Paper presented at the 21st Cambridge International Manufacturing Symposium | 28-29 September 2017, Cambridge, UK
https://www.ifm.eng.cam.ac.uk/insights/global-supply-chains/cambridge-international-manufacturing-symposium/

Is last-mile delivery only viable in densely populated centres? A preliminary cost-to-serve simulation for online grocery in the UK

Jagjit Singh Srai (jss46@cam.ac.uk)
Institute for Manufacturing, University of Cambridge, Cambridge, UK

Ettore Settanni
Institute for Manufacturing, University of Cambridge, Cambridge, UK

Abstract

This paper proposes a conceptualisation of the 'cost-to-serve' linking the economic viability of last-mile delivery. Specifically, a cost-to-serve model is developed against business-specific indicators, such as share of online sales for a given catchment/geography. The outputs are then discussed in terms of e-commerce penetration and retailer density for the designated catchment/geography. Preliminary evidence is generated by means of computer simulation for a case in online retail of fast moving consumer goods (FMCG) with a focus on the UK. To this purpose a three-stage, pragmatic simulation approach is outlined, using widely available data in order to evaluate alternative last-mile network configurations, and their associated cost-to-serve cost curves. At each stage a sub-model generates, respectively, 1) order-generating locations; 2) basket composition; and 3) last mile delivery cost. Off-the-shelf tools are employed throughout to generate and visualise key analytics, thus facilitating replicability in real-world industrial settings. Results suggest that as well as having cost benefits with increased market penetration and/or increasing the number of drops per journey, as might be predicted, there appears significant potential to narrow the spread of cost variability for a given market penetration by influencing the available locations/time/price options to online customers. The proposed approach can support similar developments besides FMCG, for example in the pharmaceutical industry as direct-to-home medicine delivery becomes a credible option.

Keywords: e-commerce; FMCG; simulation; last mile logistics; field data analysis;

1 Introduction

With the increasing importance of e-commerce for the UK's retail landscape established mass grocery retailers are now expanding their online presence, leveraging on delivery to win online retail market shares (BMI, 2017). E-commerce sales in the UK amounted to £554 billion in 2015, 40% of which were received through a website (Prescott, 2016). Online sales of product categories such as food and beverages, personal and household care, health and beauty accounted for circa £1.9 billion, whereas £19.8 billion sales took place through convenience stores and £3.6 billion through pharmacies and drug stores (Market Line, 2017).

Given the changing nature of the UK omnichannel landscape, this paper introduces recent research on the economic viability of last mile (LM) delivery in different geographies in the UK. A 'cost-to-serve curve' conceptualisation is proposed, linking the economic viability of last-mile delivery to such indicators as market penetration and population density for a given geographical location to distinguish areas which are either economic or uneconomic to serve, or might become

Table 1 –Overview of exemplar literature (non comprehensive)

Refeference	*Application*			*Boundaries*			*Methodology*		
				Downstream		Upstream			
	e-commerce	Logistics	Pharma /health	B2B	LM		OPT/ SIM	SI/ DEA	SRV/ CSR
Pagès-Bernaus et al (2017)	●			●			●		
Shapiro (2007, Ch.9)	●				●		●		
Harrington et al. (2016)	●				●				●
Sultanow et al. (2016)	●		●		●				●
Park et al. (2016)		●			●		●		
Wygonik and Goodchild (2016)		●			●		●	●	
Longo (2012)		●		●			●		
Zhuan et al (2008)			●	●			●		
Chahed et al. (2009)			●		●	●	●		
Aized and Srai (2014)			●		●		●		
Gevaers et al. (2014)		●			●				●
Farahani and Elahipanah (2008)		●		●		●	●		

Abbreviations: B2B: Business to business; LM; last mile; OPT/SIM: optimisation (network or route) or simulation; SI/DEA: statistical inference or data envelopment analysis; SRV/CSR: survey or case study research

economic to serve depending on whether shared platform between retail partners and alternative end-user payment models are adopted. To achieve this, a multi-level modelling approach driven by real-world data that are likely available to the business is outlined. The suggested approach includes estimating typical catchment areas and basket composition for location and market penetration scenarios; and setting out likely economics of supply for a set of post-codes.

Findings from the research proposed in this paper are a preliminary step towards building a 'map' of the UK where LM delivery and e-commerce home delivery in its current format becomes a viable option considering 'location' factors which can act to reduce costs, such as penetration of e-commerce and density of e-commerce providers in a given geography. While based on a case in FMCG, the findings of this research can be of interest for other sectors. In particular, similar developments towards omni-channel, more customer centric structures begin to be observed in the pharmaceutical industry , including direct-to-patient delivery.

