Accounting for temporal demand variations in retail location models

Waddington, Tom

Clarke, Graham*

Clarke, Martin

Hood, Nick

Newing, Andy

Consumer Data Research Centre and School of Geography

University of Leeds,

Leeds,

LS2 9JT

* Corresponding Author: G.P.Clarke@leeds.ac.uk

Post-review version of manuscript accepted for publication in Geographical Analysis (date of acceptance, 6th August 2018).

Abstract

This paper develops and calibrates a spatial interaction model (SIM) incorporating additional temporal characteristics of consumer demand for the UK grocery market. SIMs have been routinely used by the retail sector for location modelling and revenue prediction and have a good record of success, especially in the supermarket/hypermarket sector. However, greater planning controls and a more competitive trading environment in recent years has forced retailers to look to new markets. This has meant a greater focus on the convenience market which creates new challenges for retail location models. In this paper we present a custom built SIM for the grocery market in West Yorkshire incorporating trading and consumer data provided by a major UK retailer. We show that this model works well for supermarkets and hypermarkets but poorly for convenience stores. We then build a series of new demand layers taking into account the spatial distributions of demand at the time of day that consumers are likely to use grocery stores. These new demand layers include workplace populations, university student populations and secondary school children. When these demand layers are added to the models we see a very promising increase in the accuracy of the revenue forecasts.

1. Introduction

Spatial interaction models (SIMs) have a long history of application in geography, transport and planning and have been especially important for retail location planning and site assessment (Birkin et al 2010, 2017). This journal has been important for detailing many aspects of their development, including general review papers on SIM (Wilson 2010, O'Kelly 2010), developments with optimal zoning systems (Hagen-Zanker and Jin 2012), new models based on exploiting improved data (Fischer and Reismann 2002) and more generally the development of urban growth models which incorporate SIMs (Favaro and Pumain 2011). For retail location planning, Birkin et al (2010) and Newing et al (2015) showed how the models can be disaggregated to improve their fitness for use in the retail sector, effectively providing decision support tools around revenue and small-area market share predictions.

It can be argued that SIMs have had great success in the grocery sector in particular and are now used by many leading blue-chip retailers in their own store location planning departments (Birkin et al 2017, Reynolds and Wood 2010, Wood and Reynolds 2011). That success has largely been attained in relation to superstores or hypermarkets (in the UK this is generally taken to be stores over 10,000 sq. ft.). Today's grocery market is more challenging for retail location planners in many western countries. In particular, the rise of the convenience market has produced fresh challenges for revenue forecasts. In the UK the convenience market now makes up around 25% of total grocery sales with the leading multiple retailers gaining a greater share of that market each year (Hood et al 2016). Thus, the development of convenience stores has become a significant growth model for the large multiple firms especially in the more restricted planning regimes that have characterised the retail environment of many western countries since the mid-1990s (Wrigley 1998, Wood et al 2006). Temporal sales profiles (see section 3) show that convenience stores in particular seem to rely more heavily on non-residential demand than superstores. That is, there are peaks at, midday in workplace locations associated with lunch time demand, 5-6pm associated with the evening commute, and for some stores, peaks around lunchtime and 3-4pm when school children are out of class. In addition, certain convenience stores rely more heavily on demand based around universities and student populations. Although SIMs generally do have a good performance in the grocery market, if these demand types are not included in the models then revenue can be seriously underestimated in revenue forecasts for certain stores and localities.

Newing et al (2015) showed how more nuanced spatially and temporally disaggregated demand layers could be built into models designed for operation in tourist areas. This included a detailed assessment of tourist accommodation by small-area and how seasonal demand could be added to improve model fit throughout the year (models built purely on residential demand could only replicate store revenues well in the winter periods). That paper was novel also in the fact that it used store and loyalty card data provided by a leading UK grocery retailer. This was especially important for the process of model calibration and the estimation of various measures of store and brand attractiveness.

The aim of this paper is to add new demand layers into retail SIMs to account for a broader range of temporal variations in consumer behaviour than those captured by Newing et al (2015). These layers are related to workplace populations, university students and secondary school children (aged 11 and older in the UK). In section 2 we review progress with applied retail SIMs and then run a 'state-of-the-art' SIM <u>without</u> these new demand layers to show the performance of a highly disaggregate model for revenue prediction for both supermarket and convenience stores. In section 3 we briefly show the importance of time in retail consumer behaviour. We then add our new layers of demand in section 4 based on time of day, outlining how these new layers are constructed and the results of adding this additional demand into the models. Finally we run what-if scenarios in Section 5 to show the value of the model for contemporary store location planning.

2. Disaggregated SIMs for retail location planning

Reynolds and Wood (2010) and Wood and Reynolds (2011) show how many UK blue chip retail organisations use SIMs as important tools in store location planning (along with other techniques). Birkin et al (2017) give many illustrations of the use of these models in different retail sectors (see O'Kelly 2009 for some US examples). They show that in order to be fit for purpose, applied models often need a high degree of disaggregation. Birkin et al (2010) show how improved results can be obtained by adding the capability to handle inelastic and elastic demand estimation techniques (see also Ottensmann 1987), as well as incorporating the network effects of a firm's store portfolio on brand attractiveness and the appropriateness of the competing destinations model for certain types of retail activity (cf. Fotheringham 1983, 1986). Newing et al (2015) set out a different type of disaggregation. They examined the appropriateness of typical residential models (i.e. models where the demand is allocated to home addresses only) in tourist areas. Using data obtained from a leading UK grocery retailer they were able to show how the traditional models under-predicted sales in the tourist seasons and thus how difficult it was for retailers to estimate sales over time effectively for stores in such areas. Their model disaggregation took the form of the addition of new demand layers based on tourists staying in different types of holiday accommodation. Using partner data and survey data provided by Acxiom Ltd. (see below) they were also able to calibrate brand attractiveness more accurately (by consumer type), another important addition to applied retail location modelling.

Based on many of the studies mentioned above, we can set out a disaggregated spatial interaction model as follows:

$$S_{ij}^{gb} = A_i^g \, O_i^g W_j^{\alpha^{gb}} exp^{(-\beta^g c_{ij})} \tag{1}$$

where:

S_{ij}^{gb} represents predicted flows of expenditure between origin i and store j by household type g and store brand b.

O_i^g represents the amount of expenditure available in origin i by household type g.

$W_j^{\alpha^{gb}}$ is the measure of attractiveness of store j and α^{gb} is a power function influencing the importance of the attractiveness variable for store j by household type g and store brand b.

$exp^{(-\beta^g c_{ij})}$ is a distance deterrence factor impacting the distance travelled between origin i and retail destination j by household classification type g.

A_i^g represents the balancing factor ensuring that all demand from origin i by household g is distributed to stores within the study area which is calculated as follows:

$$A_i^g = \sum_{jb} \frac{1}{\sum_j W_j^{\alpha^{gb}} exp^{(-\beta^g c_{ij})}} \tag{2}$$

First, we applied this model to the entire grocery market in an area of the UK, West Yorkshire, incorporating consumer demand which is geo-located in relation to consumers' residential origin. Figure 1 shows all the grocery stores in the West Yorkshire study area, illustrating the concentrations around the core urban areas of Leeds, East Bradford and North Kirklees (the town of Huddersfield). The stores include both traditional superstores and convenience stores for all retailers, data provided by our partner retailer and correct as of October 2014. The 48 partner stores break down into 16 supermarkets and 32 convenience stores. This is typical of the general split for other brands too (a ratio of 1:2 is common for other major brands with a supermarket/convenience store split).

