Is competition sufficient to drive observed retail location and revenue patterns? An agent-based case study.

by

Jiaxin Zhang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Geography and Environmental Management

Waterloo, Ontario, Canada, 2020

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

This thesis consists of three chapters all of which I have been the lead author. While my supervisor (Dr. Robinson) has provided comments and edits on Chapters 1 and 3, he has taken a more collaborative role as coauthor on Chapter 2. As lead author of Chapter 2, I conceptualized the study design, conducted all computer coding and model creation, conducted all data analysis and reporting, wrote the majority of the texts, and created all figures and tables.

Abstract

Agent-based models (ABMs) have been widely used to represent and investigate complex systems and are a contemporary modelling approach used in the study of land-use and land-cover change. While many ABMs have been constructed to address research questions associated with residential land development and human choices, agricultural land transition and farmer decision-making, and transportation networks and planning, less attention has been given to improving our understanding about the drivers and agent behaviours associated with commercial and retail competition, which subsequently affects land-use change. Among existing ABMs that represent the retail system, the focus has been on understanding consumer behaviours, but the inclusion of the store competition is lacking, and most retail competition models still use a top-down modelling framework. The thesis herein provides a new contribution to retail competition literature through the development and use of a retail-competition agent-based model (RC-ABM). Utilizing previous empirical research on consumer expenditures and retail location site selection, competition for home-improvement expenditures is simulated within the home-improvement retail system in the Region of Waterloo, Ontario, Canada. Results exhibit a high level of alignment between the RC-ABM and a traditional Location-Allocation Model (LAM) in estimating a market capture and store revenue acquisition. In addition, while modelled competition itself cannot reproduce the observed spatial pattern of home-improvement stores in our study area, results from the model can be used to identify path dependencies associated with retail success generated by competition and factors affecting retail store survival. Lastly, the presented RC-ABM provides the potential to enrich future land-use and land-cover change models by better representing commercial development.

Acknowledgements

It would have been impossible to complete this thesis without the support of my advisors, colleagues, families, and friends.

I would like to thank Dr. Derek Robinson, who has dedicated a countless amount time to help me with all my academic work. He is more than a great supervisor, but also a good friend who has always encouraged me and given me confidence to overcome challenges when I experienced low points during the MSc process.

To my committee members Dr. Peter Deadman and Dr. Markus Moos, thank you for reading my thesis and providing professional feedback, and helping me to finalize this thesis.

To Prof. Dawn Parker and Dr. Yu Huang, thank you for inspiring me with your expert knowledge and helping me with the first version of my Retail Competition Agent-based Model.

To my colleagues in the Geospatial Innovation Lab, Bo, Ben, Tanya, Omar, Jenny, Erin, and Juan, thank you for all the advice, discussions, and support. You made life much easier!

To my parents and all other family members, thank you for your endless love and support!

To all my friends in the University of Waterloo (especially Lingyu, Yupu, Yujun, Yiwei, Duo, Qihang, Qiutong, and Roy), thank you for your friendships and infinite support!

Table of Contents

List of Figures

List of Tables

Chapter 1 A Primer on Modelling Retail Location Patterns

The retail sector plays a critical role in most developing (e.g., India, Guruswamy *et al.* 2005) and developed (e.g., UK, Rhodes and Brien 2014) countries (Siebers et al. 2010). For example, retail constituted over seven percent of total US economic activity every year for almost 50 years before 2000 and over six percent annually from 2000 to 2014 (Hortaçsu and Syverson 2015). Similarly, in Canada, approximately 5.5% ($97.8 billion) of the country's GDP ($1.79 trillion) was generated by the retail sector in August 2018 (Statistics Canada 2018a). In addition to its contributions to GDP, retail has been a leading sector of employment. During the 1980s, retail industries contributed over 12% of the total employment in the US, notwithstanding a slight decrease in the 21st century, it still occupies a large proportion (over 11%) of the US' total employment (Hortaçsu and Syverson 2015). In Canada, retail plays a more prominent employment role with about 15% of the total employment among all industrial sectors provided by its retail from 2015 to 2019. The importance of the retail sector to a nation's economy (GDP) and social stability (the employment rate) should not be underestimated.

Despite the valuable economic and social role of retail, retail is a vulnerable commercial sector, especially when encountering Black Swan (e.g., economic crisis, Davies 2000) events. For example, in Canada, the global financial crisis (started in 2008) led to a six percent decline in the number of retail stores from 2008 to 2011 (Statistics Canada 2020). While the number of stores increased thereafter, as of 2012, the number of stores remained less than existed prior to the financial crisis (Statistics Canada 2020). Although the Canadian retail sector experienced a sales growth from 2012, the rate of the growth was unstable, which fluctuating from 2% to 7% (CBRE 2019). In particular, Canada's annual retail sales growth rate in 2019 reached its lowest point (2.1%) since 2009. According to these data, retail plays a prominent role in the economy and social fabric of a country, but its vulnerability suggests that a more rigorous investigation into the conditions that derive success is needed. In most cases, these conditions (e.g., accessibility, market share, and competition) are a function of the retail store location. Despite a variety of approaches used to select a store location, there are novel approaches that have yet to be adopted that could benefit a variety of actors (e.g., retail companies, local economic development teams).

Historically, retail site selection commonly used heuristics (i.e., "rule of thumb") (Hernandez and Bennison 2000) embedded in corporate or real-estate agent experiential knowledge. However, more recently, efforts have been taken to formalize site location strategies to avoid failure (Moore 2005). Formalization of location decisions has become essential since several new types of retail-store formats have evolved to capture market share and ensure success (McArthur et al. 2016). Among revenue streams, e-commerce has been occupying an increasing proportion of market shares in the retail business (Nisar and Prabhakar 2017). Although the evolution of retail site selection, formats, and revenue streams may result in various operational and strategic modes, there are two goals shared by all retailers: to launch a long-living profitable business (Ghosh and MacLafferty 1987; Colla and Dupuis 2002; Eisenmann 2006) and mitigate and avoid obstacles, which are derived from both stores' internal and external environments, to achieve that goal.

Internal environmental challenges include failures caused by the store's self-management strategies, such as the business mode, staff hiring and training management, inventory or supply chain management, and so among others (e.g., Ghosh and MacLafferty 1987). A small failure with choosing a proper self-management strategy might lead the company to crumble from within (Gaskill et al. 1993). Complementing internal challenges are external environmental challenges that may impose negative pressures on a store's survival. These challenges are mainly generated by unpredictable policy changes by local governments, competitors' actions, the complex socio-demographic environment of the local markets, and emergent incidents among others (Hernandez et al. 1998; Hu and Ansell 2007). For example, many "big brand" retail companies experience expansion failure in new countries despite success or even dominance in their origin countries, e.g., Target's failure in entering the Canadian market (Hoffman and Gold 2015; Yoder et al. 2016), Best Buy's failure in entering the Chinese market (Feng 2013).

Given that the customer is the fundamental source of profit for retailers, understanding customer behaviour is a major concern for retail companies (Verhoef et al. 2009) and customer behaviors are sensitive to a store's external environments (Chen 2015). Store or brand loyalty may change due to competition-oriented actions, like new special offers (Clarke et al. 2004). Completion of a new highway may lead a customer to switch his first-ranked store to a competitor due to a change in travel time. A new store entering the local market may cause many existing stores to lose customers, especially those who live near the new store (Clarke 2000). Overall, for retailers, it can be inferred that most challenges derived from the external environment will ultimately be rendered into the stress of keeping its current customers and attracting new customers from its competitors.

On the customer side, choosing a destination store to purchase commodities is usually not simple and straightforward. Literature shows a multitude of factors affecting store evaluation and destination choice (Ghosh and MacLafferty 1987) such as: 1) customers are more likely to shop at a store which has an assortment and variation in quality of merchandise (Koelemeijer and Oppewal 1999.); 2) the overall atmosphere (Sharma and Stafford 2000; Gowrishankkar 2017), and aesthetics of a store can influence customers' expectations and shopping experience; 3) the service mode of a store comprising the attitude of staff toward customers and the quality of post-sale services are important to the perception of store image by customers (Gagliano and Hathcote 1994; Sharma and Stafford 2000); and 4) pricing, plays a critical role where for some target markets the lowest price dominates purchasing decisions, while for others it is not always the case (Grewal et al. 2009) such that some brands' success is derived by selling higher-quality goods at higher prices (Moore and Carpenter 2006). Nevertheless, *ceteris paribus*, most customers would be attracted by 5) location and convenience. A more accessible store can save a customer commuting time, commuting cost, and create a more convenient shopping experience (Grewal et al. 2009; Swoboda et al. 2013). Among these five factors, price and location are regarded as the most important drivers of retail competition (Ghosh and MacLafferty 1987). However, location has a lower fault-tolerance given that the decision is relatively irreversible, involves sunk costs, and changes carry negative social consequences.

On the one hand, a store's location is a huge and risky one-off investment from the aspect of both money and time. On the other hand, the location strategy is also a prerequisite to other retail strategies. A store's location also determines the socio-demographic characteristics (e.g., purchasing power and disposable income) of its service area, which are essential references to strategies associated with merchandise assortment and quality. Additionally, a store's location also determines its competition. Hence, it can be said that the location "unintentionally" helps the store to select both its existing customers and competitors.

Furthermore, the location is also an immutable factor to a store when trying to respond to future changes in its nearby competition context. When a competitor locates a new store near an existing competitor, the existing competitor can launch any type of strategy such as lowering prices and enriching merchandise categories, but it is difficult for a large store (e.g., Walmart) to relocate. Contrary to other strategies that are revisable, if a wrong choice of location has been initiated, its negative impacts will be perpetual until the close of the store. Consequently, the location could be treated as one of the most important factors to be finalized.

To reduce the risk of locating a store, retail companies typically use a series of methods to evaluate possible storing locations (Ghosh and MacLafferty 1987; Ken and Simmons 1993). For each possible location, the evaluation should not only help retailers to forecast a store's operational performance using available data with consideration of influences exerted by some possible future changes within the local market. It is almost impossible to make an absolutely accurate prediction about the future in any scientific field (Scriven 1959) including retailing. However, by making some assumptions, models can provide a powerful tool to simulate and predict plausible futures (Schichl 2004; Page 2018).

1. Traditional Methods and Models for Retail Site Selection

Retail site selection initially involved a simple "checklist" approach (Hernandez and Bennison 2000). Popularized in the 1960s (Ghosh and MacLafferty 1987), checklists consist of multiple factors used to evaluate (or rank) the anticipated retail performance of a potential site(s) (Ciari et al. 2008). While the factor values of interest (e.g., target customer) may vary by retail brand and sector, factor selection should be rigorous, accurate, and the data collection supporting the factor analysis is often labor- and time-consuming. Moreover, although the suitability of potential sites can be evaluated, the potential profit of each site cannot be predicted using a checklist approach and therefore the approach lacks quantitative evidence for site selection.

Similar to a checklist, an analog approach is a simple and commonly used method (Clarke et al. 2003). An analog approach estimates the performance of a future store based on a reference (i.e., existing) store. The characteristics of a new site and projected new store are matched with existing stores with similar conditions and the existing store(s) revenue and profit are assumed to be achievable at the potential location (Applebaum 1966). The simplicity and low cost of an analog approach has made it prevalent in site location. However, if there are no reference stores with similar characteristics, or the influence of some factors are not well known, then predicted and actual revenues may have substantial differences.

The incorporation of data or information from existing stores to predict potential store performance using statistical models provides a quantitative approach not captured by checklists and analogues (Clawson 1974; Olsen and Lord 1979; Dibb and Simkin 1994). Statistical models assume that a store's sales performance (dependent variable) is related to measurable factors (independent variables) in a way that can be represented by mathematical equations (e.g., linear regression, multinomial relationship), and the coefficient of each factor can be estimated based on existing stores' data (Agrawal and Schorling 1996). Statistical modelling techniques can provide accurate predictions about store profits and to some extent become the foundation for more advanced models in the future (Hernandez and Bennison 2000). Nevertheless, statistical approaches are often incorrectly used and multicollinearity among independent factors require advanced approaches (Suarez Alvarez et al. 2007; Keener 2013). Furthermore, most independent variables used in statistical retail models are estimated from aggregated data, which embeds the assumption of consumer homogeneity across a population of consumers (Ghosh and MacLafferty 1987) and the issues with what is known as the ecological fallacy (Steel and Holt 1996).

The impacts of customers' spatial distribution and transportation cost on a store's location strategy became more prominent in the 1930s (e.g., Hotelling 1929). Then, in the 1960s, the first Spatial-Interaction Model (SIM) was developed, whereby modelled customers evaluated a store based on the store size and the distance to store (Huff 1964). However, SIMs are limited in the representation of consumer heterogeneity as customers are aggregated by location (e.g., census unit). For example, customers in the same census unit are assumed to have the same value of factors (e.g., location, disposable income).

The factors included in SIMs (e.g., Huff defines store size and customers' distance to store as factors) are used to calculate the utility acquired by a customer who patronizes a store. For each customer, each store will have an expected utility value and a customer will typically choose to patronize the store with the highest utility. By using a utility function, customers' shopping patterns in different geographical areas are represented in the SIMs, enabling models to better represent customer catchment areas and estimate profits (Yrigoyen and Otero 1998; Dramowicz 2005). From the 1980s, an increasing number of SIMs, including gravity models have emerged and are used in retail site-selection (Hernandez and Bennison 2000; Drezner 2009; Clarke and Birkin 2018).

A specific version of SIM, the location-allocation model, has made tremendous contributions to exploring location-oriented retail strategies (Goodchild 1984; Russell and Urban 2010). In the location-allocation model, customers are aggregately represented as demand points, which indicate the level of demand available for geographical zones. For instance, each Census Dissemination Area has an estimated demand for goods that can be calculated and represented by its centroid. In addition to demand points, the model uses feasible sites, determined *a priori* as potential new store sites. Stores' site selection processes are constrained by an objective function, which can be modified to maximize market share (Goodchild 1984), minimize the average distance between customers and stores (Baray and Cliquet 2013), or other objectives that affect model behaviour (Yeh and Chow 1996). Then, based on the distance relationships between feasible sites (stores) and demand points (customers), customers are allocated to stores to meet the requirements of the *objective function*. In the location-allocation model, the customer behaviors are determined by the analysis of real-world customer shopping patterns and needs, and represented by *allocation rules*.

Spatial interaction models, including the location-allocation model, have advanced the quantitative modeling capabilities used to evaluate retail location problems. However, like the statistical model, the heterogeneity of stores or customers are crudely at best represented by SIMs. In these models, store or customer behaviors are derived from analyzing existing and aggregated data. The heterogeneity of customer behaviors of different geographical areas is depicted by the utility function in SIMs, e.g., customers in different zones have heterogenous shopping patterns, but individuals within the same zone are all assigned homogenous (stationary) behaviors, which deviate from reality. Although the existence of SIMs, gravity model and location-allocation models have advanced the methods applied to site location and integrated spatial data, alternative models are needed to better represent a heterogenous population of consumers, stores, their interactions, and the non-stationary and evolving decisions of both actors.

2. Complexity Science and Agent-Based Modelling: a way forward

Most traditional retail model building mechanisms follow the logic of analyzing the causal mechanisms of a phenomenon from the macro-level patterns and observations (Crooks and Heppenstall 2012; Railsback and Grimm 2019). In many cases, customers' shopping patterns can be quantified and modelled based on empirical observations, which are then used to assess the outcomes of alternative scenarios. However, in these models, individual customer behavior is "constrained" by the macro-level phenomena and they fail to represent the interactions and adaptive behaviours of heterogenous individuals. For instance, in a utility-oriented model, although the utility of Store A is maximal to a customer, he might still go to Store B because Store A is out of stock of his target goods. Such circumstances occur frequently in the real world because the retailing system comprises a multitude of heterogeneity and interactions that cannot be represented using traditional modelling approaches (Macal and North 2009).

Like Stephen Hawking who stated, "I think the next century will be the century of complexity" (Wilensky and Rand 2015), more researchers have realized the importance of taking a *complex systems* perspective to scientific inquiry. A *complex system* is conceptualized as a system comprising multiple components that interact with each other and take adaptive actions in response to those interactions (Dooley 1996). This notion of complexity is often confused with and used interchangeably with *complicated*. However, in a complicated system, which also has connections between components, there is a high degree of independence across the whole system and the system can still function in the absence of certain components (Miller and Page 2009). In contrast, components in a *complex system* are more tightly connected and losing one component can lead the whole system to collapse (Miller and Page 2009). Considering each component of a system as one dimension, the *complicated system* is dimension reducible whereas the *complex system* is not. For example, in a Retailer-Consumer system, the merchandise categories of retailers can be simplified, e.g., decrease merchandise types from 5 to 4, but "the retailer" as the component cannot be abandoned, i.e., cutting off connections between retailers and consumers makes the system meaningless.

Another important concept in the study of complex systems is emergence (Nicolis and Nicolis 2012; Janssen 2020). Most traditional models use a "top-down" logic or approach (from macro-level to meso-level and micro-level) that derives relationships from an outcome (e.g., dependent variable) based on a set of drivers (e.g., independent variables). In contrast, the complex system uses the "bottom-up" logic to build models (Srbljinović and Škunca 2003), in which only individual components' behaviors and action rules are defined without considering the potential macro-level outcomes (e.g., dependent variable) or patterns of the model. By representing and enabling individuals to autonomously and dynamically perform actions and interactions, some unpredictable patterns might emerge from the model running (Miller and Page 2009; Nicolis and Nicolis 2012). For instance, the phenomena of segregation have been shown to occur simply by having households move to be near other similar households with a very small preference threshold (Schelling 1971). In our study of retail, these patterns may have different target markets and subsequently affect different retail location choices.

Guided by a "bottom-up" model building logic, Agent-based Modelling is widely acknowledged as an appropriate and advancing technique to represent and explore problems residing within or comprising a complex system (De Marchi and Page, 2014). The approach is appealing because it can be used to represent adaptive behaviors, landscape and actor heterogeneity, and can model analytically intractable outcomes like emergence (De Marchi, 2005). In an agent-based model (ABM), the elementary component is called an agent, which has been used to represent any type of entity for which some sort of decision-making process and capacity for interaction is assigned (e.g., an individual, a household, a retail store). Aligning with complex systems, ABMs distinguish themselves from other modelling approaches in their ability to represent 1) heterogeneity, instead of using a representative individual, an average individual, or a population of individuals as a whole, each agent can be assigned different characteristics that cause it to behave differently from other agents; 2) interaction, agents can make observations and choose to interact or avoid other agents; 3) adaptation, agents can take reactive actions and change their behaviour in response to feedback generated by interactions and changing circumstances; and 4) Autonomy, agent behaviours are not prescribed, instead agents are enabled to make decisions autonomously which endogenizes the processes of interaction and adaptation in a way that can not be represented with equation based modelling approaches.

In a broad view, agent-based modeling has been applied in solving problems under both natural science, social science, and engineering contexts (Epstein and Axtell 1996; Wilensky and Rand 2015). Based on literature the analytic data of the academic database (e.g., Web of Science), thousands of ABMs have been created to investigate topics ranging from biology (Gorochowski et al. 2012), chemistry (Pogson et al. 2006), ecology (Fitzpatrick and Martinez 2012), environmental science (Schwarz and Ernst 2009), urban planning (Arsanjani et al. 2013), and so on. However, within the land systems modelling community, it can be argued that the agent-based modeling community has not given the issue of retail location development enough attention as other land use and land cover (LULC)-oriented problems. For example, under the context of urban issues, plenty of models have been created to answer questions about residential choices and decision making (Huang et al. 2014), housing market dynamics (Gauvin et al. 2013), transportation system simulation and planning (Bernhardt 2007). Under the rural or suburban contexts, topics like agricultural land use and land cover change (Millington et al. 2008), farmer decision making (Ng et al. 2011), policy analysis and impacts (Lempert 2002), forest management and risk estimation (Spies et al. 2017) are also popular to be solved using the agent-based approaches.

In comparison, possibly due to the difficulty of data acquisition (Sturley et al. 2018), retail-related data are usually confidential, which may explain why few ABMs exist that simulate any aspect of the retail system. Among the few ABMs that do exist in the domain of retail are investigations into supply-chain management problems (Verdicchio and Colombetti 2001; He et al. 2013), pricing strategies (Yousefi et al. 2011) and stores' internal management strategies (Siebers et al. 2010). Several, ABMs were developed to gain insight into customer behaviors like shopping patterns (Schenk et al. 2007) and mobilities (Vanhaverbeke and Macharis 2011). However, there are few ABMs that represent retail competition, instead the focus lies on the customer side (e.g., Sturley et al. 2018) with retail store agents rarely integrated into problem (Ciari et al. 2008). Despite acknowledging the need to model the dual representation of customers and retail stores in a process-oriented way nearly two decades ago (e.g., Baydar 2003), progress has been minor compared to advances in other fields. Instead, many retail competition models still reside in the abstract level, i.e., theoretically discuss the questions using mathematical derivations or an oversimplified model with assumptive agents (e.g., Xie and Chen 2004; Miura and Shiroishi 2018).

Where retail competition has been modelled, using an ABM approach, involved the simulation petrol pricing behaviours in petrol retail system in the United Kingdom (Heppenstall et al. 2005; Heppenstall et al. 2007; Heppenstall et al. 2013). Using retailers as the major agent, Heppenstall's model (named as "the petrol model" in the later description) demonstrated the ability of an agent-based approach to replicate realistic retail price changing patterns caused by competition and to distinguish observed differences among urban-rural regions (Heppenstall et al. 2005). The presented research within this thesis seeks to fill this gap in retail competition literature. In Chapter 2, a retail competition agent-based model (RC-ABM) is presented that was constructed to 1) evaluate the degree of alignment between an ABM representation of competition and that represented using a spatial interaction model, 2) determine if the process of competition is sufficient to simulate observed patterns of retail location and revenue acquisition, and 3) are there conditions of path dependence, whereby locations comprise stores with consistent success or failure and are their clear drivers of these outcomes. After presenting the results of RC-ABM and situating it in the broader literature, Chapter 3 identifies potential future research directions that build off the contributions of the RC-ABM.

