Logistics Sprawl: Spatial Patterns and Characteristics of New Warehousing Establishments in The Greater Toronto and Hamilton Area

by

Gagandeep Singh

A thesis submitted in conformity with the requirements for the degree of Master of Applied Science
Department of Civil Engineering
University of Toronto

© Copyright by Gagandeep Singh 2018
Logistics Sprawl: Spatial Patterns and Characteristics of New Warehousing Establishments in The Greater Toronto and Hamilton Area

Gagandeep Singh
Master of Applied Science
Department of Civil Engineering
University of Toronto
2018

Abstract

A significant sprawl in logistics facilities have been identified both in the Greater Toronto Area (GTA) and the Greater Golden Horseshoe (GGH). The objective of this thesis is to analyze the spatial patterns and characteristics of new warehousing establishments that have located in the Greater Toronto and Hamilton Area (GTHA) since 2003. The study found that the major share of increase in number of warehouses in the GTHA occurred between the years 2005 and 2006, and years 2009 and 2010 respectively. This growth illustrates the emergence of clusters of new warehouses around the Toronto Pearson International Airport, and the Canadian National and the Canada Pacific intermodal terminals. The building footprint and the property size used for development by the new warehouses indicates the emergence of large scale warehouses in the GTHA. Inward movement of warehousing establishments towards the distribution centers of major firms has been observed.
Acknowledgments

Writing acknowledgments is the most interesting part of the thesis as people who read it get a chance to glance through one’s life. Life to me means cherishing moments spent with the God’s creation including humans and nature. I am thankful to almighty for a wonderful life given to me. My life has been an interesting journey. I am thankful to everyone with whom I have ever met in my life because I have learnt a bit from everyone. I wish I could be able to help everyone the way I have been helped.

I am thankful to Prof. Matthew Roorda who gave me this opportunity to study under his supervision at a great university. Matt has been one of the most amazing person I have met in my life. He has been very supportive and helpful throughout my post-graduate studies. I owe you for helping me starting my family in Canada. I am thankful to everyone at the Map and Data Library where I spent half of my academic time. They were present all the time to answer my queries and helping me with the data. Thanks Jun Zhang for providing help with an important component of my thesis.

I am thankful to my guru ‘Sant Rampal Singh Ji’ from the core of my heart without whom I would not have been anything in my life. It is all because of my Baba Ji’s grace that I have achieved everything in my life. I am thankful to my Bdepapa ‘Ajaib Singh’ and my father ‘Sukhdev Singh’ who have been with me throughout the journey of my life. Bdepapa has been a pillar of consistent support and has been with me during all ups and downs of my life. I am thankful to my mom and masi, and my family (Amani, Preeti, Rinku, Bhabhi Ji and my wife) who faced all the hardships to keep me happy. Their support kept me focused and thus I never lost hope.

I dedicate this research to my daughter Surleen Kaur and to my nephew Bhagirath Singh.
Table of Contents

Acknowledgments .......................................................................................................................... iii

Table of Contents ........................................................................................................................... iv

List of Tables ................................................................................................................................ vii

List of Figures .............................................................................................................................. viii

List of Appendices ......................................................................................................................... xi

Chapter 1 Introduction .....................................................................................................................1

1.1 Background ..........................................................................................................................1

1.2 Research Motivation ............................................................................................................2

1.2.1 Study Area ...............................................................................................................3

1.3 Research Objectives .............................................................................................................4

1.4 Organization of the Thesis ...................................................................................................6

Chapter 2 Literature Review ............................................................................................................7

2.1 Phenomenon of Logistics Sprawl ........................................................................................7

2.2 Reasons for Logistics Sprawl .............................................................................................10

2.2.1 Land Availability and Affordability ......................................................................12

2.2.2 Proximity to Customers and Transportation Infrastructure ...............................12

2.2.3 Accessibility to Labour and Supply Chain .........................................................13

2.2.4 Regulatory Environment ................................................................................13

2.3 Implications of Logistics Sprawl .......................................................................................14

2.3.1 Impact on Urban Freight Geography .....................................................................14

2.3.2 Contribution to Vehicle Kilometers Travelled by Trucks ..................................14

2.3.3 Environmental Impacts ..................................................................................15

2.3.4 Impact on Commuting of Logistics Employment ..............................................15

iv
Chapter 3 Data ...............................................................................................................................16

3.1 NAICS 493 Warehousing and Storage .....................................................................................16
3.2 DMTI Enhanced Points of Interest (EPOI) ........................................................................18
3.3 InfoCanada Business Establishment Dataset .................................................................18
3.4 ESRI Business Analyst Application ................................................................................18
3.5 Teranet Property Parcels ..............................................................................................19
3.6 Google Maps ...................................................................................................................19
3.7 Transportation Networks .................................................................................................19

Chapter 4 Methodology .................................................................................................................20

4.1 Dataset Development .........................................................................................................20
4.2 Warehousing Sprawl ..........................................................................................................23
4.2.1 Centrographic Analysis ..........................................................................................23
4.2.2 Kernel Density Estimation .....................................................................................25
4.3 Building Footprint Calculation (Zhang, 2017) .....................................................................26
4.4 Distance Calculations from Nearest Highway, Intermodal Rail Terminals and Toronto Pearson International Airport .................................................................26

Chapter 5 Characteristics of NAICS 493 establishments ..............................................................28

5.1 Appearance and Disappearance of NAICS 493 Establishments ........................................28
5.2 Pre-Recession Period (2003-2007) ....................................................................................30
5.3 Recession Period (2008-2009) ...........................................................................................31
5.4 Post-Recession Period (2010-2013) ...................................................................................32
5.5 Distance from the Barycenter and Movement of the Barycenter ......................................33

Chapter 6 Results and Discussion ..................................................................................................35

6.1 Warehousing Sprawl ..........................................................................................................35

Chapter 7 Characteristics of New NAICS 493 Establishments .....................................................40

7.1 Spatial Distribution ..............................................................................................................40
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.2</td>
<td>Cluster Analysis</td>
<td>43</td>
</tr>
<tr>
<td>7.3</td>
<td>Building Footprint and Property Parcel Size</td>
<td>48</td>
</tr>
<tr>
<td>7.4</td>
<td>Distance from the Nearest Highway, Intermodal Rail Terminals and Toronto Pearson International Airport</td>
<td>51</td>
</tr>
<tr>
<td>Chapter 8</td>
<td>Spatial Relationships within Multi-Establishment Firms</td>
<td>57</td>
</tr>
<tr>
<td>8.1</td>
<td>Objective</td>
<td>57</td>
</tr>
<tr>
<td>8.2</td>
<td>Dataset Development</td>
<td>58</td>
</tr>
<tr>
<td>8.2.1</td>
<td>Details of Multi-Establishment Firms</td>
<td>58</td>
</tr>
<tr>
<td>8.3</td>
<td>Methodology</td>
<td>61</td>
</tr>
<tr>
<td>8.4</td>
<td>Results</td>
<td>62</td>
</tr>
<tr>
<td>8.4.1</td>
<td>Characteristics of Distribution Centers</td>
<td>62</td>
</tr>
<tr>
<td>8.4.2</td>
<td>Relationships Between Annual Revenue, and Distance Between Distribution Center and Centroid of Firms</td>
<td>64</td>
</tr>
<tr>
<td>8.4.3</td>
<td>Spatial Distribution of Distribution Centers and Centroid of Firms</td>
<td>65</td>
</tr>
<tr>
<td>8.4.4</td>
<td>Relationship Between Distribution Centers and NAICS 493 Establishments</td>
<td>71</td>
</tr>
<tr>
<td>Chapter 9</td>
<td>Conclusions and Future Work</td>
<td>72</td>
</tr>
<tr>
<td>References</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>Appendix</td>
<td>82</td>
<td></td>
</tr>
</tbody>
</table>
List of Tables

Table 1: Summary of factors determining logistics sprawl adapted from Jaller & Leticia, 2017 (Source: Jaller & Leticia, 2017) ................................................................................................................................. 11

Table 2: NAICS 493 sub-categories (Source: Statistics Canada, 2012) ........................................... 17

Table 3: Types of logistics firms studied (Source: Woudsma et al., 2016) ..................................... 17

Table 4: Number of classified warehouses and mini-warehouses in the DMTI EPOI dataset .... 21

Table 5: Annual revenue of multi-establishment firms ................................................................ 60
List of Figures

Figure 1: GTHA Census Divisions and the Greenbelt................................................................. 4

Figure 2: Number of NAICS 493 establishments in the DMTI EPOI dataset............................. 22

Figure 3: Appearance and disappearance within NAICS 493 establishments............................ 29

Figure 4: Appearance and disappearance within new NAICS 493 establishments...................... 29

Figure 5: Spatial patterns of NAICS 493 establishments during pre-recession period ............... 30

Figure 6: Spatial patterns of NAICS 493 establishments during recession period.................... 31

Figure 7: Spatial patterns of NAICS 493 establishments during post-recession period.............. 32

Figure 8: Average distance of NAICS 493 establishments from their barycenter ...................... 33

Figure 9: Movement of the barycenter of NAICS 493 establishments....................................... 34

Figure 10: Growth in NAICS 493 establishments by Census Divisions................................... 35

Figure 11: Warehousing sprawl in the GTHA using Centrographic analysis............................. 36

Figure 12: Difference in kernel density of NAICS 493 establishments...................................... 38

Figure 13: Average value of private dwellings (Source: Statistics Canada, 2011) ....................... 39

Figure 14: Distribution of property parcels with size greater than 50000 square meter.............. 39

Figure 15: Distribution of new NAICS 493 establishments...................................................... 40

Figure 16: Spatial pattern of new NAICS 493 establishments.................................................. 41

Figure 17: Kernel density of new NAICS 493 establishments.................................................. 42

Figure 18: Test of existence of warehousing establishments clusters (generated by ArcMap).... 44

Figure 19: Optimal number of warehousing clusters using Gap Statistic................................. 46
Figure 20: Clusters amongst new NAICS 493 establishments in the GTHA ......................... 47
Figure 21: Building footprint distribution of new NAICS 493 establishments ................. 48
Figure 22: Spatial distribution of new NAICS 493 establishments according to building footprint ................................................................. 49
Figure 23: Property parcel size distribution of new NAICS 493 establishments .............. 50
Figure 24: Spatial distribution of new NAICS 493 establishments according to property parcel size ............................................................................................................. 50
Figure 25: Distribution of distance of new NAICS 493 establishments from nearest highway... 51
Figure 26: Buffer around major highways in the GTHA ................................................. 52
Figure 27: Distribution of distance of new NAICS 493 establishments from CN Brampton intermodal terminal ................................................................. 53
Figure 28: Distribution of distance of new NAICS 493 establishments from CN Mississauga intermodal terminal ................................................................. 54
Figure 29: Distribution of distance of new NAICS 493 establishments from CP Vaughan intermodal terminal ................................................................. 54
Figure 30: Distribution of distance of new NAICS 493 establishments from Toronto Pearson International Airport ................................................................. 55
Figure 31: Radial rings around Toronto Pearson International Airport ......................... 56
Figure 32: Distribution of distance between distribution centers and centroid of retail establishments of multi-establishment firms ......................................................... 61
Figure 33: Land use pattern of distribution centers of multi-establishment firms ............. 62
Figure 34: Spatial distribution of distribution centers according to property parcel size ...... 63
Figure 35: Distribution of annual revenue of firms and distance between their distribution centers and centroid of retail establishments.

Figure 36: Location of distribution centers and centroid of firms in Category I.

Figure 37: Location of distribution centers and centroid of retail firms in Category II.

Figure 38: Location of distribution centers and centroid of firms in Category III.

Figure 39: Location of distribution centers and centroid of firms in Category IV.

Figure 40: Location of distribution centers and centroid of firms in Category V.

Figure 41: Spatial relationship between distribution centers and NAICS 493 establishments.
List of Appendices

Appendix A – Building Footprint Calculation (Zhang, 2017).................................................................82
Chapter 1
Introduction

1.1 Background

Warehouses and distribution centers perform a variety of logistics functions and have an important role to play in the supply chain network of any firm. There are different types of logistics facilities that require land depending on the category and objective of the facility, which can be a warehouse, distribution centre, truck terminal or intermodal facility (Mckinnon, 2009). The location of logistics facilities affects not only the activities of urban goods movement, but also the urban environment as these facilities represent major originators and receivers of freight (Aljohani & Thompson, 2016). Due to trade liberalization and increasing freight demand many cities across the world have experienced growth in the number and size (both physical size and employment size) of the warehousing industry (Bowen, 2008). Lower freight costs have allowed them to move away from the markets they serve in search for cheaper land required for their operations (Rodrigue, 2004b).