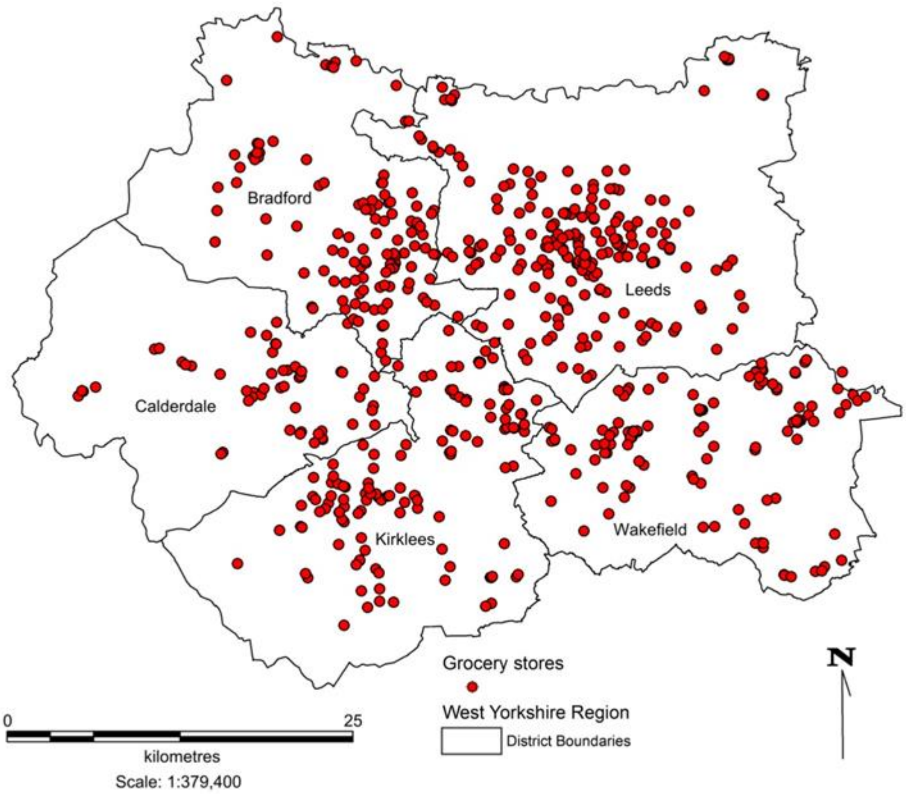

Figure 1: Study region of West Yorkshire and all stores in the models

The demand side contains our own expenditure estimates calculated at the Census Output Area (OA) level by household type (g). An OA is the lowest level of aggregation for the reporting of Census data in England and Wales, typically containing around 125 households. Our OA level estimated expenditures (in £s per week) are calculated using surveyed average household grocery spending derived from the 2014 ONS Living Costs and Food Survey (LCF). Surveyed expenditures relate to all grocery expenditures by these households as distributed across the entire retail supply side. These expenditures are reported by the 2001 Output Area Classification (OAC) (Vickers and Rees 2001), categorising households into one of 7 supergroups (see tables 1 and 2) based on their neighbourhood type. The residential demand was built into our model by aggregating OA level expenditure estimates to the Lower Layer Super Output Area (LSOA) level, a higher level of geographic aggregation enabling model calibration to take place using company loyalty card data.

The parameters of the model were calibrated against data supplied by our partner retailer using their loyalty card data and store sales records, both of which are not normally available to the academic community. Data relate to a one week period in October 2014 (12^{th} – 18^{th} inclusive), representing a typical trading week and are described fully in Waddington et al (2017). This data was used to establish observed flows and revenue figures vital in the calibration process. Observed consumer flow data were extracted from the extensive loyalty card dataset (consisting of approximately 29 million individual records).

Our calibration interaction data represent a 'choice-based sample' (O'Kelly, 1999) since these interactions (between demand side origins and our partners' stores) represent only a subset of consumers – those who have shopped at one of our partner retailers' stores in a transaction that can be attributed to a loyalty card. These data have been collected at the store level and aggregated by origin demand zone. Whilst our choice-based sample doesn't reveal observed (known) flows between origin zones and non-sampled competitor stores, our SIM enables us to make inferences about consumers' store patronage behaviours by origin zone. Our LSOA level expenditure estimates capture all available grocery expenditures by these consumers (referred to by O'Kelly (1999) as 'aggregate measure of trip production'), enabling us to calculate our partner retailers' market penetration by origin zone. Our data also identifies all competing retail destinations (in this case stores) on the supply side, enabling our modelling to account for intervening (competing) destinations at which a consumer (originating in a given demand zone) could shop.

A smaller validation subset (10 out of the 48 stores) was taken from this data and excluded from the calibration process so that they could be used as a control group after calibration to ensure that actual observed consumer behaviour has been replicated (store revenues and market penetration by origin zones) and that the model has not been artificially fitted to the observed data.

The attractiveness term $W_j^{\alpha^{gb}}$ was a combination of store size and brand attractiveness. Size has commonly been used in SIMs in the past and Fik (1988) was one of the first to illustrate the importance of brand on retail performance.. For each OAC group an alpha parameter specific to each retail brand for that consumer type was estimated. This initial segmentation of brand attractiveness using alpha has been adapted from existing research on disaggregation in Newing et al. (2014) and Thompson et al. (2012). Their research made use of extensive consumer data from the private sector provided by the research company Acxiom Ltd, which detailed household spending habits through an extensive consumer lifestyle survey of approximately 750,000 UK households. Using the research opinion poll provided by Acxiom, Thompson et al. (2012) and Newing et al (2014) demonstrated the preferences for individuals to shop at different major grocery retailers. Table 1 shows the estimated alpha values for each OAC by the major grocery brands in West Yorkshire based on that survey data.

Brand (retailer)	OAC supergroup						
	1	2	3	4	5	6	7
	Blue collar	City living	Countryside	Prospering suburbs	Constrained by circumstances	Typical traits	Multicultural
Aldi	0.9980	0.9970	1.0051	0.9987	1.0025	1.0005	0.9952
ASDA	1.0076	0.9912	0.9904	0.9970	1.0023	0.9992	1.0013
Co-Op	1.0020	0.9990	1.0157	0.9922	1.0008	1.0000	0.9894
Lidl	1.0015	0.9995	1.0066	0.9962	0.9957	0.9997	1.0091
M&S	0.9891	1.0381	0.9967	1.0066	0.9952	1.0051	1.0003
Morrisons	1.0005	0.9942	0.9997	0.9987	1.0020	1.0005	0.9990
Sainsbury's	0.9904	1.0121	1.0013	1.0088	0.9942	1.0028	0.9997
Tesco	0.9992	0.9987	1.0071	1.0010	0.9965	0.9990	0.9985
Waitrose	0.9811	1.1000	1.0061	1.0124	0.9843	1.0023	1.0068
Iceland	0.9997	0.9982	1.0058	0.9975	0.9991	1.0001	1.0021

Table 1: Alpha values used to capture brand attractiveness by consumer type. Source: Newing et al. (2014, p228).

For the estimation of the distance decay parameters average trip distance (ATD) was calculated for both observed and predicted flows. A beta value of 0.43 for the aggregate population was optimal. However as we could differentiate flows from different OAC households from the retail partner's loyalty card data we could calibrate a beta value for each OAC making a much more powerful and robust model. Table 2 shows the disaggregated beta values for each OAC. The respective beta values for each OAC were fitted by taking the value of beta when the respective ATD value was as close to 1 as possible.

Table 2 – Beta vales by OAC group for the disaggregate model.

OAC	Beta	R^2
1 – Countryside	0.32	0.79
2 – City Living	0.63	0.96
3 - Multicultural	0.68	0.91
4 – Typical Traits	0.5	0.92
5 – Prospering suburbs	0.46	0.87
6 – Constrained by circumstances	0.49	0.76
7 – Blue Collar	0.49	0.83
Overall GOF		0.84

Finally, we consider calibration and revenue estimations. The assessment of the model accuracy for revenue prediction through standard goodness-of-fit (GOF) statistics on flows reveals that some partner stores are under predicted, demonstrated through a poor fit for modelled flows of consumer expenditure between origin zones and destination stores (in £s per week). GOF assessment on flows following calibration of the model gave an R^2 of 0.76. However, revenue predictions for superstores are generally very good, with an average accuracy level of 90%. The level of accuracy of convenience stores was considerably lower with an average revenue prediction rate of only 55%. Fig 2 shows the differences in model fit between the two store types (Figure 2a shows the results for 16 supermarkets whilst Figure 2b shows the results for 32 convenience stores – these represent all the stores within the study area for the partner organisation, placed in random order in order to protect store confidentiality). It is clear that for some stores we need to reconsider the nature of demand in their catchment areas which currently may not be well represented by traditional measures of residential demand (i.e. the 'night-time' population as captured on census day). The hypothesis therefore is that for a better fitting model for today's grocery environment we need to add new demand layers, those that reflect where consumers are at different times of the day.