Chapter 2 Is competition sufficient to drive observed retail location and revenue patterns? An agent-based case study.

1. Introduction

Retail firms are a critical component of the economy. Not only does retail influence economic productivity and growth (Kosová and Lafontaine 2010; Holmberg and Morgan 2004), but it also stabilizes and catalyzes the employment market (Watson and Everett 1993; Miller et al. 2003). Over the past three decades, retail industries comprise approximately 6% of the U.S. GDP and over 10% of its labour force (Hortaçsu and Syverson 2015). A similar situation exists in Canada, whereby approximately 5.5% ($97.8 billion) of total Canadian GDP ($1.79 trillion) was generated by the retail sector in August 2018 (Statistics Canada 2018a). Complementing the economic contributions of retail is a high risk of failure, e.g., in the 1990s, over 1.5% of the U.S. retailers failed annually while 9.7% of GDP were contributed by them (Mcgurr and Devaney 1998).

Retail success is the product of a combination of business ideas, marketing and human resource management, pricing and merchandising, and business expansion and location selection strategies (Ghosh and MacLafferty 1987; Mou et al. 2018). Among these components of success, location is critically valued (Yang and Yang 2005; Reigadinha et al. 2017) since it is a prerequisite assessment processes (e.g., sales forecasting) and tied to the socio-economic and demographic characteristics (e.g., population, average income, transportation context, land value) of a store’s service area (Ghosh and MacLafferty 1987). Furthermore, location investments are fixed and sunk costs, whereby commercial land is long-term leased (Goodacre 2003) and construction and renovation costs cannot be fully recovered. Despite the criticality of location, companies underutilise analytic and computational approaches and instead prefer intuition or experience-based methods for site selection (Hernandez and Bennison 2000; Ladle et al. 2009).

Despite the lack of adoption of advanced site selection methods by retail companies (Hernandez and Bennison 2000), significant advances in site selection have been made beyond heuristics, checklists, and analog approaches (e.g., Durvasula et al. 1992). Regression (e.g., Ozuduru and Varol 2011), gravity (e.g., Merino and Ramirez-Nafarrate 2016) and spatial interaction models (e.g., Marceau and Benenson 2011), and machine learning models (e.g., Krause-Traudes et al. 2008) have all been used in association with retail site selection.

However, these approaches fail to represent many of the themes of complexity science such as heterogeneity, feedbacks, thresholds and tipping points, interaction, and emergence (e.g., Wilensky and Rand 2015).

Instead of representing store and customer heterogeneity, most traditional approaches assume homogeneous decision-making or behavioral rules interpreted from survey or census data (e.g., Nakaya et al. 2007). While mathematical-based and spatial-interaction models perform well at sales forecasting (e.g., Merino and Ramirez-Nafarrate 2016) or trade area estimation (e.g., Wang et al. 2016), they typically act on a static representation of the economy and society. In cases where time has been incorporated, research has focused on strategic pricing (Voss and Seiders 2003; e.g., Warner and Barsky 1995; Matsui 2018), product entry (e.g., Green et al 1995; Radas and Shugan 1998), or crisis response strategies (e.g., Claeys et al. 2013). A clear gap remains in dynamic site selection under changing competition landscapes (i.e., spatial patterns) with heterogeneous competitors (e.g., store sizes).

One approach capable of representing the themes of complexity science, including heterogeneous actors and landscapes, is agent-based modelling (ABM). The ABM approach is founded in the idea that by representing heterogeneous micro-level behaviours and interactions, one can simulate or grow system-level outcomes (Epstein and Axtell 1996). In an ABM, real-world actors are represented as computational agents in a one-to-one mapping (Rounsevell et al. 2012b), which facilitates participatory modelling (e.g., Zellner 2008; Voinov and Bousquet 2010) and knowledge transfer (Anzola 2019) in addition to scientific advances that cannot be achieved with equation-based modelling (Parunak et al. 1998). While ABM has been widely applied across a number of disciplines for decades (e.g., Gilbert and Troitzsch 2005, Gimblett 2002, Grimm and Railsback 2005, Robinson et al. 2007), its presence in retail research is almost nil. The exception being an ABM applied to understand the effects of pricing strategies in a competitive petrol-retail system (Heppenstall et al. 2005).

We take a step toward filling the gap in retail science by presenting an agent-based model of retail competition, which explicitly represents store competition behaviour. The model exploits previous empirical research that quantified consumer expenditures at a high spatial resolution using census dissemination areas (populations of 400-700; Robinson and Balulescu 2018) and other spatial characteristics (e.g., service areas; Caradima 2015). The combination of empirical data and the presented ABM are used in a realistic setting to simulate a local retail competition system and answer four questions: 1) What is the level of correspondence between market share and revenue acquisition for an agent-based approach compared to a traditional location-allocation-based approach? 2) To what degree can the observed store spatial pattern be reproduced by competition? 3) To what degree are their path dependent patterns of retail success? 4) What is the relationship between retail survival and the endogenous geographic characteristics of stores and consumer expenditures?

2. Methods

2.1 Study Area

The Region of Waterloo (ROW) is a medium-sized (1384 km^2) regional municipality located in Southern Ontario, Canada, which consists of 7 Census Subdivisions (CSDs) including three cities (Kitchener, Waterloo, and Cambridge), and four townships (Wellesley, Wilmot, Woolwich, and North Dumfries; Figure 1). Over the past 15 years, the population in the ROW has experienced a continuous increase with an average annual growth rate of 1.58% (Region of Waterloo n.d.a), resulting in a total population of 617,870 in 2019, which shared about 4.25% and 1.64% of the Ontario and Canada's total population, respectively (Region of Waterloo 2020). As the 10th most populous Census Metropolitan Area (CMA) in Canada, it is projected to be one of the fastest-growing regions in Ontario (Region of Waterloo n.d.b). Among 7 CSDs, the tri-cities area is densely populated where over 88.53% of the regional population resides within 23.1% of the region's areal coverage (Region of Waterloo 2020). Additionally, most of the regional population is classified as the population in regular households (96.82%) versus those in temporary residences (3.18%), which includes but is not limited to student residences and group homes (Region of Waterloo 2020).

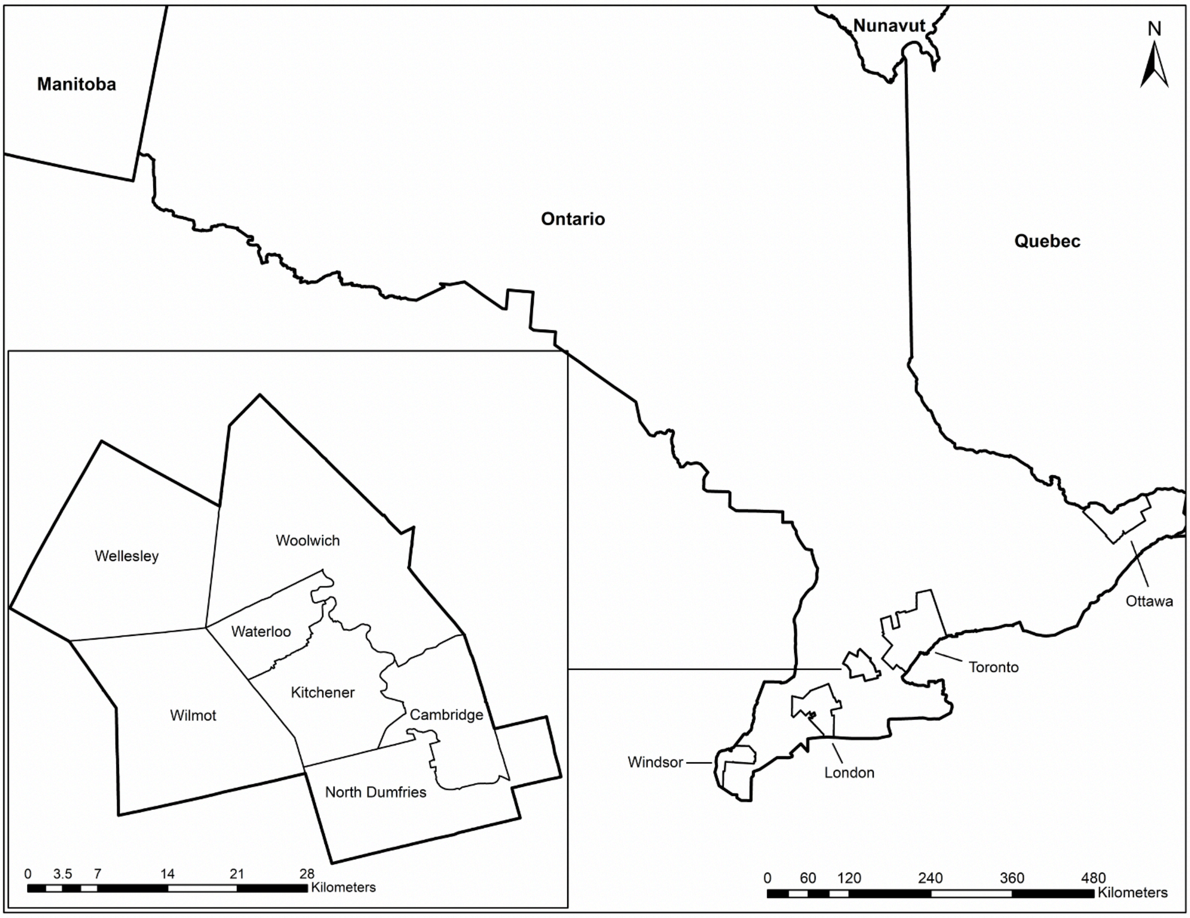


Figure 1: The Region of Waterloo (lower left inset) within Southern Ontario, Canada, and the location of 4 relative Census Metropolitan Areas (CMAs), including Ottawa, Toronto, London, and Windsor.

According to Statistics Canada (Table 36-10-0225-01), The home-improvement subsector is one of the most important retail trade subsectors which shared over 5% of the total annual consumer expenditures in Canada in recent decades. Since the 2008 economic depression, consumer expenditures on home-improvement products experienced flat growth (growth rate = 0.06%) but gradually recovered with an average annual growth rate of 4.1% since 2013. However, the home-improvement subsector is one of only two subsectors in which the number of stores quickly recovered from the depression within 4 years. Ontario owns the largest home-improvement retail market in Canada — over 38% of expenditures in home-improvement products are made by Ontarians. The strong continuous population growth in the Region of Waterloo raises the need to expand the local housing market thereby indicates a rising demand of home-improvement retail products as well (Caradima 2015).

Therefore, we choose the Region of Waterloo as a suitable study area to investigate the home-improvement retail competition system.

2.2 The Model

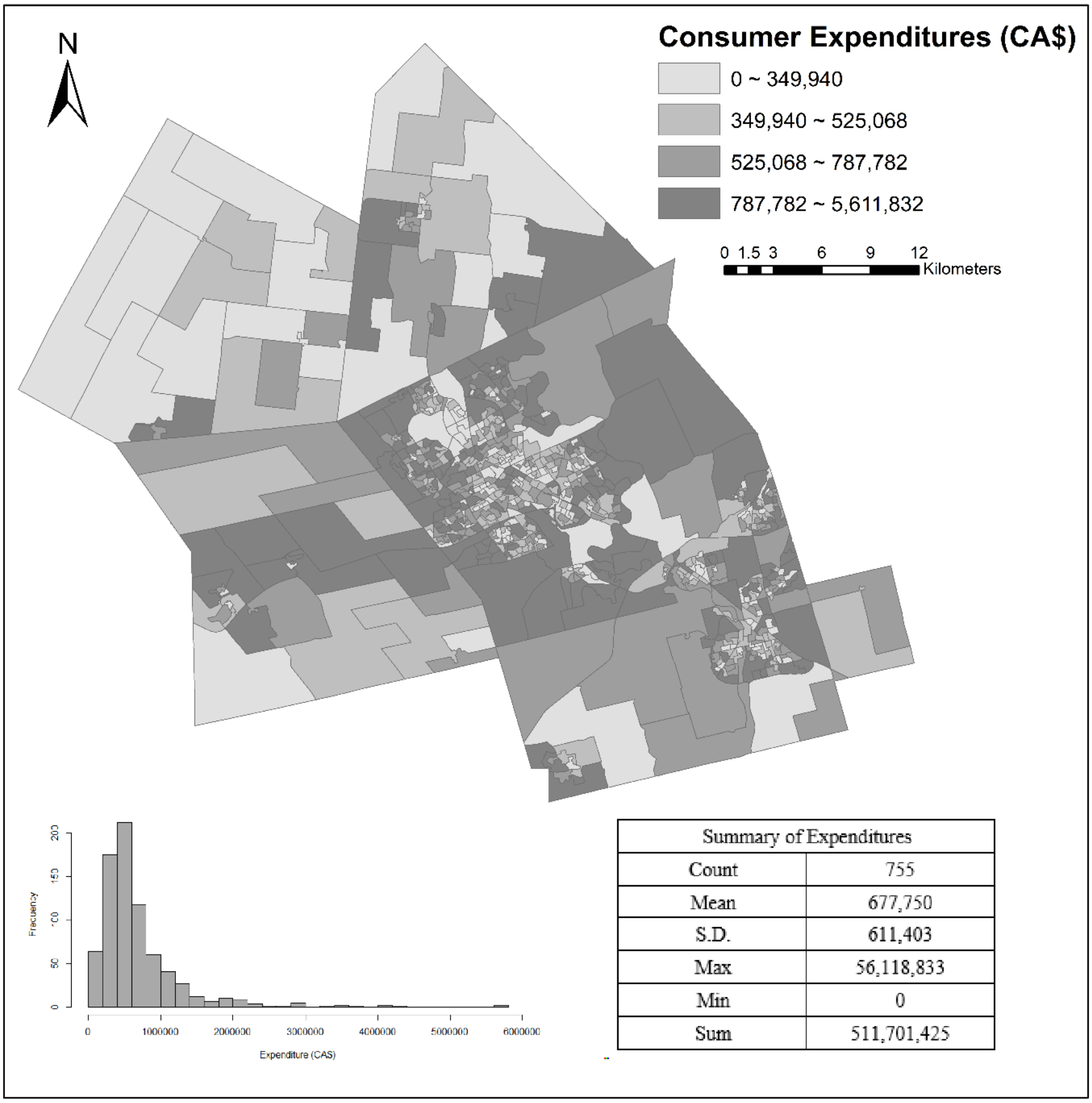
An agent-based model was created to simulate regional-scale retail market dynamics, with a focus on the role of local competition on store success and failure. In the retail competition agent-based model (RC-ABM), retail stores are represented as virtual retail store agents (RSAs) that interact with each other and their environment. The model environment, or the landscape with which RSAs interact, is represented as a spatially distributed population of consumers whose expenditures determine the success and failure of nearby stores.

Inspired by ecological Niche theory (Vandermeer 1972), and the idea of Many-Model thinking (Page 2018), we analogize the RSAs as a species and consumer expenditures as environmental resources, whereby stores compete with each other for revenue derived from consumer expenditures. While a variety of ABM modelling packages are available for model creation (e.g., Railsback et al. 2006), RC-ABM was operationalized in NetLogo. NetLogo is a software platform designed for constructing agent-based models (Wilensky 1999) with an open access and active user-community. For example, the open access model library CoMSES contains 803 shared models of which 545 (67.9%) are programmed in NetLogo (CoMSES September 16, 2020). Furthermore, NetLogo incorporates native functions for working with spatial (e.g., shapefiles) and aspatial (e.g., csv and database connections; Gaudou et al. 2017) data as well as has been coupled with other scientific software and data analysis tools such as GIS (Walker and Johnson 2019), R (Thiele and Grimm 2010), and Python (Jaxa-Rozen and Kwakkel 2018) among others.

2.2.1 The Consumer Expenditure Market

The landscape of consumer expenditures of home-improvement products is generated from census data at the census dissemination area (CDA). Here we work with home-improvement retail firms that predominantly reside within the North American Industry Classification System (NAICS) of 444, which defines retailers of “building material and garden equipment and supplies dealers” (Statistics Canada 2011). Twenty-three consumer’s home-improvement spending categories (Appendix 5, Table A5) were identified from the

National Household Survey and were aggregated and refined by household income for each of 755 CDAs in our study area (Robinson and Balulescu 2018). A CDA is the smallest geographic census unit, comprising a population of 400-700 individuals (Statistics Canada 2018b) for which their estimated values of home improvement expenditures range from \$0 to \$5,611,832 (Figure 2).



Summary of Expenditures	
Count	755
Mean	677,750
S.D.	611,403
Max	56,118,833
Min	0
Sum	511,701,425

Figure 2: The map of the consumer expenditure value among 755 CDAs in the Region of Waterloo; the histogram in the bottom left corner shows the statistical distribution of the consumer expenditures; the table in the bottom right corner contains some basic statistics information of the consumer expenditures.

The model's landscape is composed of 70272 (288 × 244) square patches with a 192m spatial resolution and a total of 37839 patches enclosed by the Region of Waterloo (ROW). A manual interpretation of residential Land Use (LU) by property parcel for the

ROW (Smith 2017; Sun and Robinson 2018) identified 137,283 residential LU parcels, whose centroids are used in the delineating consumer agent (CA) locations. Due to the mismatch between parcel sizes and patch resolution, only 7501 patches intersect parcel centroids (i.e., in many cases the area of the patch is greater than the area of a parcel). Therefore, the population of CA patches is 7501 and all other patches are null and void of expenditure or landscape information. Every CA possesses a number of home-improvement expenditures, determined by Equation (1):

$$Expenditure_{CA(i)} = \frac{Expenditure_{CDA(ji)}}{Count_{CDA(ji)}} \qquad (1)$$

where "$Expenditure_{CA(i)}$" is the total expenditures stored in *CA(i)*, "$Expenditure_{CDA(ji)}$" is the total amount of expenditures of *CDA(j)* which *CA(i)* belongs to, and "$Count_{CDA(ji)}$" is the total number of CAs reside in C*DA(j)*. Through Equation (1), the aggregated expenditures of the Region of Waterloo are distributed to individual CAs according to CAs' locations and the region's total expenditures' realistic spatial distribution (Figure 3).

Figure 3: The visualization of CAs in NetLogo; The study area is represented by patches with grey color, the CAs are patches with yellow color, and the white patches have Null data.

2.2.2 The Store Agent

Each RSA has a location, a size (square feet), and a service area. Prior to commencement of a model run, each RSA is given a position (x- and y-coordinates and which CDA it resides within) and is assigned a fixed size. A service area (SA) is assigned based on the CDA centroid and was precalculated in ArcMap using the street network.

The delineation of retail store service areas (SAs) remains an ongoing academic and industry challenge. While a variety of determinants and ranges can be found in retail literature (5 minutes to 19 minutes, Gordon and Richardson 1997; Önden et al. 2012; Caradima 2015), consumer sales data, acquired from a dominant home-improvement company, illustrated that consumers for stores in Ontario span both Canada and United States, including purchases from those living in Alaska. Despite these challenges a negative exponential distance decay function can be fit to consumer sales data showing a high degree of explanatory power. Using a large-sized SA (e.g., 19-minute drive time) and iteratively identifying services areas, their overlap, and the weighted distribution of expenditures by services area and store attractiveness for each of the 37,839 potential store locations exceeded the computational limits of NetLogo software used in this thesis. To reduce the computational load, and maintain performance and use of NetLogo, we instead used a 12-minute drive time SA and generated a service area for each of the 755 CDAs. Analysis of 23 home-improvement stores annual customer sales data found that 69.6% of customers inside their corresponding Census Metropolitan Areas (CMAs) resided within a 12-minute SA.

After initializing the location, size, and SA, an RSA receives expenditures revenue from CAs within its SA at each timestep of the model and remains or exits the landscape based on this revenue stream (see details in *Section 2.2.4*).

2.2.3 Competition

Retail store agents (RSAs) compete for revenue from consumer expenditures when a RSA's service area (SA) overlaps with the SA of one or more other RSAs (Figure 4). When *CA(i)* resides within multiple SAs in a time step, it will firstly evaluate each store's attractiveness based on the following utility function:

$$u_{ij} = {D_{ij}}^{-2} \times A_j \qquad (2)$$

where u_{ij} is the expected utility that *CA(i)* will acquire from shopping at RSA(j), D_{ij} is the standardized distance between CA*(i)* to *RSA(j)*, and A_j is the standardized size of *RSA(j)*. Equation (2) is similar to the initial version of the Huff's model (1962), which sets the exponent of D_{ij} to -2 and the exponent of A_j to +1 (Ghosh and MacLafferty 1987; Youn et al. 2012). These settings define the importance of the distance between the customer and the store as twice that of the size of the store on a consumer's evaluation of store attractiveness. We ensure this weighting by using a standardized distance and size that constrains the two variables range (0 to 1) using the following equations:

$$D_{ij} = \frac{(d_{ij} + \varepsilon)}{\sum_{j=1}^{k}(d_{ij} + \varepsilon)} \qquad (3)$$

$$A_j = \frac{a_j}{\sum_{j=1}^{k} a_j} \qquad (4)$$

whereby assuming *CA(i)* is the target of total *k* stores, d_{ij} is the Euclidean distance between *CA(i)* and *RSA(j)*, ε is a correction term to avoid having 0 in the denominator, and a_j is the actual size of *RSA(j)*. In Equation (3) and Equation (4), we obtain the standardized distance and size by dividing a store's actual distance to a CA and size by the sum of its competitors' distances to the same CA and their sizes, respectively.