Several changes have been observed in large metropolitan areas such as the increasing concentration of warehousing establishments around intermodal terminals and airports, the rise of urban distribution centers, the emergence of inland ports and terminals, polarization of logistics activities and logistics sprawl (Bowen, 2008; Cidell, 2010; Dablanc & Rakotonarivo, 2010; Dablanc & Ross, 2012). These trends amongst spatial patterns of freight and logistics establishments have changed the geography of urban freight distribution in metropolitan areas. Logistics sprawl is a phenomenon that is being observed in metropolitan areas worldwide, and is defined as the relocation of logistics facilities from urban to sub-urban areas (Dablanc & Rakotonarivo). There has been a strong evidence of logistics sprawl in the Greater Golden Horseshoe (GGH) and some evidence in the Greater Toronto Area (GTA) in Canada (Woudsma et al., 2016). The objective of this thesis is to analyze and characterize in detail the characteristics of warehousing and distribution facilities that have come up in the Greater Toronto and Hamilton Area (GTHA) since 2003 by merging information from multiple data sources. This information would help urban planners, public/private agencies with future policy framework on freight transportation, infrastructure development and land-use planning while
taking into consideration the changing spatial patterns and characteristics of these logistics facilities.

1.2 Research Motivation

Woudsma’s research found that in the Greater Toronto Area, from year 2002 to 2012, the number of logistics establishments grew by 108% and the number of warehousing establishments grew by almost 40% (Woudsma et al., 2016). Also, according to the Canadian Business Patterns (CBP) data, the number of warehousing establishments in the GTHA since 2003 grew by 38% (Statistics Canada, 2013). Dablanc claimed that out of 20 metropolitan areas studied, Toronto had the highest level of logistics sprawl (Dablanc, 2016). These facts led us to carry out a detailed analysis to develop a deeper understanding of the factors contributing to the phenomenon of logistics sprawl in the GTHA, and to evaluate the characteristics of the new warehousing facilities that have located in the GTHA. Currently, most of the freight movements in the GTHA are local, and about 89% of all freight movements are by truck (Metrolinx, 2008). One of the goals under The Big Move, the Regional Transportation Plan is to develop a multi-pronged approach to improve goods movement within the GTHA and with adjacent regions (Metrolinx, 2008). These establishments are the major truck traffic generators, and their truck travel demand changes with their changing location patterns over time thus affecting the highway system (Giuliano & Kang, 2017). Modern warehouses and distribution centers are very large facilities with sizes exceeding 500,000 square feet and due to the large freight volumes handled, they generate many consolidated freight trips (Andreoli et al., 2010). Hence, it is fundamental to analyze the characteristics of new location sites of these establishments and their land utilization (both land-use and property size) for logistics operations.

This research aims at analyzing the spatial distribution patterns of warehousing establishments that are engaged in providing secure storage facilities for farm products, household and commercial goods of any nature. These establishments offer a wide range of logistics services related to goods distribution such as labelling, breaking bulk, inventory control and management, and packaging etc.Warehouses and distribution centers are often regarded as one of the key economic agents of the freight transportation system and are fundamental to goods movement. Greater concentration of these facilities in urban areas implies greater localized impacts while decentralized trends may result in more truck traffic and additional vehicle kilometers traveled.
for deliveries. Hence, examining changes in spatial patterns of these facilities is essential for metropolitan planning, land-use decision making and for forecasting freight demand.

1.2.1 Study Area

This study focuses on the Greater Toronto and Hamilton Area (GTHA), the largest urban region in Canada with a population of 6.95 million (Statistics Canada Census, 2016) and one of the fastest growing urban regions of Canada (Statistics Canada, 2016). The GTHA comprises the Cities of Toronto and Hamilton, and the Regions of Peel, Durham, York and Halton. The logistics activities taking place in the GTHA support a diverse and dynamic economy that has a strong manufacturing base, among other transportation intensive economic activities (Woudsma et al., 2016). The GTHA is a very significant generator of goods movement activities in Canada, and efficient movement of goods and delivery of services in the GTHA are important factors in establishing the regions overall competitiveness and prosperity (Metrolinx, 2011).

Also, the GTHA is home to intermodal rail terminals of the Canadian National Railways (CN Brampton and CN Mississauga) and the Canadian Pacific Rail (CP Vaughan), and the Toronto Pearson International Airport. The Toronto Pearson International Airport is an essential component of the logistics supply chain for Canada’s export industries, acting as the link between the national ground transportation network and global markets (Toronto Pearson, 2016). The Ministry of Public Infrastructure Renewal (2006) enacted “greenbelt” legislation in 2006 (see Figure 1) to discourage outward growth into surrounding regions of the GTHA (Woudsma et al., 2016). This analysis would help to understand the role of these three intermodal terminals and the Pearson airport, major highways and the Greenbelt in location and spatial patterns of new warehousing establishments that have come up in the GTHA.
1.3 Research Objectives

This study aims to analyze the spatial patterns and characteristics of the new warehousing establishments that have located in the GTHA since 2003, and to evaluate growth in the warehousing industry in the GTHA by making a longitudinal comparison. Studies identified in the literature have only focused on analyzing the changing spatial patterns of logistics facilities and minor consideration has been given to analyze the characteristics of these facilities. Some studies used employment size of the establishments as a proxy to the size of the facility to separate large warehouses from others (Bowen, 2008; Andreoli et al., 2010). Another major objective of this research is to analyze the spatial relationships within multi-establishment firms i.e. analyzing the spatial relationship between the firms’ distribution center and other retail and logistics establishments within the same firm. This helps to analyze if location choice is a part of larger network planning decisions within multi-establishment firms. Typically, the literature focuses on the location choices of single-plant firms and little attention has been paid in location
theory to study the location choices of firms that set up multi-plants (or multi-establishments or multi-stores) (Mota, 2017). This study builds upon the research of Woudsma et al., (2016) by characterizing in detail the new warehousing establishments by fusing information from multiple available datasets. Some of the questions addressed are: Where are the new warehousing establishments located? What is the building footprint of the establishments? What is the size of the properties used for development? How far are the facilities located from the nearest highways and intermodal rail terminals? What is the relationship between the logistics facility and other retail and logistics establishments within the firm? What is the relationship between warehousing establishments and the distribution centers of major firms in the GTHA? What is the spatial relationship between NAICS 493 warehousing establishments and distribution centers of major firms in the GTHA?

New warehousing establishments, defined in the thesis, include the North American Industry Classification System - NAICS 493 Warehousing and Storage establishments that appeared in the GTHA between year 2003 and 2013. More description of the NAICS 493 warehousing establishments is provided in Section 3.1. The following activities are within the scope of this study:

- Analyze the spatial patterns of warehousing establishments from year 2003 to 2013,
- Identify new warehousing establishments in the GTHA between the years 2003 and 2013,
- Analyze the phenomenon of “warehousing sprawl”, the sub-urbanization trend of warehousing establishments in the GTHA,
- Analyze the characteristics of new warehousing establishments identified in the GTHA from 2003 to 2013 by fusing information from multiple datasets,
- Analyze the spatial relationship between a distribution center and other retail and logistics establishments within the same firm.
1.4 Organization of the Thesis

This research analyzes the spatial patterns and characteristics of NAICS 493 warehousing establishments in the GTHA, and spatial relationships within multi-establishment firms in eight chapters. Research motivation, description of the study area and objectives of the study are introduced in Chapter 1. Chapter 2 describes the phenomenon of logistics sprawl observed in metropolitan areas across the world, and in the GTHA. It also outlines studies of changing spatial patterns of logistics facilities and methods to quantify logistics sprawl. Some discussion is provided on reasons for the sub-urbanization trend of logistics activities, and policy and environmental implications of sprawling nature of firms. Chapter 3 describes the datasets used to identify and characterize new warehousing establishments, and for analyzing spatial relationships within multi-establishment firms in the GTHA. Chapter 4 describes the methodology followed for identifying and analyzing the characteristics of new warehousing establishments. It also outlines the methodology used for quantifying warehousing sprawl in the GTHA, analyzing their clustering patterns and exploring spatial relationships within multi-establishment firms. Chapter 5 describes spatial patterns of warehousing establishments in the GTHA from year 2003 to 2013. Chapter 6 describes the growth in the number of warehousing establishments in the GTHA, and the phenomenon of warehousing sprawl observed in the GTHA since 2003. Chapter 7 describes the spatial patterns and characteristics of new warehousing establishments identified in the GTHA, and conducts a cluster analysis of new warehousing establishments. Chapter 8 outlines the spatial relationships within multi-establishment firms, and relationship between NAICS 493 warehousing establishments and distribution centers of major firms in the GTHA. Chapter 9 summarizes the key outcomes of the study and provides directions for future research.
Chapter 2
Literature Review

Logistics sprawl is a phenomenon that is being observed in urban areas worldwide. This literature review will outline in detail the phenomenon of logistics sprawl observed in metropolitan areas worldwide, factors contributing to this trend of sub-urbanization and the impacts of logistics sprawl on the geography of freight in urban areas.

2.1 Phenomenon of Logistics Sprawl

Due to various factors related to regulatory land-use control and requirements for present day operations, logistics facilities are often located in primarily logistics clusters on the periphery of metropolitan areas with greater accessibility to highway networks, major airports and seaports (Aljohani & Thompson, 2016). This trend of sub-urbanization and sprawl of logistics firms has been identified in several cities in Europe (Dablanc & Rakotonarivo, 2010; Allen et al., 2012; Heitz & Dablanc, 2015, Todesco et al., 2016), North America (Bowen, 2008; Cidell, 2010; Dablanc & Ross, 2012; Dablanc et al., 2014b, Woudsma et al., 2016) and Japan (Sakai et al., 2016). Various tools and techniques have been used by researchers to analyze and quantify the changing spatial patterns of logistics establishments. Methods used to evaluate the spatial patterns of concentration and dispersion of logistics establishments in urban areas include estimating Gini indices (Cidell, 2010), Centro-graphic analysis (Dablanc & Rakotonarivo, 2010; Dablanc & Ross, 2012; Dablanc et al., 2014b, Heitz & Dablanc, 2015; Woudsma et al., 2016) and Kernel Density estimation (Heitz & Dablanc, 2015; Sakai et al., 2016). Details of the methodology used in literature for quantifying logistics sprawl has been discussed in Section 4.2.

Bowen (2008) studied the growth in number of warehouses in fifty Metropolitan Statistical Areas (MSA) in the United States between 1998 and 2005. The study found that sub-urban counties located within the MSA were experiencing faster growth of logistics activities than the central counties as these ex-urban areas offer better highway network accessibility and have readily available cheaper land. The study also found growth of logistics facilities in the central counties within the MSA that are characterized by growing populations and intensive freight sector businesses.
Cidell (2010) examined the suburbanization trend of transportation, warehousing and trucking activity in fifty metropolitan areas in the United States from 1986 to 2005. The author used the Economic Census and US County Business Patterns data, and developed Gini Indices for the metropolitan areas studied. The study observed two trends in warehousing industry in the United States: their move towards inland distribution centers and decentralization of freight activities towards sub-urban counties. Both the studies of Bowen (2008) and Cidell (2010) established evidence for sub-urbanization of logistics facilities but they did not make an attempt to determine the spatial location changes of warehouses within counties (Aljohani & Thompson, 2016).

Jakubicek and Woudsma (2011) carried out a survey based study (stakeholder interviews) to examine the relative importance of location factors for logistics firms in the GTA. The study found out set of “retaining factors”, such as access to major suppliers and customers, proximity to highways, ability to operate 24/7 hours, which will encourage firms to remain in their current location. The “push factors” included factors such as low land costs/tax rates, availability of skilled labor and open land which encourage firms to leave their current location. Verhetsel et al., (2015) in their study on Flanders region in Belgium also supported that land costs/tax rate was the most important location factor for logistics facilities.

Dablanc and Ross (2012) analyzed the spatial patterns of freight and logistics activities in Atlanta and the Peidmont Atlantic Megaregion using US Census Bureau County Business Patterns Survey data at zip-code level for year 1998 and 2008 respectively. The study found spatial decentralization of logistics facilities at the metropolitan level and found that average distance to the barycenter of logistics facilities increased by 4.5 km between 1998 and 2008. On the other hand, they observed polarization of logistics activities (or the concentration of logistics activities) at the megaregional level of Peidmont Atlantic Megaregion.

Dablanc et al., (2014) looked at the growth in geographic distribution of warehouses in both Los Angeles and Seattle metropolitan areas between 1998 and 2009 using County Business Patterns data. For the Los Angeles metropolitan area, they found that warehousing establishments has sprawled considerably with average distance of the establishments from their barycenter increased by 9.7 km. For the Seattle Metropolitan area they found that logistics facilities have not sprawled significantly and found a decrease in the average distance of the warehousing establishments from their barycenter across the two years. The reasons for sprawl in Los Angeles
are identified as differentials in land prices (ex-urban areas offering land at cheaper prices), sub-
urban areas offering connectivity to complex network of highways that allow to serve regional as
well as national economy, proximity of suburban areas to Airport (more space required by
current warehousing facilities) and actions by suburban communities to encourage growth.

Heitz and Dablanc (2015) looked at the locational patterns of logistics and warehousing facilities
both at the metropolitan scale (Ile-de-France) and at the megaregional scale (Paris Basin) of Paris
between 2000 and 2012 using SIRENE database of the French National Institute of Statistics and
Economic Studies. The study observed that warehouses have sprawled significantly in the Paris
metropolitan area and found that the distance of the warehouses to their barycenter has increased
by 10 km between the two years. For the Paris basin, they found that the distance of the
warehouses from their barycenter reduced by 55 km specifying an inward movement in the Paris
region.

