BIG DATA ANALYSIS TO EXPLORE THE CLUSTERING OF TRUCK TRIPS PATTERNS ACROSS CANADA

Ayat Hussein and Hanna Maoh[1]
University of Windsor

Introduction

Freight transportation is a major driver of economic prosperity as it has a direct impact on the performance of many economic sectors (Mathew 2009). Whilst current figures may be significantly larger, it has been estimated that transportation accounts for approximately 16% of the Canadian gross national product (GNP), and 10% for the United States GNP(Crainic and Laporte 1997). Freight transportation has a major role in determining the cost of a product. For example, in Canada, 13% of the product cost for the primary industrial sector and 11% of the transformation and production industryis determined through transportation(Crainic and Laporte 1997). Consequently, many companies across the food, commodity and consumer goods sectors struggle with defining their profit margins as transportation costs climb to nearly double the inflation rate(Johnson and Prentice, 2018). Transportation is a complex domain having to adapt rapidly to changing political, social and economic conditions and trends, hence, accurate and efficient methods and tools are needed to enhance the planning and decision-making processes associated with freight logistics. Understanding the clustering patterns of truck trip ends is an important step for freight transportation planning and optimal warehouse location.

Clustering is the classification of data into groups of similar objects allowing data to be represented in fewer categories where the data loses some fine details in order to achieve simplification and highlight patterns. Clustering is the current subject of research in several fields such as statistics, pattern recognition, and machine learning. Each group, called cluster, consists of objects that are similar between themselves and different to objects of other groups. The use of clusters as a descriptive or exploratory tool for regional economic relationships provides a more affluent, meaningful representation of local industry drivers and regional dynamics (SANDAG,2018). The purpose of the paper is to analyze the clustering of truck trip productions and attractionsby economic sector using the Kernel estimation technique. The latter is carried out in the ArcGIS Geographic Information System software.

Background

Freight transportation has become one of the most important activities to plan due to the impact of goods transportation and distribution on the performance of the economy (Crainic and Laporte 1997). Urban freight transportation planning plays a major role for cities to reach sustainability (Sönke et al. 2007). Traditional urban planning focused on passenger transport and disregarded the impacts of freight transport on daily traffic.However, given the growth in freight transportation activities over the past 20 years, sustainable city planning had to consider truck movements on the road transportation network.

To better design and analyze freight transportation systems, detailed data must be collected to identify recurring patterns and to develop predictive models. Historically, such data was derived through sample surveys of populations which can end up being labour exhaustive and expensive (Stopher, 2007). Technology, such as GPS, are now used as an alternative method to track the movement of vehicles allowing for more accurate data collection and tracking movements of trucks and cargo. In the recent past, GPS data has been used to study the movement of people between home, work, and shopping trips, using different modes of travel such as car, train, truck, or walking (Gingerich 2016). In the context of freight, the biggest advantage of having GPS data is the large volume of information gathered on the movement

[1]Presented at the 54th Annual Meetings of the *Canadian Transportation Research Forum*, May 26 - 29, 2019 at Vancouver, BC

of individual trucks. This raw data can then be processed and transferred into multiple useful forms, such as origin-destination trips, truck routes, speed, truck tours, and bottlenecks (Bernardin, 2015). Ma et al. (2011) used origin destination trips collected from GPS data between traffic analysis zones in Puget Sound, Washington, placing them in a custom software interface to provide users with freight movement patterns between the origins and destination pairs.

This paper utilizes truck trip information that were derived from GPS data to come up with truck trip origin and destination clusters that could be used for better placement of logistic hubs. Origin-destination (OD) data are a special type of route data that concern origin and destination locations without paying attention to the actual route taken. Clustering methods are popular tools when it comes to pattern recognition although they have been mainly used for the matching and clustering of complex routes. Few efforts have been made to study the clustering of the Origin-Destination trips (He et al. 2018). There are two basic types of clustering algorithms related to OD data: spatial clustering of actual trajectories and point clustering of origin or destination points. Spatial clustering of actual routes may be classified into two groups: partitioning and hierarchical clustering (Adrienko, 2011). The point clustering method for OD data was derived from the traditional OD matrix analysis which generates OD flow by counting the numbers of origin and destination points in a spatial region of interest. This can be defined subjectively by users or derived from the data by using the density-based grouping method to cluster geographically closed points acquiring the region of interest (Adrienko, 2011).