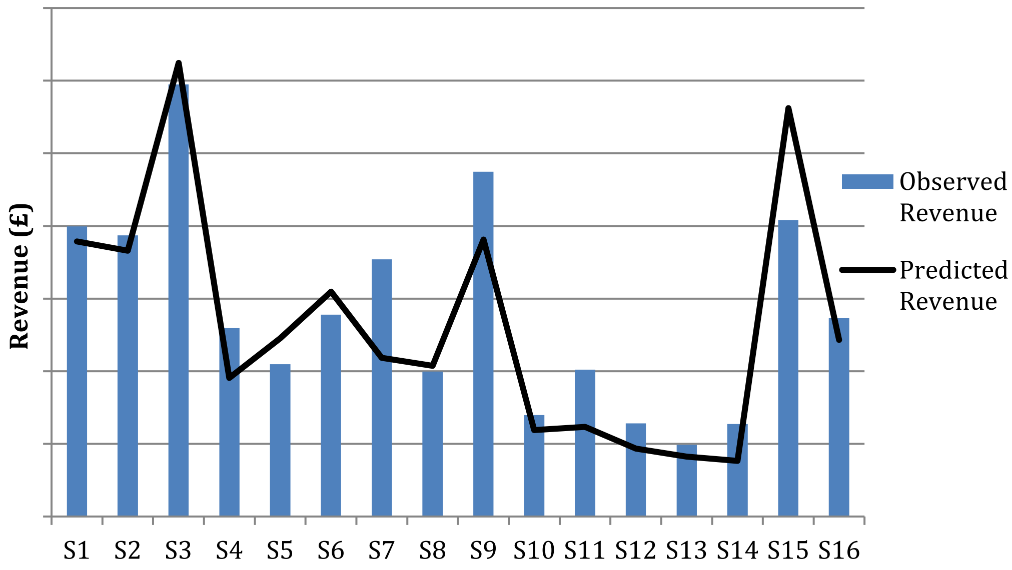

Figure 2a Observed and predicted revenue for supermarkets (S1 to S16)

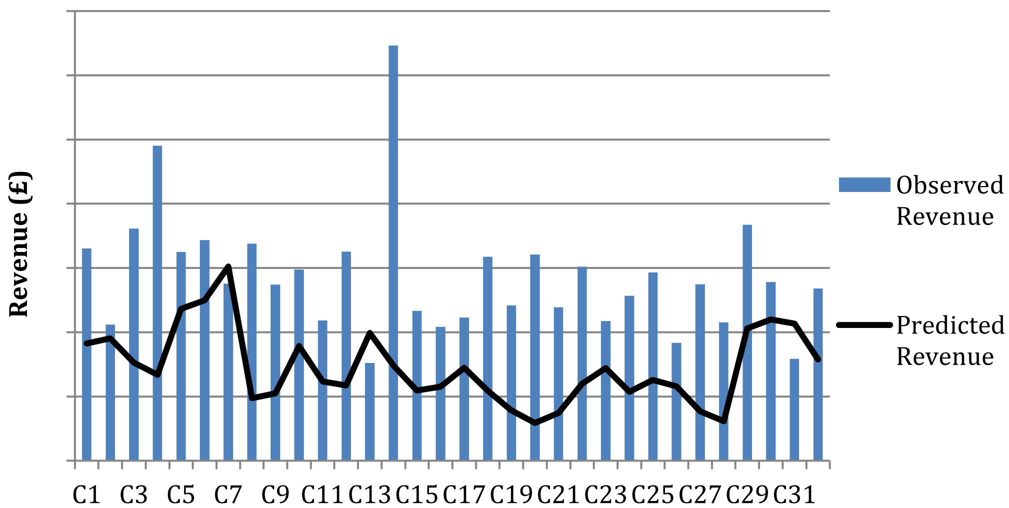

Figure 2b Observed and predicted revenue for C-Stores (C1 to C32)

3. Grocery sales by time of day

There is a limited literature to date on the importance of time of day in retail location analysis, although in urban facility location research more generally there have been examples of studies which have modelled by day of the week, especially in relation to emergency services

(Rajagopalan, et al 2008). However, for retail modelling there have been many studies in the past exploring individual choice preferences and how this could be influenced by time (a good summary appears in Timmermans et al 2002). These behavioural style models have been supplemented by attempts to include multi-purpose trip making, which by definition involves a consideration of how consumers use stores from places other than home (Mulligan 1983, 1987, O'Kelly 1981, Thill 1987, Arenze et al 1993, 2005). One of the most comprehensive analysis of retail sales over time is provided by East et al (1994). Their research, a survey of supermarket shopping habits and shopping times, offers evidence of temporal interactions and how demand varies throughout the day and over different days of the week. A second important study is that of Yun and O'Kelly (1997). Based on a two week survey of shoppers in Hamilton, Canada, they built a suite of retail destination choice models, disaggregating the models by time of the week. They showed it was important to have a weekday model and a weekend model in particular, as the latter saw people shop further from home or work given that they had less time constraints and more access to the family car (see also Hornik 1984, Baker 1996 and Schwanen 2004). More recently Waddington et al (2017) provide a more comprehensive analysis of temporal fluctuations in grocery sales across our study region, showing how sales uplifts appear in certain stores at certain times of the day, hypothesising that these demand fluctuations are driven by the presence at those times of additional demand in the local catchment area. We build on that analysis in this paper by quantifying that additional demand more precisely and adding it into a retail model.

The hypothesis that time of day is important for various types of grocery stores (and convenience stores in particular) can be demonstrated through the examination of actual store sales data provided by our partner organisation. Figure 3 shows the peaks in trading around early morning (people on route to work), lunchtime and 3-5pm when school children then workers are on their way home in all partner stores.

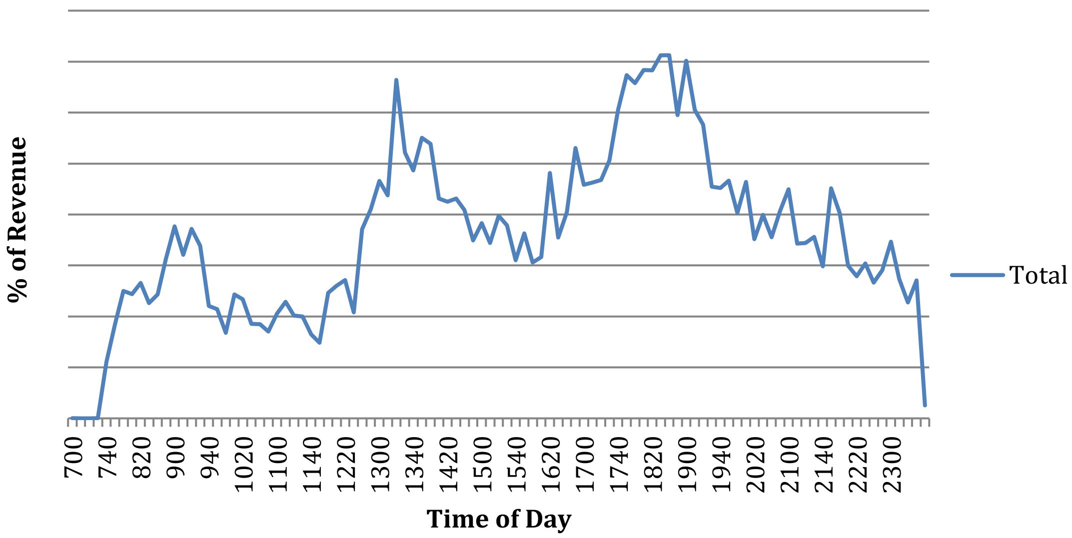

Figure 3: Average sales by time of day in the 48 partner stores in West Yorkshire (supermarkets and convenience stores: note many convenience stores in the UK open 0700-2300 whilst some supermarkets open 24/7)

Our modelling is not seeking to predict the diurnal sales fluctuations shown in Figure 3. However, these fluctuations are indicative of the composition of demand at different times of the day, with many of these sales peaks likely to be driven by non-residential demand within the vicinity of a store. These forms of additional demand are important to capture within weekly revenue predictions as they will improve model accuracy, which we demonstrate within this paper.