Woudsma et al., (2016) analyzed the spatial patterns of freight and logistics activities in the GTA
and the GGH region using DMTI’s Enhanced Points of Interest data. They addressed
methodological issues like facility identification (mis-classification of logistics facilities within
government and private data sources), usability and interchangeability of NAICS and SIC
Classification system. The study observed that logistics facilities are not sprawling significantly
in the GTA because of land required for expansion due to the presence of the Greenbelt at
periphery of the GTA. However, they observed that logistics sprawl was evident in the GGH
(Toronto Megaregion) and found that average distance of the logistics facilities from their
barycenter increased by 9.5 km between 2002 and 2012. This observation is different from the
Paris Megaregion which experienced the inward movement of logistics facilities from the outer
region of Paris Megaregion towards sub-urban areas of Paris (Aljohani & Thompson, 2016).

Todesco et al., (2016) analyzed the spatial patterns of four types of logistics firms (road
transportation, storage, courier services, postal service) in Zurich area in the period of 1995-2012
to determine the degree of logistics sprawl. They found that the mean distance to the Zurich city
center increased significantly for the storage firms (+9.5 km) and courier service firms (+7.7 km)
and these firms moved from urban cores (densely populated) toward more sub-urban areas.
While they did not observe any sprawl for transportation firms and observed the pattern of
concentration (moving from sub-urban to urban areas) for postal services firms.
Sakai et al., (2016) examined the logistics facility distribution in the Tokyo Metropolitan area (TMA) using the Tokyo Metropolitan Freight Survey (TMFS) data for the year 2003. The study found that the average distance of the logistics facilities from the urban center of Tokyo increased by 4.2 km between 1980 and 2003. They found that shortage of industrial land in inner urban areas play key role in sub-urbanization of logistic facilities. They suggested for careful monitoring of logistics facility developments through land-use control to prevent their leapfrogging developments. A long-term plan to promote efficient and sustainable city logistics should consider these logistics sprawl trends which threaten sustainability of urban areas (Lukic, 2017).

2.2 Reasons for Logistics Sprawl

Several reasons have been identified for the phenomenon of logistics sprawl. These reasons may be endogenous or exogenous to the firms. Endogenous reasons include moving to a new location as a part of the business strategy (Jakubicek & Woudsma, 2011). Relocations can form an integral part of a firm’s business strategy and occur more frequently amongst smaller establishments (Maoh & Kanaroglou, 2007). Exogenous reasons include availability of cheaper and larger plots of land, lower land tax rates, proximity to customers, and transportation infrastructure such as highways and intermodal terminals, proximity to suppliers, availability of skilled labour, regulatory environment and zoning plans (Jaller & Leticia, 2017). Table 1 shows factors identified for logistics sprawl in previous research.
Table 1: Summary of factors determining logistics sprawl adapted from Jaller & Leticia, 2017 (Source: Jaller & Leticia, 2017)

<table>
<thead>
<tr>
<th>Factors</th>
<th>Definition</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land available for expansion</td>
<td>Space that can be acquired or existing space that can be converted to intensify usage or storage capacity onsite. Zoning can affect this factor (e.g., parking requirements onsite reduces storage capacity)</td>
<td>Flexibility to expand or contract depending on the state of business. The ability to expand onsite rather than purchasing or renting a separate facility.</td>
</tr>
<tr>
<td>Number of dock doors</td>
<td>Number of dock doors</td>
<td>Appropriate number of dock doors for operation needs</td>
</tr>
<tr>
<td>Proximity to highways</td>
<td>On-road distance to the highway and time it takes to get to the highway</td>
<td>To allow for the ease of good transports by trucks</td>
</tr>
<tr>
<td>Public transit availability</td>
<td>On-road distance to public transit and time it takes to get to public transit</td>
<td>For workers (typically unskilled) to get to work</td>
</tr>
<tr>
<td>Long combination Vehicle accessibility</td>
<td>Surrounding roads and facility’s yard wide enough for long combination vehicle to maneuver</td>
<td>Infrastructure (e.g., large enough surrounding roads) available for operators to utilize long combination vehicles</td>
</tr>
<tr>
<td>Proximity to airport</td>
<td>Distance and travel time by truck to airport</td>
<td>To reduce truck drayage cost and time to airports</td>
</tr>
<tr>
<td>Proximity to sea port</td>
<td>Distance and travel time by truck to sea port</td>
<td>To reduce truck drayage cost and time to sea ports</td>
</tr>
<tr>
<td>Proximity to rail intermodal facility</td>
<td>Distance and time to rail intermodal facility</td>
<td>To reduce truck drayage cost and time to intermodal rail</td>
</tr>
<tr>
<td>Ability to operate 24 hours per day, 7 days per week</td>
<td>The ability to increase and decrease operation by time of day to avoid peak hour traffic</td>
<td>More flexibility and control of operations</td>
</tr>
<tr>
<td>Trailer parking/truck staging areas</td>
<td>Land available for staging areas and outside storage of trucks / trailers</td>
<td>A reduction in the amount of floor space required for just in time (JIT) deliveries is often offset by more land being required for outside storage, and staging areas</td>
</tr>
<tr>
<td>Telecommunication systems</td>
<td>Communication technologies between the warehouse, suppliers and customers</td>
<td>Certain regions do not have good telecommunication systems, which are a major requirement in modern economy</td>
</tr>
<tr>
<td>Quality and reliability of modes</td>
<td>Quality of the transportation services between the warehouse, suppliers and customers</td>
<td>Ability to have timely deliveries, delivery to the current location and undamaged goods</td>
</tr>
<tr>
<td>Access to customers</td>
<td>Distance and time to deliver goods to customers</td>
<td>To allow for constant and on-time deliveries</td>
</tr>
<tr>
<td>Access to suppliers</td>
<td>Distance and time to obtain goods from suppliers</td>
<td>Minimization of travelling time and distance</td>
</tr>
<tr>
<td>Customer population in surrounding area</td>
<td>Customer population in the surrounding area of the facility</td>
<td>Maximize access to customer base</td>
</tr>
<tr>
<td>Spending power of population in the surrounding area</td>
<td>Income of the population in the surrounding area of the facility</td>
<td>Maximize access to high spending customer base</td>
</tr>
<tr>
<td>Distance from competitor to customers</td>
<td>Distance from competitor to customers</td>
<td>Increase advantage over competitors</td>
</tr>
<tr>
<td>Availability of skilled workers</td>
<td>Sales, administrative staff, trained forklift drivers etc.</td>
<td>Access to essential personnel</td>
</tr>
<tr>
<td>Availability of unskilled workers</td>
<td>Workers that would have to be trained before they can be operational</td>
<td>Access to labour</td>
</tr>
<tr>
<td>Pro-business regulatory environment</td>
<td>How active municipalities are attracting business through various incentives</td>
<td>Reduced cost and have more control of operation</td>
</tr>
<tr>
<td>Zoning and construction plan</td>
<td>Different development plans, implementations and arrangements at alternative locations</td>
<td>To ensure that the zoning and regulator’s visions for the land that the facility to be built on matches the vision of the firm</td>
</tr>
<tr>
<td>Land costs/tax rates</td>
<td>Operating cost</td>
<td>Reduce operating cost</td>
</tr>
<tr>
<td>Proximity to other similar businesses</td>
<td>Logistics campuses - where similar businesses are in the same complex</td>
<td>Logistics campuses were seen as a way for companies to reduce costs</td>
</tr>
<tr>
<td>Labour costs</td>
<td>Wages, salaries etc.</td>
<td>Operating cost</td>
</tr>
<tr>
<td>Transportation costs</td>
<td>Fuel and equipment cost</td>
<td>Cost of transporting goods</td>
</tr>
<tr>
<td>Handling costs</td>
<td>Cost of good storage</td>
<td>Operating cost</td>
</tr>
</tbody>
</table>
Details of the exogenous reasons for logistics sprawl are provided in the following sections.

2.2.1 Land Availability and Affordability

Today’s warehouses are not just storage facilities; they provide a location for freight handling, sorting, consolidation, break-bulk, labelling, packaging etc. Logistics hubs function as transfer points between long distance haulage and urban distribution (Lavassani et al., 2008). Hence, with the evolution of functional diversity of the modern warehouses they require larger spaces than before for their operations. Logistics facilities such as warehouses and distribution centres tend to locate farther from urban centres where jobs and population are concentrated (Cidell, 2010; Dablanc & Rakotonarivo 2010; Dablanc & Ross, 2012; Sakai et al., 2015). Sub-urban areas generally have lower land prices and cheaper rents in comparison to the city core, and offer the provision of open land providing accessibility to important transportation infrastructure. Logistics facilities may move towards sub-urban areas because of lack of industrial land in core urban areas as observed in Tokyo, Japan (Sakai et al., 2016). Location of warehouses at the periphery of the city leads to trade-offs between rental costs and transportation costs (Combes & Cerema, 2016). However, studies provide opposing conclusions on areas where logistics sprawl is not evident, such as in Seattle. This is due to geographical conditions, land use patterns and local policies that guide logistics facilities location choices (Dablanc et al., 2014b; Woudsma et al., 2016).

2.2.2 Proximity to Customers and Transportation Infrastructure

Population size, spending power of customers, and proximity to those customers have been identified as important factors for facility location in retail distribution (Vlachopolou et al., 2001; Jakubicek & Woudsma, 2011). Thus, highways and airport transportation networks are an important factor in location choice of logistics facilities (Jakubicek & Woudsma, 2011; Dablanc & Rakotonorivo, 2010; Dablanc et al., 2014b). Transportation infrastructure such as highways and intermodal terminals provide access to broader markets and better freight transfer facilities. Some studies provide contrasting results, for example, a Paris case study of parcel and express transport facilities found that proximity to customers is not an important factor (Dablanc & Rakotonorivo, 2010).
2.2.3  Accessibility to Labour and Supply Chain

Some regions offer geographical advantage to firms in terms of labour markets, and accessibility to both suppliers and customers (Hesse, 2008). Many firms cluster together in space to increase productivity, and to take advantage of agglomeration economies, localized pooling of skilled workers resulting from co-location and from knowledge spillovers (Brown et al., 2013). Geographical industrial clustering of establishments increases the propensity of remaining in the same location or decrease mobility (Maoh & Kanaroglou, 2007). Jakubicek & Woudsma (2011) found that availability of skilled labour and ability of the logistics facilities to operate 24 hours per day, seven days per week are two of the most important factors in locating warehouses.

2.2.4  Regulatory Environment

Government policies play an important role in location choice of logistics firms. Government policies can either promote or constrain the development of logistics facilities and freight terminals. Land use, environmental/safety regulations and taxation policy can either constrain or advantage certain locations (Jaller & Leticia, 2017). Dablanc & Rakotonorivo (2010) showed that local governments give explicit consideration to logistics activities either for the jobs they create, or due to their adverse impacts on communities. Jakubicek & Woudsma (2011) identified the need for municipalities to give more attention to improving the operational environments of logistics firms, and the need to better understand industry requirements.
2.3 Implications of Logistics Sprawl

Logistics facilities generate jobs and economic development, but sub-urbanization of these facilities has some negative consequences. Logistics sprawl has the potential to increase commercial vehicle kilometers travelled for urban pick-up and delivery, and thus cause an environmental impact through additional greenhouse gas emissions. Logistics sprawl can result in large scale warehouse development that consumes large tracts of land at the fringe, can increase the contribution of freight to congestion, and can negatively affect delivery times. Various impacts of logistics sprawl have been discussed in the following sections.

2.3.1 Impact on Urban Freight Geography

Changing spatial patterns of logistics facilities have the potential to alter the structure of urban freight geography. Warehouses and distribution centers have been recognized as major freight generators and attractors which attract increasing level of trucking activities and layout of the roads may not be well suited for heavy trucks (Allen et al., 2012). Presence of large logistics facilities can result in significant wear and tear of the roads accompanied by traffic congestion (Cidell, 2015). In comparison to the logistics facility itself urban freight activities impact the overall built environment over a larger spatial extent (Hesse, 2008). Hesse (2008) also raised concern on encroachment between freight transport and other urban land uses because of the conflict arising due to heavy truck activities. The sub-urbanization trend of logistics facilities has the potential to push the boundaries of urban areas as observed in Paris and Melbourne where large number of these facilities are being located in the sub-urban areas (Heitz & Dablanc, 2015; O’Connors & Parsons, 2011). Lack of land-use control can result in uncoordinated and disorganized growth in logistics facilities which then have a poor integration into the urban environment as observed in Paris (Heitz & Dablanc, 2015).

2.3.2 Contribution to Vehicle Kilometers Travelled by Trucks

Logistics sprawl has resulted in greater distances travelled by trucks to serve customers (either final customer or retail establishment) located in urban areas. Mega- distribution centers located in sub-urban areas that serve metropolitan and regional markets, have to travel greater distance between the distribution center and the receivers for making deliveries in the urban areas (Andreoli et al., 2010). Dablanc (2013) highlighted that freight trucks have to travel more
distance in urban areas as logistics facilities are becoming much larger in size and are serving more businesses and households than they used to do. Empirical study done on Paris found an increase of ten kilometers per terminal for delivering goods due to sub-urbanization of logistics facilities (Dablanc & Rakotonarivo, 2010).