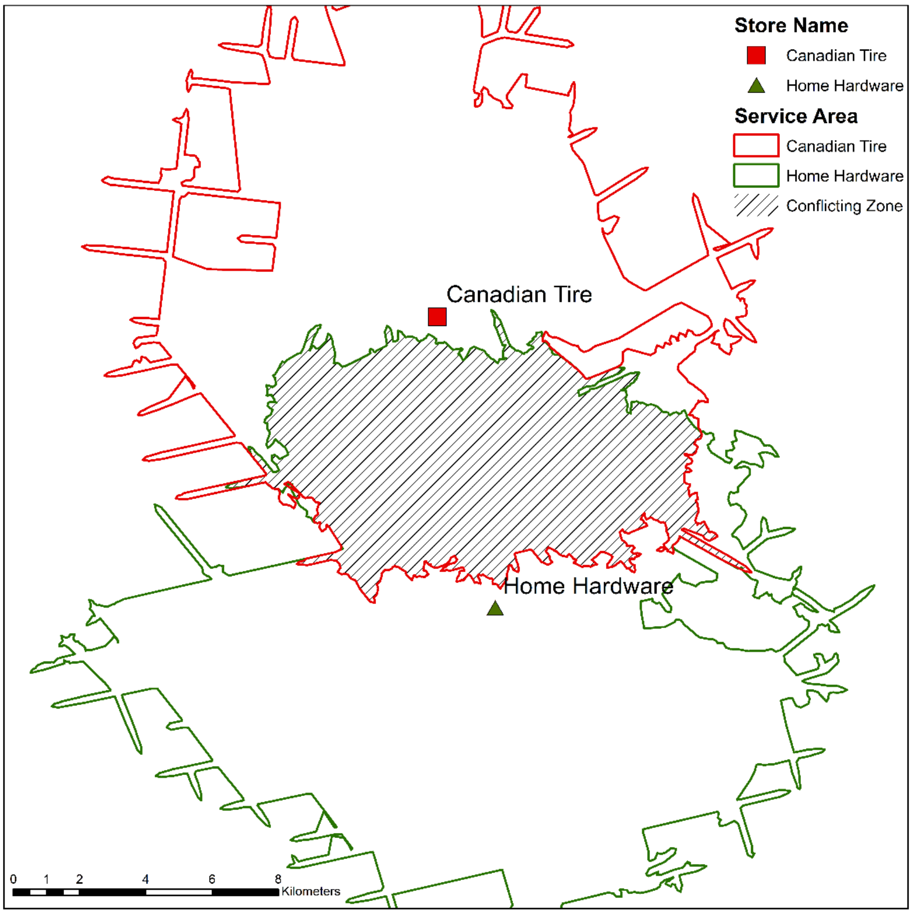


Figure 4: An example of competitors having the overlapping SAs

With all utilities of potential destination stores generated, *CA(i)* will calculate the proportion of expenditures it will spend in each destination store using the following probabilistic function:

$$P_{ij} = \frac{u_{ij}}{\sum_{j=1}^{k} u_{ij}} \qquad (5)$$

where u_{ij} is the utility score of *CA(i)* to *RSA(j)*, $\sum_{j=1}^{k} u_{ij}$ is the sum of utility scores of *CA(i)* to all destination stores, and P_{ij} is the probability of *CA(i)* to choose *RSA(j)*. The probability value determines the number of *expenditures that CA(i) sends to RSA(j)* as follows:

$$E_{ij} = P_{ij} \times E_i \qquad (6)$$

where E_{ij} is the money that *CA(i) sends to RSA(j)*, and E_i is the total available expenditures of *CA(i)*.

Overall, the RSA's pressure of competition is derived from its choice of location and size. An RSA's location determines the number of its competitors and the maximum amount of expenditure resources provided by consumers within its service area. In addition, a store's size defines its maximum level of attraction in comparison with competitors. The combined effects of these two factors (location and size) generate a complex competition system where an unpredictable pattern of RSA's success and failure emerge, e.g., a large store with relatively good location may fail whereas a small store in a consumer-deficiency has possibility to succeed.

2.2.4 Retail Store Agent Exiting Rule

The conditions of store closure or market exit are difficult to capture in the real-world. Low or unprofitable stores may remain open for a variety of strategic reasons that include being the first mover to an area and establishing brand presence (Patterson 1993), location with high visibility and marketing appeal (rule of thumb), avoidance of bad publicity, long-term speculation, and for competitive purposes (Mayadunne et al. 2018) among other reasons. Furthermore, the lack of store operations costs data, due to the proprietary nature of business, also limits one's ability to identify the conditions of closure.

In the absence of store closure information, we used a simple heuristic rule to define store closure and market exit based on relative store performance. To evaluate store performance, we first calculate the revenue density (i.e., sales per square metre) for an RSA as follows:

$$RD_j = \frac{E_j}{a_j} \qquad (7)$$

where E_j is the gross revenue and a_j is the size of *RSA(j)*. The revenue density offsets the larger gross revenues acquired by larger stores by the higher operational costs (e.g., labour, utilities, maintenance) by incurred larger stores. Hence, the revenue density requires larger stores to have higher gross revenues relative to smaller stores to cover their costs.

Then, we use a probabilistic index, which we define as the exiting ratio, to determine a RSA's probability of exiting the market:

$$ER_j = \left(1 - \frac{r_j}{N_{RSAs}} - \frac{a_j}{\sum a_j}\right) \times 100 \qquad (8)$$

where ER_j is the exiting ratio of RSA(j); r_j is rank of RSA(j)'s revenue density among all RSAs, following an incremental sequence; N_{RSAs} is the total number of stores on the landscape; $\sum a_j$ is the sum of all store sizes. The exiting ratio is a percentage probability value, using this ratio, stores with lower revenue density have higher possibilities to exit the market. We added an error term "$\frac{a_j}{\sum a_j}$" to adjust the ratio because larger stores typically are less likely than small stores to exit the market considering the huge investments and the company's reputation (e.g., Lai et al. 2016). The ER_j is only assigned to RSAs in a timestep if the RSA's revenue density is among the lowest 10% of all RSA revenue densities.

2.3 Validation

Model validation is an important step in the accreditation of a model (Balci 1994). However, our lack of access to proprietary data (sales, costs, store closure) constrained our ability to conduct a standard comparison between model outputs and observational data (e.g., Bianchi et al. 2007). In lieu of these data, we validate the RC-ABM using a model-to-model comparison approach (Klügl 2008) with the widely used Location-Allocation Model (LAM), known as Huff's Model, as implemented in ArcGIS v.12.4 (ArcMap n.d.a).

The LAM is a spatial interaction model that simultaneously selects optimal locations for new stores or facilities while allocating demand for those facilities (Goodchild 1984). LAMs have been applied and validated across a variety of areas of interest, which include but are not limited to retail site selection (e.g., Goodchild 1984), public service facility planning (e.g., Rahman and Smith 2000), supply chain management (e.g., Rabbani et al. 2020), humanitarian logistics (e.g., Paul and Wang 2019). In a LAM, locations of the demand points, competitor points, and potential facility points are determined prior to the simulation.

A LAM requires a set of demand points to allocate demand (i.e., consumer expenditures) to stores. We initialize the LAM with the same CA locations as demand points weighted by their available expenditures on home improvement products. A second requirement of a LAM is a set of competitor points, which is given as the observed locations of home-improvement retailers in 2014 (Appendix 1, Figure A1). These same locations are also used for the RSAs in the RC-ABM for comparison. Finally, like the RC-ABM, the LAM is parameterized with the same 12-minute maximum drive-time and the same transportation network as the RC-ABM.

While we have tried to create similar parameters and initial conditions between the RC-ABM and the LAM, there is a clear structural difference in the allocation of consumer expenditures to retail stores. Although both models use the Huff model to determine the allocation of the amount of consumer expenditures to retail stores, the RC-ABM includes an additional parameter (ε in Equation (3)) to ensure that the denominator is not equal to zero whereas the LAM function does not include this parameter (ArcMap n.d.b). To evaluate the effect of this functional change, we sweep five values (10, 1, 0.1, 0.01, 0.001) and assess the impact of those changes in our comparison with the LAM.

To determine the level of correspondence between the RC-ABM and LAM we use the replication standard (Axtell et al. 1996). The replication standard defines the following three levels of increasing correspondence between two models: relational equivalence (RE), distributional equivalence (DE), and numerical identity (NI). RE indicates that two models produce the same relationship between a resultant variable and alternative internal variable, DE indicates that the distributions of results derived from two models are statistically identical, and NI indicates that the numeric results of both models are exactly the same (Axtell et al. 1996).

To evaluate the level of replication between the RC-ABM and the LAM we first test for RE by fitting a linear regression between store revenues (model result) and store sizes (one model’s input variable) and if the coefficients in both regressions have the same sign (i.e., both positive or negative) then RE was achieved. To test for DE, we apply the Kolmogorov-Smirnov test (KS test), which is a statistical test to compare the similarity of two results based on their distributions (Lopes et al. 2007); if the distributions of store revenues calculated by two models are statistically the same then DE was achieved. Finally, if DE was achieved and two models’ results are numerically the same then NI was achieved.

2.4 Computational Experiments

We initialize the model with the observed distribution of consumer expenditure data (*Section 2.2.1*) and randomly distribute retail store agents (RSAs) within the study area. In each timestep of a model run, the consumer agents (CAs) transfer their expenditures to target RSAs using a probabilistic utility function (Equation (6)). After all CAs have transferred their expenditures (i.e., made their retail purchases), each RSA evaluates whether to remain in the market or exit. If an RSA decides to exit the market, a new RSA of the same size is generated (i.e., enters the market) and is randomly located within the study region. Using this approach, our computational experiments preserve the observed distribution of stores and store sizes.

Each model run consists of 20 timesteps, representing a time span of 20 years. At the end of a model run, the pattern of store locations; the total market share captured by all stores of each timestep; the location (CDA id) of both dead stores and stores that survived for 15 consecutive years; and duration, cumulative revenue, and average revenue of each store are recorded for analysis. Considering the model's stochastic behavior, a Monte Carlo approach is used to generate a 1000 model runs and capture the variation in potential model outcomes for subsequent analysis.

2.4.1 Replication of Observed Patterns

Using the existing distribution of observed home-improvement stores to populate the RC-ABM, we first answer the question to what degree can competition reproduce the observed spatial pattern and revenue of home-improvement retail stores? To assess the similarity between the observed and modelled spatial patterns we use the Spatial Point Pattern test (SPPT) (Andresen and Malleson 2011), which compares the similarity of two spatial point patterns based on a geographic unit (e.g., CDA).

The SPPT produces an S-index, which is a measurement of the similarity between two point patterns with a range of values from 0 (no similarity) to 1 (identical). Using an S-index threshold of >= 0.8 to indicate two patterns are similar (Boivin and de Melo 2019), we calculate the percentage of model runs that replicate the observed store spatial pattern. In addition to comparing the spatial pattern of RC-ABM generated store locations to the observed pattern, we also compare the amount of market share captured in each of our 1000 model runs against an estimated observed market share. The results of this comparison enable us to evaluate how optimal the observed pattern is with respect to capturing market share and how many other configurations are capable of capturing more. Lastly, we compare the similarity in the distribution of RSA revenues using Kolmogorov-Smirnov and Spearman's rank correlation coefficient tests. Again, we then determine the percentage of model runs that produce similar RSA revenues to the observed distribution.

2.4.2 Identifying Path Dependence

The repeated random placement of stores and process of competition can be used to identify path dependencies associated with patterns of store success. Using the 1000 model runs we answer the question, what is the spatial distribution of areas that guarantee retail success and failure? To answer this question, we identify invariant and variant regions at the CDA level and then compute a number of simple landscape metrics (largest patch size, number of patches, and total area) to describe their spatial pattern.

We identify a Successful Location in a single model run as a CDA that has a minimum of one store residing within it that persists for a minimum of 15 timesteps (>= 75% of the model run). We then identify locations as successfully invariant as those locations where the number of model runs for which there is a Successful Location occurs more than is expected by chance. The number of expected successes (48) is calculated as the probability of any one store (among n=48) locating in one of the 755 CDAs (0.0635%) times the number of model runs (1000), times a success threshold that we arbitrarily define as 75% to coincide with our 15 time-step threshold. Therefore, we expect any CDA to obtain an average of 48 successful stores. Other CDAs that have no stores persist for 15 timesteps are considered Unsuccessful Locations and are unsuccessful invariant regions. All other areas are considered variant regions.

2.4.3 Characteristics for Success

Given the generation of a 1000 model runs and assessment of the distribution of revenue (*Section 2.3*), market share (*Section 2.4.1*), and invariant and variant regions (*Section 2.4.2*), we evaluate what is the relative influence of store (size) and geographic characteristics (average nearest neighbour distance, and average number of competitor stores with overlapping service areas) as well as consumer expenditures (consumer expenditure density) on RSA survivability?

Under the complex context of the competition system, a store’s revenue may not always have a monotonically positive correlation with its longevity, e.g., stores with lower revenues may exist longer. Hence, we define the RSA survivability in two dimensions, the duration of survival (DS) and the cumulative revenue (CR), and respectively analyze the influence of store, geographic and consumer characteristics on each dimension. Moreover, we combine two dimensions into one using the store average revenue (AR), which is generated by dividing CR by DS.

The DS, CR, and AR are calculated as follows:

$$DS_{RSA(i)} = T\_Death_{RSA(i)} - T_{Enter_{RSA(i)}} + 1 \qquad (9)$$

$$CR_{RSA(i)} = \sum_{j=T_{Enter_{RSA(i)}}}^{T\_Death_{RSA(i)}} R_{ij} \qquad (10)$$

$$AR_{RSA(i)} = \frac{CR_{RSA(i)}}{DS_{RSA(i)}} \qquad (11)$$

where $DS_{RSA(i)}$ is the duration of survival of *RSA(i)*, $T\_Death_{RSA(i)}$ is the timestep when *RSA(i)* exits the market, and $T_{Enter_{RSA(i)}}$ is the timestep when *RSA(i)* enters the market; $CR_{RSA(i)}$ is the cumulative revenues of *RSA(i)* during its lifetime, R_{ij} is the revenue of *RSA(i)* at the timestep *j*; and $AR_{RSA(i)}$ is the average revenue of *RSA(i).*

An *RSA(i)*'s competitor is defined as other stores whose service areas are overlapping with the *RSA(i)*'s service area. The number of competitors (NC) of an RSA could be calculated and set to be an attribute of each RSA.

The CDA's consumer expenditure density (CED) is calculated using Equation (12):

$$CED_{CDA(i)} = \frac{Expenditure_{CDA(i)}}{S_{CDA(i)}} \qquad (12)$$

where $CED_{CDA(i)}$ is the CED of *CDA(i)*, $S_{CDA(i)}$ is the size (square feet) of *CDA(i)*. $CED_{CDA(i)}$ is a index by which the CDA's consumer expenditure is normalized by the CDA size (in square feet).

By setting the $DS_{RSA(i)}$, $CR_{RSA(i)}$, and $AR_{RSA(i)}$ as dependent variables, we use the Ordinary Least Square (OLS) regressions to quantify the influence of store size, NC, and CED on the store survivability. Prior to the OLS regression analyses, we use the Min-Max normalization (Scikit-Learn n.d.) to rescale both dependent and independent variables from 0 to 1, thereby ensuring the coefficient values can be constrained within a reasonable magnitude (from -1 to 1).

3. Results

3.1 Validation

We initiated the process of validation by evaluating if we could obtain Relational Equivalence (i.e., the same relationship) between inputs (store size) and outputs (store revenue) for the RC-ABM as acquired for the LAM. In our assessment of validation, we initialized both models with the same observed spatial pattern and distribution of store sizes as well as did not allow stores to exit and enter the market (*Section 2.3*). A linear regression was used to conduct this evaluation and determine if the correlation between store size and revenue was the same between the two models. Results showed a significant positive linear correlation between store size and revenue between the RC-ABM and LAM models ($R^2 > 0.7$, coefficient value > 0, p-value < 0.05; Table 1). Moreover, our sensitivity analysis of ε, which was included in the RC-ABM and does not exist within the LAM, showed only minor with ε < 1 – from 2.8% (when ε = 0.001) to 3.4% (when ε = 0.1). These results statistically demonstrate relational equivalence between the RC-ABM and LAM.

Table 1: Coefficient values from linear regressions between store revenue (dependent variable) and store size (independent variable) generated by LAM and 5 RC-ABMs with different ε settings

Model Name	LAM	RC-ABM				
		ε = 10	**ε** = 1	**ε** = 0.1	**ε** = 0.01	**ε** = 0.001
coefficient	+ 169.5	+ 200.9	+ 181.9	+ 175.2	+ 174.4	+ 174.3
p-value	2.963e-14	2.2e-16	1.083e-14	4.886e-14	5.832e-14	5.938e-14
R^2	0.71	0.79	0.72	0.71	0.70	0.70

We extended our model-to-model comparison as part of our validation process by evaluating if the RC-ABM was able to achieve Distributional Equivalence with the LAM. This evaluation was conducted using a Kolmogorov-Smirnov (KS) test, which has a null hypothesis that the results of two models (in our case store revenues) have the same distribution of values. Using the LAM as the standard sample and the revenue generated by each of the 5 RC-ABMs as the replicated sample, we found no significant difference between the RC-ABM distributions and that produced by the LAM (Table 2). Results show that p-values of 5 KS tests are all higher than 0.96, and as the ε value decreases to 1, the p-value increases to 0.997. Considering most alpha values used to reject the null hypothesis (significant differences) are 0.05 or less, we conclude that Distributional Equivalence was obtained between the revenues derived by the RC-ABM and the LAM.

With Distributional Equivalence achieved we looked to determine if Numerical Equivalence was possible. Detailed results of each store's revenues by the LAM and the RC-ABMs (Appendix 2) illustrates that Numerical Equivalence could not be achieved. Our inability to achieve Numerical Equivalence is not surprising given that hardware (Wilensky and Rand 2007), rounding (Polhill et al. 2005), and sequences of actions have also been shown to inhibit the ability to achieve Numerical Equivalence (Axtell et al. 1996). Despite our inability to obtain exact numerical equivalence, we demonstrated that alignment between the RC-ABMs and the location-allocation model (LAM) was achieved, which provides justification for the use of the RC-ABM for further scientific investigation. In addition, through our validation process we demonstrated that the RC-ABM results are not sensitive to ε as values spanning from 0.001 to 10 showed little-to-no difference in results and all ε achieved Distributional Equivalence. Nevertheless, since the fitness of linear regression by RC-ABM ($R^2 = 0.71$) is the same as the value acquired for the LAM ($R^2 = 0.71$) when ε is equal to 0.1, we use a value of ε = 0.1 for all subsequent computational experiments.

Table 2: Results of Kolmogorov-Smirnov (KS) tests between store revenue predictions by the LAM and 5 RC-ABMs with different ε settings

	LAM vs. RC-ABM by Store Revenues				
ABM *ε* value	***ε*** = 10	***ε*** = 1	***ε*** = 0.1	***ε*** = 0.01	***ε*** = 0.001
KS-value	0.104	0.083	0.083	0.083	0.083
p-value	0.960	0.997	0.997	0.997	0.997

3.2 Replication of Observed Patterns

The Spatial Point Pattern test (SPPT) was used to evaluate the degree of point pattern similarity between RC-ABM simulated patterns and observed patterns. The SPPT may be operationalized as either the standard or robust S-index. The standard S-index considers all areal units (e.g., CDA) regardless of whether stores reside within. In contrast, the robust S-index only uses areal units that have at least one store within. Our Monte Carlo approach produced 1000 spatial patterns of store locations that were each compared to the observed pattern using both standard and robust S-indices. The standard S-index results showed reproduction of the observed pattern under the CDA-based context, whereby all 1000 simulations (100%) yielded patterns with a standard S-index greater than 0.8, our similarity threshold value. Further strengthening these results, the variation of similarities remains at a very low level (mean = 0.95, S.D. = 0.00, maximum = 0.96, minimum = 0.95; Table 3).

In contrast, the robust S-index yielded no values greater than our similarity threshold value of 0.8. The contradiction of results from the two indices suggest the differences may be due to an imbalance between the number of stores (48 stores) and number of areal units (755 CDAs). In each spatial pattern, at least 94% (707 of 755) of CDAs have no stores. Hence, the similarity level is significantly improved by the inclusion of unoccupied CDAs, which resulted in a high standard S-index.

Table 3: Mean, standard deviation (S.D.), maximum, and minimum of SPPT standard and robust S-index values between the observed spatial pattern and 1000 simulated patterns derived from the Monte Carlo experiment; And the percentage of patterns that have a S-index that is over 0.8.

	Based on CDA		**Based on Hexagon**	
	# of CDAs: 755		# of Hexagons: 50	
	Standard S-index	Robust S-index	Standard S-index	Robust S-index
Mean	0.95	0.57	0.74	0.30
S.D.	0.002	0.021	0.033	0.096
Maximum	0.96	0.63	0.82	0.57
Minimum	0.95	0.50	0.66	0.06
% over 0.8	100.00	0.00	8.60	0.00

To further investigate the reproducibility of spatial store patterns, we uniformly tessellated the study area with 40 km^2 hexagons (Figure 5) and conducted the same SPPT analysis using the hexagon boundaries. Results from this evaluation revealed that only 8.6% of the patterns generated by the RC-ABM had a degree of similarity with the observed pattern above our similarity threshold using the Standard S-index (Table 3). Nevertheless, results using the Robust S-index again produced no simulated patterns that corresponded to the observed patterns. Therefore, we conclude that competition alone does not capture a sufficient amount of variation in home improvement retail market to generate the observed spatial distribution of store locations.