It is known that residential populations (as defined by the Census of Population) can be a poor representation of actual daytime populations due to the levels of variation in population movements, both spatially and temporally, that individuals make at certain times of the day (Bell 2015, Martin et al 2015). Models which do not account for this, focusing on a residential populous only, will therefore be more limited in scope and accuracy. This is likely to be particularly relevant for the convenience market given that many of these stores are located in catchments which have low residential populations. Instead, their catchments often contain major workplaces or are in proximity to educational establishments such as universities or secondary schools and where there may be limited residential populations. We next add a workplace model and then turn attention to estimating other forms of new demand associated with educational establishments.

4. Adding a workplace model

The process of adding in work-based consumers is in itself not an entirely new idea within location analysis (Birkin et al 2010, Berry et al 2016). However, it has rarely been done through adding a new model of flows from workplace areas, largely due to poor data availability in the past. This has been recognised as a major data limitation around the production of population statistics and consequently, following extensive input, public consultation and academic involvement, the ONS in the UK released a new workplace based geography termed 'workplace zones' (WPZs) for the reporting of workplace population statistics from the 2011 census (Martin et al., 2013, Mitchell, 2014). WPZs are designed to supplement residential geographies and are fashioned from Census Output Areas (OAs). In some instances existing OAs had very small workplace populations and so these were merged to form much larger WPZs, such as in very rural locations. Where an OA exceeded a workplace population of 625 workers, OAs were subdivided to create multiple WPZs (except in cases where this would lead to a single employer representing the entire workforce of a WPZ). WPZs allow a far more detailed disaggregation of workplace populations (when compared to reporting of workplace populations by residentially specified geographies), particularly in the central location types in which branded convenience stores are found.

Similar to the residential based SIM in section 2, a customised and disaggregated SIM specifically designed to operate using workplace geographies was developed:

$$S_{kj}^{bf} = A_k O_k^z W_j^{\alpha bf} \exp^{-\beta c_{kj}} \tag{3}$$

where:

S_{kj}^{bf} represents predicted expenditure between workplace zone k and store j by brand b and store format f. Store format f differentiates between store formats of major retailers based on whether store j is a supermarket or a convenience store.

O_k^z represents total grocery expenditure in workplace zone k

where:

$$O_k^z = T_k^p \gamma \tag{4}$$

and;

T_k^p is the total workforce population in workplace zone k (including those working from home or near their home, those travelling to work from another location within the model and those traveling to work from outside of the modelled region).

γ represents the value of expenditure per person spent in the grocery sector in proximity to the workplace

$W_j^{\alpha bf}$ is the measure of attractiveness of store j raised to the power of α^{bf} $where$ α^{bf} is a power function determining the importance of attractiveness variables for store j by brand b and by store format f.

$exp^{(-\beta c_{kj})}$ is a distance deterrence factor impacting the distance travelled between workplace origin k and retail destination j.

A_k represents the balancing factor controlling competition in the model ensuring that all demand from origin k is distributed to stores within the study area and is defined as:

$$A_k = \sum_j \frac{1}{\sum_j W_j^{\alpha^{bf}} exp^{(-\beta c_{kj})}} \tag{5}$$

A major issue is the estimation of demand in each WPZ. Some workers will spend nothing in local convenience stores whilst others may buy both lunch and evening supplies there. Anecdotal evidence gained from industry representatives and location analysts at our partner organisation suggested that daytime spending equates on average to the value of £5 per person per week. This includes expenditure on lunch bought from supermarkets or convenience stores in proximity to their workplace, and top-up shopping carried out during the working day or immediately after work. The loyalty card data also revealed that very few customers spend more than £10-£15 per person visit and the majority of sales per person visit are much lower than that. This reassured us that we did not need to reduce spend in the residential model in order to account for any major expenditure in WPZs (a form of double counting). If we apply a £5 per person per week value to the working population in our study area the total weekly workplace derived expenditure is approximately £5.1 million, 8.4% of total weekly grocery expenditure. The workplace model needed to have a high beta value (the model parameter used to control the rate of distance decay) to reflect the high likelihood that consumers will shop in very close proximity to their workplace, driven by typically short lunchbreaks and a high number of intervening opportunities to purchase food and drink (cafes, takeaways, fast food restaurants etc.)

The resulting model, combining modelled flows from both the residential and workplace SIM (representing an initial refined time-of-day fit) is expressed as:

$$S_{ij}^{gbf} + S_{kj}^{bf} = (A_i^g O_i^{rg} W_j^{\alpha^{gbf}} \exp^{-\beta^g c_{ij}}) + (A_k O_k^z W_j^{\alpha bf} \exp^{-\beta C_{kj}})$$

(6)

Where the variables are as before.

Figures 4a and 4b detail the revenue predictions from this extended model. Revenue predictions are again compared to observed store revenues for our partner's stores. A comparison of accuracy between the residential model and the refined model indicates that there has been a marked improvement in model fit. Average accuracy has risen by over 30%. The biggest improvements in revenue predictions are unsurprisingly observed in the convenience stores. It is clear that the disaggregation of demand and improved time-of-day fit of population distribution has overall had a positive impact upon the accuracy of the revenue predictions. The improvements through this approach demonstrate the success of combining separate models that are temporally informed. However, further temporal refinement will be necessary as some stores still remain outside an acceptable accuracy threshold (typically within +/- 10% of observed revenue) suggesting that additional demand types are unaccounted for.

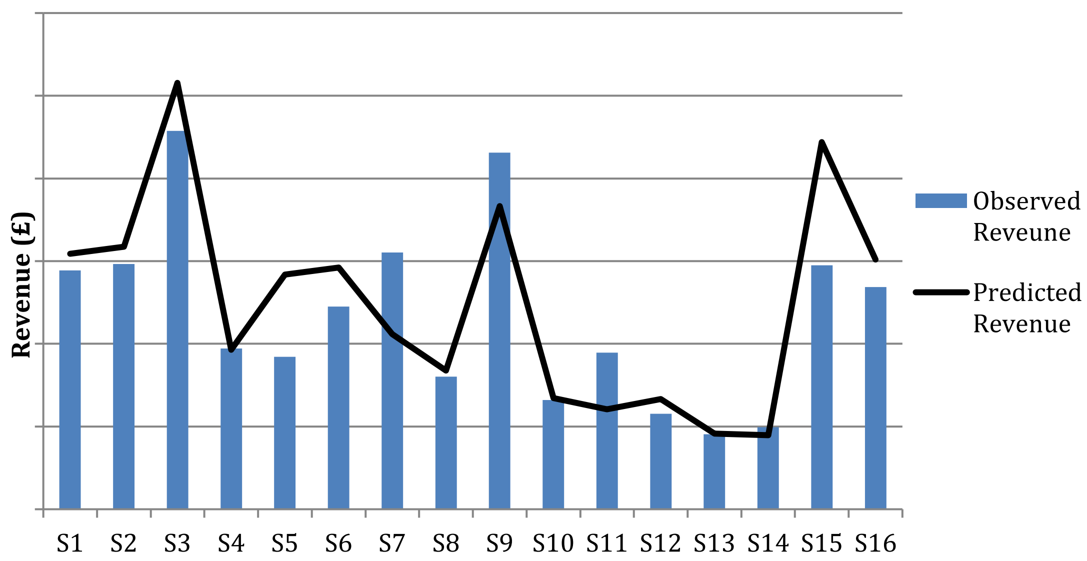

Figure 4(a) Observed and predicted revenue following incorporation of a workplace population for supermarkets (S1-S16)