Similarly, Guo et al. (2012) presented an approach to group spatial points into clusters, derive statistical summaries, and visualize spatiotemporal mobility patterns. Another approach for the spatial clustering of the OD pairs as presented by Mao et al. (2016) is to base it on the traffic grid separation to determine spatiotemporal patterns in job housing and urban commuting balance. This may generate clusters using different sized traffic grids to adjust to different density surfaces and classify threshold values from the statistics of OD clusters. He et al. (2018) presents a new Simple Line Clustering Method (SLCM) designed to determine the fastest route for every OD trip within a certain radius. This simple line clustering method (SLCM) aggregates OD lines into small spatial clusters of sufficient size to reveal the spatial characteristics in terms of movement.

Logistics planning is essential to reduce extra expenses, avoid damaging goods and/or missed deadlines. Logistics planning should be a priority for any business that relies on shipped goods to provide ways to reduce transportation costs including finding an optimal location to ensure timely planning (Adam, 2017). According to Ruriani (2014), choosing optimal warehouse locations based on transportation efficiency is more important than leasing rates or tax incentives. According to this study, centralizing the transportation locations is profitable where, a single focused location is beneficial analytically, and strategically. Lin et al. (2019) emphasize the importance of determining the location and size of distribution centers by solving a location-allocation problem. The authors note that "*a reasonable location and appropriate size is a guarantee for considerable profits*".

Methods of Analysis

The Study area and Data sets

The data used for analysis in this paper represent truck trips for September 2014. These trips were analyzed and generated from a large sample of GPS pings, which depicted the movement of 22,865 trucks owned by 449 Canadian carriers. The granular GPS data was acquired by Transport Canada and lent to the Cross-Border Institute at the University of Windsor for analysis (Gingerich et al. 2016). When the granular GPS data were processed, Canada and the United states were divided into zones to determine the location of truck trip origins and destinations. On the Canadian side, the delineations representing the census divisions were utilized whereas on the US side, the Metropolitan statistical areas (MSAs) and countieswere used (see Figure 1). The processed data resulted in the origin and destination information indicating the zone where a truck trip started and where it ended. The trip records also included the exact longitude and latitude coordinates of the trip starts and ends. These coordinates were used to assign the trip ends to the nearest business establishment location. The industry for which the nearest business belongs to was then assigned to the trip. In total, the trips were classified into 12 classes that included 11 general industries and 1 unclassified group. The 11 industries included: agriculture, communication, construction, finance, manufacturing, mining, retail, services, transportation, wholesale, and public administration. Trip ends that were not in close proximity to an establishment or those that were tied to establishments with an unclassified industry were assigned to the unclassified group. The type of trips associated with each industry and its detailed structure is included in the SIC Canadian code provided by stats Canada (http://www.statcan.gc.ca.).

