means that such problems have now largely been resolved. The

prototype implementation of our TTSA algorithm would similarly require further optimisation before being applied nationally.

Additionally, the impact of simplifying the X-equation in the prototype implementation presented here is unknown, although the initial results suggest that appropriate trade-offs are being reached. Whilst there is considerable practical experience underpinning the thresholds used in the TTWA algorithm for delineating local labour markets, the development of appropriate TTSA criteria thresholds for retail conurbations will require further experimentation across a wider range of regions and data sets.

## **4.5 Future Research**

An immediate priority for future work on the TTSA algorithm is improvement in the prototype software's performance. Some progress has already been made in this area by Boots UK Limited who devised three alternative simplifications of the algorithm. It is also recommended that future work should investigate and refine the trade-off between TTSA size and self-containment, as embodied in the 'X'-equation of (Coombes, 2000). This would improve the sophistication with which the most poorly configured proto-TTSA is identified during algorithm runs, so that it can be subsequently broken up and reconstituted.

In evaluating the TTSA algorithm relative to existing alternative zone delineation approaches, we examined the locations of stores sharing a large number of customers in common. However, there may be potential to develop further independent metrics for evaluating the coherence of retail conurbations. It would also be of interest to see how the algorithm would perform when applied to other retailers and their loyalty card datasets. As discussed in detail earlier in Section 4.2, some users may wish conurbations not to be fragmented and to form as whole entities. The suitability and adaptability of other automated zone design tools (such as 'AZTool' (Cockings et al., 2011)) and other methods (such as the modularity and geographical weighting approach recently presented by Farmer and Fotheringham (2012)) could also be explored. The algorithm should also be applied to other scenarios in reinforcing the findings from this research.

## 5. Conclusions

This research has demonstrated the relevance of the TTWA-algorithm to the delineation of retail conurbations. A simplified version of the TTWA algorithm (the TTSA algorithm) has been implemented to produce retail conurbations for two regions in the UK. These retail conurbations provide perfect tessellation and full scene coverage while maximising intra-zonal spend flows between customers' places of residence and retail stores. The generic approach enables the methodology to be applied to any retailer in any country with sufficient coverage and quantity of stores and flow matrices, preferably based on bank card transactions or loyalty data.

Innovative methods developed in academia can often go missed in the commercial sector. However, academic researchers should also be aware that locational planning performed by retailers relies on a mixture of codified and tacit-based knowledge, as opposed to modelling removed from practical contexts (Wood and Browne, 2007; Wood and Tasker, 2008). Academic conceptualisations of spatial design applied to retail modelling remain largely absent from literature (Wood and Browne, 2007). This paper has shown that the well documented TTWA-algorithm has the potential to be of benefit to the retail sector. The combination of an improved approach to the design of retail conurbations together with a prototype implementation, provide both conceptual and practical advances for businesses in their location-analysis modelling and the guidance of decision-making processes.

## 6. Acknowledgements

This research would not have been possible without the help of Boots UK Limited who provided the data and performed much of the pre-processing. The ESRC 'Engaging with Retail' initiative provided funding for collaboration between the retailing and academic communities, which led to this research. Finally, the authors wish to thank Mark Bass, Senior Commercial Analyst at The Insight Group, for his helpful comments on a draft of this manuscript. All views are, of course, the authors' own.

## 7. References

- Alvanides, S., Openshaw, S., Duke-Williams, O., 2000. Designing zoning systems for flow data. In GIS and Geocomputation (Eds P. Atkinson and D. Martin) pp. 115-134 London: Taylor and Francis.
- Byrom, J.W., Bennison, D.J., Hernández, T., Hooper, P.D., 2001. The use of geographical data and information in retail locational planning. J. Target. Meas. Anal. Mark. 9, 219–229.
- Casado-Diaz, J., 1996. Local labour markets: preliminary analysis of the Valencian case. Rev. Estud. Reg. 45, 129 155.
- Casado-Díaz, J.M., Coombes, M.G., 2005. The Delineation of 21st Century Local Labour Market Areas (LLMAs), in: Proceedings of the 8th Nectar Conference, Las Palmas CG.