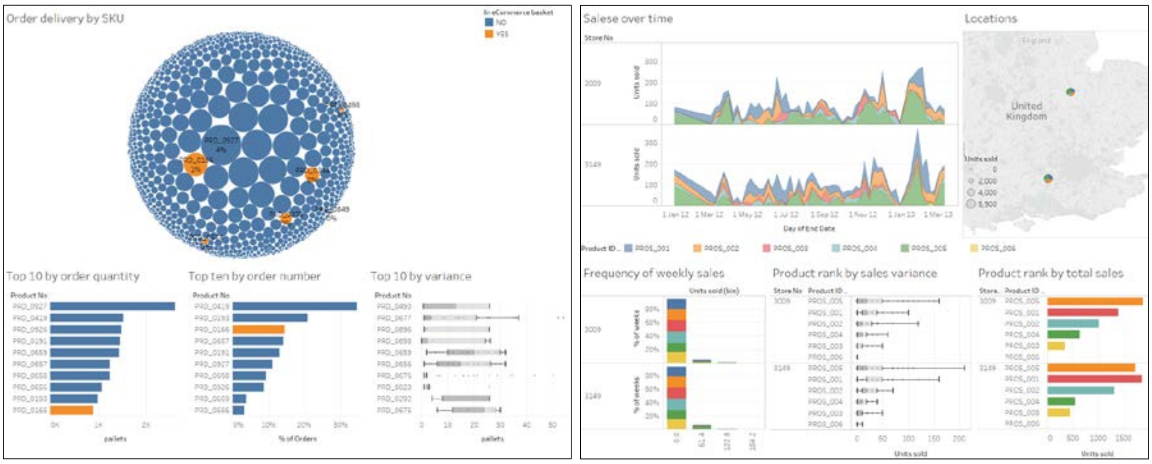
In the following sections, the methodology and data used in this research are introduced, followed by an illustration of preliminary findings generated through simulation for selected locations. Discussion of these finding leads to a conceptualisation of possible 'economic regions'. A closing section wraps up the research presented in this paper while pointing out its limitations and linking to future research agendas.

2 Material and methods

In principle, the analysis of near real-time streams of electronic data and the underpinning IT infrastructures play a major role in modelling e-commerce based supply chains – see Siddiqui and Raza (2015) for an overview. Without aiming to a detailed account of the literature, Table 1 summarises selected references with a focus on areas of application, boundary and scope of the analysis, and methodology employed. In the literature considered here, computer-aided supply chain optimisation and simulation is widely implemented to generate transportation optimisation scenarios for selected "Milk runs" (Downstream) or, alternatively, inventory profiles and location decisions for distribution facilities (Upstream). However, seldom is the analysis of real-world data available to businesses made explicit and included explicitly in the broader optimisation/simulation exercise, especially when it comes to understanding the demand signal coming from online sales.

This limitation is addressed here by outlining an approach to evaluate alternative supply network configurations, and the associated cost-to-serve informed by real-world data on an exemplar basket of fast moving consumer goods (FMCG) sold though e-commerce retail channels in the UK provided by industry partners. The proposed approach is based on the general principle of minimum cost flow through a network (Williams, 1998), and can be summarised as follows:

1. Harmonisation, visualisation and descriptive/prescriptive analysis of business data excerpts on both e-commerce retail (downstream) and B2B delivery (upstream) activities;
2. Set up and implementation of a computer-aided supply chain simulation to generate:



(a) *(b)*

Figure 1 –visual summaries of B2B order data (b); and e-commerce sales for selected locations (b)

- 2.1. Transportation optimisation scenarios for selected "Milk runs" (Downstream);
  - 2.1.1. Household order location simulation (Model A);
  - 2.1.2. Simulation of basked/order composition (Model B);
  - 2.1.3. Transportation cost and service time estimate (Model C);
- 2.2. Inventory profiles for selected delivery operations (Upstream).

3. Set out likely economics of supply for exemplar locations whereby cost-to-serve is conceptualised as a function of market penetration, for a given geographical location.

3 Findings

3.1 Time series data analysis

Preliminary analysis was carried out on the following data excerpts provided by industrial partners: 1) over 60k itemised B2B orders delivered from an FMCG manufacturer's DC to a mass retailer's DC involving circa 1.1k stock-keeping units (SKUs) over 2 months; and 2) 97.7k aggregated weekly online sales across 300+ stores over 53-weeks for 6 pre-selected SKUs.

A dashboard-type visual summary of these data is shown in Figure 1. The visualisation of B2B delivery data in Figure 1a contextualises the relative importance of those SKUs for which online sales data is collected in relation to the broader context of inventory replenishment. Although not shown here, further analysis was performed to investigate co-occurrence and associations in B2B order delivery leading to discovering patterns in purchasing behaviour. Conversely, due to the weekly aggregation of online sales data, and the focus on pre-selected SKUs rather than on individual orders, the analysis carried out at the store level could only be descriptive in nature. Figure 1b provides some insights into the distribution and variability of online sales, for use as constraints in the modelling activities described below. For illustrative purposes, Figure 1b shows only 2 geographies, as these will be further analysed in the following sub-sections.