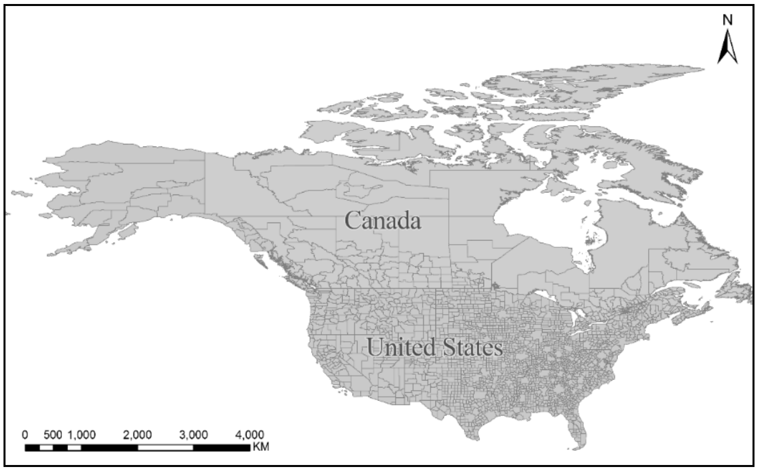

Figure 1: Canada and the US Zones

Kernel Density

In this Study, the kernel density estimation method was used to visualize the clustering of trucking trip ends for each of the 11 industries presented as well as the unclassifiable group. The kernel estimation was performed in ArcGIS 10.6. Kernel mapping, also known as heat maps, is one of the method available in the Spatial Analyst extension of ArcGIS. It is usually used to express the geographic clustering of a phenomenon, where it shows the locations of higher densities of geographic entities (Dempsey 2012). Kernel maps are created by directly using a point pattern or by using an attribute of the point pattern. The purpose of kernel estimation is to create a continuous surface known as a density surface. When calculating a density surface, three main parameters will affect the results: cell size, bandwidth or search radius, and type of calculation. Given that the output is a raster file, the cell size will determine the coarseness of the resulting density surface. A larger cell size will result in a coarse surface whereas a smaller cell size will result in a smoother surface. The second parameter to set is the bandwidth which is the area around each cell the GIS software will factor into the density calculation. A small radius value will restrict density patterns to the immediate area of the point event whereas a larger radius will allow the density patterns to become generalized. The third parameter is the type of calculation used in interpolating the density surface. The simplest calculation is a straightforward count of features within a search radius. From the density surface created, a visualization of the areas of highest density or "hot spots" are presented using a gradient. In this paper, we use kernel estimations to visualize the clustering of the point pattern produced by trucking trip endsbased on the exact longitude and latitude of each trip for a given industry to create the density surfaces.

Results and Discussion

The derived truck trip generations (productions and attractions) from the processed GPS data for Canada and the United States were analyzed by examining the clustering patterns based on the type of industry the tripsbelong to. Agriculture related trips were highly clustered all around in comparison to other industries. The analysis suggests that Montreal and Toronto/Peel regions have higher agricultural truck trip actives. Other highly clustered regions in terms of trips associated with the Agriculture industry include Cambridge, Chicago, San Diego, California, and Altona, Manitoba, as shown in Figure 2.

Trips associated with the communication industry are found to be a little less clustered than the ones associated with the agriculture industry although the highest clusters are still more noticeable in the eastern side of North America, mainly, in the Toronto/Peel region. High clusters are occurring around the Michigan-Ontario border extending all the way to the Toronto and Chicago regions. Smaller clusters can be found all over the map including, San Diego, Ohio, Texas, Calgary as well as North Carolina. It is interesting to see a noticeable amount of activity at the Northern Rockies and Yukon territory, as shown through the kernel map presented in Figure 3.