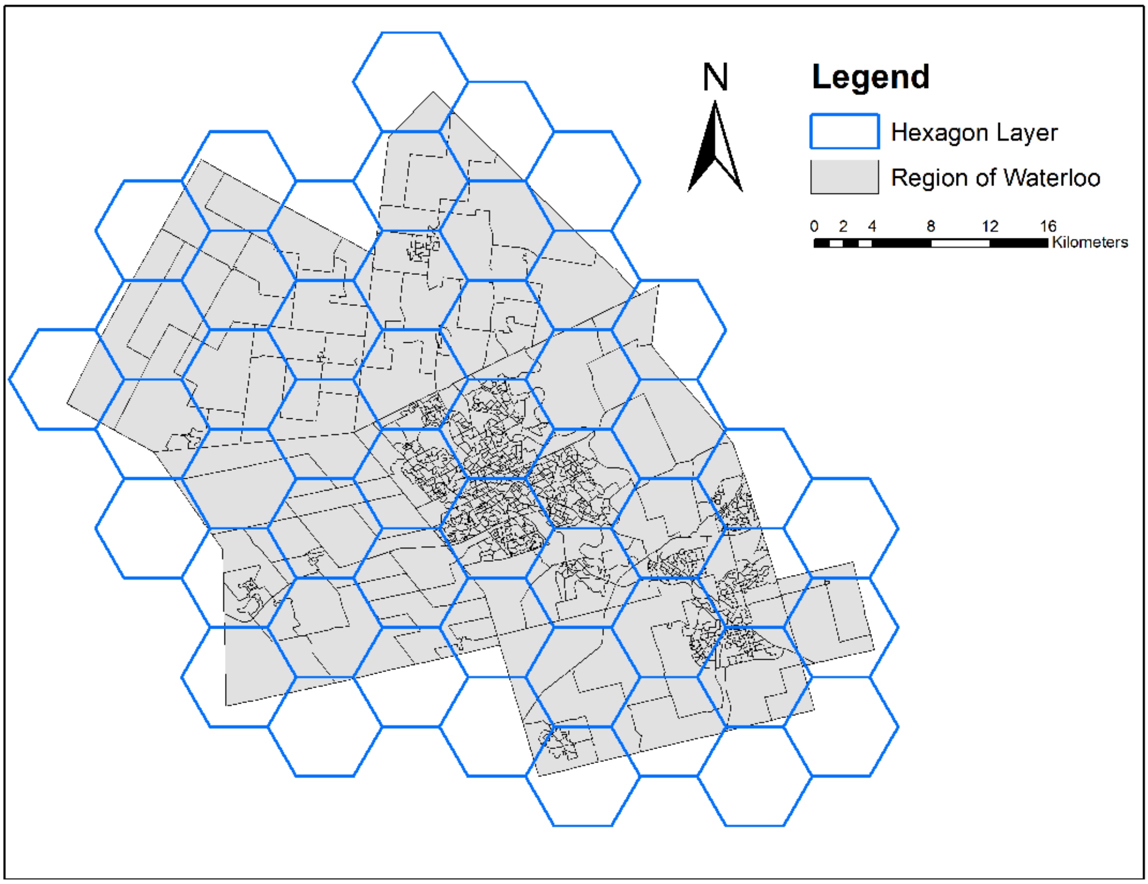


Figure 5: Map of the artificially generated hexagons based on the Region of Waterloo boundary

Despite the inability of the RC-ABM to replicate the observed retail location patterns, we sought to evaluate the degree to which the RC-ABM captured home-improvement market share relative to the observed pattern. However, since market share data are proprietary and unavailable, we generated market share capture (MSC) from the observed store locations using the LAM. Simulated MSC was calculated at the end of each of the 1000 runs by the RC-ABM (MSC1) as well as for each time step for each model run (MSC2, comprising 20000 spatial patterns; Table 4; Appendix 3, Figure A3). The observed store pattern captured more of the market (MSC 99.92%) than the mean of both our simulated measurements of MSC1 97.52% and MSC2 97.81% (Table 4; Appendix 3, Figure A3). The standard deviations of MSC1 (1.49%) and MSC2 (1.43%) are small and the minimum values are over 92% market capture.

Given the similar market share capture of the simulated spatial patterns to the observed, we evaluated the ability of the RC-ABM to reproduce the distribution of store revenues. However, like the market share data, store revenue data are proprietary and unavailable. In lieu of these data, store revenues assigned by the LAM were used. We compared the RC-ABM generated store revenue data to the LAM data using the Spearman Rank Correlation test and the Komogorov-Smirnov (KS) test. The hypothesis when using the Spearman's Rank Correlation is that the simulated store's revenue does not covary with the rank of observed revenues. Results of these tests show that all simulated patterns produce a similar distribution of store revenues with the observed pattern, e.g., 0% of Spearman's p-values and 100% of KS p-values are over 0.05 (Table 5). Although the RC-ABM was not able to reconstruct the spatial pattern of home-improvement stores within our study area, the RC-ABM was able to reproduce the observed distribution of store revenues.

The combination of market share capture, as an aggregate measurement of retail store coverage, and rank order distribution of store revenues illustrates that regardless of simulated spatial patterns, the study area home-improvement market is likely saturated.

Table 4: Mean, Standard Deviation (S.D.), maximum, and minimum of the total market share captured (MSC) at only the final timestep and the MSC of all appeared patterns; And the percentage of MSCs that is lower than the observed MSC (99.92%).

	Total Market Share Captured (%)	
	MSC 1 (only final timestep)	MSC 2 (all patterns appeared)
Mean	97.52	97.81
S.D.	1.49	1.43
Maximum	99.96	99.98
Minimum	92.92	92.76
% lower than 99.92%	99.70	99.52

Table 5: Mean and Standard Deviation (S.D.) of 1000 Spearman's test's rho and p-value, and KS test's statistics and p-value; And the percentage of results that have a p-value that is over 0.05.

	Spearman's Test		**Kolmogorov-Smirnov (KS) Test**	
	rho	p-value	statistics	p-value
Mean	0.83	2.1E-10	0.11	0.92
S.D.	0.04	1.9E-09	0.02	0.10
% over 0.05	N/A	0.00	N/A	100.00

3.3 Identifying Path Dependence

Our analysis of path dependence involved segmenting patterns of retail location into those that had a persistent store located for > 75% of a model run (i.e., 15 years; Successful Invariant), those that had no persistent stores (Unsuccessful Invariant), and those that had variable store presence (Variant). Overlaying these areas generated from 1000 model runs illustrates that a small percentage of the CDAs are invariant (34.44%) relative to those that are Variant (65.56%). Breaking down the invariant CDAs into those that are successful and unsuccessful locations for store longevity, we found that 96.92% invariant CDAs were Successful Invariant and only 3.08% were Unsuccessful Invariant. These results suggest that all CDAs have the potential to house a successful home-improvement retail store. However, there are a limited number of CDAs that are more likely to guarantee longevity.

An assessment of the areal representation of the CDAs further emphasizes the differentiation between Invariant and Variant regions. The proportion of the study area classified as invariant comprises on 9.01% compared to 90.99% classified as variant (Table 6). The wider margin in areal coverage, compared to CDA counts, occurs because 251 of the 252 Successful Invariant CDAs reside within the cities of Kitchener, Waterloo, and Cambridge (KWC). These CDAs are densely populated and therefore relatively small in area (Figure 6). In contrast, the less densely populated and larger CDAs around the periphery of these three connected cities are either Unsuccessful Invariant or Variant (Figure 6, Table 6). An evaluation of the Unsuccessful Invariant CDAs suggests that edge effects are likely playing a role since five of the eight unsuccessful CDAs are located along the border of the study area.

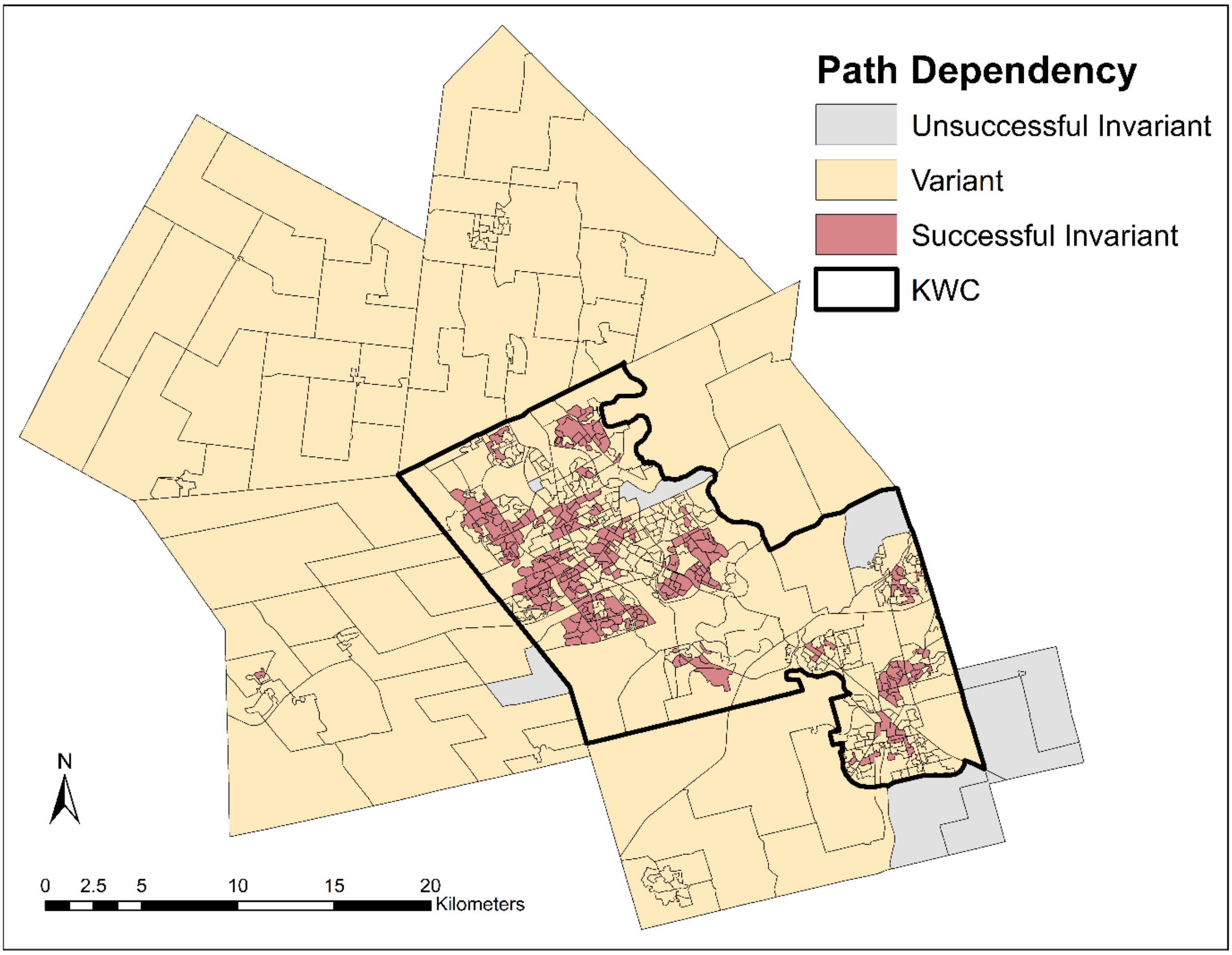


Figure 6: Census dissemination areas (CDAs) classified as Successful Invariant, Unsuccessful Invariant, and Variant within the Region of Waterloo, Ontario, Canada. The subregion of Kitchener, Waterloo, and Cambridge (KWC) represents an urban corridor within which 88.53% of the population of the Region of Waterloo.

Looking beyond the composition of Successful Invariant, Unsuccessful Invariant, and variant regions, the configuration of these classes shows that, within the Region of Waterloo, the Unsuccessful Invariant region has a low level of fragmentation (i.e., few number of patches) with an intermediate mean patch size (Table 6). In contrast, the Successful Invariant class is the most fragmented of the three classes comprising 31 patches, the smallest average patch size, smallest largest patch, and covers the smallest area. Thirty of the 31 patches reside within the KWC area, leaving just 1 small successful patch (16.74 ha) in the peripheral regions (Figure 6, Table 6). In contrast, Variant patches are few, but have the largest mean patch size, largest patch, and occupy the largest area within both the region and KWC. The standard deviation in patch size for the Variant class is substantially larger than the Successful and Unsuccessful Invariant classes because it is composed almost solely by a single patch comprising 99.93% of the Variant area (Table 6, Figure 6).

Our invariant and variant mappings identify a clear path dependency in the KWC corridor, whereby it is highly likely that store located in the Invariant Successful patches have the opportunity for success and longevity. This outcome can be partly explained by the composition of the corridor which includes 88.53% of the Region's population and 87.3% (119844 over 137283) of the Region's residential land-use parcels despite covering only 23.1% (31912.91ha over 138428.78 ha) of the Region's area. The path dependency results utilize the same number and distribution of store sizes as were observed. Therefore, when combined with the previous finding that the region is saturated, it suggests that new entrants must find strategic locations within Successful Invariant regions along with other competitive strategies (e.g., pricing and service) to achieve success.

Table 6: Summary of the composition and configuration of Unsuccessful Invariant, Successful Invariant, and Variant census dissemination areas (CDAs) in the Region of Waterloo (ROW), the subregion of Kitchener, Waterloo, and Cambridge (KWC), and the peripheral region.

	Region of Waterloo			KWC			Peripheral Region		
Region Area (ha)	Area = 138428.78			Area = 31912.91			Area = 106515.87		
Total CDA Numbers	n = 755			n = 655			n = 100		
	Unsuccess	Success	Variant	Unsuccess	Success	Variant	Unsuccess	Success	Variant
Number of CDAs	8	252	495	3	251	401	5	1	94
Percent of CDAs (%)	1.06	33.38	65.56	0.46	38.32	61.22	5	1	94
Total Area (ha)	7552.96	4924.16	125951.66	1358.72	4907.42	25646.77	6194.24	16.74	100304.9
Percent Area	5.46	3.56	90.99	4.26	15.38	80.36	5.82	0.02	94.17
Largest Patch Size (ha)	5551.8	1382.94	125860.16	916.48	1382.94	25555.26	5551.8	16.74	100304.9
Mean Patch Size (ha)	1510.59	158.84	25190.33	452.91	163.58	5129.35	3097.12	16.74	100304.9
S.D. Patch Size (ha)	2041.06	284.81	50334.91	359.81	288.32	10212.96	2454.68	0	0
Number of Patches	5	31	5	3	30	5	2	1	1

3.4 Characteristics for Success

We conducted three regression analyses to identify the relative influence of store size, number of competitors, nearest neighbour (competitor) distance, and consumer expenditure density (CED) on the duration of survival (DS), cumulative revenue (CR), and average revenue (AR) of a store. Our results found that the DS was negatively affected by increasing store size in the RC-ABM (coef.size = -0.73, Table 7). This result contrasts with what we expected and what others have shown, i.e., that big-box stores negatively impact smaller individual or chain stores (e.g., Haltiwanger et al. 2010). The negative impact of size could have been an artifact of our representation of revenue; however, store size has relatively no effect on CR and has a positive effect on AR. We then evaluated the existing rule (*Section*

2.2.4) and quantified the frequency of large versus small stores existing the market. Using a threshold of 60,000 square feet, whereby the 12 stores larger than the threshold are considered large stores and the 36 stores less than the threshold are considered small stores, we identified that large stores more frequently exit the market relative to small stores.

While store size had the largest effect on DS, our regression analysis also found that the number of competitors significantly influenced the DS of a store. While a higher number of competitors increases the DS of a store (coef.NOC = 0.43), it has little effect on the cumulative and average store revenue. However, DS of a store decreases as competitor distance increases (coef.NND = -0.07), which corroborates similar findings in the literature that retail stores benefit by locating nearer to competitors (e.g., Schmidt and Lee 1979, cited in Karande and Lombard 2005). Correspondingly, the NND also has a negative impact on the cumulative and average revenues of a store (Table 7).

Among our four independent variables, CED held the strongest positive influence on duration of survival and the cumulative revenue of a store. This outcome also corroborates literature that suggests that the most important factor driving retail success is high consumer purchasing power.

Table 7: The R^2, coefficient value and p-value of the intercept and 4 independent variables (store size, Nearest Neighbour Distance (NND), Number of Competitors (NOC), and Consumer Expenditure Density (CED)) derived from 3 OLS regressions between 3 dependent variables (Duration of Survival (DS), Cumulative Revenue (CR), and Average Revenue (AR)) and 4 independent variables.

Dependent Variable	**R^2**	**Intercept**		**Independent Variables**							
				Size		NOC		NND		CED	
		coef.	p-value	coef.	p-value	coef.	p-value	coef.	p-value	coef.	p-value
Duration of Survival (DS)	0.46	0.27	0.00	-0.73	0.00	0.43	0.00	-0.07	0.00	0.52	0.00
Cumulative Revenue (CR)	0.04	0.02	0.00	-0.01	0.00	0.07	0.00	-0.02	0.00	0.18	0.00
Average Revenue (AR)	0.75	0.04	0.00	0.36	0.00	0.02	0.00	-0.20	0.00	0.15	0.00

Collectively the four independent variables were significant for all three independent variables. The goodness of fit, as represented by R^2, was highest for explaining the variance in the average revenue of stores ($R^2 = 0.75$) followed by the cumulative revenue of stores ($R^2 = 0.46$). The linear regression had the least (and low) explanatory power for cumulative revenue ($R^2 = 0.04$). In contemplation as to why this was the case, we realized that the inclusion of all stores regardless of their DS likely obfuscated the relationship between cumulative revenue and our independent variables.

We subsequently conducted an iterative regression analysis starting with all stores from all 1000 model runs (Table 8). Then we systematically removed stores based on their DS to evaluate changes in model fit and the influence of the independent variables on cumulative revenue. A total of 17 regressions were conducted, with model fit (i.e., R^2) increasing from $R^2 = 0.04$, when all stores with all DS are included, to $R^2 = 0.93$ when only stores that survived for 17 timesteps are analysed. Hence, it can be inferred that the OLS model has a better performance in investigating the relationship between CR and independent variables of stores with a longer lifespan.

Based on Table 8, we find that when the instant-exit stores (duration = 1) are considered, store size has little influence on cumulative revenues (coef.Size = -0.01). After eliminating the instant-exit stores, a positive impact of store size exists and continuously becomes stronger, i.e., the coef.Size increases from 0.09 to 0.78, and the store duration increases (Table 8). A similar situation is present with our NND variable, whereby exclusion of the instant-exit stores changes the NND to positively affect cumulative revenues. In contrast, the positive effect of NOC is reduced as we eliminate stores that have a short time in the market and becomes negative when a store's duration is greater than 5 years in the market. Ultimately, we can see that stores lasting over 5 timesteps can have a higher cumulative revenue by keeping away from their competitors (coef.NOC < 0 and coef.NND > 0, Table 8). While obtaining a foothold in a dense and high-expenditure area is always a better choice for stores seeking to accumulate more revenue, stores which persist for a longer term are less dependent on such type of advantages, i.e., coef.CED decreases from 0.18 to 0.04 as duration increases (Table 8).

Table 8: The R^2, coefficient value and p-value of the intercept and 4 independent variables (store size, Nearest Neighbour Distance (NND), Number of Competitors (NOC), and Consumer Expenditure Density (CED)) derived from OLS regressions between Cumulative Revenue (CR) as the dependent variable and 4 independent variables under different requirement of store durations.

OLS Version	# of data	R^2	Intercept		Independent Variables							
					Size		NOC		NND		CED	
			coef.	p-value	coef.	p-value	coef.	p-value	coef.	p-value	coef.	p-value
All results	131258	0.04	0.01	0	-0.01	0	0.07	0	-0.02	0	0.18	0
Duration > 1	89360	0.09	0.01	0.01	0.09	0	0.06	0	0.09	0	0.16	0
Duration > 2	75455	0.2	0.02	0	0.16	0	0.04	0	0.17	0	0.13	0
Duration > 3	66886	0.31	0.04	0	0.22	0	0.02	0	0.17	0	0.12	0
Duration > 4	60401	0.41	0.05	0	0.27	0	0	0.24	0.17	0	0.1	0
Duration > 5	55250	0.5	0.06	0	0.32	0	-0.01	0.04	0.17	0	0.09	0
Duration > 6	50882	0.59	0.07	0	0.37	0	-0.02	0	0.18	0	0.08	0
Duration > 7	47235	0.65	0.08	0	0.41	0	-0.03	0	0.18	0	0.07	0
Duration > 8	43917	0.71	0.09	0	0.46	0	-0.05	0	0.19	0	0.07	0
Duration > 9	41085	0.76	0.09	0	0.5	0	-0.05	0	0.19	0	0.07	0
Duration > 10	38507	0.8	0.09	0	0.54	0	-0.06	0	0.19	0	0.06	0
Duration > 11	36174	0.84	0.1	0	0.58	0	-0.06	0	0.2	0	0.06	0
Duration > 12	34026	0.86	0.1	0	0.62	0	-0.07	0	0.2	0	0.05	0
Duration > 13	32074	0.89	0.1	0	0.66	0	-0.07	0	0.21	0	0.05	0
Duration > 14	30332	0.9	0.1	0	0.69	0	-0.08	0	0.21	0	0.04	0
Duration > 15	28654	0.91	0.1	0	0.72	0	-0.08	0	0.22	0	0.04	0
Duration > 16	27043	0.92	0.1	0	0.75	0	-0.08	0	0.22	0	0.04	0
Duration > 17	25455	0.93	0.1	0	0.78	0	-0.09	0	0.23	0	0.04	0

Without considering other factors like the management strategy and store expenses, we can simply evaluate the store's average gravitational pull and attraction of consumer expenditures reflected by the AR. According to Table 7, a larger store locating in a densely-expenditure (higher CED) area has a higher potential to absorb more money. Moreover, store size is the most important factor and its influence on AR is over twice of alternative three factors (coefficient value, Table 7). Similar to results of the DS, the retail clustering effect (closer to competitors) can be an accelerating factor for stores to earn more average revenues (negative coef.NND and positive coef.NOC, Table 7).