3.2 Model A: simulation of customer 'catchment' area

In the absence of detailed household data associated with online sales, such as customer locations, the customers 'catchment' for a retail store has to be estimated by simulation as described, for example, in Schätter (2016, Ch. 6).

The proposed model to simulate order generating locations ('model A') is summarised pictorially in Figure 2. The top half of the figure shows publicly available data from the Office of National Statistics (ONS) on population density and postcodes with reference to Lower-layer Super Output

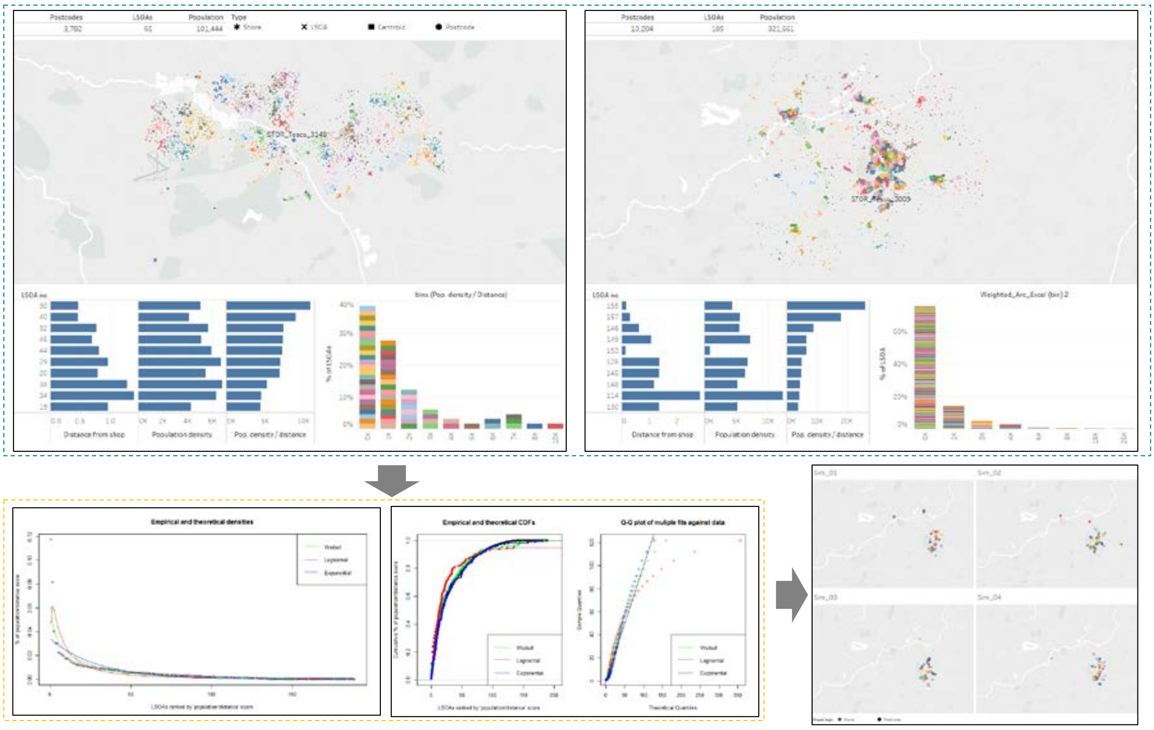


Figure 2 – simulation of potential order-generating locations based on public domain data on population density and distance from store in selected geographies combined with assumptions on market penetration

Areas (LSOA) for the chosen locations (ONS, 2016). In the chosen locations there are, respectively, over 3k and 10k postcodes which are serve here as a proxy for order generating locations. The two bar charts underneath each map in the top half of Figure 2 show, respectively the ranking and binning of each location's LSOA according to the ratio between population density and geographical distance from a given store. While the former is given, the latter can be estimated either as the centre of gravity of the postcodes assigned to it, or by computing an equivalent number of centroids/clusters means independently of the original attribution provided by the ONS. This information is used to fit and sample a theoretical density function so that those LOSAs that score higher are more likely to be generating an order (bottom-left corner in Figure 2). Individual postcodes within a sampled LSOA are assumed to have the same probability of generating an order.