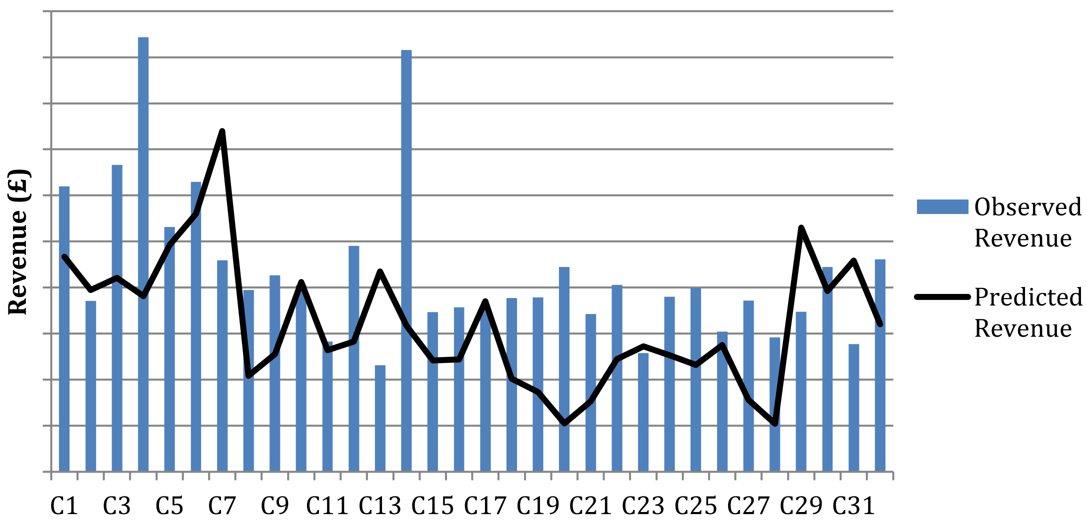

Figure 4b observed and predicted revenue following incorporation of a workplace population for convenience stores (C1-C32)

5. Adding additional demand layers related to populations in proximity to their place of education

i) University students

Although students do appear in the UK Census of Population, counts of student households are often believed to underestimate student consumption patterns (Smith, 2014). Student properties often have many bedrooms as students favour shared lets (Hubbard, 2009, Rugg et al., 2002). Whilst students' consumption patterns sometimes demonstrate that shopping activities are undertaken together, food is more commonly purchased on an individual basis (Ness et al., 2002). Therefore, student expenditure estimates should be based on the total student population for an area rather than based on a count of households (as is usual in traditional SIMs). In many cities this can be considerable: in Leeds alone there are over 50000 university students.

Devine et al (2006) suggest students favour stores close to their home and that they prefer to shop at major retail firms. Consequently, brand attractiveness (weighted by alpha in the model) will be higher for major retailers to account for this behaviour. Similarly, 60% of students were observed to walk to grocery stores (Devine et al., 2006), with many not having access to private transport. Hence, the distance deterrent factor will be higher than for typical residential behaviour to account for the tendency to shop locally and the reduced mobility of students. However, there is a second component of university student demand: the daytime campus population. This can be estimated using the results of Charles-Edwards and Bell (2013). They estimated that by midday 80% of students are on campus (Australian universities). For students on campus, the distance deterrence will be very high. Analysis of loyalty card data and store revenue data for stores close to universities in West Yorkshire appears to demonstrate very little impact on revenue for stores over 500 meters away from the presence of a sizeable university campus (at all times of the day).

The individual weekly spend is reported to equate to approximately £20-£30 on average per student per week (Ness et al., 2002, ZenithOptimedia, 2016, Save the student, 2016). Research indicates that approximately 45% of students purchased food on or near campus multiple times during the week (Pelletier and Laska, 2013). Following the same framework as workplace expenditure a daytime expenditure of £3.25 per person per week was assigned to university students located on campus during the day. This was reallocated from the available student household spend.

ii) Secondary school students

A number of convenience stores show a strong uplift in trade around 3-4pm which is likely to be a result of school children leaving schools in proximity to the stores (such stores also have a sales uplift at lunchtime too, although not as strong as for workplace locations). Evidence suggests that school students are relatively immobile and that they appear to operate within relatively small catchments surrounding their school, typically within an 800m radius when consuming food and snacks (Caraher et al., 2014). Thus the model needs to have a high beta value for this consumer group also. The purchasing habits of school pupils are focused on the purchasing of snacks such as crisps, drinks and sweets (Caraher et al., 2014). Based on the typical food and drink products bought, school students are unlikely to favour any grocery retail brand over another. As major supermarket stores and local independents, typically both stock the range of the snack based products primarily purchased by school pupils, the subsequent store attractiveness for school based demand will be neutral for all brands, making distance the primary factor affecting store choice.

The estimated school based demand expenditure was based on the analysis of two recent consumer research surveys. The first survey reported the average pocket money received by children and the second survey reported teenage spending habits. Following the 2016 survey of teenagers, data suggests that teens spend approximately 20% of their available money on snacks (Piper Jaffyay & Co., 2016). The most recently available survey of national child pocket money suggests that the average child in the UK receives £6.20 per week (Mortimer et al., 2015).

Following this logic, secondary school students spend on average approximately £1.25 of their pocket money on food. However, Caraher et al. (2014) reported that school pupils spent approximately £1.75 on snacks such as crisps, sweets and drinks during the school day, based on UK data from 2005. Thus, a value of around £1.50 seems to be sensible. Although these sums of money are small overall, they will be important for those stores in close proximity to secondary schools, many of which have in excess of 1000 pupils.

iii) Summary

Table 3summarises the contribution of the new demand layers. Traditional household refers to residential demand without any additional layers, based only on household expenditure estimates. Although the additional demand total generated by temporal demand types is a relatively low proportion of the regional total, the important point is that the spatial variations in this additional demand is crucial for improving turnover estimates in certain stores.

Table 3 - Total expenditure estimates for demand groups in West Yorkshire (per week)

Demand group	Expenditure estimate (£ Millions)
Traditional household	**61.46**
Residential	54.24
University student	2.47
Campus based	0.27
School based	0.19
Workplace	5.18
Daytime total	**62.35**
Daytime Difference	0.89

6. New model and its results

Following the disaggregation using the new demand layers , the new SIM (which incorporates the demand types shown in table 1) is expressed below.

Individual demand is now expressed as follows: $h \in [h_1, h_2, h_3, h_4]$

where: h_1 is the total university student demand from residences h_2 is the total available campus based university student demand, h_3 is the total school based demand, and h_4 is the total daytime residential demand by household type g. The model can be written as:

$$S_{ij}^{hbf} + S_{kj}^{bf} = \sum_h [(A_i^h O_i^h W_j^{\alpha^{hbf}} \exp^{-\beta^h c_{ij}})] + (A_k O_k^z W_j^{\alpha bf} \exp^{-\beta C_{kj}}) \tag{6}$$

and the variables are as before.

After applying this model there are an increased number of stores (2 supermarkets and 20 convenience stores) with a predicted revenue within an acceptable performance threshold (within +/- 10% of observed revenue), suggesting an improvement in the predictive capability of the model. The average store prediction for **all** stores is now 90% of observed store sales. For supermarkets, the average store prediction is 103% and convenience stores 83%, which is a noticeable improvement on the first, residential only model (55%). Figs 5a and b show the results for supermarkets and convenience stores with the final model. The results for convenience stores now show a much improved fit. Store C14 for example, now predicts better as it is located in a student area. It is of course still far from perfect. Store C4 remains a poor fit. This store is located on the approach to a major rail station in West Yorkshire. Formally building in transport flows to capture the catchment area of such a store remains an interesting future research project. In addition, it is important to note that other factors not in the model may be influencing sales. For example, management efficiency or the presence of major leisure facilities may (partly) explain some of the residuals.