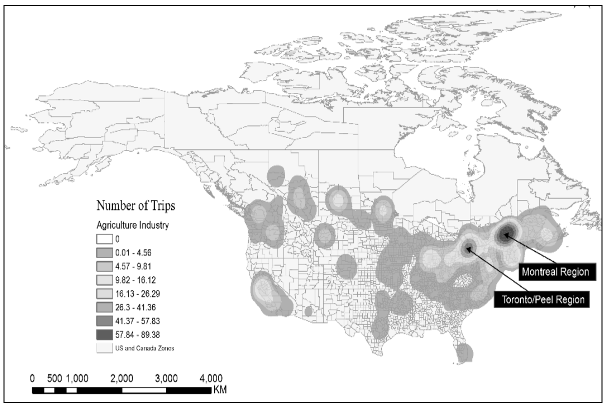

Figure 2:Trips Associated with the Agriculture Industry

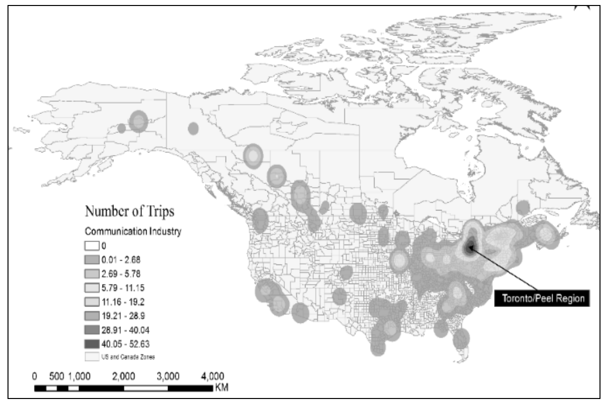

Figure 3: Trips Associated with the Communication Industry

A similar pattern is seen for the trips associated with the Construction Industry kernel map (Figure 4) where the Toronto region is the most clustered along with the Montreal Region. However, as in the previous cases, the territories do not have any activities. Edmonton and Winnipeg are noticeably clustered when it comes to Construction with little to no activity in the rest of North America as illustrated in Figure 4. The clustering map for the trips associated with the Financial sector (Figure 5) suggests a more dispersed pattern all-around North America aside from the Canadian Territories. Despite the general dispersed pattern, it is important to note that some areas are more clustered than others. For example, the Toronto region has the most activity, followed by Montreal and Columbus regions. Furthermore, Edmonton, Ohio, and Philadelphia are also busy although not as clustered.

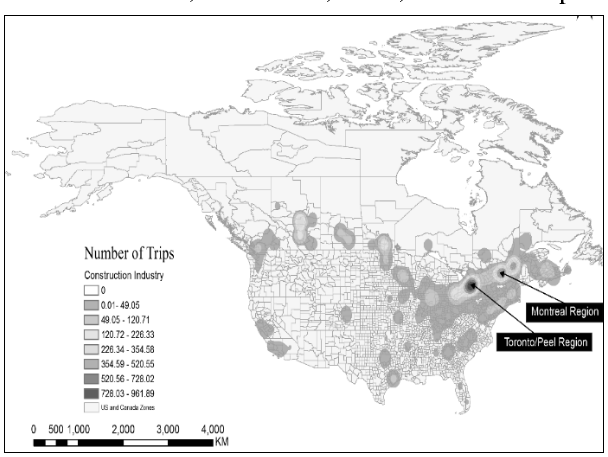

Figure 4: Trips associated with the construction Industry

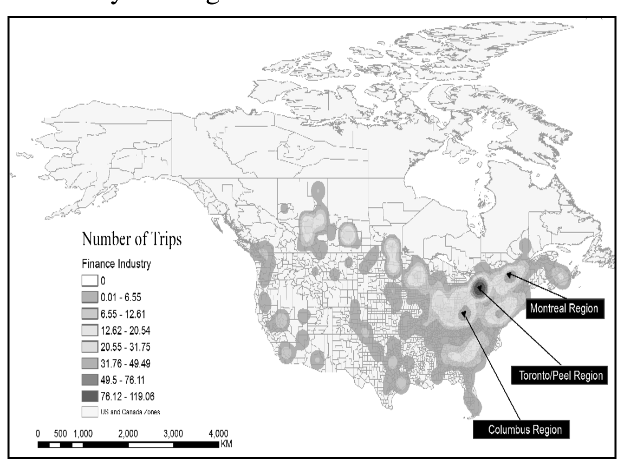

Figure 5: Trips associated with the Financial Industry

Trips associated with Manufacturing activities, which involves plants, factories, mills or any establishments that transfers materials and substances into new products, happen to cluster the most in the eastern region of North America, specifically in Michigan and Ontario. As the kernel map shows, the highest clusters are presumably occurring in the Toronto and Montreal regions. Other interestingly clustered regions for this industry include Niagara, Chicago, Philadelphia, and Edmonton as illustrated by Figure 6. On the other hand, truck trips associated with the Mining industry were the only ones not clustered in the Toronto region but rather more towards the Northern West of Canada. More specifically, the highest clusteringcan be seen in Calgary, Alberta and Huston with a smaller cluster in the Montreal and Oklahoma regions. Figure 7 illustrates the specific clustering patterns of the trucking trips associated with mining.