The bottom-left part of Figure 2 shows the output of 4 iterations whereby 19 postcodes are randomly sampled, representing approximated order-generating locations, and hence drops/stops in the LM delivery. A rough estimate of the number of ordering locations/drops is based on the following assumptions: a bi-weekly order frequency for the SKUs of interest; and a market penetration of 7%. For the location with 3,782 postcodes (upper left part of Figure 2) this leads to the following estimate: $\text{drops per day} = \text{households} \times \frac{\text{market penetration}}{\text{order frequency}} = 3{,}782 \times \frac{0.07}{14} \cong 19$. For the location with 10,204 postcodes (upper-right part of Figure 2) the same reasoning leads to estimating 50 drops per day. However, this number is unrealistically high considering that the variability of daily sales for each SKU derived by crude approximation form the weekly data in Figure 1b is roughly the same for both locations despite the considerable difference in terms of number of postcodes.

In the absence of better evidence, the choice of keeping or rejecting the assumption that areas with higher population density and closer to a point of sale are more likely to generate an online order remains a subjective one.

3.3 *Model B: simulation of basket composition*

In the previous sub-section, a number of order-generating locations to be visited during a LM delivery was simulated based on market penetration, population density and distance from a given store. The next step is to estimate the order composition to be delivered at each drop ('model B').

Model B is necessary in the absence of details on orders and SKUs per order in the e-commerce sales data. Given that only aggregated weekly sales data are available for a pre-selected number of SKUs, the modelling process is as follows. First, express the variability in daily sales for each SKU as triangular distributions. In the absence of daily data the min, max, and most likely value in each distribution were roughly approximated as 1/7 of the min, max and median weekly sales (Figure 1b). Next, each distribution is to randomly generate a daily sales cap by SKU consistent with actual sales data. The daily sales cap thus generated serves as a constraint in setting up a binary programming model seeking to allocate an amount between 0 and 1 of each SKU to each one of the baskets (drops) determined by model 'A' as described earlier. The procedure is iterated four times to match the four random scenarios shown in Figure 2.

Key assumptions in model B is that the variability in daily sales can be obtained from weekly data by linear approximation; and that it is admissible to build a fictitious basket consisting only of those 6 SKUs for which data is available. The value of reasoning at the individual SKU level rather than considering an unspecified basket, is twofold: it provides a link between downstream LM delivery model and upstream inventory simulation models; and it provides useful insights for the allocation of compartmentalised space in temperature-controlled vehicles.

3.4 Model C: vehicle routing and journey cost estimation

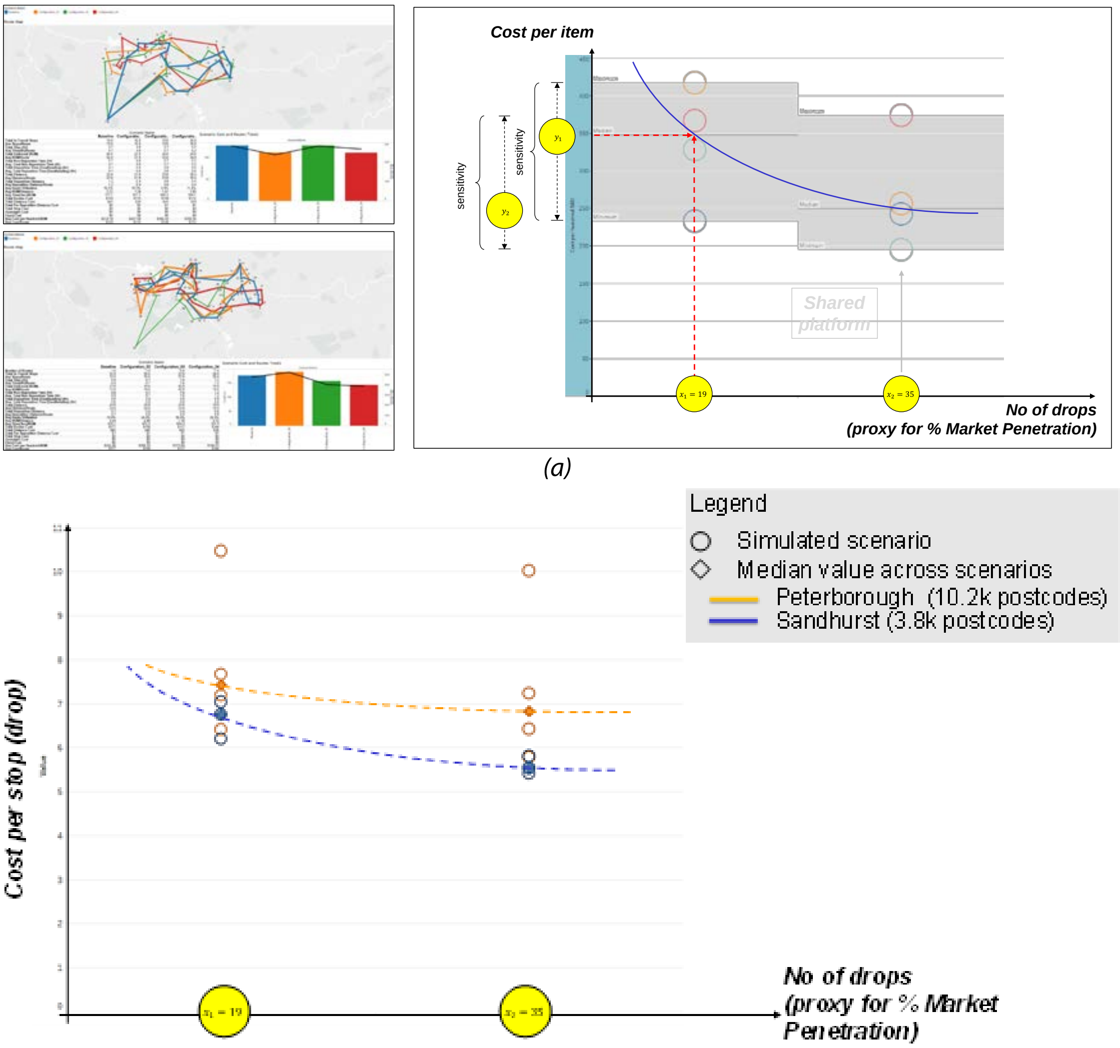
The insights generated by models A (drops/site locations) and B (shipments over time) provide inputs to the LM transportation optimisation problem referred to here as 'model '.