2.3.3 Environmental Impacts

Relocation of logistics facilities in sub-urban areas result in additional distance to be travelled by delivery trucks and this can result in negative environmental impact through Greenhouse Gas emissions and pollution. The study on Paris found that changes in the location patterns of parcel and express parcel terminals added 400 vehicle-km every day and thus 1,500 tonnes per year of CO₂ emissions (Dablanc & Rakotonarivo, 2010). One criticism of much of the relevant literature regarding this impact is that most studies alleged that logistics sprawl have contributed negatively to the urban environment primarily based on analysis of previous studies without empirically measuring the impact of warehousing locational shifts on the urban environment (Aljohani & Thompson, 2016). Sakai et al., (2016) argued the detrimental effects of sub-urbanization of logistics facilities as several factors relevant to individual facilities (such as type of vehicles used, time of operation, shipment origins and destinations etc.) needs to be taken into consideration.

2.3.4 Impact on Commuting of Logistics Employment

The relocation of logistics facilities to sub-urban areas affect the patterns and commuting modes for those employed in the logistics industry (Aljohani & Thompson, 2016). The location of logistics facilities in sub-urban areas can increase the employee’s dependency on automobile and result in longer commute as these areas may not be well serviced by public transportation (UN-Habitat, 2013, pp. 65). The lack of access to public transportation services to these logistics facilities might contribute to a smaller local labour pool as employees who lack access to personal vehicle might have difficulty travelling from urban areas to these logistics facilities (US Department of Transport: Federal Highway Administration, 2012).
Chapter 3
Data

Description of various datasets used for identifying and characterizing new warehousing establishments, and analyzing the spatial relationships within multi-establishments firms is provided in the following sections.

3.1 NAICS 493 Warehousing and Storage

The North American Industry Classification System (NAICS) is the standard developed by statistical agencies of Canada, Mexico and the United States to classify businesses based on their economic activity. NAICS was adopted in 1997 to replace the Standard Industry Classification (SIC) system and is designed to provide common definitions of the industrial structure of the three countries and a common statistical framework to facilitate the analysis of the three economies (Statistics Canada, 2017). The structure of NAICS is hierarchical and uses a six-digit numbering system to classify different levels of any sector. The first two digits designate the sector, the third digit designates the subsector, the fourth digit designates the industry group and the fifth digit designates the industry, and sixth digit is used to designate national industries (Statistics Canada, 2017).

This study focuses on businesses classified as NAICS 493 Warehousing and Storage, which is a subsector of NAICS 48-49 Transportation and Warehousing. This sub-sector comprises establishments primarily engaged in operating general merchandise, refrigerated and other warehousing and storage facilities, and also includes third party warehouses serving retail chains and wholesalers (Statistics Canada, 2017). The four sub-categories of NAICS 493 warehousing establishments are shown in Table 2. The sub-sector includes both public and contract warehousing establishments.
Table 2: NAICS 493 sub-categories (Source: Statistics Canada, 2012)

<table>
<thead>
<tr>
<th>NAICS Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>49311</td>
<td>General Warehousing and Storage</td>
</tr>
<tr>
<td>49312</td>
<td>Refrigerated Warehousing and Storage</td>
</tr>
<tr>
<td>49313</td>
<td>Farm Product Warehousing and Storage</td>
</tr>
<tr>
<td>49319</td>
<td>Other Warehousing and Storage</td>
</tr>
</tbody>
</table>

This focus is consistent with other studies which focus exclusively on NAICS 493 Warehousing and Storage as a proxy for all logistics facilities (Dablanc et al., 2014). Table 3 shows the type of logistics firms studied and the corresponding data sources used by researchers.

Table 3: Types of logistics firms studied (Source: Woudsma et al., 2016)

<table>
<thead>
<tr>
<th>Author</th>
<th>Firm Classification</th>
<th>Data Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bowen, 2008</td>
<td>NAICS 493 – Warehousing and Storage</td>
<td>US County Business Patterns</td>
</tr>
<tr>
<td>Cidell, 2010</td>
<td>NAICS 48&amp;49 – Transportation and Warehousing</td>
<td>US County Business Patterns</td>
</tr>
<tr>
<td>Dablanc et al., 2013</td>
<td>NAICS 493- Warehousing and Storage</td>
<td>US County Business Patterns</td>
</tr>
<tr>
<td>Sivitanidou, 1996</td>
<td>Survey of Firms</td>
<td>Survey of Warehousing Firms</td>
</tr>
<tr>
<td>Dablanc &amp; Ross, 2012</td>
<td>NAICS 493 – Warehousing and Storage</td>
<td>US County Business Patterns</td>
</tr>
</tbody>
</table>

Several datasets are used for analyzing the characteristics of new NAICS 493 warehousing establishments in the GTHA, and for analyzing spatial relationships within multi-establishment firms. The datasets used are DMTI Spatial Inc. Enhanced Points of Interest (EPOI) dataset, InfoCanada Business Establishments dataset, Environmental Systems Research Institute (ESRI) Business Analyst Online App (data sourced from Environics Analytics), the Teranet property parcel dataset, Google Maps and DMTI transportation networks. New warehousing establishments that have been established in GTHA since 2003 are identified using the DMTI EPOI dataset, while the remaining datasets are used to assess the additional characteristics and
spatial relationships within logistics facilities. Detailed description of each of these datasets is provided in the following sections.

3.2 DMTI Enhanced Points of Interest (EPOI)

The primary data used for analysis is the DMTI Spatial Inc. Enhanced Points of Interest (EPOI) dataset. The EPOI dataset is a national database of over 1 million Canadian business and recreational points of interest that has been engineered using CanMap Streetfiles (DMTI Spatial Inc., 2013). The dataset contains information on the name, location, contact, geographic coordinates, primary and secondary industry classification, and common address flag. The EPOI datasets for the years 2003 to 2013 are used. Only the primary industry classification is considered for analysis. The EPOI dataset for each year is directly imported into ArcMap and the concerned warehousing businesses within the boundary of GTHA are selected.

3.3 InfoCanada Business Establishment Dataset

The InfoCanada Business Establishment dataset provides information on the business name, location, contact, geographic coordinates, primary industry classification both NAICS and SIC Code, description of industry classification, year business appeared in yellow pages, annual sales volume and number of employees. The InfoCanada dataset is used to extract information on location of distribution centers for major firms in the GTHA. Major firms are identified using their annual revenue as described in Section 8.3. Despite the classification of NAICS 493 establishments given by Statistics Canada, the warehouses and distribution centres of large retailers such as Walmart, Canadian Tire, Costco, Metro, Sears, and Best Buy are not classified under NAICS 493 in the EPOI dataset. These facilities are classified under NAICS 42399004 Other Miscellaneous Durable Goods Merchant Wholesalers in InfoCanada Business Establishment dataset and in ESRI Business Analyst App (Section 3.4).

3.4 ESRI Business Analyst Application

ESRI Business Analyst is an online application that provides information on the business name, location, geographic coordinates, primary industry classification both NAICS and SIC Code, annual sales volume and number of employees. The application uses data sourced from Environics Analytics. This application is used to extract information on the location of distribution centers of major firms in the GTHA that could not be identified in the InfoCanada
dataset. Distribution centers of only multi-establishment firms are identified using the above two datasets to serve the methodology used for analyzing spatial relationships within multi-establishment firms as described in Section 8.3.

3.5 Teranet Property Parcels

The Teranet property parcel database provides information about Ontario’s estimated 4 million land parcels and has been assembled by the Ontario Ministry of Natural Resources, the Municipal Property Assessment Corporation (MPAC), and Teranet Enterprises Inc. (Ontario Parcel, 2002). The Ontario Parcel database contains information on the parcels, such as parcel length and parcel area. This dataset is used to determine the property area used by the new warehousing establishments in the GTHA. The parcel dataset is directly imported into ArcGIS and the new warehousing establishments identified in the DMTI EPOI dataset are joined spatially to parcels to determine the property areas for the warehouses.

3.6 Google Maps

Aerial photographs from Google Maps are used to identify the warehouse building footprints. The building image is identified using the geographic coordinates of the establishments. The size of the building footprint of new warehousing establishments are estimated using Gray-Level Co-occurrence Matrix (texture-based image segmentation). Details of the methodology are provided in Appendix A. Also, Google Maps are used to identify the location of distribution centers of major firms in the GTHA for analyzing spatial relationships within multi-establishment firms. Details have been provided in Section 8.2.

3.7 Transportation Networks

DMTI transportation networks are used to assess the distance of the new warehousing establishments from major highways, from the three intermodal rail terminals and the Pearson International Airport. Canadian National (CN) and Canadian Pacific (CP) railways both have intermodal rail terminals in Toronto region. CN Rail has two intermodal terminals located in Brampton and Mississauga (CN Rail) and CP Rail has one intermodal terminal facility in Vaughan (CP Rail). Network distances from the new warehousing establishments to the intermodal terminals and the Pearson airport are obtained using Google Maps Distance Matrix API.
4.1 Dataset Development

The methodology to develop annual warehousing establishment datasets, and to identify new warehousing establishments is discussed in the following sections.

Step 1: Identification of mini-warehouses

As noted by Woudsma et al., (2016) mini-warehouses and self-storage units are mis-classified under SIC 422 Public Warehousing and Storage. They should have been classified under NAICS 53113 Lessors of mini warehouses and self-storage units. The EPOI dataset is refined to include only warehousing and storage establishments. The mini-warehouses are identified in the dataset using keyword searches on the name of the establishments. Keywords used are mini, self, storage, public and movers. They are removed from the datasets after cross checking them by visiting their websites or other online business directories. Despite the availability of NAICS classifications in the EPOI dataset from 2010 onwards, the SIC 422 classification is used to distinguish warehouses and mini-warehouses to maintain consistency across datasets for each analysis year. Table 4 shows the total number of SIC 422 classified businesses, warehouses and mini-warehouses identified in the EPOI dataset for each year. Table 4 shows that many mini-warehouses are classified under NAICS 493 and are incorrect classifications.
## Table 4: Number of classified warehouses and mini-warehouses in the DMTI EPOI dataset

<table>
<thead>
<tr>
<th>Year</th>
<th>Total Number of SIC 422 warehouses</th>
<th>Number of warehouses (correctly classified)</th>
<th>Number of mini-warehouses (incorrectly classified and removed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>318</td>
<td>173</td>
<td>145</td>
</tr>
<tr>
<td>2004</td>
<td>320</td>
<td>173</td>
<td>147</td>
</tr>
<tr>
<td>2005</td>
<td>230</td>
<td>156</td>
<td>74</td>
</tr>
<tr>
<td>2006</td>
<td>422</td>
<td>226</td>
<td>196</td>
</tr>
<tr>
<td>2007</td>
<td>489</td>
<td>255</td>
<td>234</td>
</tr>
<tr>
<td>2008</td>
<td>492</td>
<td>258</td>
<td>234</td>
</tr>
<tr>
<td>2009</td>
<td>417</td>
<td>235</td>
<td>182</td>
</tr>
<tr>
<td>2010</td>
<td>664</td>
<td>229</td>
<td>435</td>
</tr>
<tr>
<td>2011</td>
<td>684</td>
<td>245</td>
<td>439</td>
</tr>
<tr>
<td>2012</td>
<td>722</td>
<td>251</td>
<td>471</td>
</tr>
<tr>
<td>2013</td>
<td>722</td>
<td>251</td>
<td>471</td>
</tr>
</tbody>
</table>

### Step 2: Conversion from SIC 422 to NAICS 493 classification

Establishments in the EPOI dataset until the year 2009 use the SIC industry classification, while datasets from year 2010 to 2013 have both SIC and NAICS industry classification. Industry classifications must be coordinated to make a longitudinal comparison. Since the definitions for SIC warehousing businesses and NAICS warehousing businesses are the same, we assume direct conversion between SIC 422 and NAICS 493. The number of NAICS 493 establishments from 2003 to 2013, according to the DMTI EPOI dataset is shown in Figure 2.
Figure 2: Number of NAICS 493 establishments in the DMTI EPOI dataset

Step 3: Longitudinal comparison for identifying new warehousing establishments

The EPOI datasets for the years 2003 to 2013 are used to identify the new warehousing establishments that have located in the GTHA. A total of 161 new warehousing establishments are found between 2003 and 2013. A longitudinal comparison is made across each year starting with 2003 to analyze the spatial distribution of the new warehousing establishments, and the dynamics of appearances, disappearances, and relocations of the establishments over the eleven-year period. Details of the methodology are described in the section below.

Let $X_n$ be the set of warehousing establishments for year $n$, where $2003 \leq n \leq 2013$.

Let $A_{(n-1, n)}$ be a set of new warehousing establishments between that have appeared between years $n-1$ and $n$. An establishment is considered to be new in year $n$ if it is present in $X_n$ but not in $X_{n-1}$ or in the datasets of the years prior to year $n-1$. If the establishment is found in datasets of any year prior to year $n-1$, then it is considered to be a missing data and is added to the years between the year from which it is found until year $n-1$. 
Let $D_{(n-1, n)}$ be a set of warehousing establishments that have disappeared between years $n-1$ and $n$. An establishment is considered to have disappeared in year $n$ if it is present in $X_{n-1}$ but not in $X_n$ provided the establishment does not appear in the datasets of the years later than year $n$. If the establishment is found in datasets of any year after $n$, then it is considered as a missing data and is added to the years between $n$ and the year where it reappears in the dataset. An establishment is considered to have relocated in year $n$ if it is present in both the years $X_{n-1}$ and $X_n$ but has a different address.