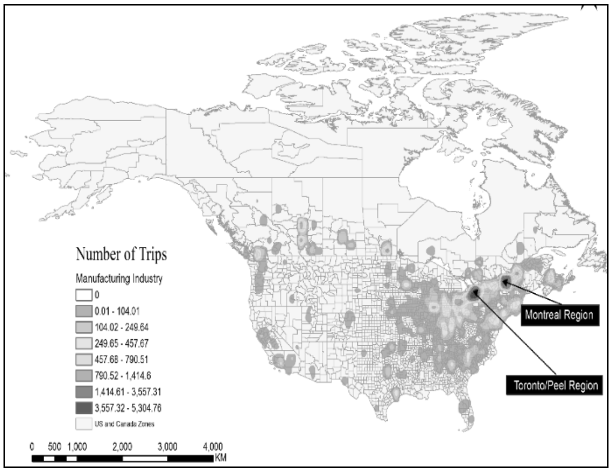

Figure 6: Trips Associated with the Manufacturing Industry

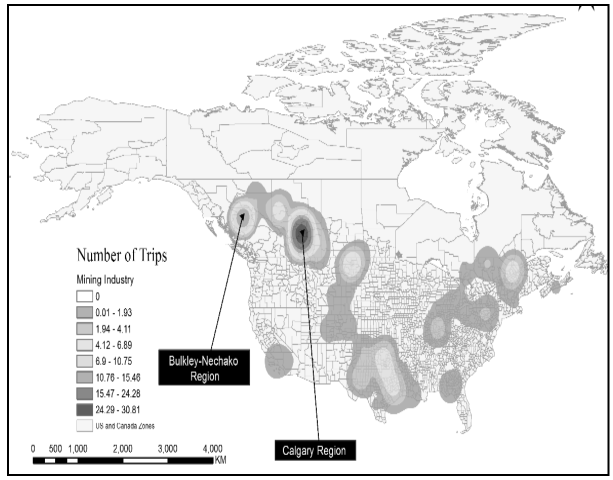

Figure 7: Trips associated with the Mining Industry

Trip activities associated with the Public Administration sector occur the most in Montreal, and the center of Quebec as depicted by the heat map in Figure 8. Clustering can also be seen in the Toronto/Peel region. Less clustering of Public Administration trips can also be observed for Queens, Halifax, Columbus and Maine along the US-Canada border. Truck trips associated with the Retail Trade industry seem to be clustered across North America in most of the populated areas. However, the highest clustering is observed Toronto/Peel and Montreal although some clustering can also be observed all around the map, specifically in Essex and Philadelphia areas, as shown in Figure 9.