Like most vehicle routing problems (as in some references in Table 1), model C can be reduced to a so called travelling salesman problem, that is, a problem of finding the order in which a number of customer locations must be visited so that the total distance covered is minimised. The travelling salesman problem can be interpreted as a problem of finding the shortest Hamiltonian circuit – a closed walk that traverses every vertex exactly once – of the corresponding complete weighted graph were, vertexes represent customers locations and weights are distances between any two locations (Deo, 1974). In practice, this problem is typically formulated as an integer programming model (Williams, 1998).

For the specific case considered here, 'models C' was set up and implemented using off-the-shelf software (Llamasoft Supply Chain Guru 8.4). The software, which is treated from now on as a 'black box', requires input data summarised below (each store is modelled as a separate problem):

- Products: the correspond to the pre-selected SKUs, and related attributes such as monetary value/price (to calculate inventory holding costs and revenues) as well as weight and volume (to calculate space utilisation and weight/cubic-based transport costs);
- Sites: these correspond to the locations of relevant stores and order-generating (customer) zones as per 'model A' described earlier;
- Shipments profile: these correspond to the order quantities generated for each product/customer/scenario combination through 'model B'. For the sake of simplicity, all the orders were assigned the same order date/time (assuming daily online orders received within a certain time window are aggregated), while the due date is specified as the next day;
- Transportation Asset: a range of features associated with the available vehicle or vehicle fleet. Exemplar value for such features, either physical (e.g. capacity constraints) or economic (e.g., cost per mile) were estimated based on interviews with industrial partners as well as by analysis of the academic literature, as specified elsewhere (Doetsch, 2017).
- Transportation policies: the software tool provides a range of pre-set policies. Relevant policies for the case considered here include "Full Truckload", which aggregates product bundles for shipment until the transportation asset's capacity in terms of weight (or cubic volume) fill level is reached; and "Pooled Inbound", which consolidates into the same transportation asset various product bundles inbound into a specific site (Watson, Lewis, Cacioppi, and Jayaraman, 2013; Ch, 6).

Figure 3(a), upper-left corner, shows the optimal last-mile routes segments obtained from running the transportation optimisation for one of the selected locations, in four scenarios



(a)

(b)

Figure 3 – Cost-to-serve simulation for exemplar locations

corresponding to different order-generating locations/order compositions. The range of key performance indicators obtained in each scenario with regards to the delivery, as well as the overall cost and service time are also shown.

The overall modelling procedure is repeated assuming an increased market penetration leading to a shift of the estimated number of drops from 19 to 35 (Figure 3(a), bottom-left corner). The sales increase might be thought of as either an increase in the daily sales volume for the store considered, or as the result of multiple stores pooling together online orders and delivering through a shared vehicle/fleet. The simulated LM cost-to-serve per item is plotted against the number of drops resulting from different levels of market penetration in the left-hand side of Figure 3(a).

The cost per journey is consistently higher in the 35 drops case (which appears reasonable given the largely linear relationships between the input cost parameters and the distance or duration of delivery). The median cost 'per stop' or 'per item' seems to decrease as market penetration increase (respectively £6.7 to £5.5/drop, and £347 to £249/hundreds SKUs). For the sake of conceptualisation Figure 3(a) shows a hypothetical curve passing through two points corresponding to the median costs per hundreds SKU obtained.