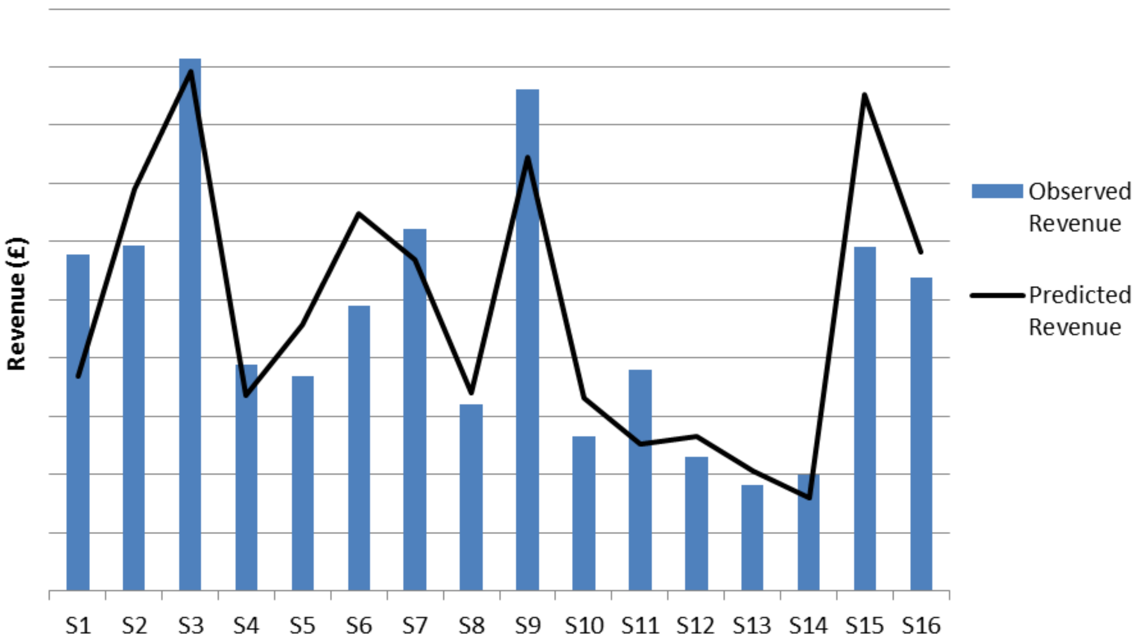

Figure 5a Observed and predicted revenue for supermarkets with the final model (S1-S16)

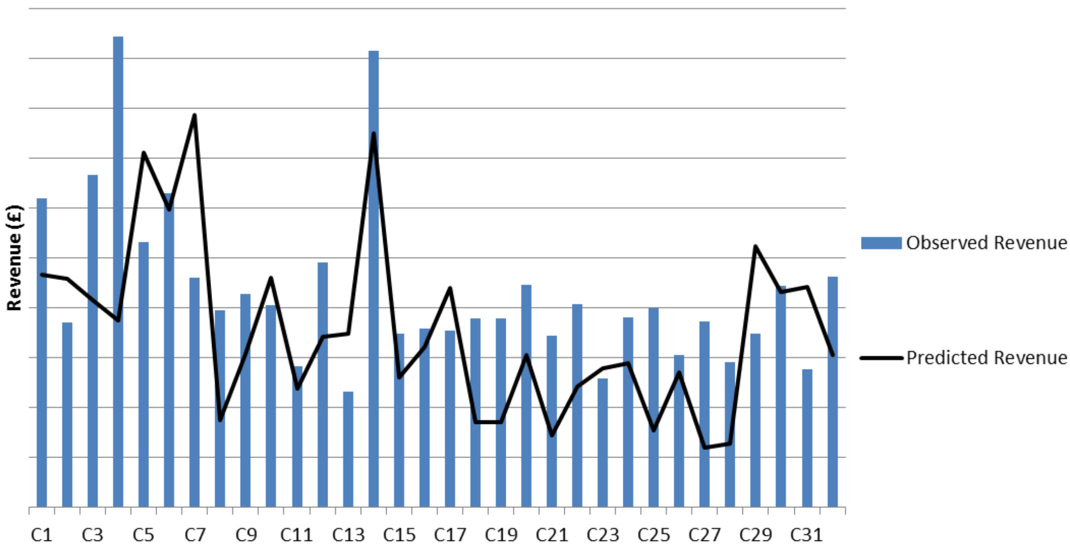

Figure 5b Observed and predicted revenue for C-Stores with the final model (C1-C32)

7. What-if modelling

An important test of the model is that it should not only produce more accurate predictions for existing stores. It must also be useful for future scenario impact assessment. In this final section we test the model with two very different types of what-if analysis. The first explores the opening of new stores in very different parts of the study region (where the new demand layers are important). Second, we test the impact of the university holiday period on revenue predictions.

i) New store revenue modelling

The aim of this section is to apply the new spatiotemporal SIM in a typical commercial situation (new store development) and perform the activity of store revenue estimation which would typically support store location planning. The aim is to estimate the revenues of four hypothetical new stores and to demonstrate the increased insight that can be achieved by using the new demand layers . Three of the new stores used for 'what-if' analysis (convenience stores) were based on actual proposed store openings. The final store (a supermarket) was located in an area with a limited presence of the major retailers (table 4).

Location	Format	Area Description	Store size (sq. ft.)
Major city centre	C-Store	Major city centre containing mixed retail, employment and civic land use, plus major transport interchanges.	1500
Urban suburb	Supermarket	Urban suburb containing predominantly residential land use with good transport links to employment centres.	10000
Small town	C-Store	Town centre store in a small town containing mixed residential, leisure, retail and light industrial land uses.	2500
High density urban residential	C-Store	High density residential housing, predominantly private rented.	2500

Table 4 - Summary of new stores

Table 5 shows an indicator of store performance (Trade Intensity, TI) at each of the proposed stores for the residential only model and then for each demand type in the the new demand layers model.

Store Location	Residential only	Daytime Residential	School	Visitors	Student	Campus	Work	Total
Major city centre	20.09	15.91	0.00	0.23	2.64	0.04	6.01	24.84
Urban suburb	15.58	13.97	0.00	0.00	0.25	0.00	0.23	14.45
Small town	16.06	14.41	0.14	0.00	0.53	0.00	4.44	19.53
High density urban residential	28.30	14.85	0.00	0.00	21.37	1.55	2.50	40.27

Table 5 - Total weekly store trade intensity estimates (£ per sq. ft. per week) for residential only model (left) **and for spatiotemporal model (by demand type)** (right) **for 'new' stores**

Aside from the more detailed demand distributions resulting from using the new demand layers , from a practical point of view the insight generated by the disaggregation of store revenue by demand type can have beneficial implications for store operations. For instance, understanding that a certain proportion of revenue will likely be generated by a particular demand type (whose behaviour and temporal activity have previously been discussed), such as students at the 'High density urban residential' store or workers at the 'Major city centre' and the 'Small town' stores and can influence staffing rotas or the types of products on offer. These are all necessary considerations that store and head office management will consider for all new stores. The novelty of the new SIM is that it provides not only more detailed revenue predictions (as well as improved forecasts based on the evidence for existing stores) but also insight into these issues at the initial planning stage, allowing retailers to prepare and make better informed decisions about store operations based on the new demand layers and subsequent sales fluctuations that has previously not been possible with more traditional SIMs.

ii) Impact of student holidays

The region of West Yorkshire contains 12 universities and around 160 000 students. In many instances students are only in their place of study during term time, normally returning home during the holiday periods. This can have a serious impact upon expenditure levels with a potential decline in store sales experienced at a local level. Furthermore, the longevity of these lower sales periods can be extensive. For example, the University of Leeds had 21 weeks of non-term time in 2016 (University of Leeds, 2017). In stores that cater to a large proportion of students, these periods of reduced demand could result in long periods of low revenue generation, resulting in stores 'under-performing' for long periods and a considerable decline in store performance. Table 6 shows the outputs of store revenue predictions simulating the long-term temporal impact of university term times. Weekly store revenue is simulated both during term-time and during holidays to demonstrate the immediate impact on store sales on a weekly scale (the new demand layer of university students is effectively removed during the holiday periods). This is then scaled to illustrate the potential decline in sales resulting from holiday periods throughout the year. This is supplemented with the potential decrease in store performance (trade intensity - TI), in addition to an accumulated total of sales that stores could potentially lose out on during the holiday periods.