4. Discussion

4.1 Modelling the Retail Competition System

In recent decades, the demand to investigate wicked problems (Game et al. 2014) and understand the interactions and behaviours driving complex systems (Miller and Page 2009) has fostered a growing use of agent-based modelling (ABM) applications to advance science (An et al. unpublished). However, due in part to the difficulty of acquiring and collecting proprietary retail data (Sturley et al. 2018), there remains a gap in our understanding of retail site selection and subsequently behaviour models of retail location. Our retail competition agent-based model (RC-ABM) provides an initial foundation for the exploration of retail competition through a variety of perspectives that is not possible using equation-based models and traditional spatial interaction models.

A model-to-model validation approach was used in the absence of empirical data (Klügl 2008) and demonstrated relational equivalence between the RC-ABM and a widely used location-allocation model (LAM). By demonstrating a high level of correspondence between the two models, we accredit the fundamental behaviour of the RC-ABM to that literature and to the best of the authors knowledge provide the first comparison of an ABM approach to a GIS model.

While the formation of the observed spatial store pattern is a long-term accumulation of complicated and continuous agent interactions in the retail system, we randomly distributed stores and enabled their spatial pattern and revenue acquisition to dynamically evolve solely based on spatial competition. Using a Monte Carlo approach, we simulated retail competition over twenty time-steps and compared the final generated pattern over 1000 model runs and found that competition alone was insufficient to replicate the observed spatial pattern of store locations. Despite this finding, we did not interrogate the RC-ABM at all timesteps (i.e., 19000 additional spatial patterns) due to the temporal requirements of exporting spatial patterns from NetLogo. Therefore, it is possible that the observed patterns were achieved but did not remain stable until the end of the model run. Further investigation among intermediate timesteps of the RC-ABM as well as the duration of longevity of the observed pattern, which was taken for a single year (2014), would provide additional evidence to support or refute the ability for competition alone to generate the observed spatial patterns.

Despite differences in modelled versus observed spatial patterns of store locations, the RC-ABM demonstrated that the model produced similar levels of market capture and distributional equivalent levels of store revenues. The level of market capture achieved suggests that the region's home-improvement market is likely saturated, and the level of market capture should be noted by local economic development initiatives and potential market entrants. However, the entry of a single big-box chain can significantly change the context of competition (Gonzalez-Benito 2005), thereby reshaping the patterns of home-improvement retail.

In our evaluation of store success, we explored the concept of path dependence as represented by invariant and variant regions similar to Brown et al. (2005). Aside from literature focusing on a firm's strategic management behaviors (e.g., Lamberg and Tikkanen 2006), to the best of the author's knowledge, the relationship between path dependence and retail behaviour is void from the literature. As a step to filling this gap, we identified regions that were Successful Invariant (as those comprising one or more stores that survived for 15 or more years) and Unsuccessful Invariant (as those areas with no stores surviving 15 or more years) across all 1000 model runs. Variant regions comprised locations that had some stores last 15 years but not in all 1000 model runs.

Our results showed that approximately one-third (34.44%) of the CDAs in the study area are path-independent within the retail competition landscape. The location of path independency almost exclusively resides within the Kitchener-Waterloo-Cambridge cities corridor and the Successful Invariant CDAs are configured as multiple small clusters. New entrants or policies to encourage new entrants would have increased success despite higher levels of competition by locating in these areas. Furthermore, our analysis of drivers of duration of survival and cumulative and average store revenues showed that locating closer to a competitor and having more competitors in proximity increased the duration of survival and cumulative and average store revenues.

An artifact of restraining our analysis to the Region of Waterloo is demonstrated by the presence of Unsuccessful Invariant regions primarily along the regional boundary. Edge effects typically occur when arbitrary finite boundaries are used (Griffith 1983), but they can also occur due to geographic boundaries (e.g., coastal boundaries, Chen 2017). To coincide with our data and software limitations we did not add a guard zone (e.g., Chen 2017) or complicate the model with some type of spatial weighting (Kenter and Elhorst 2012).

4.2 GIS and ABM Integration

The presented research validated the behaviour of an agent-based model (ABM) against a location-allocation model (LAM) that is widely used within the GIScience community (e.g., Mestre et al. 2015). The RC-ABM also utilized geospatial data (the residential land use parcel data) to more accurately position consumer agents, similar to dasymetric modelling approaches (e.g., Briggs et al. 2007). While it has been argued that the generative (i.e., bottom-up) approach of ABMs (Gilbert and Troitzsh 2005) align more closely to real-world systems relative to statistical and system dynamics approaches (Wilensky and Rand 2015), the integration of GIS data and spatial behaviors (e.g., searching, topology, distance weighted interactions) are typically required to move from an existence proof (Waldrop 1993) to high-fidelity models (e.g., Gehlot et al. 2019). Movement along this spectrum toward higher fidelity data informed ABMs may also enable ABMs to overcome their historical challenges associated with calibration and validation (e.g., Ligtenberg et al. 2010).

The integration of agent-based modelling and Geographic Information System (GIS) commenced decades ago with GRASS GIS (Westervelt 1999) and has continued to rise as an integrated approach for scientific inquiry (Sengupta and Sieber 2007) with contributions from Gimblett (2002), Brown et al. (2005), and Robinson and Brown (2009) among others. While the integration of GIS and ABM has not achieved ubiquitous support and use, the integration is imperative to extend the static data management, analysis, and visualization capabilities of GISystems to incorporate time and process, which have historically been difficult to represent in GISystems (Andris et al. 2018). Furthermore, the anthropocentrism of ABM also emphasizes the role of humans in shaping and responding to natural processes and their configurations in the landscapes within which humans interact and are supported.

With the increase in spatial functions and methods of analysis in R (e.g., spatstat, Baddeley and Turner 2013), python (e.g., PySAL, Rey and Anselin 2010), and C++ (e.g., GDAL, Warmerdam 2008), as well as the availability to work with open source GISystems (e.g., QGIS, Graser 2016), GIS and ABM integration will comprise an assemblage of libraries rather than interacting platforms (e.g., Agent Analyst; Choi and Lee 2009). This movement is required since the most widely used ABM development platforms (e.g., NetLogo, Wilensky 1999; Repast, Collier 2003) have support for GIS data (Liebert et al. 2008), but lack the fluidity of working with spatial data in R or Python and are restricted in the size of data with which their models can operate (e.g., Sturley et al. 2018). Contemporary tools like PyNetLogo (Jaxa-Rozen 2018) and RNetLogo (Thiele 2014) provide wrapper functions to facilitate a more flexible integration between spatial data and processes with ABMs, the workflow and knowledge requirements remain inhibiting to novice modellers and coders.

In addition to the benefits of integration of GIS and ABM for scientific applications, are insights into data collection and representation. While the former is relatively straightforward, whereby the incorporation of process and behaviour identifies a gap in our empirical data and collection approaches, the latter may contribute to theory and how data is represented in the future. For example, a historical challenge among users of spatial data resides in the selection of how to conceptualize spatial data model (e.g., field- and object-based, Bian 2000; relative and absolute space, Couclelis 1997). However, there are unknown representations of data, yet to be discovered, that may fuse these models with process representations to generate a data model that is more continuous, fluid, and overcomes the spatio-temporal limitations of current GISystems.

Lastly, it is also worth noting that ABMs not only offer a descriptive modelling approach but have also been used as a solution modelling approach known as software agents (e.g., Sengupta and Sieber 2007). In this case, there are a variety of standard spatial data functions derived from equation-based approaches that could be compared to agent-based solutions and with observations similar to algorithmic comparisons (e.g., slope, Jones 1998).

4.3 Representing Commercial Land-use Behaviours in Land Use Models

Land use change (LUC) modelling is widely used across a variety of disciplines to understand human-environment interactions (e.g., da Silva et al. 2016), evaluate the influences of land transitions on local ecological function (e.g., carbon storage, Robinson 2009) and Earth system processes (e.g., climate change) (Dale 1997), inform policy makers and increase their capacity for decision making (e.g., Zellner 2007), and plays a critical role among a variety of other social, economic, and environmental scientific investigations. While statistical models of land-use change have been the historical approach of choice (e.g., Arowolo and Deng 2018), contemporary LUC models have used cellular automata (CA, Stevens and Dragićević 2007) and agent-based modelling (ABM, Rounsevell et al. 2012a) approaches to advance the science of LUC modelling.

Despite the role of commercial land development as a driver of employment (Fragkias and Geoghegan 2010), economic development (e.g., Sun et al. 2016), and residential land development (e.g., Jjumba and Dragićević 2012). Commercial land development is one of the most important LUCs in the past decades (Verburg 2004a); however, few models have been built to explicitly simulate actor behaviours within the commercial system, which is a clear gap within the land use modelling community. A large proportion of existing LUC studies, which take commercial land or development into consideration, quantify the impacts of driving factors from socio-economic, demographic, or geographic disciplines on the possibility of commercial land development using statistical models (e.g., multinomial logit, Fragkias and Geoghegan 2010; logistic regression, Braimoh and Onishi 2007). However, these approaches lack the behavioural representation required to evaluate different perturbations to the system such as thresholds induced by policy (e.g., Gollnow and Lakes 2014), non-linear feedbacks induced by changing development densities (e.g., Marshall et al. 2016), or changing societal preferences due to social norms and learning (e.g., Xu et al. 2020).

To overcome the limitations of statistical approaches CA and ABM-based LUC models have been created, some of which consider commercial land use agents, to answer questions such as the influence of different planning policies on LUC patterns (Jjumba and Dragićević 2012), neighbourhood interactions or characteristics of LUC (Verburg et al. 2004b), and measurements of urban sprawl (Sun et al. 2007) and urban renewal (Zheng et al. 2015). Many of these models are spatial explicit and have included the interactions between commercial and alternative land use classes, however, the internal competition between commercial agents has rarely been accounted (e.g., Bone et al. 2011).

The RC-ABM, as a spatially explicit model focused on evaluating retail competition within a realistic landscape has the capacity to be expanded to address a host of new types of questions. For example, the timing of one or more store entries into the market, how format (small versus big box) enables success and changes with market maturation, and how land use change affects store revenue acquisition over time have little-to-know presence in the literature. Furthermore, characteristics on both the consumer and retail actors can be deepened by representing social norms, preferences, and loyalty on the consumer side as well as dynamic service areas, pricing strategies, and market segmentation or micromarketing on the retail store side. Lastly, while the presented research was operationalized at the census metropolitan level, frameworks for generating large scale behavioural models that act out across large spatial extents (e.g., provincial and national) are currently under development (e.g., Murray-Rust et al. 2014) and can facilitate regional assessments of retail competition and success.

5. Conclusion

The RC-ABM was operationalized for the Region of Waterloo using local and observed consumer expenditures and store location information. Our results could provide utility to existing home-improvement stores, alternative retail companies aiming to enter the market, and local-economic development teams. For example, our invariant-variant maps (*Section 3.3*) and quantification of drivers of success (*Section 3.4*) could be used as a tool for existing home-improvement stores to qualitatively evaluate the relative impact of their location against their merchandise assortment, advertising, and store atmosphere (Ghosh and MacLafferty 1987). New market entrants may use those same outputs to identify market saturation and that entry will require highly competitive pricing, sales, and marketing to out-compete established stores. Finally, local-economic development teams, may use the aforementioned outputs to identify complementary firms and economic opportunities rather than put effort into a saturated home-improvement market.

The bottom-up individual-based structure of the RC-ABM has the capacity to dynamically monitor changes in market share, revenue acquisition, competitive stress, and store entry and exit in every timestep, which cannot be done using equation-based modelling (Parunak et al. 1998). Therefore, the RC-ABM is capable of generating a large volume of data that can be analyzed to better understand the drivers of retail success and the impacts of location based and other retail strategy decisions of a host store as well as of its competitors. An interesting output of the approach is the endogenous generation of retail clustering, which has both theoretical (e.g., Hotelling 1929) and empirical support (e.g., Sevtsuk 2014). In development of this simple model, we balanced parsimony and realism (Manson 2007). However, a multitude of extensions are possible, such as dynamic generation of consumer expenditures. To facilitate these extensions the RC-ABM has been made available on COMSES.net (Zhang and Robinson 2021) for others to further advance the science associated with land use change and retail location

Chapter 3 Future Directions

1. Integrated Urban System Modelling

The urban system can be conceptualized as a complex system (Barros and Sobreira 2002) comprising a variety of land uses and complicated human activities. The interactions among different land users (due in part to their different land-use types) and the unpredictability of human behaviours create a system of actors and interactions that is challenging to accurately represent in a computational model (Carley 2009) and very difficult to validate against observational data (Shelton et al. 2018). Since the 1970s, researchers have developed different modelling techniques to gain insight about urban systems and through this process increase decision making capacity associated with policy generation and master planning (e.g., Long et al. 2009). While Chapter 2 presented a retail competition agent-based model (RC-ABM) and highlighted the extensions of an ABM approach over a traditional spatial interaction model, the focus was situated strictly on competitive site selection behaviour. As previously noted, behavioural models of retail and commercial actors are not typically represented in ABMs. However, the RC-ABM may be integrated into other land-use change models or other modelling frameworks focused on the urban system (e.g., ILUTE, Miller et al. 2004; MATSim, Horni et al. 2012).

An integrated modelling framework, utilizing an agent-based approach empirically informed with substantial data, is the Integrated Land Use, Transportation, Environment (ILUTE) model, which is a microsimulation modelling framework developed by a research team at the University of Toronto (Perveen et al. 2017). According to Miller et al. (2004) and Salvini and Miller (2005), in the ILUTE model, interactions between residential and commercial housing agents are emphasized. These agents are represented abstractedly, whereby, a residential agent could represent a household or an individual and the commercial agent could be a firm. Agent behaviours and decision-makings are represented by encapsulating other modelling methods (e.g., random utility choice models) and statistically analyzing the empirical data. Moreover, agent interactions are established by incorporating realistic spatial transportation data (i.e., road networks, public transit lines, and commuting data), which increases the fidelity of the ILUTE model and enables its use for real-world applications.

The ILUTE model has been used to simulate urban commercial real estate market dynamics and validated against empirical data (Rosenfield et al. 2013). However, given that the home-improvement retail market serves a variety of residential activities including home renovation, decoration, and construction which are correlated with the housing market (e.g., Balulescu 2015) and the current version of the ILUTE model does not consider the home-improvement commercial land use, the incorporation of the RC-ABM into the ILUTE could increase the breadth of applications and the types of research questions the framework can answer. For example, in the ILUTE model, if the residential agents determine to enter the housing market, they will have a chance to successfully purchase a new house and relocate. The RC-ABM can be utilized as a sub-model. The commercial retailers (home-improvement retailers) could interact with the residential agents, and the geographical characteristics (location, size) of the retailer agents should be represented on the model's virtual landscape. Inclusion of retail behaviour would endogenize a critical driver of home selection and market valuation related to shopping amenities as well as enable the ILUTE to represent many of the concepts formalized in planning literature such as leap frogging (Andris et al. 2018).

While inclusion of the RC-ABM within can be mostly preserved in the ILUTE framework, consideration of integration suggests several improvements that would enable new investigations into how retail affects and is affected by land use change and transportation. For example, the housing market offers the potential to generate a typology of home-improvement retail agents associated with those making recent purchases versus those remaining at their home. In this example, the new home purchases in older neighbourhoods could have increased allocation of their income to home improvement purchases relative to those acquiring new-build homes. Moreover, as the residential agents in the ILUTE model also are heterogenous in vehicle ownerships, which means their tolerance of the distance to stores might be constrained by their mobilities (Ghosh and MacLafferty 1987). Therefore, residential agents could have different ranges of maximal acceptable commuting distance, e.g., a longer distance for consumers with vehicle and a shorter distance for those without vehicle (Ghosh and MacLafferty 1987), to better address the consumer's behaviors in estimating the potential stores' attractiveness.

Overall, we believe that the ILUTE model's structure and its advanced design in integrating urban land use with transportation networks offers an opportunity to better understand the effects of large-scale retail competition on transportation and land use change as well as how these processes affect retail success.

2. RC-ABM – Model Design

There are two major components of the RC-ABM, the store and consumer agents. To advance the model towards a more wholistic representation of commercial land use change and activities would not only fill a gap in large-scale land use change modelling, but it would also enable new scientific contributions to business and retail sciences. In the remainder of this section, I note areas that could move the RC-ABM to a more wholistic representation of a commercial or retail agent and in doing so touch on some of the shortcomings of the initial version of the model presented in Chapter 2.

Retail store agents. The store agent, as the driving force of the competition process, can be regarded as the most important part in the RC-ABM. In Chapter 2, the store agent comprises only two characteristics: store size and geographic location. A more sophisticated representation of store characteristics may expand the opportunity for model use. For example, store format has been realized as an important factor that affects consumer behaviours. Many studies evaluate the effect of asymmetric inter- and intra-store format competition (e.g., Gonzalez-Benito et al. 2005, Cleeren et al. 2010). Classification of store format is not a difficult task as store size is commonly used to classify store formats as big-box stores or small-format stores (e.g., Balulescu 2015). After grouping store agents by formats, more complicated behavioral rules can be applied to describe a more realistic relationships between consumer and store agents. For example, consumer agents have heterogeneous price expectations to different formats (e.g., size) of stores (Koschmann and Isaac 2018).

Another potential area for improvement of store agent behaviour is associated with comes from its financial and operation strategy. In Chapter 2, retail store agents receive money from consumer agents and stay or exit the market based on a probabilistic function (Equation 8) of revenue density rank and store size. In the real-world, the store's exiting decision is a composition of complex considerations and is not determined by only the financial incompetence that extend beyond financial performance and include strategic decisions, qualitative reasoning, and for retail chains regional representation and success. Therefore, retail companies may not close a store even it has experienced revenue losses in successive years due to other executive considerations such as first to market, generating customer loyalty, avoiding bad publicity, anticipated future revenues, and among other reasons, for marketing purposes associated with visibility. Therefore, it is possible to represent a more accurate and nuanced decision-making strategy associated with a store's market exist decision.

The first solution is to improve upon the representation of the market existing rule deployed by retail store agents. The financial well-being of a store could be better represented by including net revenue (gross revenue - expense) as the determinant factor of store failure because it captures store profits and losses than the revenue density. We opted not to take this approach in Chapter 2 because store expense data like labor and utility costs are difficult to acquire because they are highly confidential and there is poor availability and lack of archival data (e.g., Thomas et al. 1998). In the absence of these data, alternative solutions may be possible with the creation of spatio-temporal data about store closures, openings, and other site and situation characteristics. With these data it may be possible to thresholds for competition and expenditure that when crossed lead to store closure similar to how Robinson et al. (2012) estimated the number of households associated with low-, medium-, and high-density residential land uses. With these spatio-temporal data, the relationship between a store's closure and its site and situation characteristics could be analyzed using advanced statistical methods (e.g., machine learning) to develop an existing rule that accounts for more drivers of closure than was incorporated in the RC-ABM.

Consumer agents. The consumer agent is where the store revenue comes from, therefore the allocation of consumer agents and their expenditures are critical to the performance of RC-ABM. The current version of RC-ABM distributes consumer expenditure, estimated at the CDA level, evenly to all residential land use (LU) parcels that reside within it. This simplification could be ameliorated by at least two methods. Firstly, the residential LU parcel data could be further classified as low-, medium-, and high-density residential parcels based on the residential building type and parcel size (Sun and Robinson 2018), which means the number of expenditures of each consumer agent could also be apportioned by sub-residential land-use type. Specifically, the high-density residential parcels typically have much more consumers to live within than the medium- and low-density parcels. Therefore, consumer agents in different types of residential parcels could have diverse weights of the expenditure allocation. Through this approach, the heterogeneity of consumer agents can be better represented, and the entire consumer market landscape might be re-established.

In addition to reclassifying consumer agents using the LU parcels, the high-precision residential land cover (LC) data could be helpful for a more accurate spatial distribution of consumer expenditures. Unlike the residential LU parcel data which contains not only residential buildings but other LCs like greenspace (Sun and Robinson 2018), the LC data can exclusively tell us where and what type the residential building is, which may lead to a higher accuracy of consumer expenditure allocations. However, this approach is recommended for studies with relatively small spatial scales because accurately extracting residential LC data from aerial images is a labor-intense and time-consuming task, which might be significantly challenging for large-scale land use modelling.

If the RC-ABM is further developed to include not only the home-improvement but other types of retailers like grocery stores, the original locations of consumers might include both residential and workplaces (Sturley et al. 2018). Redistributing a portion of consumer expenditures from residential locations to workplace zones (e.g., Heppenstall et al. 2013) could represent trip chaining (e.g., Brooks et al. 2008) and other related concepts that could better reflect the potential impact of store location choice on revenue generation.