Catton, N., 2002. Redefining Territories: The Functional Regions. Paris Fr. OECD Publ.

- Cockings, S., Harfoot, A., Martin, D., Hornby, D., 2011. Maintaining existing zoning systems using automated zone-design techniques: methods for creating the 2011 Census output geographies for England and Wales. Environ. Plan.-Part 43, 2399.
- Coombes, M., 2000. Defining locality boundaries with synthetic data. Environ. Plan. 32, 1499-1518.
- Coombes, M., Bond, S., 2008. Travel-to-Work Areas: the 2007 review. Off. Natl. Stat. UK.
- Coombes, M., Casado Díaz, J.M., 2007. The evolution of local labour market areas in contrasting regions. Pap. Present. 45th Congr. Eur. Reg. Sci. Assoc. ERSA Amst. 23-28th August 2005.

Coombes, M., ONS, 1998. 1991-based Travel-to-Work Areas. Office for National Statistics London.

- Coombes, M., Stillwell, J., Duke-Williams, O., Dennett, A., 2010. Defining labour market areas by analysing commuting data: innovative methods in the 2007 review of travel-to-work areas. Technol. Migr.
   Commut. Anal. Spat. Interact. Data Appl. 227–241.
- Coombes, M.G., Green, A.E., Openshaw, S., 1986. An efficient algorithm to generate official statistical reporting areas: the case of the 1984 travel-to-work areas revision in Britain. J. Oper. Res. Soc. 943–953.
- Coombes, M.G., Openshaw, S., 1982. The use and definition of travel-to-work areas in Great Britain: Some comments. Reg. Stud. 16.
- Dennis, C., Marsland, D., Cockett, W.A., 2001. The mystery of consumer behaviour: market segmentation and shoppers' choices of shopping centres. Int. J.
- Drezner, T., 1994. Optimal continuous location of a retail facility, facility attractiveness, and market share: an interactive model. J. Retail. 70, 49–64.

- Duque, J.C., Church, R.L., Middleton, R.S., 2011. The p-Regions Problem. Geogr. Anal. 43, 104–126.
- ESRC, 2012. "Economic and Social Research Council: Engaging with Retail". Retrieved 20/07, 2012, from http://www.esrc.ac.uk/funding-and-guidance/collaboration/working-with-business/index.aspx.

Eurostat, 1992. Study on Employment Zones. Publ. Eur. Communities Luxemb.

Farmer, C.J., Fotheringham, A.S., 2012. Network-based functional regions. Environ. Plan. 43, 2723–2741.

- Flórez-Revuelta, F., Casado-Díaz, J.M., Martínez-Bernabeu, L., 2008. An evolutionary approach to the delineation of functional areas based on travel-to-work flows. Int. J. Autom. Comput. 5, 10–21.
- Frey, W.H., Speare, A., 1995. Metropolitan areas as functional communities, in: Metropolitan And Nonmetropoolitan Areas: New Approaches To Geographic Definition. Working Paper No. 12. Washington DC: US Census Bureau.
- Goodchild, M., Haining, R., Wise, S., 1992. Integrating GIS and spatial data analysis: problems and possibilities. Int. J. Geogr. Inf. Syst. 6, 407–423.
- Goodman, J.F.B., 1970. The definition and analysis of local labour markets: some empirical problems. Br. J. Ind. Relations 8, 179–196.
- Guagliardo, M.F., 2004. Spatial accessibility of primary care: concepts, methods and challenges. Int. J. Heal. Geogr. 3, 3.
- Guy, C.M., 1998. Classifications of retail stores and shopping centres: some methodological issues. GeoJournal 45, 255–264.
- Hernandez, T., Bennison, D., 2000. The art and science of retail location decisions. Int. J. Retail Distrib. Manag. 28, 357–367.
- Huff, D.L., 1964. Defining and estimating a trading area. J. Mark. 34–38.
- Mackaness, W.A., Chaudhry, O.Z., 2011. Automatic Classification of Retail Spaces from a Large Scale Topographic Database. Trans. GIS 15, 291–307.
- Marble, D.F., Gou, Z., Liu, L., Saunders, J., 1997. Recent advances in the exploratory analysis of interregional flows in space and time. Innov. GIS 4, 75–88.
- Martínez, L.M., Viegas, J.M., Silva, E.A., 2009. A traffic analysis zone definition: a new methodology and algorithm. Transportation 36, 581–599.
- Nakanishi, M., Cooper, L.G., 1974. Parameter estimation for a multiplicative competitive interaction model: least squares approach. J. Mark. Res. 303–311.