The results obtained through the procedure described above for both stores locations are plotted in Figure 3(b). It can be noted that, the approximated cost-per-drop curve plotted for the

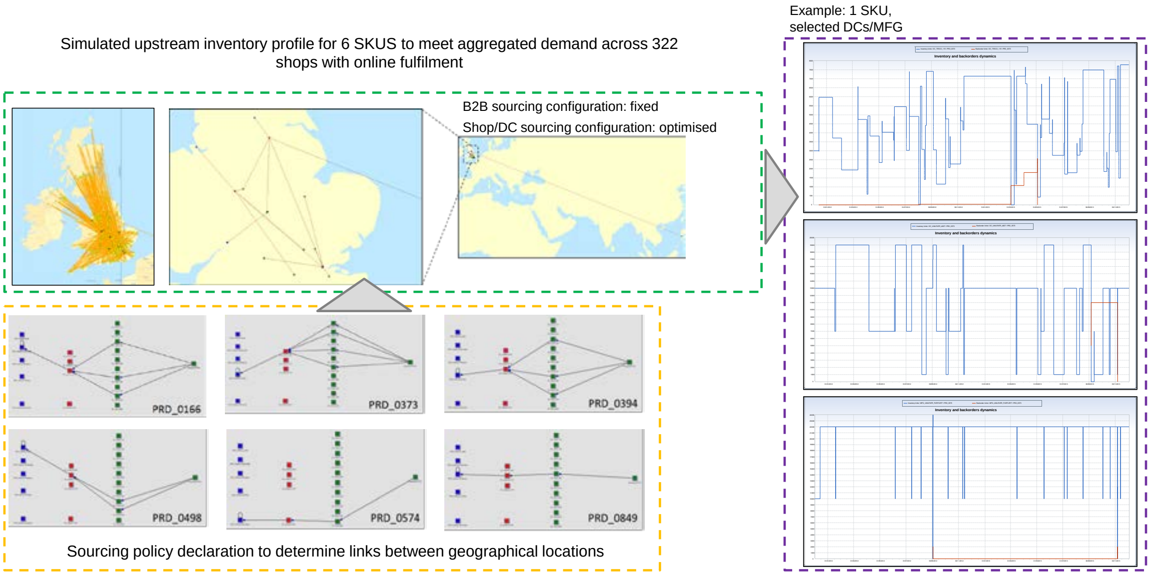


Figure 4 – data-driven upstream inventory simulation linked to online demand

location with the higher number of postcodes lies systematically below the same curve plotted for the location with lower number of postcodes.

It is worth noting that the following simplifications were made in model C at this stage, despite the practical relevance of the aspects involved: 1) the number of drops is equivalent to the number of vehicle stops; 2) additional trips due to missed deliveries are not considered.

3.5 Upstream inventory profile simulation

In simulating of LM delivery cost described above, online sales at individual stores acted as the point of focus. The aggregate online sales across all the stores with online fulfilment provide a total demand signal that can be used to simulate the inventory profiles at multiple echelons of the upstream distribution system – from the manufacturer's facilities and DCs through to the retailer's DCs supplying individual stores with the relevant SKUs.

Single sourcing relationships between manufacturing sites, DCs and retailer's DCs could be identified from documentation shared by industrial partners, for each pre-selected SKUs considered in the downstream LM delivery model. Multiple sourcing relationships were assumed between the retailer's DCs and stores. The resulting network structure is shown on the left-hand side of Figure 4. To select between multiple sourcing relationships, where necessary, a network optimisation procedure based on transportation costs (see e.g. Watson et al., 2013 Ch. 6) was implemented using Llamasoft Supply Chain Guru 8.4. The transportation cost coefficients and policies were estimated using business-specific data.

To simulate how the supply chain dynamically responds to customer orders being placed over time horizon, lead times in manufacturing and distribution were introduced (e.g., business' estimates on the duration of picking, loading and unloading operations), and the inventory policies at each echelons were declare (mostly subjective assumptions about reorder points, replenishment levels and initial inventories for each SKU). An exemplar inventory profiles generated by simulation for one of the SKUs at the retailer's DC, at the manufacturer's DC and at the manufacturing site is shown on the right-hand side of Figure 4.