Agent interactions. Compared to traditional modelling approaches, one of the major advantages of agent-based modelling is its ability to represent interactions among agents. There exist three types of agent interactions in the RC-ABM, including the consumer-store, consumer-consumer, and store-store interactions, that might be developed in the subsequent models.

The consumer agents interact with store agents by evaluating each store's attractiveness and proportionally spending money at different stores as a function of their distance and size, which collectively represent a store's attractiveness to potential consumers. The influence of distance is set to twice that of store size based on settings from the original Huff's model (Huff 1962). However, if we treat the weights on distance and store size as adjustable parameters then they can be calibrated to our study area (e.g., Suárez-Vega et al. 2015), which may enhance model performance and its application to alternative case studies in other similar regions in North America. Moreover, instead of only using the outlet size and distance, the parking lot size, merchandise average price, and the store atmosphere may also become potential supplement factors to be considered in the utility function.

Distance interactions between consumer and store agents are constrained to those consumers residing within a 12-min drive-time service area (SA). However, for simplicity, the RC-ABM uses Euclidean distance, rather than street-network distance, which was chosen due to the relatively large number of agents, complicated transportation network, and the limited running capacity of NetLogo. By integrating RC-ABM and some transportation network ABMs (e.g., Démare et al. 2017), it is possible to achieve the use of realistic commuting distances and travel impedances at the cost of run time and computational load. Therefore, such attempts should be started from small-scale studies.

An additional interactive action that could be incorporated between consumer and store agents involves representing consumer loyalty. With an agent-based approach, modelers can create agents that have learning behaviors (e.g., Kirman 2011), like brand loyalty shopping. Incorporating brand loyalty creates an opportunity to examine to what degree store success is correlated with different levels of consumer loyalty shopping behaviors and if a brand is first to market. Specifically, the consumer agent can be programmed to have experiential learning and dynamically adjust its preferences for different stores based on its experiences. To achieve this, alternative techniques like machine learning methods can be utilized as an embedded model (Abdulkareem et al. 2019) to enhance the consumer agent decision-making rules.

A better store performance (higher product quality and better advertising strategies) may increase the consumer's brand loyalty (Bell and Mgbemena 2018) (positive feedbacks from store-consumer interactions), whereas the consumer-consumer interaction (word-of-mouth) may decrease consumers' loyalty and makes their behaviors more complex and unpredictable (Zhang and Zhang 2007). The RC-ABM does not account for interactions among consumer agents, but could, by adding the effect of neighborhood similarity or social norms. Technically, the utility function (Equation 2, Chapter 2), used to estimate store's attractiveness could be extended by adding a neighborhood similarity score as follows:

$$utility = f(D, A, NS) \qquad (13)$$

where *D* is the inverse distance, *A* is the store size, and *NS* is the neighborhood similarity index.

The store-store interaction is indirectly defined in the RC-ABM, e.g., through the retail competition process. However, we do not construct any adaptive behaviors related to the competition, which might become another potential research objective: the comparison between effects of store's different adaptive behaviors on the resultant spatial distribution of stores. Specifically, three types of store's adaptive behaviors can be defined: (1) stores randomly select locations; (2) stores prefer locations which are away from competitors; (3) stores choose locations which are near to competitors. Based on these, we can artificially define store agents with different behavioral rules and evaluate whether some emergent patterns can be generated from the model run.

Overall, it is my hope that the RC-ABM is extended in the directions proposed in this chapter or alternative directions by others such that it may become a more advanced and flexible tool that is used to solve broader research questions.

References

Abdulkareem, S.A., Mustafa, Y.T., Augustijn, E.W. and Filatova, T., 2019. Bayesian networks for spatial learning: a workflow on using limited survey data for intelligent learning in spatial agent-based models. *Geoinformatica*, *23*(2), pp.243-268.

Agrawal, D. and Schorling, C., 1996. Market share forecasting: An empirical comparison of artificial neural networks and multinomial logit model. *Journal of Retailing*, *72*(4), pp.383-408.

An, L., Grimm, V., Robinson, D., Sullivan, A., Turner II, B.L., Wang, Z., Ye, X., Huang, R., Lindvist, E., Liu, J., Malleson, N., Heppenstall, A., Vincenot, C., Tang, W., and Radchuk, V., (in preparation). The new science of agent-based complex systems in an era of data science and artificial intelligence.

Andresen, M.A. and Malleson, N., 2011. Testing the stability of crime patterns: Implications for theory and policy. *Journal of Research in Crime and Delinquency*, *48*(1), pp.58-82.

Andris, C., Liu, X. and Ferreira Jr, J., 2018. Challenges for social flows. *Computers, Environment and Urban Systems*, *70*, pp.197-207.

Anzola, D., 2019. Knowledge transfer in agent-based computational social science. *Studies in History and Philosophy of Science Part A*, *77*, pp.29-38.

Applebaum, W., 1966. Methods for determining store trade areas, market penetration, and potential sales. *Journal of marketing Research*, *3*(2), pp.127-141.

ArcMap, n.d.a. *Location-allocation analysis*, viewed 20 November 2020, <https://desktop.arcgis.com/en/arcmap/latest/extensions/network-analyst/location-allocation.htm>.

ArcMap. n.d.b. *How Original Huff Model works*, viewed 20 November 2020, <https://desktop.arcgis.com/en/arcmap/10.3/tools/business-analyst-toolbox/how-original-huff-model-works.htm#:~:text=The%20Huff%20Model%20is%20an,and%20attractiveness%20of%20competing%20sites>.

Arowolo, A.O. and Deng, X., 2018. Land use/land cover change and statistical modelling of cultivated land change drivers in Nigeria. *Regional environmental change*, *18*(1), pp.247-259.

Arsanjani, J.J., Helbich, M. and de Noronha Vaz, E., 2013. Spatiotemporal simulation of urban growth patterns using agent-based modeling: The case of Tehran. *Cities*, *32*, pp.33-42.

Axtell, R., Axelrod, R., Epstein, J.M. and Cohen, M.D., 1996. Aligning simulation models: A case study and results. *Computational & mathematical organization theory*, *1*(2), pp.123-141.

Baddeley A, Turner R. 2013. Spatstat: spatial point pattern analysis, model-fitting, simulation, tests. *R Foundation for Statistical Computing*, Vienna, Austria.

Balci, O., 1994. Validation, verification, and testing techniques throughout the life cycle of a simulation study. *Annals of operations research*, *53*(1), pp.121-173.

Balulescu, A.M., 2015. *Estimating retail market potential using demographics and spatial analysis for home improvement in Ontario* (Master's thesis, University of Waterloo).

Baray, J. and Cliquet, G., 2013. Optimizing locations through a maximum covering/p-median hierarchical model: Maternity hospitals in France. *Journal of Business Research*, *66*(1), pp.127-132.

Barros, J. and Sobreira, F., 2002. City of Slums: self-organisation across scales. *UCL Working Papers Series*, *55*, pp.1-10.

Baydar, C., 2003. Agent-based Modeling and simulation of store performance for personalized pricing. *Conference Paper*.

Bell, D. and Mgbemena, C., 2018. Data-driven agent-based exploration of customer behavior. *Simulation*, *94*(3), pp.195-212.

Bernhardt, K., 2007. Agent-based modeling in transportation. *Artificial Intelligence in Transportation*, *72*.

Bian, L., 2000. Object-oriented representation for modelling mobile objects in an aquatic environment. *International Journal of Geographical Information Science*, *14*(7), pp.603-623.

Bianchi, C., Cirillo, P., Gallegati, M. and Vagliasindi, P.A., 2007. Validating and calibrating agent-based models: a case study. *Computational Economics*, *30*(3), pp.245-264.

Bone, C., Dragicevic, S. and White, R., 2011. Modeling-in-the-middle: bridging the gap between agent-based modeling and multi-objective decision-making for land use change. *International Journal of Geographical Information Science*, *25*(5), pp.717-737.

Braimoh, A.K. and Onishi, T., 2007. Spatial determinants of urban land use change in Lagos, Nigeria. *Land use policy*, *24*(2), pp.502-515.

Briggs, D.J., Gulliver, J., Fecht, D. and Vienneau, D.M., 2007. Dasymetric modelling of small-area population distribution using land cover and light emissions data. *Remote sensing of Environment*, *108*(4), pp.451-466.

Brooks, C.M., Kaufmann, P.J. and Lichtenstein, D.R., 2008. Trip chaining behavior in multi-destination shopping trips: A field experiment and laboratory replication. *Journal of Retailing*, *84*(1), pp.29-38.

Brown, D.G., Page, S., Riolo, R., Zellner, M. and Rand, W., 2005. Path dependence and the validation of agent-based spatial models of land use. *International journal of geographical information science*, *19*(2), pp.153-174.

Caradima, B., 2015. Multi-criteria suitability analysis and spatial interaction modeling of retail store locations in Ontario, Canada.

Carley, K.M., 2009. Computational modeling for reasoning about the social behavior of humans. *Computational and Mathematical Organization Theory*, *15*(1), pp.47-59.

CBRE, 2019. *Canada Retail Market Report H1 2019*, viewed 15 November 2020, <https://www.cbre.ca/en/research-and-reports/Canada-Retail-Market-Report-H1-2019>.

Chen, S.C., 2015. Customer value and customer loyalty: Is competition a missing link?. *Journal of retailing and consumer services*, *22*, pp.107-116.

Chen, X., 2017. Take the edge off: A hybrid geographic food access measure. *Applied Geography*, *87*, pp.149-159.

Choi, J. and Lee, J., 2009. 3D geo-network for agent-based building evacuation simulation. In *3D geo-information sciences* (pp. 283-299). Springer, Berlin, Heidelberg.

Ciari, F., Löchl, M. and Axhausen, K.W., 2008. Location decisions of retailers: An agent-based approach. *Arbeitsberichte Verkehrs-und Raumplanung*, *499*.

Claeys, A.S., Cauberghe, V. and Leysen, J., 2013. Implications of stealing thunder for the impact of expressing emotions in organizational crisis communication. *Journal of Applied Communication Research*, *41*(3), pp.293-308.

Clarke, I., 2000. Retail power, competition and local consumer choice in the UK grocery sector. *European Journal of Marketing*.

Clarke, I., Hallsworth, A., Jackson, P., Kervenoael, R.D., Perez-del-Aguila, R. and Kirkup, M., 2004. Retail competition and consumer choice: contextualising the. *International Journal of Retail & Distribution Management*, *32*(2), pp.89-99.

Clarke, I., Mackaness, W. and Ball, B., 2003. Modelling intuition in retail site assessment (MIRSA): making sense of retail location using retailers' intuitive judgements as a support for decision-making. *The International Review of Retail, Distribution and Consumer Research*, *13*(2), pp.175-193.

Clarke, M. and Birkin, M., 2018. Spatial Interaction Models: From Numerical Experiments to Commercial Applications. *Applied Spatial Analysis and Policy*, *11*(4), pp.713-729.

Clawson, C.J., 1974. Fitting Branch Locations, Performance Standards, and Marketing Strategies to Local Conditions: A Straightforward Approach to Establishing Goals and Reevaluating Marketing Strategies for Individual Branch Locations. *Journal of Marketing*, *38*(1), pp.8-14.

Cleeren, K., Verboven, F., Dekimpe, M.G. and Gielens, K., 2010. Intra-and interformat competition among discounters and supermarkets. *Marketing science*, *29*(3), pp.456-473.

Colla, E. and Dupuis, M., 2002. Research and managerial issues on global retail competition: Carrefour/Wal-Mart. *International Journal of retail & distribution management*.

Collier, N., 2003. Repast: An extensible framework for agent simulation. *The University of Chicago's Social Science Research*, *36*, p.2003.

Couclelis, H., 1997. From cellular automata to urban models: new principles for model development and implementation. *Environment and planning B: Planning and design*, *24*(2), pp.165-174.

Crooks, A.T. and Heppenstall, A.J., 2012. Introduction to agent-based modelling. In *Agent-based models of geographical systems* (pp. 85-105). Springer, Dordrecht.

da Silva, R.F., Batistella, M. and Moran, E.F., 2016. Drivers of land change: Human-environment interactions and the Atlantic forest transition in the Paraíba Valley, Brazil. *Land Use Policy*, *58*, pp.133-144.

Dale, V.H., 1997. The relationship between land-use change and climate change. *Ecological applications*, *7*(3), pp.753-769.

Davies, K., 2000. The Asian economic recession and retail change: the implications for retailer strategies in Asia. *The International Review of Retail, Distribution and Consumer Research*, *10*(4), pp.335-353.

De Marchi, S. and Page, S.E., 2014. Agent-based models. *Annual Review of political science*, *17*, pp.1-20.

De Marchi, S., 2005. *Computational and mathematical modeling in the social sciences*. Cambridge University Press.

Démare, T., Bertelle, C., Dutot, A. and Lévêque, L., 2017. Modeling logistic systems with an agent-based model and dynamic graphs. *Journal of Transport Geography*, *62*, pp.51-65.

Dibb, S. and Simkin, L., 1994. Retail Store Performance Model the Impact of Own-branch 'Friendly Fire'. *International Transactions in Operational Research*, *1*(4), pp.479-487.

Dooley, K., 1996. Complex adaptive systems: A nominal definition. *The Chaos Network*, *8*(1), pp.2-3.

Dramowicz, E., 2005. Retail trade area analysis using the Huff model. *Directions Magazine, Jul*, *2*.

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

Durvasula, S., Sharma, S. and Andrews, J.C., 1992. Storeloc: A retail store location model based on managerial. *Journal of Retailing*, *68*(4), p.420.

Eisenmann, T.R., 2006. Internet companies' growth strategies: determinants of investment intensity and long-term performance. *Strategic Management Journal*, *27*(12), pp.1183-1204.

Epstein, J.M. and Axtell, R., 1996. *Growing artificial societies: social science from the bottom up*. Brookings Institution Press.

Feng, J.B., 2013. What's Up with US Big-Box Retailers in China? The Cases of The Home Depot and Best Buy. *China Currents, 12*(2).

Fitzpatrick, B. and Martinez, J., 2012. Agent-based modeling of ecological niche theory and assortative drinking. *Journal of Artificial Societies and Social Simulation*, *15*(2), p.4.

Fragkias, M. and Geoghegan, J., 2010. Commercial and industrial land use change, job decentralization and growth controls: a spatially explicit analysis. *Journal of Land Use Science*, *5*(1), pp.45-66.

Gagliano, K.B. and Hathcote, J., 1994. Customer expectations and perceptions of service quality in retail apparel specialty stores. *Journal of Services Marketing*.

Game, E.T., Meijaard, E., Sheil, D. and McDonald-Madden, E., 2014. Conservation in a wicked complex world; challenges and solutions. *Conservation Letters*, *7*(3), pp.271-277.

Gaskill, L.R., Van Auken, H.E. and Manning, R.A., 1993. A factor analytic study of the perceived causes of small business failure. *Journal of small business management*, *31*, pp.18-18.

Gaudou, B., Lang, C., Marilleau, N., Savin, G., Coyrehourcq, S.R. and Nicod, J.M., 2017. Netlogo, an open simulation environment. In *Agent-based Spatial Simulation with NetLogo, Volume 2* (pp. 1-36). Elsevier.

Gauvin, L., Vignes, A. and Nadal, J.P., 2013. Modeling urban housing market dynamics: can the socio-spatial segregation preserve some social diversity?. *Journal of Economic Dynamics and Control*, *37*(7), pp.1300-1321.

Gehlot, H., Zhan, X., Qian, X., Thompson, C., Kulkarni, M. and Ukkusuri, S.V., 2019. A-RESCUE 2.0: A High-Fidelity, Parallel, Agent-Based Evacuation Simulator. *Journal of Computing in Civil Engineering*, *33*(2), p.04018059.

Ghosh, A. and MacLafferty, S.L., 1987. *Location strategies for retail and service firms*. Lexington Books.

Gilbert, N. and Troitzsch, K., 2005. *Simulation for the social scientist*. McGraw-Hill Education (UK).

Gimblett, H.R. ed., 2002. *Integrating geographic information systems and agent-based modeling techniques for simulating social and ecological processes*. Oxford University Press.

Gollnow, F. and Lakes, T., 2014. Policy change, land use, and agriculture: The case of soy production and cattle ranching in Brazil, 2001–2012. *Applied Geography*, *55*, pp.203-211.

Gonzalez-Benito, O., 2005. Spatial competitive interaction of retail store formats: modeling proposal and empirical results. *Journal of Business Research*, *58*(4), pp.457-466.

Gonzalez-Benito, O., Munoz-Gallego, P.A. and Kopalle, P.K., 2005. Asymmetric competition in retail store formats: Evaluating inter-and intra-format spatial effects. *Journal of Retailing*, *81*(1), pp.59-73.

Goodacre, A., 2003. Operating lease finance in the UK retail sector. *The International Review of Retail, Distribution and Consumer Research*, *13*(1), pp.99-125.

Goodchild, M.F., 1984. Lacs: A location-allocation mode for retail site selection. *Journal of retailing*, *60*(1), pp.84-100.

Gordon, P. and Richardson, H.W., 1997. Are compact cities a desirable planning goal?. *Journal of the American planning association*, *63*(1), pp.95-106.

Gorochowski, T.E., Matyjaszkiewicz, A., Todd, T., Oak, N., Kowalska, K., Reid, S., Tsaneva-Atanasova, K.T., Savery, N.J., Grierson, C.S. and Di Bernardo, M., 2012. BSim: an agent-based tool for modeling bacterial populations in systems and synthetic biology. *PloS one*, *7*(8).

Gowrishankkar, V., 2017. A strategic importance of store ambience for retaining customers. *International Journal of Management Research and Reviews*, *7*(2), p.129.

Graser, A., 2016. *Learning Qgis*. Packt Publishing Ltd.

Green, D.H., Barclay, D.W. and Ryans, A.B., 1995. Entry strategy and long-term performance: Conceptualization and empirical examination. *Journal of marketing*, *59*(4), pp.1-16.

Grewal, D., Levy, M. and Kumar, V., 2009. Customer experience management in retailing: An organizing framework. *Journal of retailing*, *85*(1), pp.1-14.

Griffith, D.A., 1983. The boundary value problem in spatial statistical analysis. *Journal of regional science*, *23*(3), pp.377-387.

Grimm, V. and Railsback, S.F., 2005. *Individual-based modeling and ecology* (Vol. 8). Princeton university press.

Guruswamy, M., Sharma, K., Mohanty, J.P. and Korah, T.J., 2005. FDI in India's Retail Sector: More Bad than Good?. *Economic and Political Weekly*, pp.619-623.

He, Z., Wang, S. and Cheng, T.C.E., 2013. Competition and evolution in multi-product supply chains: An agent-based retailer model. *International Journal of Production Economics*, *146*(1), pp.325-336.

Heppenstall, A.J., Evans, A.J. and Birkin, M.H., 2005. A hybrid multi-agent/spatial interaction model system for petrol price setting. *Transactions in GIS*, *9*(1), pp.35-51.

Heppenstall, A.J., Evans, A.J. and Birkin, M.H., 2007. Genetic algorithm optimisation of an agent-based model for simulating a retail market. *Environment and Planning B: Planning and Design*, *34*(6), pp.1051-1070.

Heppenstall, A.J., Harland, K., Ross, A.N. and Olner, D., 2013. Simulating Spatial Dynamics and Processes in a Retail Gasoline Market: An Agent-Based Modeling Approach. *Transactions in GIS*, *17*(5), pp.661-682.

Hernandez, T. and Bennison, D., 2000. The art and science of retail location decisions. *International Journal of Retail & Distribution Management*.

Hernandez, T., Bennison, D. and Cornelius, S., 1998. The organisational context of retail locational planning. *GeoJournal*, *45*(4), pp.299-308.

Hoffman, A. and Gold, N., 2015. *Target corp's tarnished reputation: Failure in Canada and a massive data breach*. Rotterdam School of Management, Erasmus University.

Holmberg, S.R. and Morgan, K.B., 2004. Retail marketing channel franchise failure: A strategic management perspective and longitudinal analysis. *Journal of Marketing Channels*, *11*(2-3), pp.55-76.

Horni, A., Ciari, F. and Axhausen, K.W., 2012. Coupling customers' destination choice and retailers' location choice in MATSim. *Arbeitsberichte Verkehrs-und Raumplanung*, *808*.

Hortaçsu, A. and Syverson, C., 2015. The ongoing evolution of US retail: A format tug-of-war. *Journal of Economic Perspectives*, *29*(4), pp.89-112.

Hotelling, H., 1929. Stability in Competition. *Economic Journal*. 39, pp.41-57.

Hu, Y.C. and Ansell, J., 2007. Measuring retail company performance using credit scoring techniques. *European Journal of Operational Research*, *183*(3), pp.1595-1606.

Huang, Q., Parker, D.C., Filatova, T. and Sun, S., 2014. A review of urban residential choice models using agent-based modeling. *Environment and Planning B: Planning and Design*, *41*(4), pp.661-689.

Huff, D., 1962. Determination of intra-urban retail trade areas. *publication of Retail Estate Research Program*.

Huff, D.L., 1964. Defining and estimating a trading area. *Journal of marketing*, *28*(3), pp.34-38.

Janssen, M.A., 2020. *Introduction to Agent-based modeling: with applications to social, ecological, and social-ecological systems.*

Jaxa-Rozen, M. and Kwakkel, J.H., 2018. Pynetlogo: Linking netlogo with python. *Journal of Artificial Societies and Social Simulation*, *21*(2).