- Openshaw, S., 1977. A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling. Trans. Inst. Br. Geogr. 459–472.
- Openshaw, S., Rao, L., 1995. Algorithms for reengineering 1991 Census geography. Environ. Plan. 27, 425– 446.
- Papps, K., Newell, J., 2002. Identifying Functional Labor Market Areas in New Zealand: A Reconnaissance Study Using Travel-to-Work Data. IZA Discuss. Pap. No 443 Febr. 2002.
- Reynolds, J., Wood, S., 2010. Location decision making in retail firms: evolution and challenge. Int. J. Retail Distrib. Manag. 38, 828–845.
- Roy, J.R., Thill, J.-C., 2004. Spatial interaction modelling. Springer, Fifty Years of Regional Science: Advances in Spatial Science.
- Sforzi, F., Openshaw, S., 1997. The procedure to identify local labour market areas. Sist. Locali Lav. F Sforzi Rome ISTAT 235 – 242.
- Smart, M.W., 1974. Labour market areas: uses and definition. Pergamon Press.
- Spence, N.A., Taylor, P.J., 1970. Quantitative methods in regional taxonomy. Prog. Geogr. 2, 1-64.
- Wood, S., Browne, S., 2007. Convenience store location planning and forecasting–a practical research agenda. Int. J. Retail Distrib. Manag. 35, 233–255.
- Wood, S., Tasker, A., 2008. The importance of context in store forecasting: the site visit in retail location decision-making. J. Target. Meas. Anal. Mark. 16, 139–155.
- Yusuf, S.A., Georgakis, P., Nwagboso, C., 2011. Review of modelling, visualisation and artificial intelligent methodologies for built environment applications. Built Hum. Environ. Rev. 3.

# Table captions:

 Table 1: Intra-zone and inter-zone retail hierarchy links for Zones-of-Dominance and TTSA delineated

 conurbations in the Newcastle Study Region

	Intra-zone link (crosses no	Inter-zone link (crosses one or
	conurbation boundaries)	more conurbation boundaries)
Zones-of-Dominance	59	10
TTWA Algorithm	66	3

Table 1

# **Figure captions:**

Figure 1: The Process for Delineating TTSAs (adapted from Coombes et al, 2010)

Figure 2: Self-containment Calculation Components for an example Proto-TTSA (labelled A)

Figure 3: Applying the 'X' equation to ensure TTSAs meet set criteria (adapted from Coombes and Bond, 2008)

Figure 4: Values of 'X' Equation Components for Retail Conurbations

Figure 5: Retail conurbations for the Newcastle Study Region, derived using (A) a Zones-of-Dominance approach and (B) the TTSA algorithm

Figure 6: Retail conurbations for Northern Ireland, derived using (A) a Zones-of-Dominance approach and (B) the TTSA algorithm

Figure 7: Self-Containment and Flow Size Box Plots for Original and TTSA-derived Retail Conurbations for the Newcastle Study Region



Figure 1



Figure 2



Figure 3



Figure 4



Figures 5A and 5B



Figures 6A and 6B



Figure 7