### 4.2 Warehousing Sprawl

Different methods are available for analyzing the spatial distribution of logistics facilities. In this study, the degree of warehousing sprawl in the GTHA is quantified using two techniques, centrographic analysis and kernel density estimation. Details of these techniques are provided in the following sections.

#### 4.2.1 Centrographic Analysis

Centrographic analysis has been widely used to assess the distribution of logistics facilities, and to address planning and policy issues of freight and logistics activities in several metropolitan areas (Dablanc & Ross, 2012, Dablanc et al., 2014b, Heitz & Dablanc, 2015, Sakai et al., 2016). The purpose of centrographic analysis is to find an elliptical graph that best fits a point distribution, and to derive statistics from the graph that describe spatial distribution (Brown and Holmes, 1971). Steps followed in the centrographic analysis technique are as follows:

**Step 1: Calculate the barycenter of establishments**

The barycenter (the mean centre) of all warehousing establishments for each year is calculated using their projected geographic coordinates in ArcMap. The projected coordinate system used for analysis is the Universal Transverse Mercator (UTM) Zone 17 N. The barycenter of the warehousing establishments for each year is calculated using the following equation (Yeates, 1974).

$$
\overline{X}_w = \frac{\sum_{i=1}^{n} x_i}{n}, \quad \overline{Y}_w = \frac{\sum_{i=1}^{n} y_i}{n}
$$

(1)
where,

$\bar{X}_w = $ north-south coordinate of the barycenter

$\bar{Y}_w = $ east-west coordinate of the barycenter

$x_i = $ north-south coordinate of warehousing establishment $i$

$y_i = $ east-west coordinate of warehousing establishment $i$

$n = $ total number of warehousing establishments in the region

**Step 2: Obtain the Standard Deviational Ellipse (SDE)**

Standard Deviational Ellipse (SDE) is a measure of dispersion from the barycenter. SDE summarizes the spatial characteristics and spread of the warehousing establishments. Comparing the area and orientation of the SDE for different years provides an indication of decentralization and direction of movement of the establishments.

\[
S_C = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{X}_w)^2 + \sum_{i=1}^{n}(y_i - \bar{Y}_w)^2}{n}}
\]  

(2)

where,

$S_C = $ standard distance from the barycenter

$\bar{X}_w = $ north-south coordinate of the barycenter

$\bar{Y}_w = $ east-west coordinate of the barycenter

$x_i = $ north-south coordinate of warehousing establishment $i$

$y_i = $ east-west coordinate of warehousing establishment $i$
n = total number of warehousing establishments in the region

**Step 3: Calculate average distance of all establishments from the barycenter**

Average distance of all the warehousing establishments from the barycenter is calculated for each year separately using equation (3). This helps to analyze the deconsolidation trend of warehousing establishments by measuring the changes in the average distance of establishments from their barycenter across different years.

\[
d = \frac{\sum_{i=1}^{n} \sqrt{\left(\overline{X}_w - x_i\right)^2 + \left(\overline{Y}_w - y_i\right)^2}}{n}
\]  

(3)

Where,

\(d\) = average distance between the barycenter and warehousing establishments

\(\overline{X}_w\) = north-south coordinate of the barycenter

\(\overline{Y}_w\) = east-west coordinate of the barycenter

\(x_i\) = north-south coordinate of establishment \(i\)

\(y_i\) = east-west coordinate of establishment \(i\)

\(n\) = total number of warehousing establishments in the region

**4.2.2 Kernel Density Estimation**

Kernel Density Estimation (KDE) is another method to visualize spatial patterns of logistics facilities and to identify hot spots or clusters of establishments (Heitz & Dablanc, 2015, Sakai et al., 2016). KDE produces a smooth, continuous surface where each location in the study area is assigned a density value irrespective of arbitrary administrative boundaries (Carlos et al., 2010). KDE calculates the density of features in the neighbourhood around each point and a smoothly curved surface is fitted over the features. The area of the neighbourhood is defined by the search
radius. The surface value is highest at the location of the point and diminishes with increasing
distance from the point, reaching zero at the search radius distance from the point (ESRI). The
estimated kernel density $f(x)$ at any point $x$ is given by the following equation (Silverman, 1986).

$$f(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)$$ (4)

where,

$n$ = total number of establishments in the area

$h$ = search radius

$x - x_i$ = distance between the concerned establishment and all other establishments

$K$ = quadratic kernel function defined as

$$K(x) = \frac{3}{4} (1 - x^2), \quad |x| \leq 1 \text{ and } K(x) = 0, \quad x > 1 \quad \text{and} \quad \int K(x) dx = 1$$

For this study of the GTHA, a bandwidth of 3 km and cell size of 500m X 500m is chosen.

4.3 Building Footprint Calculation (Zhang, 2017)

Aerial images used for building footprint calculations are obtained using Google Static Maps
API. Google Static Maps API enables users to extract and download a Google Maps image. A
python-based tool is developed to acquire aerial photographs using geographic coordinates of the
warehousing establishments at an appropriate scale. A geocode adjustment is required when the
geographic coordinates in EPOI dataset do not centre on the warehouse building. Gray level Co-
ocurrence Matrix (texture based image segmentation) (Haralick et al., 1973) is used for building
footprint calculations. Details of the methodology are provided in Appendix A.

4.4 Distance Calculations from Nearest Highway, Intermodal Rail
Terminals and Toronto Pearson International Airport

Distance of the new warehousing establishments from the nearest highway, the three intermodal
terminals described in Section 3.7, and the Toronto Pearson International Airport are calculated
using Google Maps Distance Matrix API. This service provides the shortest travel time distance and travel time for a matrix of origins and destinations, based on the recommended route between start and end points.
Chapter 5
Characteristics of NAICS 493 establishments

In this chapter, spatial characteristics of NAICS 493 warehousing establishments from year 2003 to 2013 are provided. Section 5.1 describes the appearance and disappearance of warehousing establishments across each year from 2003 and 2013. Sections 5.2 to 5.4 show the location of barycenter and SDE of warehousing establishments for the pre-recession (2003-2007), recession (2008-2009) and post-recession (2010-2013) periods.

5.1 Appearance and Disappearance of NAICS 493 Establishments

Figure 3 shows the appearance and disappearance of warehousing establishments across each year in the GTHA. This examination helps to understand the growth in the warehousing industry in the GTHA. Significant growth in the number of warehousing establishments occurred between 2005-2006 (72 new establishments), and between 2006-2007 (29 new establishments). During the height of the recession period (2008-2009) the number of closures (36 establishments) exceeded the number of new establishments (13). And in the immediate aftermath of the recession (2009-2010) there was a large turbulence in the warehousing sector. From 2009 to 2010, 96 warehousing establishments disappeared while 90 establishments appeared. 18 of these were relocations. The trend of growth resumed at a slower rate after the recession in the period from 2010 to 2012 (28 new establishments over these two years). Of the 132 warehousing establishments that disappeared during the recessionary period (2008-2010), many were young establishments. In total, 73 of these establishments had started in the period from 2003 to 2010.

Using longitudinal comparison as described in Section 4.1, 161 new warehousing establishments have been identified to be located in the GTHA between 2003 and 2013. Figure 4 shows the appearance and disappearance of warehousing establishments within the 161 new NAICS 493 warehousing establishments in the GTHA. A total of 239 new warehousing establishments came up in the GTHA since 2003 but 78 of them disappeared with major disappearances happening during recessionary period (2008-2010).
Figure 3: Appearance and disappearance within NAICS 493 establishments

<table>
<thead>
<tr>
<th>Year</th>
<th>Appeared</th>
<th>Disappeared</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003-2004</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2004-2005</td>
<td>2</td>
<td>19</td>
</tr>
<tr>
<td>2005-2006</td>
<td>72</td>
<td>2</td>
</tr>
<tr>
<td>2006-2007</td>
<td>29</td>
<td>0</td>
</tr>
<tr>
<td>2007-2008</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>2008-2009</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>2009-2010</td>
<td>90</td>
<td>36</td>
</tr>
<tr>
<td>2010-2011</td>
<td>18</td>
<td>96</td>
</tr>
<tr>
<td>2011-2012</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>2012-2013</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

Figure 4: Appearance and disappearance within new NAICS 493 establishments

<table>
<thead>
<tr>
<th>Year</th>
<th>Appeared</th>
<th>Disappeared</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003-2004</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>2004-2005</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>2005-2006</td>
<td>72</td>
<td>0</td>
</tr>
<tr>
<td>2006-2007</td>
<td>29</td>
<td>0</td>
</tr>
<tr>
<td>2007-2008</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>2008-2009</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>2009-2010</td>
<td>90</td>
<td>21</td>
</tr>
<tr>
<td>2010-2011</td>
<td>18</td>
<td>52</td>
</tr>
<tr>
<td>2011-2012</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>2012-2013</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>
5.2 Pre-Recession Period (2003-2007)

Figure 5 shows the barycenter and SDE of NAICS 493 establishments during pre-recession period. During this period, the area of the ellipses increased by 11.4%. The average distance of the warehousing establishments from their barycenter increased by 0.74 km during this period. A significant change in size as well as directional orientation of SDE can be observed between 2005-2006 as this was a time when a large number of warehousing establishments appeared in the GTHA (refer to Figure 3).

Figure 5: Spatial patterns of NAICS 493 establishments during pre-recession period
5.3 Recession Period (2008-2009)

Figure 6 shows the barycenter and SDE of NAICS 493 establishments during the recession period. During this period, the area of the ellipses increased slightly by 1.8%. The average distance of the warehousing establishments from their barycenter increased slightly by 0.28 km. During this period, significant number of warehousing establishments disappeared along with some appearances, thus shifting orientation of the SDE.

Figure 6: Spatial patterns of NAICS 493 establishments during recession period
5.4 Post-Recession Period (2010-2013)

Figure 7 shows the barycenter and SDE of NAICS 493 establishments during the post-recession period. During this period, the area of the ellipses remained approximately the same (reduction of 0.07%). The average distance of the warehousing establishments from their barycenter decreased very slightly by 0.11 km. Some growth in the number of warehousing establishments occurred between 2010-2011 (approx. 7%) and 2011-2012 (approx. 2.5%).

Figure 7: Spatial patterns of NAICS 493 establishments during post-recession period
5.5 Distance from the Barycenter and Movement of the Barycenter

Figure 8 shows the increase in the average distance of the warehousing establishments from their barycenter from 2003 to 2013. The average distance of the establishments from the barycenter increased by 1.3 km between the year 2003 and 2013. The periods of greatest sprawl, according to this measure, were from 2004 to 2006 (pre-recession) and from 2008 to 2010 (during the recession). These are the years in which the greatest number of warehousing establishments appeared and disappeared in the GTHA (refer to Figure 3). Average distance of the warehousing establishments from their barycenter did not increase further from 2010 to 2013.

![Figure 8: Average distance of NAICS 493 establishments from their barycenter](image)

Figure 9 shows the movement of the barycentre of the warehousing establishments from 2003 to 2013. Overall, the barycentre moved 0.6 km north between 2003-2013. However, the movement of the barycentre from year to year does not show any clear pattern of movement in one direction and so no clear conclusion can be made about the spatial pattern of warehousing establishments.
Figure 9: Movement of the barycenter of NAICS 493 establishments
Chapter 6
Results and Discussion

6.1 Warehousing Sprawl

According to the DMTI EPOI dataset, there has been an increase of 45% in the number of warehousing establishments in the GTHA from 2003 to 2013 (refer to Table 4). Figure 10 shows the percentage increase in the NAICS 493 establishments across the census divisions in the GTHA. The Region of Peel, a major centre of logistics activities in Ontario is found to have the highest growth in both the number and percent change of warehousing establishments. The City of Toronto is found to have the second largest absolute growth of approximately 20 warehousing establishments, and the Region of Durham and the City of Hamilton experienced percentage growth of over 40%.

Figure 10: Growth in NAICS 493 establishments by Census Divisions
Figure 11 summarizes warehousing sprawl in the GTHA between 2003 and 2013 using centrographic analysis. During this period area of the SDE increased by 10.5%, and the average distance of the warehousing establishments from their barycentre increased by 1.3 km. As noted earlier, the outward sprawl has been steadily increasing (Figure 3), but the movement of the barycentre from year to year has not shown any consistent trend, so this result is not conclusive. This finding is consistent with previous research which found that the logistics firms in the GTA are not sprawling significantly because of a lack of land required for expansion because of the Greenbelt (Woudsma et al., 2016).

Figure 11: Warehousing sprawl in the GTHA using Centrographic analysis

Centrographic analysis provides information only on the overall spatial distribution of establishments, and the approximate directional orientation. However, it is important to analyze the specific areas where the growth in warehousing establishments between 2003 and 2013 in the GTHA has occurred. Kernel density estimation helps to analyze the specific regions where
warehousing establishments have appeared, or regions where establishments have disappeared between 2003 and 2013.