Jiaxin Zhang, Derek Robinson (2021, January 03). "Retail Competition Agent-based Model" (Version 2020.11.27). *CoMSES Computational Model Library*. Retrieved from: https://www.comses.net/codebases/1d3db47b-ee37-4fec-873a-fc2a587ca363/releases/2020.11.27/

Jjumba, A. and Dragićević, S., 2012. High resolution urban land-use change modeling: Agent iCity approach. *Applied Spatial Analysis and Policy*, *5*(4), pp.291-315.

Jones, K.H., 1998. A comparison of algorithms used to compute hill slope as a property of the DEM. *Computers & Geosciences*, *24*(4), pp.315-323.

Karande, K. and Lombard, J.R., 2005. Location strategies of broad-line retailers: an empirical investigation. *Journal of Business Research*, *58*(5), pp.687-695.

Keener, M.H., 2013. Predicting The financial failure of retail companies in the united states. *Journal of Business & Economics Research (JBER)*, *11*(8), pp.373-380.

Ken, J. and Simmons, J., 1993. *Location, Location, Location, Analyzing The Retail Environment*. Nelson Canada.

Kenter, P. and Elhorst, J.P., 2012. The initial value problem in spatial econometrics.

Kirman, A., 2011. Learning in agent-based models. *Eastern Economic Journal*, *37*(1), pp.20-27.

Klügl, F., 2008, March. A validation methodology for agent-based simulations. In *Proceedings of the 2008 ACM symposium on Applied computing* (pp. 39-43).

Koelemeijer, K. and Oppewal, H., 1999. Assessing the effects of assortment and ambience: a choice experimental approach. *Journal of Retailing*, *75*(3), pp.319-345.

Koschmann, A. and Isaac, M.S., 2018. Retailer categorization: how store-format price image influences expected prices and consumer choices. *Journal of Retailing*, *94*(4), pp.364-379.

Kosová, R. and Lafontaine, F., 2010. Survival and growth in retail and service industries: Evidence from franchised chains. *The Journal of Industrial Economics*, *58*(3), pp.542-578.

Krause-Traudes, M., Scheider, S., Rüping, S. and Meßner, H., 2008. Spatial data mining for retail sales forecasting. In *11th AGILE International Conference on Geographic Information Science* (pp. 1-11).

Ladle, J.K., Stiller, D. and Stiller, D., 2009. Retail site selection: a new, innovative model for retail development. *Cornell Real Estate Review*, *7*(1), p.17.

Lai, Y., Saridakis, G., Blackburn, R. and Johnstone, S., 2016. Are the HR responses of small firms different from large firms in times of recession?. *Journal of Business Venturing*, *31*(1), pp.113-131.

Lamberg, J.A. and Tikkanen, H., 2006. Changing sources of competitive advantage: cognition and path dependence in the Finnish retail industry 1945–1995. *Industrial and Corporate Change*, *15*(5), pp.811-846.

Lempert, R., 2002. Agent-based modeling as organizational and public policy simulators. *Proceedings of the National Academy of Sciences*, *99*(suppl 3), pp.7195-7196.

Liebert, K.D., Earnest, D.C. and Tolk, A., 2008, April. Using GIS vector data to build virtual environments for agent based models. In *Proceedings of the 2008 Spring simulation multiconference* (pp. 45-52).

Ligtenberg, A., van Lammeren, R.J., Bregt, A.K. and Beulens, A.J., 2010. Validation of an agent-based model for spatial planning: A role-playing approach. *Computers, Environment and Urban Systems*, *34*(5), pp.424-434.

Long, Y., Mao, Q. and Dang, A., 2009. Beijing urban development model: Urban growth analysis and simulation. *Tsinghua Science and Technology*, *14*(6), pp.782-794.

Lopes, R.H., Reid, I.D. and Hobson, P.R., 2007. The two-dimensional Kolmogorov-Smirnov test.

Macal, C.M. and North, M.J., 2009, December. Agent-based modeling and simulation. In *Proceedings of the 2009 Winter Simulation Conference (WSC)* (pp. 86-98). IEEE.

Manson, S.M., 2007. Challenges in evaluating models of geographic complexity. *Environment and Planning B: Planning and Design*, *34*(2), pp.245-260.

Marceau, D.J. and Benenson, I. eds., 2011. *Advanced Geo-Simulation Models*. Bentham Science Publishers.

Marshall, S., Sharley, D., Jeppe, K., Sharp, S., Rose, G. and Pettigrove, V., 2016. Potentially toxic concentrations of synthetic pyrethroids associated with low density residential land use. *Frontiers in Environmental Science*, *4*, p.75.

Matsui, K., 2018. When and what wholesale and retail prices should be set in multi-channel supply chains?. *European Journal of Operational Research*, *267*(2), pp.540-554.

Mayadunne, S., Johar, M. and Saydam, C., 2018. Competitive store closing during an economic downturn. *International Journal of Production Economics*, *199*, pp.162-178.

McArthur, E., Weaven, S. and Dant, R., 2016. The evolution of retailing: A meta review of the literature. *Journal of Macromarketing*, *36*(3), pp.272-286.

Mcgurr, P.T. and Devaney, S.A., 1998. A retail failure prediction model. *The International Review of Retail, Distribution and Consumer Research*, *8*(3), pp.259-276.

Merino, M. and Ramirez-Nafarrate, A., 2016. Estimation of retail sales under competitive location in Mexico. *Journal of Business Research*, *69*(2), pp.445-451.

Mestre, A.M., Oliveira, M.D. and Barbosa-Póvoa, A.P., 2015. Location–allocation approaches for hospital network planning under uncertainty. *European Journal of Operational Research*, *240*(3), pp.791-806.

Miller, E.J., Hunt, J.D., Abraham, J.E. and Salvini, P.A., 2004. Microsimulating urban systems. *Computers, environment and urban systems*, *28*(1-2), pp.9-44.

Miller, J.H. and Page, S.E., 2009. *Complex adaptive systems: An introduction to computational models of social life*. Princeton university press.

Millington, J., Romero-Calcerrada, R., Wainwright, J. and Perry, G., 2008. An agent-based model of Mediterranean agricultural land-use/cover change for examining wildfire risk. *Journal of Artificial Societies and Social Simulation*, *11*(4), p.4.

Miura, M. and Shiroishi, H., 2018, June. Agent-Based Modeling and Simulation for Two-Dimensional Spatial Competition. In *KES International Symposium on Agent and Multi-Agent Systems: Technologies and Applications* (pp. 67-75). Springer, Cham.

Moore, M. and Carpenter, J., 2006. The effect of price as a marketplace cue on retail patronage. *Journal of Product & Brand Management*.

Moore, M., 2005. Towards a confirmatory model of retail strategy types: an empirical test of Miles and Snow. *Journal of Business Research*, *58*(5), pp.696-704.

Mou, S., Robb, D.J. and DeHoratius, N., 2018. Retail store operations: Literature review and research directions. *European Journal of Operational Research*, *265*(2), pp.399-422.

Murray-Rust, D., Brown, C., van Vliet, J., Alam, S.J., Robinson, D.T., Verburg, P.H. and Rounsevell, M., 2014. Combining agent functional types, capitals and services to model land use dynamics. *Environmental modelling & software*, *59*, pp.187-201.

Nakaya, T., Fotheringham, A.S., Hanaoka, K., Clarke, G., Ballas, D. and Yano, K., 2007. Combining microsimulation and spatial interaction models for retail location analysis. *Journal of Geographical Systems*, *9*(4), pp.345-369.

Ng, T.L., Eheart, J.W., Cai, X. and Braden, J.B., 2011. An agent-based model of farmer decision-making and water quality impacts at the watershed scale under markets for carbon allowances and a second-generation biofuel crop. *Water Resources Research*, *47*(9).

Nicolis, G. and Nicolis, C., 2012. *Foundations of complex systems: emergence, information and predicition*. World Scientific.

Nisar, T.M. and Prabhakar, G., 2017. What factors determine e-satisfaction and consumer spending in e-commerce retailing?. *Journal of Retailing and Consumer Services*, *39*, pp.135-144.

Olsen, L.M. and Lord, J.D., 1979. Market area characteristics and branch bank performance. *Journal of Bank Research*, *10*(2), pp.102-110.

Önden, İ., Şen, C.G. and Şen, A., 2012. Integration of integer programming with GIS analyzing abilities for determining the convenience levels of retail stores. *Procedia-Social and Behavioral Sciences*, *62*, pp.1144-1149.

Ozuduru, B.H. and Varol, C., 2011. Spatial statistics methods in retail location research: A case study of Ankara, Turkey. *Procedia Environmental Sciences*, *7*, pp.287-292.

Page, S.E., 2018. *The model thinker: What you need to know to make data work for you*. Hachette UK.

Parunak, H.V.D., Savit, R. and Riolo, R.L., 1998, July. Agent-based modeling vs. equation-based modeling: A case study and users' guide. In *International Workshop on Multi-Agent Systems and Agent-Based Simulation* (pp. 10-25). Springer, Berlin, Heidelberg.

Patterson, W.C., 1993. First-mover advantage: the opportunity curve. *Journal of Management Studies*, *30*(5), pp.759-777.

Paul, J.A. and Wang, X.J., 2019. Robust location-allocation network design for earthquake preparedness. *Transportation research part B: methodological*, *119*, pp.139-155.

Perveen, S., Yigitcanlar, T., Kamruzzaman, M. and Hayes, J., 2017. Evaluating transport externalities of urban growth: a critical review of scenario-based planning methods. *International Journal of Environmental Science and Technology*, *14*(3), pp.663-678.

Pogson, M., Smallwood, R., Qwarnstrom, E. and Holcombe, M., 2006. Formal agent-based modelling of intracellular chemical interactions. *Biosystems*, *85*(1), pp.37-45.

Polhill, J.G., Izquierdo, L.R. and Gotts, N.M., 2005. The ghost in the model (and other effects of floating point arithmetic). *Journal of Artificial Societies and Social Simulation*, *8*(1).

Rabbani, M., Hosseini-Mokhallesun, S.A.A., Ordibazar, A.H. and Farrokhi-Asl, H., 2020. A hybrid robust possibilistic approach for a sustainable supply chain location-allocation network design. *International Journal of Systems Science: Operations & Logistics*, *7*(1), pp.60-75.

Radas, S. and Shugan, S.M., 1998. Seasonal marketing and timing new product introductions. *Journal of Marketing Research*, *35*(3), pp.296-315.

Rahman, S.U. and Smith, D.K., 2000. Use of location-allocation models in health service development planning in developing nations.

Railsback, S.F. and Grimm, V., 2019. *Agent-based and individual-based modeling: a practical introduction*. Princeton university press.

Railsback, S.F., Lytinen, S.L. and Jackson, S.K., 2006. Agent-based simulation platforms: Review and development recommendations. *Simulation*, *82*(9), pp.609-623.

Region of Waterloo, 2020. *Year-End 2019 Population and Household Estimates for Waterloo Region*, viewed 20 November 2020, <https://www.regionofwaterloo.ca/en/regional-government/resources/Census/DOCS_ADMIN-3260012-v3-PDL-CPL-20-11__Year-End_2019_Population_and_Household_Estimates_for_Waterloo_Region.pdf>.

Region of Waterloo, n.d.a. *Population*, viewed 20 November 2020, <https://www.regionofwaterloo.ca/en/regional-government/population.aspx#>.

Region of Waterloo, n.d.b. *Demographics*, viewed 20 November 2020, <https://www.regionofwaterloo.ca/en/doing-business/demographics.aspx>.

Reigadinha, T., Godinho, P. and Dias, J., 2017. Portuguese food retailers–Exploring three classic theories of retail location. *Journal of Retailing and Consumer Services*, *34*, pp.102-116.

Rey, S.J. and Anselin, L., 2010. PySAL: A Python library of spatial analytical methods. In *Handbook of applied spatial analysis* (pp. 175-193). Springer, Berlin, Heidelberg.

Rhodes, C. and Brien, P., 2014. The retail industry: statistics and policy. *House of Commons Library Briefing Paper, available at: http://researchbriefings. files. parliament. uk/documents/SN06186/SN06186. pdf (accessed 13th August, 2018)*.

Robinson, D.T. and Balulescu, A.M., 2018. Comparison of methods for quantifying consumer spending on retail using publicly available data. *International Journal of Geographical Information Science*, *32*(6), pp.1061-1086.

Robinson, D.T. and Brown, D.G., 2009. Evaluating the effects of land-use development policies on ex-urban forest cover: An integrated agent-based GIS approach. *International Journal of Geographical Information Science*, *23*(9), pp.1211-1232.

Robinson, D.T., 2009. *Effects of Land-Use Policy, Forest Fragmentation, and Residential Parcel Size on Land-Cover and Carbon Storage in Southeastern Michigan* (Doctoral dissertation).

Robinson, D.T., Brown, D.G., Parker, D.C., Schreinemachers, P., Janssen, M.A., Huigen, M., Wittmer, H., Gotts, N., Promburom, P., Irwin, E. and Berger, T., 2007. Comparison of empirical methods for building agent-based models in land use science. *Journal of land use science*, *2*(1), pp.31-55.

Robinson, D.T., Murray-Rust, D., Rieser, V., Milicic, V. and Rounsevell, M., 2012. Modelling the impacts of land system dynamics on human well-being: Using an agent-based approach to cope with data limitations in Koper, Slovenia. *Computers, Environment and Urban Systems*, *36*(2), pp.164-176.

Rosenfield, A., Chingcuanco, F. and Miller, E.J., 2013. Agent-based housing market microsimulation for integrated land use, transportation, environment model system. *Procedia Computer Science*, *19*, pp.841-846.

Rounsevell, M.D., Pedroli, B., Erb, K.H., Gramberger, M., Busck, A.G., Haberl, H., Kristensen, S., Kuemmerle, T., Lavorel, S., Lindner, M. and Lotze-Campen, H., 2012a. Challenges for land system science. *Land use policy*, *29*(4), pp.899-910.

Rounsevell, M.D., Robinson, D.T. and Murray-Rust, D., 2012b. From actors to agents in socio-ecological systems models. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1586), pp.259-269.

Russell, R.A. and Urban, T.L., 2010. The location and allocation of products and product families on retail shelves. *Annals of Operations Research*, *179*(1), pp.131-147.

Salvini, P. and Miller, E.J., 2005. ILUTE: An operational prototype of a comprehensive microsimulation model of urban systems. *Networks and spatial economics*, *5*(2), pp.217-234.

Schelling, T.C., 1971. Dynamic models of segregation. *Journal of mathematical sociology*, *1*(2), pp.143-186.

Schenk, T.A., Löffler, G. and Rauh, J., 2007. Agent-based simulation of consumer behavior in grocery shopping on a regional level. *Journal of Business research*, *60*(8), pp.894-903.

Schichl, H., 2004. Models and the history of modeling. In *Modeling languages in mathematical optimization* (pp. 25-36). Springer, Boston, MA.

Schmidt, C.G. and Lee, Y., 1979. INTER-METROPOLITAN ANALYSIS OF RETAIL LOCATIONAL PATTERNS AND COMPETITIVE PROCESSES. *Social Science Journal*, *16*(1), pp.43-52.

Schwarz, N. and Ernst, A., 2009. Agent-based modeling of the diffusion of environmental innovations—An empirical approach. *Technological forecasting and social change*, *76*(4), pp.497-511.

Scikit-Learn, n.d. *sklearn.preprocessing.MinMaxScaler*. viewed 20 November 2020, <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>.

Scriven, M., 1959. Explanation and prediction in evolutionary theory. *Science*, *130*(3374), pp.477-482.

Sengupta, R. and Sieber, R., 2007. Geospatial agents, agents everywhere... *Transactions in GIS*, *11*(4), pp.483-506.

Sevtsuk, A., 2014. Location and agglomeration: The distribution of retail and food businesses in dense urban environments. *Journal of Planning Education and Research*, *34*(4), pp.374-393.

Sharma, A. and Stafford, T.F., 2000. The effect of retail atmospherics on customers' perceptions of salespeople and customer persuasion:: An empirical investigation. *Journal of Business Research*, *49*(2), pp.183-191.

Shelton, R.E., Baeza, A., Janssen, M.A. and Eakin, H., 2018. Managing household socio-hydrological risk in Mexico city: A game to communicate and validate computational modeling with stakeholders. *Journal of environmental management*, *227*, pp.200-208.

Siebers, P.O., Aickelin, U., Celia, H. and Clegg, C.W., 2010. Simulating Customer Experience and Word-Of-Mouth in Retail—A Case Study. *Simulation*, *86*(1), pp.5-30.

Smith, A.K., 2017. *An evaluation of high-resolution land cover and land use classification accuracy by thematic, spatial, and algorithm parameters* (Master's thesis, University of Waterloo).

Spies, T.A., White, E., Ager, A., Kline, J.D., Bolte, J.P., Platt, E.K., Olsen, K.A., Pabst, R.J., Barros, A.M., Bailey, J.D. and Charnley, S., 2017. Using an agent-based model to examine forest management outcomes in a fire-prone landscape in Oregon, USA. *Ecology and Society*, *22*(1).

Srbljinović, A. and Škunca, O., 2003. An introduction to agent based modelling and simulation of social processes. *Interdisciplinary Description of Complex Systems: INDECS*, *1*(1-2), pp.1-8.

Statistics Canada, 2011. *2011 National Household Survey: Data tables*, viewed 20 November 2020, < https://www12.statcan.gc.ca/nhs-enm/2011/dp-pd/dt-td/Rp-eng.cfm?LANG=E&APATH=3&DETAIL=0&DIM=0&FL=A&FREE=0&GC=0&GID=0&GK=0&GRP=0&PID=105852&PRID=0&PTYPE=105277&S=0&SHOWALL=1&SUB=0&Temporal=2013&THEME=96&VID=0&VNAMEE=&VNAMEF=>

Statistics Canada, 2018a. *Canada GDP by Industry Report*, viewed 20 November 2020, <https://ycharts.com/indicators/reports/canada_gdp_by_industry>.

Statistics Canada, 2018b. *Dissemination area: Detailed definition*, viewed 20 November 2020, < https://www150.statcan.gc.ca/n1/pub/92-195-x/2011001/geo/da-ad/def-eng.htm>

Statistics Canada, 2020. *Archived - Retail store financial estimates, inactive* viewed 20 November 2020, <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=2010001001>.

Statistics Canada. Table 36-10-0225-01. Detailed household final consumption expenditure, provincial and territorial, annual (x 1,000,000); DOI: https://doi.org/10.25318/3610022501-eng

Steel, D.G. and Holt, D., 1996. Analysing and adjusting aggregation effects: the ecological fallacy revisited. *International Statistical Review/Revue Internationale de Statistique*, pp.39-60.

Stevens, D. and Dragićević, S., 2007. A GIS-based irregular cellular automata model of land-use change. *Environment and Planning B: Planning and Design*, *34*(4), pp.708-724.

Sturley, C., Newing, A. and Heppenstall, A., 2018. Evaluating the potential of agent-based modelling to capture consumer grocery retail store choice behaviours. *The International Review of Retail, Distribution and Consumer Research*, *28*(1), pp.27-46.

Suarez Alvarez, L., Díaz Martín, A.M. and Casielles, R.V., 2007. Relationship marketing and information and communication technologies: Analysis of retail travel agencies. *Journal of travel Research*, *45*(4), pp.453-463.

Suárez-Vega, R., Gutiérrez-Acuña, J.L. and Rodríguez-Díaz, M., 2015. Locating a supermarket using a locally calibrated Huff model. *International Journal of Geographical Information Science*, *29*(2), pp.217-233.

Sun, B. and Robinson, D.T., 2018. Comparison of Statistical Approaches for Modelling Land-Use Change. *Land*, *7*(4), p.144.

Sun, C., Sun, C., Yang, Z., Zhang, J. and Deng, Y., 2016. Urban land development for industrial and commercial use: a case study of Beijing. *Sustainability*, *8*(12), p.1323.

Sun, H., Forsythe, W. and Waters, N., 2007. Modeling urban land use change and urban sprawl: Calgary, Alberta, Canada. *Networks and spatial economics*, *7*(4), pp.353-376.

Swoboda, B., Berg, B., Schramm-Klein, H. and Foscht, T., 2013. The importance of retail brand equity and store accessibility for store loyalty in local competition. *Journal of Retailing and Consumer Services*, *20*(3), pp.251-262.

Thiele, J.C. and Grimm, V., 2010. NetLogo meets R: Linking agent-based models with a toolbox for their analysis. *Environmental Modelling & Software*, *25*(8), pp.972-974.

Thomas, R.R., Barr, R.S., Cron, W.L. and Slocum Jr, J.W., 1998. A process for evaluating retail store efficiency: a restricted DEA approach. *International Journal of Research in Marketing*, *15*(5), pp.487-503.

Vandermeer, J.H., 1972. Niche theory. *Annual review of Ecology and Systematics*, pp.107-132.

Vanhaverbeke, L. and Macharis, C., 2011. An agent-based model of consumer mobility in a retail environment. *Procedia-Social and Behavioral Sciences*, *20*, pp.186-196.

Verburg, P.H., van Eck, J.R.R., de Nijs, T.C., Dijst, M.J. and Schot, P., 2004a. Determinants of land-use change patterns in the Netherlands. *Environment and Planning B: Planning and Design*, *31*(1), pp.125-150.

Verburg, P.H., de Nijs, T.C., van Eck, J.R., Visser, H. and de Jong, K., 2004b. A method to analyse neighbourhood characteristics of land use patterns. *Computers, Environment and Urban Systems*, *28*(6), pp.667-690.

Verdicchio, M. and Colombetti, M., 2001. Commitments for agent-based supply chain management. *ACM SIGecom Exchanges*, *3*(1), pp.13-23.

Verhoef, P.C., Lemon, K.N., Parasuraman, A., Roggeveen, A., Tsiros, M. and Schlesinger, L.A., 2009. Customer experience creation: Determinants, dynamics and management strategies. *Journal of retailing*, *85*(1), pp.31-41.