4 Discussion

The findings illustrated in the previous section provide some preliminary evidence linking the economic viability of LM delivery to indicators such as share of online sales and catchment served in a given geography. The findings where contrasted with an initial conceptualisation whereby:

- LM 'cost-to-serve' is a decreasing function of market penetration;
- High population density locations being economic to serve as van loading journey drops enable to achieve target cost for a given market penetration;
- Low population density locations are either uneconomic, or require shared platform between retail partners, or end-user payment models

While only a limited number of market penetration scenarios was considered, the results seem to be partly support by the approximated curves plotted in the bottom-half of Figure 3. This suggests that cost-to-serve is also impacted by the following 'location' factors which can act to reduce costs (i.e. lower the profile of the cost – penetration curves):

- High levels of e-Commerce sales in a given geography – Penetration of e-Commerce);
- Low density of e-Commerce providers – enabling higher e-market share for the service provider

Furthermore, results suggest that as well as having cost benefits with increased market penetration and/or increasing the number of drops per journey, as might be predicted, there appears significant potential to narrow the spread of cost variability for a given market penetration by influencing the available locations/time/price options to online customers.

Conceptualizing these possible scenarios leads to the following possible 'economic regions' as

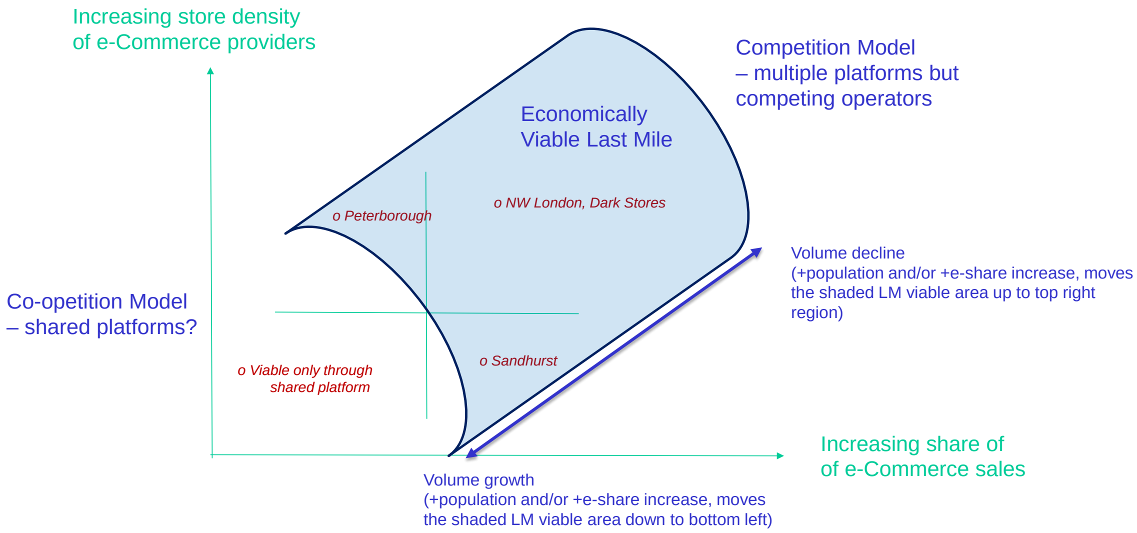


Figure 5 – Impact of store density on the economics of last mile delivery

shown pictorially in Figure 5. The proposed conceptualisation takes into account the changing nature of UK omni-channel landscape, potentially leading to a 'map' of the UK where LM Delivery and e-Commerce home delivery in its current format becomes a viable option. To support the proposed conceptualisation, analysis of representative locations for each quadrant in Figure 5 is required.

5 Conclusion

This paper uses visual data analytics and simulation to evaluate the economic viability of last-mile delivery (downstream), and the upstream repercussions in terms of multi-echelon inventory profiles for a case in online retail of fast moving consumer goods in the UK. With the aid of a numerical example underpinned by industry data, a three-stage, pragmatic simulation approach is outlined. Results generated at each modelling stage with the aid of off-the-shelf tools were illustrated, including: 1) order-generating locations; 2) basket composition; and 3) last mile delivery cost.

The findings suggest a 'cost-to-serve curve' conceptualisation linking to business-specific indicators, such as share of online sales and catchment served in a given geography, as well as sector-level indicators for the same geography, such as e-commerce penetration and retailers density. The proposed approach has a number of limitations. Some modelling steps were necessary in the absence of data on households order detail, introducing crude assumptions regarding for example daily demand for a given SKU. Such models may become unnecessary as further, more granular data becomes available. Using off-the-shelf tools, the inventory simulation and transportation optimisation algorithms were treated as a black box. Finally, several activities upstream of an e-ecommerce vehicle routing system were left outside the boundaries of the proposed analysis (for example, the electronic information exchange in terms of Web-based order entry and data management). Despite its limitations, this work provide initial impulse to research aimed at conceptualizing through simulated scenarios a 'map' of the UK where LM Delivery and e-Commerce home delivery in its current format becomes a viable option.

Future research will explore options for reducing variability in cost-to-serve, perhaps by influencing consumer behaviour through reduced locations and time options and price incentives. Extensions of the application of the modelling approach suggested in this paper should also be explored, as it could support the evaluation of options for similar developments in other sectors, such as direct-to-home medicine delivery.