Figure 12 shows the difference in the kernel densities of NAICS 493 establishments between 2003 and 2013. In strict terms, “sprawl” means relocations of activities in the sub-urban areas accompanied by a decrease in their concentration in the central areas (Heitz & Dablanc, 2015). The blue regions on the map indicate the areas where there is an increase in the density of warehousing establishments, while red regions on the map indicate the areas where there has been a decrease in the density of the establishments. Figure 12 shows that growth in number of warehousing establishments is most prominent in the Region of Peel near Toronto Pearson International Airport, the CN intermodal yard in Mississauga and the CN intermodal yard in Brampton. This growth area is also located near the junction points of Highway 401, Highway 407, Highway 410 and Highway 427 thus connecting them to broader markets. Another cluster of new warehousing establishments is found near the junction of Highway 400 and Highway 407 and some growth is observed in the City of Hamilton.

Dark red regions on the map indicate the areas where there has been a drop in the number of warehousing establishments. One of the prominent red coloured areas is the junction of Highway 427, Queen Elizabeth Way (QEW) and Gardiner Expressway east in southern Etobicoke. This region used to be the location of Obico Yard, a small intermodal terminal operated by CN Rail which was closed in 2012 and operations were consolidated at the Vaughan intermodal terminal (Lu, 2012). There are couple of red patches in Vaughan, Markham and the City of Toronto indicating that warehousing establishments have either disappeared or relocated from those locations. Analysis suggests that there have been 83 closures of warehousing establishments between 2003-2013 which include 17 relocations (approx. 21%), and these relocations have primarily occurred towards Brampton and Mississauga.
Figure 13 shows the distribution of average value of private dwellings in 2011 at the level of Census Tracts (Statistics Canada, 2011). Figure 14 shows the distribution of property parcels in the GTHA with size greater than 50,000 m². These two figures show that major growth in new warehousing establishments occurred around the Toronto Pearson Airport, and census subdivisions of Brampton, Mississauga and Milton as these regions have lower dwelling prices and offer larger land parcels as compared to the City of Toronto. These findings are consistent with the literature that warehousing establishments are concentrating in regions with rich transportation infrastructure, cheap and open land and intense freight sector businesses (Bowen, 2008; Cidell 2010; Heitz & Dablanc, 2015; Woudsma et al., 2016).
Figure 13: Average value of private dwellings (Source: Statistics Canada, 2011)

Figure 14: Distribution of property parcels with size greater than 50000 square meter
Chapter 7  
Characteristics of New NAICS 493 Establishments

7.1 Spatial Distribution

Using the EPOI datasets 161 new warehousing establishments are found to have been located in the GTHA since 2003. Figure 15 shows the distribution of new NAICS 493 warehousing and storage establishments. Amongst the new NAICS 493 warehousing establishments, there has been a significant increase in the number of General Warehousing and Storage (NAICS 49311) followed by NAICS 49312 Refrigerated Warehousing and Storage. NAICS 49311 establishments provide storage facilities for retail stores and handle goods in containers using equipment such as forklifts, racks, etc (Statistics Canada, 2012). NAICS 49312 establishments provide public and contract warehouse and storage services, using equipment designed to keep goods frozen or refrigerated, and include establishments engaged in the storage of furs for trade (Statistics Canada, 2012). This increase in the number of NAICS 49311 and NAICS 49312 could be attributed to the fact that there has been a significant growth in the consumer products industries and in consumption in that time period. The Canadian e-commerce industry has grown substantially in terms of volume and breadth of goods and services being purchased (CIRA, 2016).

Figure 15: Distribution of new NAICS 493 establishments
Figure 16 shows the geographic locations, the barycentre and the SDE of the new warehousing establishments from 2003 to 2013. The barycentre of the new warehousing establishments is located at the edge of Toronto Pearson International Airport and average distance of all the new establishments from their barycentre is found to be 17.7 km. The orientation of the SDE suggests that the new warehousing establishments are spreading out in north-east and south-west direction of the GTHA.

Figure 16: Spatial pattern of new NAICS 493 establishments

Figure 17 shows kernel density of the new warehousing establishments. The maximum density of the new warehousing establishments is found in the vicinity of Toronto Pearson International Airport, CN Mississauga, CN Brampton, and at the intersection of Highway 400 and Highway 407. It is evident from the map that there is another cluster of new warehousing establishments located in the City of Hamilton, on the south side of Hamilton Harbour, which has historically been the home of steel-making and other heavy industry. There are some clusters locating south-
west along Queen Elizabeth Way (QEW) starting from Mississauga through the town of Oakville until Burlington. Also, there are several new establishments located east along Highway 401 in Scarborough, Ajax, Whitby and Oshawa. Thus, we can say that the new warehousing establishments are locating in close proximity to intermodal rail terminals and at the junctions of major highways which is also evident from the findings in Section 6.1 (refer to Figure 12).

Figure 17: Kernel density of new NAICS 493 establishments
7.2 Cluster Analysis

Many firms cluster together in space because clustering influences productivity, and firms can take advantage of agglomeration economies, localized pooling of skilled workers resulting from co-location and from knowledge spillovers (Brown & Rigby, 2013).

A formal test is conducted to statistically verify the existence of warehouse clusters. We define the null hypothesis that the warehousing industry in the GTHA area is spatially randomly distributed or follows a complete spatial randomness (CSR). The test on CSR relies on the calculations of two types of nearest neighbour distance for each point, the observed nearest neighbour distance and the expected nearest neighbour distance. The expected mean nearest neighbour (NN) distance is derived from a hypothetical complete random distribution. NNR is the ratio of the observed mean distance \((D_O)\) to the expected mean distance \((D_E)\) as shown below (ESRI).

\[
NNR = \frac{D_O}{D_E}
\]  

\((5)\)

\(D_O\) is calculated using the equation below, where \(d_i =\) distance between feature i and its nearest neighbouring feature, and \(n=\) total number of features

\[
D_O = \frac{\sum_i d_i}{n}
\]  

\((6)\)

\(D_E\) is calculated using the equation below, given the assumption that the total number of features \((n)\) are completely randomly distributed in a minimum rectangle area \((A)\) that covers all the points.

\[
D_E = 0.5/\sqrt{n/A}
\]  

\((7)\)

If the resulting value of NNR is much less than 1, then existence of spatial clusters can be concluded with a high confidence, while value of NNR close to 1 implies spatial dispersion as shown in Figure 17. Z-score, for the statistic is calculated using the equation below. The larger the Z-score value, the higher confidence to reject the CSR or to support clustering.
\[ Z = \frac{(D_0 - D_E)}{SD} \]  

Where, \[ SD = \frac{0.26136}{\sqrt{n^2/A}} \]

The resulting NNR and Z-score suggest a clustering pattern with over 99% confidence level. As shown in Figure 18, the NNR value of 0.45 falls in the range of "clustered", and the Z-score of -13.27 rejects the null hypothesis of CSR at more than 99% confidence level.

![Figure 18: Test of existence of warehousing establishments clusters (generated by ArcMap)](image)

After analyzing the presence of clusters of warehousing establishments in the GTHA we also analyze how many clusters are present and where are they present. The Gap statistic is used to
identify the optimal number of warehouse clusters. Gap statistic (Tibshirani et al., 2001) works by comparing the within-cluster dispersion of any clustering algorithm with that expected under an appropriate reference null distribution. Let 

\[ D_r = \sum_{i,j \in C_r} d_{ij} \]  \hspace{1cm} (9)

be the sum of the pairwise distances for all points in cluster \( r \), and set 

\[ W_k = \sum_{r=1}^{k} \frac{1}{2n_r} D_r \]  \hspace{1cm} (10)

So, \( W_k \) is the pooled within-cluster sum of squares around the cluster means. The logic is to standardize the graph of \( \log(W_k) \) by comparing it with its expectation under an appropriate null reference distribution of the data. The gap statistic is defined as 

\[ \text{Gap}_n(k) = E_n^*[\log(W_k)] - \log(W_k) \]  \hspace{1cm} (11)

Where \( E_n^* \) denotes expectation under a sample size \( n \) from the reference distribution. Then we define the decision rule as the smallest \( k \) value that satisfies the condition of 

\[ \text{Gap}(k) \geq \text{Gap}(k+1) - s_{k+1} \]  \hspace{1cm} (12)

is considered the optimal number. The optimal number of clusters occurs at the solution with the largest local or global gap value. Here, \( s_{k+1} \) represents the normalized standard deviation of within-cluster point distances when the number of clusters equals \( k+1 \).

As shown in Figure 19, there are six optimal clusters of new warehousing establishments according to the gap statistic.
Figure 19: Optimal number of warehousing clusters using Gap Statistic
Figure 20 shows the location of these warehousing establishments clusters. As significant growth in the new warehousing establishments occurred in vicinity of the Toronto Pearson International Airport, three clusters occur around the airport. The first cluster is located towards north-east and south-east of the airport in the City of Toronto (red colour), the second cluster (light blue) is located north-west of the airport in Brampton, and the third (yellow) is located towards south-west of the airport in Mississauga. An additional three clusters have been identified in the GTHA, one of them extending from Oakville towards the City of Hamilton. The second extends north-west from eastern part of City of Toronto towards Newmarket. The last cluster extends from Ajax towards Oshawa.

![Map of clusters](image)

**Legend**
- Toronto Pearson International Airport
- Census sub-divisions GTHA
- Cluster
  - 1
  - 2
  - 3
  - 4
  - 5
  - 6

**Figure 20:** Clusters amongst new NAICS 493 establishments in the GTHA
7.3 Building Footprint and Property Parcel Size

Today’s warehouses are not just storage facilities; they provide a location for freight handling, sorting, consolidation and break-bulk, labeling and packaging activities etc. Logistics hubs function as transfer points between long distance haulage and urban distribution (Lavassani et al., 2008). Hence, with the evolution of functional diversity of the warehouses they require larger spaces than before for their operations. Figure 21 shows the distribution of new warehousing establishments according to their building footprint. The average building footprint of the new warehousing establishments is found to be approximately 12,000 m². Figure 22 shows the spatial distribution of new warehousing establishments according to building footprint. Out of 161 new warehousing establishments in the GTHA, 5 establishments are found to have property size greater than 50,000 m². A significant number of new warehousing establishments with building footprint greater than 25,000 m² has been found to be located in the vicinity of the Toronto Pearson International Airport, and some being in the City of Hamilton (refer to Figure 19).

![Figure 21: Building footprint distribution of new NAICS 493 establishments](image-url)
One of the major reasons identified for the sub-urbanization trend of warehouses in metropolitan areas is the requirement for open and cheap land. All the 161 new warehousing establishments are spatially joined to the Teranet’s property parcel data using ArcGIS. The average parcel size used by new warehousing establishments for development is found to be approximately 40,000 m². Figure 23 shows the distribution of new warehousing establishments according to the size of the property parcel used for development. Figure 24 shows the spatial distribution of new warehousing establishments according to the property parcel size. According to the map, a significant number of new warehousing establishments with a property parcel size greater than 25,000 m² are found to be located in the vicinity of Toronto Pearson International Airport and City of Hamilton. Figures 17 and 19 indicate the emergence of the City of Hamilton as another important cluster of warehousing activities in the GTHA after Toronto Pearson Airport.
Figure 23: Property parcel size distribution of new NAICS 493 establishments

Figure 24: Spatial distribution of new NAICS 493 establishments according to property parcel size
7.4 Distance from the Nearest Highway, Intermodal Rail Terminals and Toronto Pearson International Airport

Modern warehouses tend to locate closer to highways which provide access to global trade routes and broader markets. The average distance of the new warehousing establishments from the nearest highway is found to be 2.6 km. Figure 25 shows the distribution pattern of new warehousing establishments according to their distance from the nearest highway. The figure shows that most establishments are located within a distance of 2-3 km from the nearest highway.

![Figure 25: Distribution of distance of new NAICS 493 establishments from nearest highway](image)

Figure 26 shows a buffer of 2.6 km around major highways in the GTHA. About 70 per cent of the new warehousing establishments are located within this average distance from the nearest highway.
Figure 26: Buffer around major highways in the GTHA

The Canadian National Railway Company (CN Rail) and the Canadian Pacific Railway (CP Rail) are two Canadian Class I freight railways that provide transportation services including rail services, intermodal container services, trucking and supply chain solutions, warehousing and distribution, freight forwarding, intermodal shipping and logistics solutions. Hence, the warehousing establishments are expected to locate in proximity to these intermodal terminals to have access to their logistics services. The average distance from the CN Brampton and CN Mississauga terminals is found to be 24.2 km and 24.0 km, respectively. The average distance of all new warehouses from the CP Vaughan terminal is found to be 31.1 km.

Figures 27, 28 and 29 show the distribution of distance of new warehousing facilities from the CN Brampton, CN Mississauga and CP Vaughan intermodal terminals, respectively. The distance distribution to all the intermodal terminals is almost same with maximum number of establishments located within 0-10 km (from CN Brampton), and within 10-20 km (from CN
Mississauga and CP Vaughan). Also, the maps indicate that there are a small number of warehousing establishments that are located more than 50 km from all the intermodal terminals.