Immediately obvious from the scenario is that all stores experience some impacts on store revenue to some degree. This is shown via the 'Sparklines' on table 6 (final column), which fluctuate throughout the year with varying degrees of magnitude. Of course, those closest to universities will suffer the most: store 'Store H' experiences the biggest temporal effect from the sample of stores, losing 30% of its predicted store sales and experiencing a fall in TI of over £10 per sq. ft. per week during the holiday periods. For this store, this represents a significant change in store performance with the store potentially underperforming for at least 1/3 of the year. However, the total decline in potential sales are highest in stores 'A' and 'D'. In these instances, the sales driven by student demand only represent 8% and 6% of total store sales respectively. However, the annual impact in terms of the potential volume in sales lost during the holiday periods is high (~£1.5 and ~£1.2 million respectively) and thus may represent a far more important commercial issue for the retailer as a business.

	Term time - weekly average			Holiday time - weekly average		Accumilated impact of holidays	Anuual impact from holiday periods
Store	Student driven sales (%)	Student driven sales (£)	Total Store revenue	Change in TI (£s per sq ft per week)	Total Store revenue	Total potential drop in sales (£)	Academic year starting September
A	8	71,067	891,193	-1.12	820,126	1,492,407	
B	2	7,816	334,964	-0.26	327,148	164,136	
C	1	8,670	567,976	-0.23	559,306	182,070	
D	6	56,410	744,193	-1.30	687,783	1,184,610	
E	8	5,015	46,648	-2.80	41,633	105,315	
F	4	2,577	59,754	-1.01	57,177	54,117	
G	2	576	23,850	-0.37	23,274	12,096	
H	30	14,714	34,095	-10.35	19,381	308,994	
I	6	2,145	32,197	-1.62	30,052	45,045	
J	2	772	28,818	-0.59	28,046	16,212	

Table 6: Temporal scenario model predicting the impact of long-term temporal fluctuations in student demand on store level sales

8. Conclusions

SIMs have a good record of successful application in the grocery market, especially for estimating revenues and market shares for larger supermarket stores (Birkin et al 2010, 2017). However, the last few decades have witnessed the rapid expansion of the convenience market in many developed countries and the same location models have been shown to perform less well for these types of stores (Birkin et al., 2002, Hood et al., 2015, Wood and Browne, 2007). When analysing newly available commercial data on store revenue by time of day it is apparent why - sales uplifts appear at certain peak times which suggest that they are driven by specific non-residential demand types which have a unique spatial and temporal distribution. These forms of demand will be missed using a conventional 'residential only' model which may fail to capture a morning uplift in expenditure just prior to the start of work, midday peaks at lunch time and evening peaks at stores in proximity to workplaces. For those stores near secondary schools there is often also a 3-4pm sales uplift. Although a number of other studies have identified the more complex nature of consumer behaviour in grocery shopping which might drive these various uplifts (i.e. Arentze et al 1993, 2005, Baker 1996, Mulligan 1983, 1986, O'Kelly 1981) there are few studies which have attempted to build new demand layers to handle the complex nature of local demand over time more explicitly.

In order to improve the predictive capabilities of SIMs it is clear that it is important to rethink the spatial distribution of demand at different times of the day. After first showing the poor results for convenience stores using a conventional SIM we built a number of new demand layers to be incorporated into the models to supplement residential demand – a workplace model to account for sales from workers (especially important for city centre stores), a school demand layer (to try and replicate the 3-4pm sales uplift in certain stores) and a university student demand layer to help produce better fits between actual and predicted revenue for stores with many students in their catchment areas (which was a problem for all stores located in student areas). The development of this new SIM created greater accuracy overall but especially for convenience stores. The benefits of the research lie not only around greater accuracy of revenue predictions (so vital for location planning exercises). Modelling sales at different times of the day allow retailers to optimise product placement, ranging, availability and staffing rotas accordingly. The ability to forecast these store-level trading considerations at the new store planning stage enables retailers to optimise store design in order to meet the needs of the anticipated trading characteristics, rather than having to react to these characteristics post-opening.

In the final section of the paper the new spatiotemporal model was applied to two location planning scenarios, first for the estimation of revenues for potential new stores to demonstrate the increased level of information that is gained from using the new demand layers compared to a residential only model. In the new store revenue estimation scenarios, four new store locations were modelled and for each new store it has been possible to show how the different demand layers can be important in each case. The message is that retailers need to consider these spatiotemporal components, because if left ignored there is a potential to overlook or underrepresent the profitable nature of a potential new-store location. The findings of the second scenario, regarding the impact of fluctuating student demand, show the loss of sales when students are away on holiday on a store-by-store basis. Operationally, retailers may be able to run promotions at these points in the year when student numbers are down to increase the sales driven by other demand types as a means of balancing the potential revenue lost by fluctuating student demand. They certainly need to factor loss of sales at key times when taking predictions to get board approval in the planning stage.

Thinking beyond the specific UK context which is the focus of the models in this paper, the take home messages about capturing different types of demand leading to improvements in revenue forecasting are applicable much more broadly. The retail landscapes in many other countries also contain a mix of different grocery retail settings and store types which are driven by varying customer bases and customer missions. For example, in the US, large warehouse like stores are supplemented by smaller 711 type convenience stores and whilst calibration and consumer behaviour may be different in different contexts, taking account of all demand origins is important to support the business case for new and existing stores. Similarly, these new demand layers may be important in other areas of retail operations. Although time of the day might not be so important when ordering groceries on-line in operational terms, there may well be an increase in click and collect opportunities in proximity to workplaces and transport interchanges which heightens the need for this form of disaggregate spatial modelling accounting for non-residential demand.

In summary, the evidence presented in this paper demonstrates the increases in accuracy and additional level of detail possible when using the new demand layers for grocery store revenue estimation .It is another step forward in producing more accurate models for both academic and industry usage.

References

ARENTZE, T.A., BORGERS, A. & TIMMERMANS, H.J.P., 1993. A model of multi-purpose shopping trip behavior. *Papers in Regional Science*, *72*(3), 239-256.

ARENTZE, T.A. & TIMMERMANS, H.J.P., 2005. An analysis of context and constraints-dependent shopping behaviour using qualitative decision principles. *Urban Studies*, *42*(3), 435-448.

BAKER, R. G. 1996. Multi-purpose shopping behaviour at planned suburban shopping centres: a space-time analysis. *Environment and Planning A, 28*, 611-630.

BELL, M. 2015. Demography, time and space. *Journal of Population Research, 32,* 173-186.

BERRY, T., NEWING, A., BRANCH, K. & DAVIES, D. 2016. Using workplace population statistics to understand retail store performance. *The International Review of Retail, Distribution and Consumer Research,* 1-21.

BIRKIN, M., CLARKE, G.P. & CLARKE, M. 2010. Refining and operationalizing entropy-maximizing models for business applications. *Geographical Analysis, 42,* 422-445.

BIRKIN, M., CLARKE, G.P. & CLARKE, M. 2017. *Retail location planning in an era of multi-channel growth,* London, Routledge.

CARAHER, M., LLOYD, S., & MADELIN, T. 2014, The 'school foodshed': schools and fast food outlets in a London borough, *British Food Journal, 116*, 472-493

CHARLES-EDWARDS, E. & BELL, M. 2013. Estimating the service population of a large metropolitan university campus. *Applied Spatial Analysis and Policy, 6,* 209-228.

DEVINE.P, LLOYD, K. & GRAY A, 2006, University student food attitudes and behaviour, Final Report, Northern Ireland Social and Political Archive, Belfast.