Voinov, A. and Bousquet, F., 2010. Modelling with stakeholders. *Environmental Modelling & Software*, *25*(11), pp.1268-1281.

Voss, G.B. and Seiders, K., 2003. Exploring the effect of retail sector and firm characteristics on retail price promotion strategy. *Journal of Retailing*, *79*(1), pp.37-52.

Waldrop, M.M., 1993. *Complexity: The emerging science at the edge of order and chaos*. Simon and Schuster.

Walker, B. and Johnson, T., 2019, March. NetLogo and GIS: A Powerful Combination. In *CATA* (pp. 257-264).

Wang, Y., Jiang, W., Liu, S., Ye, X. and Wang, T., 2016. Evaluating trade areas using social media data with a calibrated huff model. *ISPRS International Journal of Geo-Information*, *5*(7), p.112

Warmerdam, F., 2008. The geospatial data abstraction library. In *Open source approaches in spatial data handling* (pp. 87-104). Springer, Berlin, Heidelberg.

Warner, E. J., & Barsky, R. B. (1995). *The Timing and Magnitude of Retail Store Markdowns: Evidence from Weekends and Holidays. The Quarterly Journal of Economics, 110(2), 321–352.* doi:10.2307/2118442

Watson, J. and Everett, J., 1993. Defining small business failure. *International Small Business Journal*, *11*(3), pp.35-48.

Westervelt, J.D., 1999. Modeling mobile individuals in dynamic landscapes. *International Journal of Geographical Information Science*, *13*(3), pp.191-208.

Wilensky, U. 1999. NetLogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Wilensky, U. and Rand, W., 2007. Making models match: Replicating an agent-based model. *Journal of Artificial Societies and Social Simulation*, *10*(4), p.2.

Wilensky, U. and Rand, W., 2015. *An introduction to agent-based modeling: modeling natural, social, and engineered complex systems with NetLogo*. Mit Press.

Xie, M. and Chen, J., 2004. Studies on horizontal competition among homogenous retailers through agent-based simulation. *Journal of Systems Science and Systems Engineering*, *13*(4), pp.490-505.

Xu, T., Gao, J., Coco, G. and Wang, S., 2020. Urban expansion in Auckland, New Zealand: a GIS simulation via an intelligent self-adapting multiscale agent-based model. *International Journal of Geographical Information Science*, pp.1-24.

Yang, J. and Yang, C., 2005, June. The retail stores' competitive location problem with retail regional saturation. In *Proceedings of ICSSSM'05. 2005 International Conference on Services Systems and Services Management, 2005.* (Vol. 2, pp. 1511-1516). IEEE.

Yeh, A.G.O. and Chow, M.H., 1996. An integrated GIS and location-allocation approach to public facilities planning—An example of open space planning. *Computers, Environment and Urban Systems*, *20*(4-5), pp.339-350.

Yoder, S., Visich, J.K. and Rustambekov, E., 2016. Lessons learned from international expansion failures and successes. *Business Horizons*, *59*(2), pp.233-243.

Youn, M.K., Kim, W., Kim, P.J., Lee, S.Y., Sim, K.Y. and Kim, Y.O., 2012. Retail sales forecast analysis of general hospitals in Daejeon, Korea, using the Huff model. *African Journal of Business Management*, *6*(3), pp.971-977.

Yousefi, S., Moghaddam, M.P. and Majd, V.J., 2011. Optimal real time pricing in an agent-based retail market using a comprehensive demand response model. *Energy*, *36*(9), pp.5716-5727.

Yrigoyen, C.C. and Otero, J.V., 1998. Spatial interaction models applied to the design of retail trade areas.

Zellner, M.L., 2007. Generating policies for sustainable water use in complex scenarios: an integrated land-use and water-use model of Monroe County, Michigan. *Environment and Planning B: Planning and Design*, *34*(4), pp.664-686.

Zellner, M.L., 2008. Embracing complexity and uncertainty: the potential of agent-based modeling for environmental planning and policy. *Planning Theory & Practice*, *9*(4), pp.437-457.

Zhang, T. and Zhang, D., 2007. Agent-based simulation of consumer purchase decision-making and the decoy effect. *Journal of business research*, *60*(8), pp.912-922.

Zheng, H.W., Shen, G.Q., Wang, H. and Hong, J., 2015. Simulating land use change in urban renewal areas: A case study in Hong Kong. *Habitat International*, *46*, pp.23-34.

Appendices

Appendix 1 – The observed spatial pattern of home-improvement stores in the study area

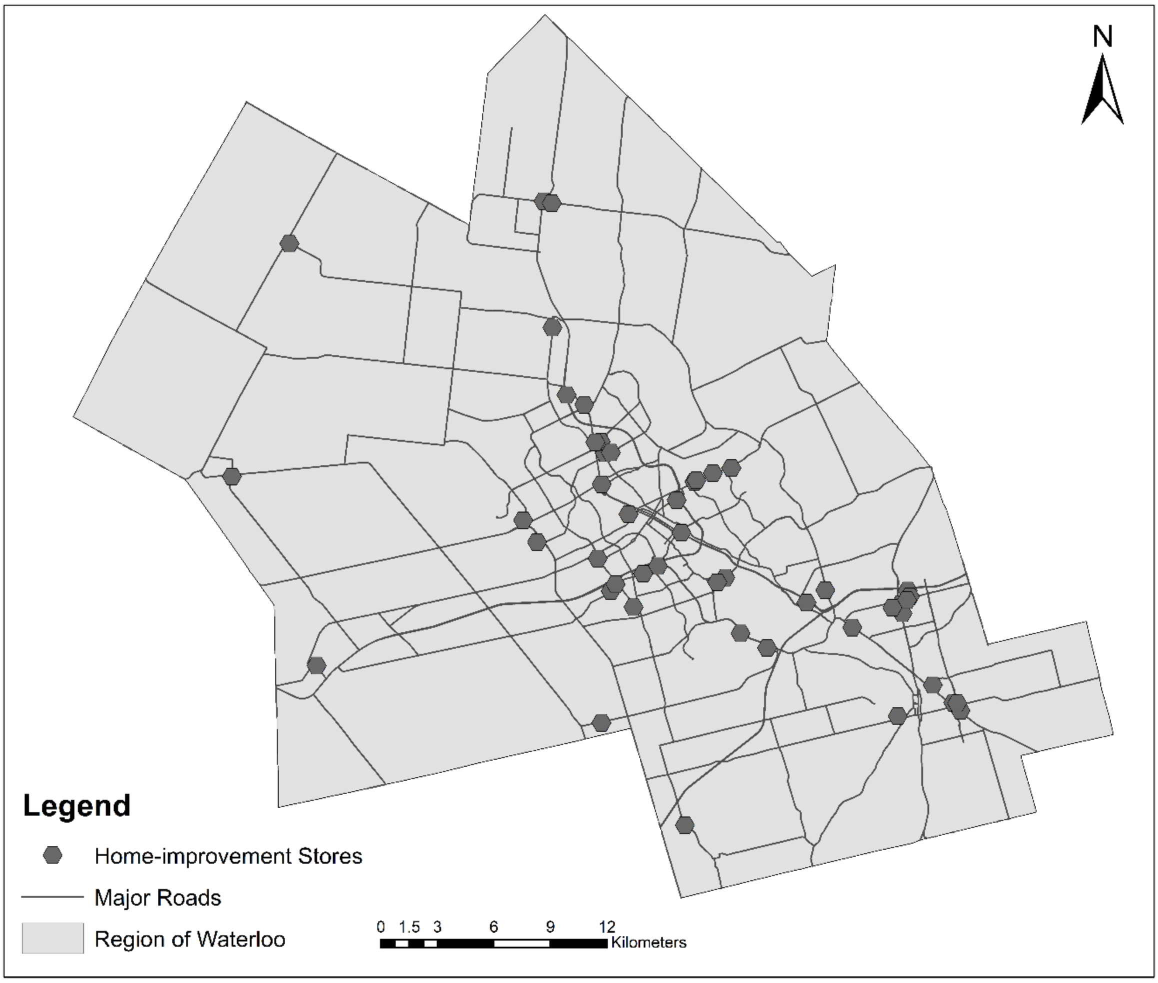


Figure A1: The observed spatial pattern of the 48 home-improvement retailers in the Region of Waterloo

Appendix 2 – Results of the store revenue estimation generated by the Location-allocation Model (LAM) and 5 RC-ABMs

Table A2: The simulated revenue of 48 home-improvement stores generated by the LAM and 5 RC-ABMs with different **ε** *value settings in the Model Validation experiment*

Store ID	Store Size	LAM	RC-ABM				
			ε = 10	ε = 1	ε = 0.1	ε = 0.01	ε = 0.001
13153	14463.29	5626198.51	5629768.59	5715979.33	5723165.17	5723841.07	5723908.22
13409	6102.91	2498211.55	1138502.88	1847836.18	2403141.06	2476555.1	2484083.57
13421	13200	5705874.2	4370116.45	5425848.5	5922058.99	5983297.9	5989547.85
16571	13655.88	7764529.93	5095864.98	5087239.21	5084268.54	5083947.63	5083915.33
17320	132469.08	22747610.45	19445596.08	16080575.43	15370991.59	15291320.07	15283265.16
17591	34520.12	11286100.68	9671084.17	10503664.01	10895593.26	10942835.73	10947640.03
17600	4188.3	2258160.83	1368701.3	1706317.24	1979807.48	2016762.36	2020563.76
17662	11587.11	3382389.2	2710069	4325799.04	4960703.8	5036334.07	5044019.91
17739	4920.13	2552114.86	1186203.06	2428906.7	2995489.74	3067301.03	3074650.29
17850	90608.69	28298535.39	26625734.15	26360058.53	25954434.8	25902539.76	25897225.41
18109	33651.98	16443944.28	12549353.26	15079309.14	15834150.91	15919101.18	15927694.28
18140	31736.14	8768561.61	6783721.06	8529030.67	8980054.67	9030444.18	9035537.61
18529	20147.08	7292054.66	6655857.83	7712830.43	7853507.49	7867447.08	7868832.01
18798	28713.75	11520577.25	9155971.27	9028923.09	8970512.99	8963152.04	8962414.47
19230	105211.48	17393201.98	17031807.07	15386312.78	14990404.76	14943478.17	14938706.6
19451	106302.93	32102198.19	32825293.93	31217982.61	30583263.48	30506917.7	30499416.62
20012	4313.15	8732293.78	9867648.57	10236166.23	10287191.4	10292420.59	10292944.71
20329	107156.32	15438157.67	33287356.77	31174607.5	30345521.01	30245662.19	30235519.47
20584	3735.56	1898131.81	1372577.28	1837941.37	2170137.89	2216996.39	2221849.1
21103	80000	19384937.92	24166651.76	25275056.49	25313537.19	25314892.02	25315002.54
22742	139880.91	26683840.95	24484512.94	25289385.98	25094129.05	25065022.32	25062007.09
22778	26553.69	7947885.76	7342550.82	7445107.13	7415115.38	7409956.54	7409414.75
22837	18007.34	3854752.49	3682657.55	3692746.09	3734035.45	3738088.22	3738491.01
23293	32266.76	6597894.71	5888023.51	6175567.63	6182302.79	6183090.71	6183153.54
23299	93124.08	21530991.48	27756397.73	26722644.93	25998079.13	25908310.27	25899153.12
23449	5290.83	1196905.29	746394.91	889926.14	934945.38	939724.53	940204.73
23493	128474.66	13175109.65	20715677.35	19212139.74	18857104.08	18816951.08	18812949.76
23646	11674.83	3259855.73	3679932.92	3795681.29	3771962.81	3769537.53	3769301.96
24027	12398.72	5367605.55	3873757.06	3976366.8	4207983.77	4236768.58	4239666.9
24037	14890.66	6087414.29	4746047.22	4773077.56	4800046.39	4805614.23	4806229.41
24042	5637.81	3262836.57	2043755.94	2834342.53	3326504.59	3390277.57	3396817.88
24059	27510.02	6949762.19	5293722.31	6497449.33	6779289.42	6808727.07	6811680.46
24095	36174.72	8177345.91	8058948.15	8491435.61	8555141.38	8561891.72	8562572.62
24101	15824.46	2880450.18	2679396.84	2519015.21	2491759.21	2489404.14	2489179.43
24116	46855.42	9623605.45	8268647.58	8273143.48	8105722.4	8083300.73	8081012.43
24153	6773.53	1592043.41	1928676.36	1873428.24	1870270.22	1869988.24	1869960.39
24187	50408.16	6441679.65	8266587.57	7757723.97	7678666.47	7667603.11	7666343.93
24191	24090.23	7106672.69	6484818.04	6331472.93	6215044.4	6199527.75	6197933.02
24195	59146.28	22081057.22	18956261.78	22219888.25	22767503.26	22824175.58	22829859.28
24199	55520.64	17095320.02	13248832.3	13856304.29	13875013.88	13874250.21	13874145.23
24235	119969.19	31295076.26	32656749.28	30108852.49	29625631.3	29573513.9	29568275.37
32021	5329.98	1983988.54	1632715.96	1548290.48	1603904.5	1608525.15	1608640.78
34019	94000	15664389.18	15323096.13	13957879.33	13621726.9	13583232.06	13579375.29
34103	40109.71	6665651.72	6534399.03	6268178.37	6297662.04	6304023.8	6304708.36
34114	112190.15	12924599.51	16768101.67	12928058.47	12207799.18	12127885.77	12119813.69
34240	17267.66	3426280.29	4241618.47	3899399.75	3865350.03	3862605.67	3862341.51
34251	44753.57	11610846.36	10760545.23	11040668.62	11014580.81	11010096.97	11009631.92
34273	57851.29	15722493.78	14445463.72	14107608.73	13930957.4	13908828.13	13906567.04

Appendix 3 – The distribution of the two types of the market share captured (MSC1 and MSC2) generated by RC-ABMs

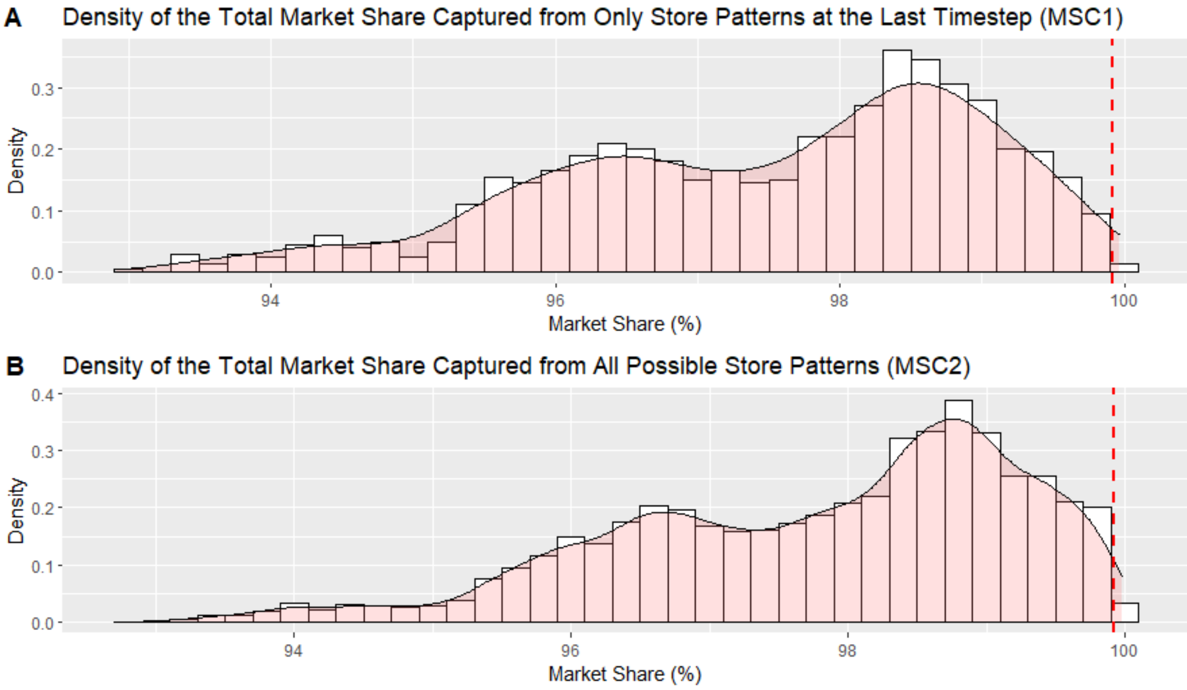


Figure A3: The plot of kernel density estimate based on the total market share captured by (A) store patterns produced by only the last timestep of each model run (1000 patterns) and (B) all store patterns appeared during all 20 timesteps in each of the 1000 model runs (20000 patterns); The red dotted line indicates the observed MSC generated by the LAM

Appendix 4 – Supplementary results of the path dependency analysis

The Region of Waterloo has 7 Census Subdivisions (CSDs), including three cities (Kitchener, Waterloo, and Cambridge) and four townships (Woolwich, Wellesley, Wilmot, and North Dumfries). In addition to the analysis of path dependency in Section 3.3, we have done a further analysis based on each of the 7 CSDs. According to Table 7.4.1, results exhibit a higher variation of path dependence patterns in tri-cities where all types of path dependencies (Unsuccessful Invariant, Successful Invariant, and Variant) occur. Moreover, stores' success can hardly be guaranteed in either of the 4 townships, where over 96% of CDAs and over 99% of areal coverages are occupied by Unsuccessful Invariant or the Variant regions. In contrast, stores randomly residing within Kitchener or Waterloo have the highest probability of success (i.e., higher percentage in number of Successful Invariant CDAs and areal coverage).

Table A4. Summary of the composition and configuration of Unsuccessful Invariant, Successful Invariant, and Variant census dissemination areas (CDAs) in the City of Kitchener, Waterloo, and Cambridge, and in the Townships of Wellesley, Woolwich, Wilmot, and North Dumfries.

	City of Kitchener			City of Waterloo			City of Cambridge		
Region Area (ha)	Area = 13836.14			Area = 6508.51			Area = 11568.26		
Total CDA Numbers	n = 313			n = 153			n = 189		
	Unsuccess	Success	Variant	Unsuccess	Success	Variant	Unsuccess	Success	Variant
Number of CDAs	1	144	168	1	61	91	1	46	142
Percent of CDAs (%)	0.32	46.01	53.67	0.65	39.87	59.48	0.53	24.34	75.13
Total Area (ha)	402.83	2789.15	10644.15	39.41	1350.72	5118.37	916.48	767.54	9884.24
Percent Area	2.91	20.16	76.93	0.61	20.75	78.64	7.92	6.63	85.44
Largest Patch Size (ha)	402.83	1382.94	10572.42	39.41	628.56	5047.25	916.48	321.96	9884.24
Mean Patch Size (ha)	402.83	199.23	2661.04	39.41	168.84	1706.12	916.48	76.75	9884.24
S.D. Patch Size (ha)	0	372.48	4567.67	0	214.17	2362.57	0	95.78	0
Number of Patches	1	14	4	1	8	3	1	10	1
	The Township of Wellesley			**The Township of Woolwich**			**The Township of Wilmot**		
Region Area (ha)	Area = 27848.11			Area = 32979.87			Area = 26599.74		
Total CDA Numbers	n = 19			n = 35			n = 29		
	Unsuccess	Success	Variant	Unsuccess	Success	Variant	Unsuccess	Success	Variant
Number of CDAs	0	0	19	0	0	35	1	1	27
Percent of CDAs (%)	0	0	100	0	0	100	3.45	3.45	93.1
Total Area (ha)	0	0	27848.11	0	0	32979.87	642.43	16.74	25940.57
Percent Area	0	0	100	0	0	100	2.42	0.06	97.52
Largest Patch Size (ha)	0	0	27848.11	0	0	32979.87	642.43	16.74	25940.57
Mean Patch Size (ha)	0	0	27848.11	0	0	32979.87	642.43	16.74	25940.57
S.D. Patch Size (ha)	0	0	0	0	0	0	0	0	0
Number of Patches	0	0	1	0	0	1	1	1	1
	The Township of North Dumfries								
Region Area (ha)	Area = 19088.15								
Total CDA Numbers	n = 17								
	Unsuccess	Success	Variant						
Number of CDAs	4	0	13						
Percent of CDAs (%)	23.53	0	76.47						
Total Area (ha)	5551.8	0	13536.35						
Percent Area	29.09	0	70.91						
Largest Patch Size (ha)	5551.8	0	13536.35						
Mean Patch Size (ha)	5551.8	0	13536.35						
S.D. Patch Size (ha)	0	0	0						
Number of Patches	1	0	1						

Appendix 5 – Home-Improvement Spending Categories

Table A5. Consumer's home-improvement spending categories in Ontario used to generate consumer expenditures by Census Dissemination Area (CDA) (Table 2, Average household spending on home improvement spending categories in Ontario, 2011, source Robinson and Balulescu, 2018, p.1067)

Spending categories	2011 ($)
Tenants' repairs and improvements	26
Repairs and maintenance for owned living quarters	424
Other expenses for owned vacation homes and other secondary residences	110
Household cleaning supplies and equipment ***(3 sub-categories)***	241
Nursery and greenhouse stock, cut flowers, decorative plants and planting seeds	272
Fertilizers, herbicides, insecticides, pesticides, soil and soil conditioners	75
Other household supplies	109
Furniture	708
Rugs, mats and underpadding	55
Other household furnishings (curtains, mirrors and picture frames)	90
Household appliances ***(7 sub-categories)***	443
Other household equipment ***(4 sub-categories)***	550