**Figure 27: Distribution of distance of new NAICS 493 establishments from CN Brampton intermodal terminal**
Figure 28: Distribution of distance of new NAICS 493 establishments from CN Mississauga intermodal terminal

Figure 29: Distribution of distance of new NAICS 493 establishments from CP Vaughan intermodal terminal
Figure 30 shows the distribution of new warehousing establishments from the Toronto Pearson International Airport. Toronto Pearson International Airport is the busiest airport by cargo tonnage in Canada and supports significant cargo services (Metrolinx, 2011). The average distance of the new establishments from the Pearson airport is found to be 20.6 km. Maximum number of warehousing establishments are located within a distance of 5-10 km from the Pearson airport.

Figure 30: Distribution of distance of new NAICS 493 establishments from Toronto Pearson International Airport

Figure 31 shows three rings around the Toronto Pearson International Airport with radius of 20 km, 40 km, and 60 km respectively. About 95 per cent of the new warehousing establishments are located within a radial distance of 60 km from the Toronto Pearson International Airport.
Figure 31: Radial rings around Toronto Pearson International Airport
Chapter 8
Spatial Relationships within Multi-Establishment Firms

8.1 Objective

An establishment is defined as a fixed physical location where economic activity occurs, and a firm may have one establishment (single-unit establishment) or many establishments (a multi-unit firm) (Census Bureau, 2015). One of the objective’s of this research is to analyze the spatial relationships within multi-establishment firms i.e. analyzing the spatial relationship between the logistics facility and other retail and logistics establishments within the same firm. The location choices of multi-establishment firms are likely to differ from those of single-establishment firms as each establishment’s decision influences the decisions of other establishments in a corporate network (Mota, 2017). This analysis is carried out to understand if location choice is part of a larger network planning decision within multi-establishment firms.

In particular, emphasis is placed on analyzing the spatial relationship between location of the firms distribution center(s) and their retail establishments in the GTHA. The analysis tries to provide an answer to the following questions. How do firms locate their distribution center(s) with respect to their retail stores? Are the distribution centers located close or far away from the retail stores? Do firms make a trade-off between transportation costs and land prices while making a location choice for their distribution center(s)? Do firms locate their distribution center(s) in proximity to distribution center(s) of other firms? What are the characteristics (property size, land-use pattern) of the land parcels where the distribution centers are located?

Despite the classification of NAICS 493 establishments given by Statistics Canada, the warehouses and distribution centres of large retailers such as Walmart, Canadian Tire, Costco, Metro, Sears, and Best Buy are not classified under NAICS 493 in the DMTI EPOI dataset. EPOI dataset provides no information on the location of distribution centers/warehouses of these firms except for Costco’s distribution center which has been categorized under NAICS 541614 Process, Physical Distribution and Consulting Services. Some of these logistics facilities have been identified in InfoCanada Business Establishments dataset and ESRI’s Business Analyst Online Application (data sourced from Environics Analytics) and they have been classified under NAICS 42399004 Other Miscellaneous Durable Goods Merchant Wholesalers. Even datasets
available from InfoCanada and Environics Analytics are not exhaustive and information on some of the distribution centers have been obtained from Google Maps.

### 8.2 Dataset Development

The retail sector has played an increasingly vital role in the Canadian marketplace, given its contribution to the Canadian economy and its impact on consumers and their consumption activities (Industry Canada, 2011). In 2011, the sector generated $457.4 billion in retail sales and represented approximately 12% of the Canadian workforce (Statistics Canada, 2012c). For this analysis, major retailers in Canada have been selected on the basis of their annual revenue from a report on Canadian retail industry by Industry Canada named *Canada’s Changing Retail Market* (Industry Canada, 2011). In order to expand the dataset, other retailers have been selected by searching for their distribution centers on Google Maps. The distribution centers used in analysis cater the need of various sub-sectors of the retail industry such as furniture and home furnishing, electronics and appliances, home improvement, grocery/food, pharmacy and personal care, clothing and accessory, hobby store, general merchandise and miscellaneous.

Information on the location of distribution centers (in the GTHA) of different retailers have been obtained from InfoCanada Business Establishments dataset, ESRI’s Business Analyst Online Application and Google Maps. The DMTI EPOI dataset for year 2013 is used for identifying the location of retail stores. The analysis is mostly descriptive in nature because of data limitations as none of the data sources provide complete information on which retail stores are served by which of the distribution centers. The study makes an assumption that retail stores are served by the distribution centers which have same name as that of the parent firm. In reality, it could be the case that retailers use third party logistics (3PL) service providers exclusively, or use 3PL in addition to their own distribution centers for catering the needs of their stores.

#### 8.2.1 Details of Multi-Establishment Firms

The multi-establishment firms studied have been categorised into 5 different categories based on their annual revenue. Annual revenue of the firms is obtained from the Industry Canada report (Industry Canada, 2011) from annual reports of the firms available online, and other financial websites. Table 5 shows the annual revenue of the different multi-establishment firms studied
arranged in descending order. The different categories of the firms are defined in the following section:

1. **Annual Revenue greater than $10,000 million**

   This category includes the following firms: Wal-Mart Stores Inc., Sobeys, Costco Canada Inc., Metro Inc., Praxair, and Shoppers Drug Mart Inc.

2. **Annual Revenue between $5,000 million - $10,000 million**

   This category include the following firms: Canadian Tire Corp. Ltd., Rona Inc., The Home Depot Inc., Best Buy Co., Lowe’s Companies Inc., and Sears Holdings Corp.

3. **Annual Revenue between $1000 million - $5,000 million**

   This category include the following firms: Maple Leaf Foods Inc., Staples Inc., Ikea, Wolseley Express, and Canada Bread Company Ltd.

4. **Annual Revenue between $500 million - $1000 million**

   This category include the following firms: Gap Inc., Toys R US, The Beer Store, Sleep Country Canada Inc.

5. **Annual Revenue less than $500 million**

   This category includes the following firms: Roots Ltd., Indigo Books & Music Inc., and The Brick Ltd.
### Table 5: Annual revenue of multi-establishment firms

<table>
<thead>
<tr>
<th>Name of the Firm</th>
<th>Annual Revenue ($mil)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walmart</td>
<td>23551 (Industry Canada, 2011)</td>
</tr>
<tr>
<td>Sobeys</td>
<td>16055 (Industry Canada, 2011)</td>
</tr>
<tr>
<td>Costco</td>
<td>13867 (Industry Canada, 2011)</td>
</tr>
<tr>
<td>Metro</td>
<td>11431 (Industry Canada, 2011)</td>
</tr>
<tr>
<td>Praxair</td>
<td>10534 (Praxair, 2016)</td>
</tr>
<tr>
<td>Shoppers</td>
<td>10459 (Industry Canada, 2011)</td>
</tr>
<tr>
<td>Canadian Tire</td>
<td>8437 (Industry Canada, 2011)</td>
</tr>
<tr>
<td>Rona</td>
<td>6800 (Industry Canada, 2011)</td>
</tr>
<tr>
<td>The Home Depot</td>
<td>6426 (Industry Canada, 2011)</td>
</tr>
<tr>
<td>Best Buy</td>
<td>6023 (Industry Canada, 2011)</td>
</tr>
<tr>
<td>Lowes</td>
<td>5910 (Financial Post, 2015)</td>
</tr>
<tr>
<td>Sears</td>
<td>5717 (Industry Canada, 2011)</td>
</tr>
<tr>
<td>Maple Leaf Foods</td>
<td>3332 (Maple Leaf Foods, 2016)</td>
</tr>
<tr>
<td>Staples</td>
<td>3040 (Industry Canada, 2011)</td>
</tr>
<tr>
<td>Ikea</td>
<td>2170 (Ikea, 2017)</td>
</tr>
<tr>
<td>Wolseley</td>
<td>1480 (Wolseley, 2013)</td>
</tr>
<tr>
<td>Canada Bread</td>
<td>1453 (Canada Bread, 2013)</td>
</tr>
<tr>
<td>Gap Inc</td>
<td>911 (Gap Inc, 2016)</td>
</tr>
<tr>
<td>Toys R Us</td>
<td>792 (Newswire, 2017)</td>
</tr>
<tr>
<td>The Beer Store</td>
<td>700 (Beer Store, 2016)</td>
</tr>
<tr>
<td>Sleep Country</td>
<td>524 (Sleep Country, 2016)</td>
</tr>
<tr>
<td>Roots</td>
<td>174 (Manta, 2016)</td>
</tr>
<tr>
<td>Chapters Indigo</td>
<td>102 (Indigo, 2017)</td>
</tr>
<tr>
<td>The Brick Ltd.</td>
<td>67 (Newswire, 2013)</td>
</tr>
</tbody>
</table>
8.3 Methodology

Methodology for analyzing the spatial relationships within multi-establishment firms is mostly descriptive in nature due to data limitations. For each firm, distance between the distribution center and mean center (or centroid) of all other establishments within the same firm is calculated. Figure 32 shows the distribution of Euclidian distance between the distribution center of each firm and the centroid of its retail establishments. The average distance between the distribution centers and centroid of their stores is found to be 22.25 km. Roots Canada ltd is found to have the distribution center located closest (in the City of Toronto) to its retail establishments while Sleep Country has its distribution center located farthest (in the City of Hamilton). Some description on the distribution of land-use pattern and property parcel size used by these distribution centers is also provided. Maps showing spatial distribution of the distribution centers and centroids of the firms’ retail establishments are described in the following sections.

![Distribution of distance between distribution centers and centroid of retail establishments of multi-establishment firms](image-url)
8.4 Results

8.4.1 Characteristics of Distribution Centers

A total of 24 distribution centers of multi-establishment firms have been identified from different data sources and from internet search. Figure 33 shows the distribution of land-use pattern of these distribution centers using DMTI land-use data. 56% of these distribution centers are located on resource and industrial land while 44% are located on open areas.

![Figure 33: Land use pattern of distribution centers of multi-establishment firms](image)

All the 27 distribution centers are spatially joined to the Teranet’s property parcel data using ArcGIS to obtain the property parcel size. The average parcel size used by these distribution centers for development is found to be approximately 140507 m². Figure 34 shows the spatial distribution of the distribution centers according to the size of the property parcel used for development. All the distribution centers with property parcel size greater than 100000 m² are located at the periphery of the City of Toronto in the Region of Peel and Halton. The average property parcel size of the distribution centers on open areas and on resource/industrial land are very similar. The distribution centers located on open areas on average have a property size of 133450 m² while those located on resource and industrial land on average have a property size of 132964 m².
Figure 34: Spatial distribution of distribution centers according to property parcel size
8.4.2 Relationships Between Annual Revenue, and Distance Between Distribution Center and Centroid of Firms

Figure 35 shows the distribution of the annual revenue of the multi-establishment firms, and the distance between their distribution center and centroid of the retail stores. Figure shows that the firms with higher annual revenue tend to have shorter distance between the distribution center and centroid of their retail stores. This mean that as the size of the firm grows or if the firm opens up new establishments, they tend to locate those establishments closer to their distribution center in order to minimize transportation costs. It was found that overall store-to-distribution center distances for Walmart decreased over time indicating that store location in close proximity to a distribution center is one of the development strategy of Walmart in the United States (Rice et al., 2016). A simple correlation analysis indicates that annual revenue of the firms and distance between the distribution center and centroid of their establishments are negatively correlated (correlation coefficient = -0.10396). Even after removing the two outliers (with highest distance and highest revenue), the correlation coefficient has a negative value (-0.011).

Figure 35: Distribution of annual revenue of firms and distance between their distribution centers and centroid of retail establishments
8.4.3 Spatial Distribution of Distribution Centers and Centroid of Firms

Spatial distribution of distribution centers of the above mentioned multi-establishment firms in relation to the centroid of their retail establishments has been discussed in the following sections. The spatial distribution patterns are described in terms of the classification of the multi-establishment firms based on their annual revenue provided in Section 8.2.1. For these multi-establishment firms, the centroids of their retail establishments are located in the west end of the City of Toronto while almost all the distribution centers are located on the outskirts of the City of Toronto. There are several important factors that influence the location choice of a distribution center but our analysis suggests that these multi-establishment firms are making a trade-off between transportation costs and land prices while locating their distribution centers in the GTHA. As the case with the new NAICS 493 warehousing establishments, almost all the distribution centers are located in the Census Sub-Divisions of Brampton, Mississauga and Milton. These regions have lower dwelling prices compared to the City of Toronto, and also offer the provision of open land, provide access to major highways, proximity to intermodal terminals and the Pearson airport as discussed in section 6.1. Distribution centers of these multi-establishments firms are located in proximity to each other and this could be because of the benefits coming from agglomeration of economies. Findings on the spatial patterns for different categories of multi-establishment firms has been discussed in the following sections.
8.4.3.1 Firms with Annual Revenue greater than $10,000 million

In this category of multi-establishment firms only Costco Canada Inc. has two distribution centers while the remaining firms have one distribution center to serve their stores. Costco recently opened a second distribution center in Vaughan (York Region, 2017). The distribution centers of these multi-establishment firms are located on an average distance of 17.93 km from the centroid of their retail establishments. Only Metro has its distribution center facility located in the City of Toronto and all other firms have their distribution centers located at the periphery of the City of Toronto. Figure 36 shows location of the distribution centers and centroids of these multi-establishment firms.