Acknowledgment
Support from the Advanced Manufacturing Supply Chain Initiative (Grant No. 35708-233529, ReMediES – Reconfiguring Medicines End-to-end supply) is gratefully acknowledged

References
[BMI] Business Monitor International Ltd. (2017). United Kingdom Consumer and Retail Report - Q4 2017. *BMI Country Industry Reports*. Retrieved from https://search.proquest.com/docview/1923824358?accountid=9851

Aized, T., and Srai, J. S. (2014). Hierarchical modelling of Last Mile logistic distribution system. *The International Journal of Advanced Manufacturing Technology*, *70*(5), 1053–1061.

Chahed, S., Marcon, E., Sahin, E., Feillet, D., and Dallery, Y. (2009). Exploring new operational research opportunities within the Home Care context: the chemotherapy at home. *Health care management science*, *12*(2), 179-191

Deo, N. (1974). *Graph theory: With applications to engineering and computer science*: Prentice-Hall.

Doetsch, S. B. (2017). *Exploring structural changes to the downstream UK Pharmaceutical Supply Chain* (MPhil). Cambridge, UK: University of Cambridge.

Farahani, R. Z., and Elahipanah, M. (2008). A genetic algorithm to optimize the total cost and service level for just-in-time distribution in a supply chain. *International Journal of Production Economics*, *111*(2), 229–243.

Gevaers, R., Van de Voorde, Eddy, and Vanelslander, T. (2014). Cost Modelling and Simulation of Last-mile Characteristics in an Innovative B2C Supply Chain Environment with Implications on Urban Areas and Cities. *Eighth International Conference on City Logistics 17-19 June 2013, Bali, Indonesia*, *125*, 398–411.

Harrington, T. S., Singh Srai, J., Kumar, M., and Wohlrab, J. (2016). Identifying design criteria for urban system 'last-mile' solutions – a multi-stakeholder perspective. *Production Planning and Control*, *27*(6), 456–476.

Longo, F. (2012). Sustainable supply chain design: An application example in local business retail. *Simulation*, *88*(12), 1484–1498.

Market Line. (2017). [Dataset] *Market Data Analyitics: Value by distribution channel*. Available at http://advantage.marketline.com.proxy.jbs.cam.ac.uk/PremiumTools

ONS [Office for National Statistics]. (2016). [*Dataset] Lower Super Output Area Population Density, Mid-2015.* Available at: *https://www.ons.gov.uk*

Pagès-Bernaus, A., Ramalhinho, H., Juan, A. A., and Calvet, L. (2017). Designing e-commerce supply chains: a stochastic facility–location approach. *International Transactions in Operational Research* [in press] DOI: 10.1111/itor.12433

Park, H., Park, D., and Jeong, I.-J. (2016). An effects analysis of logistics collaboration in last-mile networks for CEP delivery services. *Transport Policy*, *50*, 115–125.

Prescott, C. (2016). *The value of e-commerce sales and the level of use by UK businesses of information and communication technologies*. Available at https://www.ons.gov.uk/

Schätter, F. (2016). *Decision support system for a reactive management of disaster-caused supply chain disturbances*. Karlsruher: Karlsruher institut für technologie (kit) publishing

Shapiro, J. F. (2007). *Modeling the supply chain* (2nd ed). *Duxbury applied series*. Belmont, CA: Thomson-Brooks/Cole.

Siddiqui, A. W., and Raza, S. A. (2015). Electronic supply chains: Status and perspective. *Computers and Industrial Engineering*, *88*, 536–556.

Sultanow, E., Chircu, A., and Chircu, F. (2016). E-commerce Adoption: A Revelatory Case Study in the German Oncology Pharmaceutical Supply Chain. In *Proceedings of the Conference on Information Systems Applied Research* (ISCAP), Las Vegas, NV (USA).

Watson, M., Lewis, S., Cacioppi, P., and Jayaraman, J. (2013). *Supply chain network design: Applying optimization and analytics to the global supply chain*. Upper Saddle River, N.J.: FT Press.

Williams, H. P. (1998). *Model building in mathematical programming* (3rd ed.). *A Wiley-Interscience publication*. Chichester: Wiley.

Wygonik, E., and Goodchild, A. V. (2016). Urban form and last-mile goods movement: Factors affecting vehicle miles travelled and emissions. *Transportation Research Part D: Transport and Environment*. [in press] doi:10.1016/j.trd.2016.09.015

Zhuan, W., Qinghua, Z., bo, Y., and Wenwen, H. (2008). 4/R/I/T distribution logistics network 0-1 programming model and application. *Computers and Industrial Engineering*, *55*(2), 365–378.