EAST, R., LOMAX, W., WILLSON, G. & HARRIS, P. 1994. Decision making and habit in shopping times. *European Journal of Marketing, 28,* 56-71.
FAVARO, J.M. & PUMAIN, D., 2011. Gibrat revisited: an urban growth model incorporating spatial interaction and innovation cycles. *Geographical Analysis*, *43*(3), 261-286.
Fik, T.J., 1988. Spatial competition and price reporting in retail food markets. *Economic Geography*, *64*(1), pp.29-44.
FISCHER, M.M. & REISMANN, M., 2002. A methodology for neural spatial interaction modeling. *Geographical Analysis*, *34*(3), 207-228
FOTHERINGHAM, A.S., 1983. A new set of spatial-interaction models: the theory of competing destinations. *Environment & Planning A, 15(1),* 15-36.
FOTHERINGHAM, A.S., 1986. Modelling hierarchical destination choice. *Environment and planning A*, *18*(3), 401-418.
GALE, C.G., SINGLETON, A., BATES, A.G. & LONGLEY, P.A., 2016. Creating the 2011 area classification for output areas (2011 OAC). *Journal of Spatial Information Science*, *12*, 1-27.
HAGEN-ZANKER, A. & JIN, Y., 2012. A new method of adaptive zoning for spatial interaction models. *Geographical Analysis*, *44*(4), 281-301.
HOOD, N., CLARKE, M. & CLARKE, G. P. 2016. Segmenting the growing UK convenience grocery market for store location planning. *International Journal of Retailing, Distribution and Consumer Research. 26*(2), 113-136
HORNIK, J. 1984. Subjective vs. objective time measures: a note on the perception of time in consumer behavior. *Journal of Consumer Research*, 615-618.
HUBBARD, P., 2009, Geographies of studentification and purpose built student accommodation: leading separate lives? *Environment and Planning A, 41,* 1903-1923
LEW, A. & MCKERCHER, B., 2006. Modeling tourist movements: a local destination analysis. *Annals of Tourism Research*, *33*(2), 403-423.
MARTIN, D., COCKINGS, S. & HARFOOT, A. 2013. Development of a geographical framework for census workplace data. *Journal of the Royal Statistical Society: Series A* (Statistics in Society), 176, 585-602.
MARTIN, D., COCKINGS, S. & LEUNG, S. 2015. Developing a flexible framework for spatiotemporal population modeling. *Annals of the Association of American Geographers, 105*, 754-772.
MITCHELL, B. 2014. *Workplace Zones: A new geography for workplace statistics*, Office of National Statistics, Titchfield
MORTIMER, C., BARRATT, C., & CODY C, 2015, Annual Halifax pocket money survey, in H.P.Team (ed) C*hildren's pocket money falls for the second year in a row,* Loyds Banking Group, Halifax.
MULLIGAN, G. F. 1983. Consumer demand and multipurpose shopping behavior. *Geographical Analysis, 15*, 76-81.
MULLIGAN, G. F. 1987. Consumer travel behavior: extensions of a multipurpose shopping model. *Geographical Analysis, 19,* 364-375.
NESS, M., GORTON, M. & KUZNESOF, S. 2002. The student food shopper: segmentation on the basis of attitudes to store features and shopping behaviour. *British Food Journal, 104,* 506-525.
NEWING, A., CLARKE, G.P. & CLARKE, M. 2013. Identifying seasonal variations in store-level visitor grocery demand. *International Journal of Retail & Distribution Management, 41*, 477-492.
NEWING, A., CLARKE, G. P. & CLARKE, M. 2015. Developing and applying a disaggregated retail location model with extended retail demand estimations. *Geographical Analysis, 47*, 219-239.
O'KELLY, M. E. 1981. A model of the demand for retail facilities, incorporating multistop, multipurpose trips. *Geographical Analysis, 13*, 134-148.
O'KELLY, M. E. 1999. Trade-area models and choice-based samples: Methods. *Environment and Planning A, 31*, 613-627.

O'KELLY M.E. 2009. Applied retail location models using spatial interaction tools, in A.S. Fotheringham and P.A. Rogerson (editors) *The SAGE Handbook of Spatial Analysis*, (SAGE Publications, London), 419-442.
O'KELLY, M.E., 2010. Entropy-based spatial interaction models for trip distribution. *Geographical Analysis*, *42*(4), 472-487
ORRELL, S., GREENWOOD, C. & LEWIS, R. 2015. *GB Day Visits 2015 Survey*. TNS: Visit England, Visit Scotland, Visit Wales.
OTTENSMANN, J.R., 1997. Partially constrained gravity models for predicting spatial interactions with elastic demand. *Environment and Planning A*, *29*(6), 975-988.
PELLETIER, J.E. & LASKA, M.N. 2013. Campus food and beverage purchases are associated with indicators of diet quality in college students living off campus. *American Journal of Health Promotion*, *28*(2), 80-87.
PIPER JAFFYAY & CO. 2016. Taking stock with teens. Available: http://www.piperjaffray.com/3col.aspx?id=4035 [Accessed 11/11/2016].
RAJAGOPALAN, H.K., SAYDAM, C. AND XIAO, J., 2008. A multiperiod set covering location model for dynamic redeployment of ambulances. *Computers & Operations Research*, *35*(3),.814-826.
REYNOLDS, J. & WOOD, S. 2010. Location decision making in retail firms: evolution and challenge. *International Journal of Retail & Distribution Management, 38*, 828-845.
RUGG, J., RHODES, D. & JONES, A. 2002. Studying a niche market: UK students and the private rented sector. *Housing Studies*, 17, 289-303.
SAVE THE STUDENT. 2016. The National Student Money Survey 2016. Available: http://www.savethestudent.org/money/student-money-survey-2016.html#earning [Accessed 11/11/2016].
SCHWANEN, T. 2004. The determinants of shopping duration on workdays in the Netherlands. *Journal of Transport Geography, 12*, 35-48.
SMITH, C. W. 2014. Households and household composition in England and Wales: 2001-11 [Online]. ONS. Available: http://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/articles/householdsandhouseholdcompositioninenglandandwales/2014-05-29 [Accessed].
THILL, J.C., 1986. A note on multipurpose and multistop shopping, sales, and market areas of firms. *Journal of Regional Science, 26(4),* 775-784.
UNIVERSITY OF LEEDS. 2017. Term Dates [Online]. University of Leeds Website: University of Leeds. Available: http://www.leeds.ac.uk/homepage/17/academic_year [Accessed 16/3/ 2017].
VICKERS, D. & REES, P. 2006. Introducing the area classification of output areas. P*opulation Trends,* 125**,** 15.
WADDINGTON, T.B., CLARKE, G.P., CLARKE, M. AND NEWING, A., 2017. Open all hours: spatiotemporal fluctuations in UK grocery store sales and catchment area demand. *The International Review of Retail, Distribution and Consumer Research*, 1-26.
WILSON, A.G., 2010. Entropy in Urban and Regional Modelling: Retrospect and Prospect. *Geographical Analysis*, *42*(4), 364-394.
WOOD, S., LOWE, M. & WRIGLEY, N., 2006. Life after PPG6—recent UK food retailer responses to planning regulation tightening. *International Review of Retail, Distribution and Consumer Research*, *16*(1), 23-41.
WOOD, S. & REYNOLDS, J., 2011. Managing communities and managing knowledge: strategic decision making and store network investment within retail multinationals. *Journal of Economic Geography*, *12*(2), 539-565.
WRIGLEY, N., 1998. PPG6 and the contemporary UK food store development dynamic. *British Food Journal*, *100*(3), 154-161.
YUN D-S, & O'KELLY M.E (1997) Modeling the day-of-the-week shopping activity and travel patterns, *Socio-Economic Planning Sciences, 31(4),* 307-319.
ZENITHOPTIMEDIA 2016. Student Living Index 2016. http://personal.natwest.com/personal/life-moments/student-living-index.html: Natwest.