Figure 36: Location of distribution centers and centroid of firms in Category I
8.4.3.2 Firms with Annual Revenue between $5,000 million - $10,000 million

In this category of multi-establishment firms only Canadian Tire Corp. Ltd. has two distribution centers while rest of the firms have one distribution center to serve their stores. The distribution centers of these multi-establishment firms are located on an average distance of 24.25 km from the centroid of their retail establishments. Figure 37 shows location of the distribution centers and centroids of these multi-establishment firms.

Figure 37: Location of distribution centers and centroid of retail firms in Category II
8.4.3.3 Firms with Annual Revenue between $1,000 million - $5,000 million

In this category of multi-establishment firms all the firms have one distribution center to serve their stores. The distribution centers of these multi-establishment firms are located on an average distance of 17.62 km from the centroid of their retail establishments. Figure 38 shows location of the distribution centers and centroids of these multi-establishment firms.

Figure 38: Location of distribution centers and centroid of firms in Category III
8.4.3.4 Firms with Annual Revenue between $500 million - $1000 million

In this category of multi-establishment firms all the firms have one distribution center to serve their stores. The distribution centers of these multi-establishment firms are located on an average distance of 35.81 km from the centroid of their retail establishments. Figure 39 shows location of the distribution centers and centroids of these multi-establishment firms.

Figure 39: Location of distribution centers and centroid of firms in Category IV
8.4.3.5 Firms with Annual Revenue less than $500 million

In this category of multi-establishment firms all the firms have 1 distribution center to serve their stores. The distribution centers of these multi-establishment firms are located on an average distance of 12.32 km from the centroid of their retail establishments. Figure 40 shows location of the distribution centers and centroids of these multi-establishment firms.

Figure 40: Location of distribution centers and centroid of firms in Category V
8.4.4 Relationship Between Distribution Centers and NAICS 493 Establishments

Figure 41 shows the spatial distribution of the distribution centers of multi-establishment firms, NAICS 493 warehousing establishments in year 2003 and 2013 respectively. It is found that distance between the mean center of distribution centers of multi-establishment firms and NAICS 493 establishments decreased slightly by 0.86% while surface area of the intersecting Standard Deviational Ellipses increased by 3.22% between 2003 and 2013. This shows that NAICS 493 warehousing establishments have moved closer towards the distribution centers of these multi-establishments between 2003 and 2013. It could be because of the reason that these distribution centers may be using 3PL for their logistics activities. Cidell (2010) also observed such a trend of inward movement of warehouses towards distribution centers in metropolitan areas in the United States.

Figure 41: Spatial relationship between distribution centers and NAICS 493 establishments
Chapter 9
Conclusions and Future Work

This thesis analyzes and characterizes the spatial patterns of the new NAICS 493 warehousing establishments that located in the GTHA in the years from 2003 to 2013, and analyzes the spatial relationships between distribution centers and retail stores of multi-establishment firms. New warehousing establishments are identified by making a longitudinal comparison across each year. The characteristics of the new warehousing establishments are analyzed by fusing information from multiple data sources. Spatial relationships within multi-establishment firms have been analyzed based on distance between the distribution center and centroid of firms’ retail establishments. Major findings of the study are provided as follows.

Growth in the warehousing industry in the GTHA from 2003 to 2013 did not happen evenly over time. The largest share of the new warehousing establishments appeared between 2004-2006, and 2008-2010. A large number of warehousing establishments appeared and disappeared even during recession period (2008-2009). The barycentre of the warehousing establishments shifted 0.6km north from 2003 to 2013, however, the movement of the barycentre from year to year does not show any clear pattern of movement in one direction and so no clear conclusion can be made about the spatial pattern of warehousing establishments. The average distance of warehousing establishments from the barycentre increased during the pre-recession (2003-2007) and the recession period (2008-2009), but did not change in the post-recession period (2010-2013).

Spatial patterns of new warehousing establishments indicate that the sub-urbanization trend observed is not obvious in the GTHA. The growth is occurring mostly near the Toronto Pearson International Airport; along Hwy 401, Hwy 407, Hwy 410, and Hwy 427; and near the three intermodal terminals in the GTHA. Although there has been growth in the warehousing establishments across all the census divisions of the GTHA, our analysis suggests that the logistics industry in the GTHA is fairly monocentric with its epicentre located in Peel Region near the Toronto Pearson International Airport. Building footprint and property parcel size of warehousing establishments suggest that large tracts of land around the Toronto Pearson International Airport are being used by establishments for their economic activities. Analysis suggest that big multi-establishment firms are making a trade-off between transportation costs
and land prices while locating their distribution centers in the GTHA. The distance between the distribution center and centroid of retail establishments is found to have a negative correlation with the firm’s annual revenue. NAICS 493 warehousing establishments have been found to move closer to distribution centers of the multi-establishment firms between 2003 and 2013.

In future research, first a location choice model for a distribution center in the GTHA could be estimated provided sufficient data is available for analysis. A theoretical framework of location choice model for a distribution center has been provided and can be estimated using a Multinomial Logit model. The model would help in analyzing the factors that affect the location choice of the distribution center in relation to retail establishments within the same firm. The model would take into consideration the location characteristics such as population density, land prices, number of NAICS 493 establishments and distribution centers in the geographic area. The model will also take into account transportation access measures include distance of distribution centers to the nearest highway, intermodal terminals and airport. Firm’s characteristics such as distance of distribution center to the centroid of retail establishments, annual revenue of the firms and number of employees can also be used as explanatory variables.

Second, it would be fruitful to conduct interviews with commercial real-estate brokers to identify the key criteria for identifying suitable locations for locating a warehousing establishment. The outcomes of this data collection exercise could provide new information about the factors affecting warehouse location choice, which could be compared against those attributes assessed in this study. Such information could help to identify the locations where warehousing establishments may locate in the future, and could help to analyze effects of different policies or regulatory environments on location choice of warehousing establishments.
References


Liu, H., Li, F., & Lu, H. (2010, October). Imaging air quality evaluation using definition metrics and detrended fluctuation


Manta. (2016). Roots Canada Ltd. Retrieved from

https://www.manta.com/ic/mtqp6dz/ca/roots-canada-ltd


http://www.mapleleaffoods.com/investors-information/annual-reports/


https://www.thestar.com/business/2012/07/18/cp_rail_to_close_etobicoke_terminal_eliminating_30_jobs.html


Zhang, J. (2017). Identify characteristics of new freight establishments due to logistics sprawl in GTHA sing aerial images and ArcGIS.
Appendix A – Building Footprint Calculation (Zhang, 2017)

Gray level Co-occurrence Matrix (texture based image segmentation) (Haralick et al., 1973) and K-means Clustering (colour based image segmentation) (Dhanachandra et al., 2015) are two commonly used techniques for image processing. In the present study, the Gray level Co-occurrence Matrix technique is used for building footprint calculations. Details of the methodology are provided in the following section.

Step 1: Acquiring aerial images

Aerial images used for building footprint calculations are obtained using Google Static Maps API. Google Static Maps API enables users to embed a Google Maps image on a web page, creates URL parameters sent through a standard HTTP request and returns the map as an image. A python-based tool is developed to acquire aerial photographs using geographic coordinates of the warehousing establishments. A scale of 1: 4513 and an image size of 640×640 (maximum resolution for free use) are appropriate to view a building and its nearby surroundings with adequate detail. A few establishments with exceptionally large property footage, a scale of 1: 9028 is chosen in order to observe the entire property. Geocode adjustment is required when the geographic coordinates in EPOI dataset are different from the geographic coordinates for the address of warehousing establishments.

The images are downloaded and the image contrast, image blur and image noise are checked using MATLAB to assess the quality of the image using following equations. Image contrast can represent the change of colour or luminance in the air. Following equation provides a means to estimate the contrast level of an aerial image (Liu et al., 2010). The method uses average of Michelson contrast of stochastic blocks to approach a global estimation. Although the selection of sample blocks is stochastic, Liu et al. (2010) suggests that more image block near the centre position should be considered because the contents in the centre of an image are generally more important than the content near edges.

\[ M_{ICL} = \frac{1}{N} \sum_{k=1}^{N} \left[ \frac{I_k^{\text{max}} - I_k^{\text{min}}}{I_k^{\text{max}} + I_k^{\text{min}}} \right] \]
Where,

$M_{ICL} = \text{Image contrast level}$

$I_k^{max} = \text{maximum gray value of } K^{th}\text{ block}$

$I_k^{min} = \text{minimum gray value of the } K^{th}\text{ block}$

$N = \text{number of sample blocks.}$

Following equation provides an estimate of the image blur level (Liu et al., 2010).

$$M_{IEBL} = \max_{i\in\Theta}\{\arctan\left[\frac{I(x_{i1}, y_{i1}) - I(x_{i2}, y_{i2})}{w_i}\right]\}$$

Where,

$M_{IEBL} = \text{Image blur level}$

$\Theta$ is a set of image blocks

$I(x_{i1}, y_{i1})$ and $I(x_{i2}, y_{i2})$ are the gray values of the $i^{th}$ image block at two edges, and

$w_i$ is the width between edge-spread points $I(x_{i1}, y_{i1})$ and $I(x_{i2}, y_{i2})$.

Except for image contrast level and blur level, image noise level reflects how an image is contaminated by noise. Following equation provides an estimate of the image noise level, and takes not only image noise (variance) but also image brightness (mean) into consideration (Liu et al., 2010).
Where,

$$M_{\text{INL}} = K_1 \times \log \left( K_2 \times \frac{255}{\hat{\mu}} \right) \times \min_{k \in \Theta} \sqrt{\sum_{i=1}^{N} (I_{ik} - \hat{\mu}_k)^2}$$

$M_{\text{INL}}$ = Image noise level

$\Theta$ is a set of image blocks

$K_1 = 1.8, K_2 = 1.1,$

$N =$ number of image blocks,

$\hat{\mu}$ and $\hat{\mu}_k$ are the means of gray level of entire image and sample image block.

**Step 2: Building footprint calculation using Gray-Level Co Occurrence Matrix technique**

Image segmentation is the classification of an image into different groups. Texture plays a vital role in detecting various objects in an image. An aerial image generally covers a large view field but with low details. If the area with the same or similar backscattering properties, the gray level of this area will give out be the same or similar values (Liu et al., 2016; Jiao & Dong, 2013). Thus, it is difficult to extract useful information solely rely on gray levels. Haralick et al., (1973) proposed a computable texture analysis approach based on gray-level spatial dependencies, which is also named as Gray-level co-occurrence matrix (GLCM). GLCM is an effective means to extract image information based on image texture feature, which is also known as texture-based image segmentation. It is a statistic-based estimation of the two-order matrix image (Jiao & Dong, 2013). Five GLCM texture features including angle second moment (energy) index, contrast index, correlation index, entropy index, and inverse difference moment (homogeneity) index will be utilized for texture analysis, which are defined as follows (Chen et al., 2016):

$$GLCM_{\text{ASM}} = \sum_i \sum_j P(i,j)^2$$
\[ GLCM_{CON} = \sum_{i} \sum_{j} (i,j)^2 P(i,j) \]

\[ GLCM_{COR} = \frac{\sum_{i} \sum_{j} (i,j) P(i,j) - \mu_{x}^{GLCM} \mu_{y}^{GLCM}}{\sigma_{x}^{GLCM} \sigma_{y}^{GLCM}} \]

\[ GLCM_{ENT} = -\sum_{i} \sum_{j} P(i,j) \log[P(i,j)] \]

\[ GLCM_{IDM} = \sum_{i} \sum_{j} \frac{P(i,j)}{1 + (i - j)^2} \]

Where,

\[ GLCM_{ASM} = GLCM \text{ angle second moment index} \]

\[ GLCM_{CON} = GLCM \text{ contrast index} \]

\[ GLCM_{COR} = GLCM \text{ correlation index} \]

\[ GLCM_{ENT} = GLCM \text{ entropy index} \]

\[ GLCM_{IDM} = GLCM \text{ inverse distance moment index} \]

\[ P(i,j) = GLCM \text{ of the original image} \]

\[ \mu_{x}^{GLCM}, \mu_{y}^{GLCM}, \sigma_{x}^{GLCM} \text{ and } \sigma_{y}^{GLCM} \text{ are the means and the variances of } P_{x} (i) \text{ and } P_{y} (j), \]
And here $P_x(i) = \sum_k p(i,k)$ and $P_y(j) = \sum_k p(k,j)$

Haralick et al. pointed out that four significance directions of $0^\circ, 45^\circ, 90^\circ, \text{and} 135^\circ$ are supposed to be considered for analyzing texture features as shown in Figure A1. The average value of texture features of each point represents the texture features of the image.

![Resolution cells in 0, 45, 90, 135 degrees](Figure A1: Resolution cells in 0, 45, 90, 135 degrees (Source: Haralick et al., 1973))

After acquiring the aerial image, samples of three different textures are extracted from an image: structure, paved area and mud. For each patch, a GLCM with a horizontal offset of 5 is computed. Next, dissimilarity and correlation which are two main features of GLCM matrices are computed. The process was completed in Python, and them the image can be used for building footprint calculations.