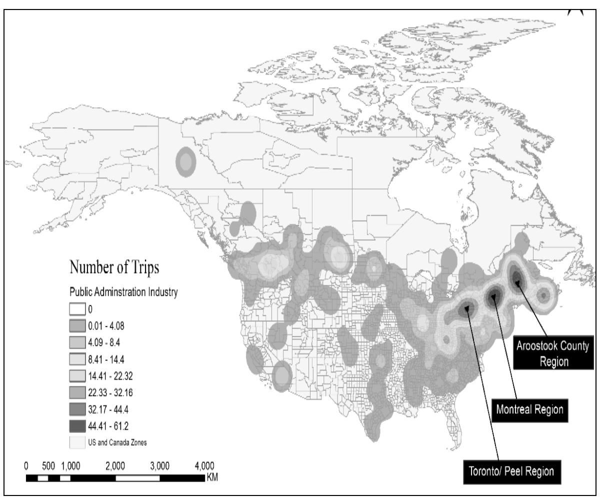

Figure 8: Trips Associated with the Public Administration Industry

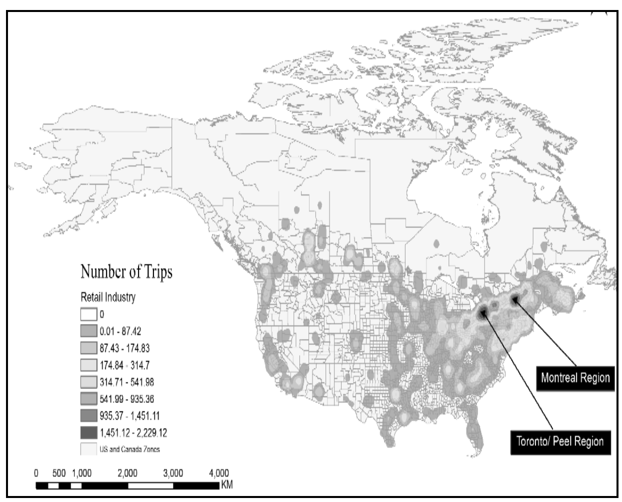

Figure 9: Trips Associated with the Retail Industry

As in the Retail Trade industry, the clustering of the Services industry related trips is most visible in the Toronto/Peel region with other regions exhibiting lesser clustering such as Montreal and the Dearborn/Windsor regions (Figure 10). The kernel map of the truck trips associated with the

Transportation sector displays the highest clusters in the Toronto/Peel, and Montreal/ Quebec regions. Clustering is also observed for Ohio, Illinois, Ottawa, and Calgary (Figure 11).

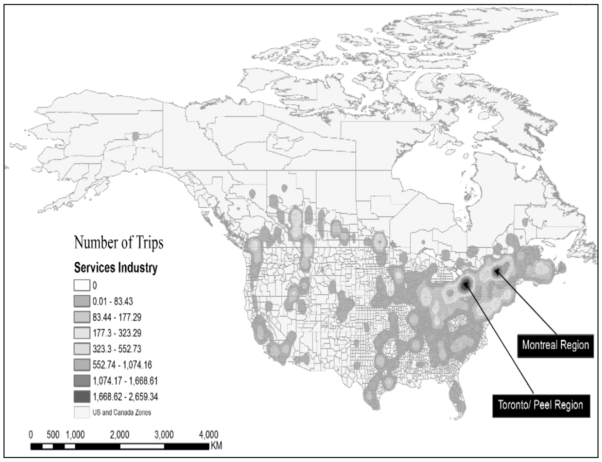

Figure 10: Trips associated with the Services Industry

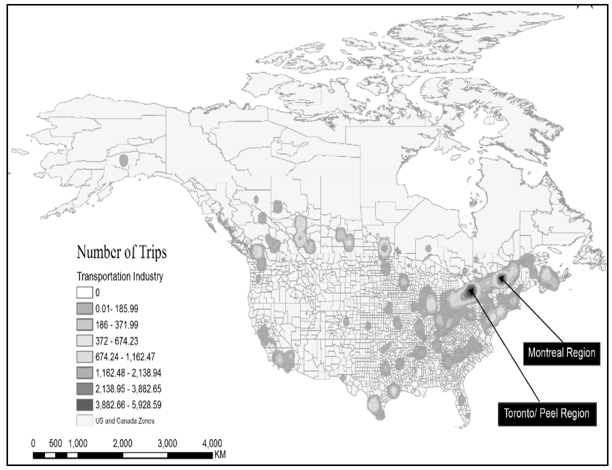

Figure 11: Trips associated with the Transportation Industry

The highest clustering of truck trips in the case of the wholesale sector occurs in the big Canadian cities specifically Toronto, Montreal and Ottawa regions, although Dearborn, Moncton and Chicago are also highly clustered as Figure 12 illustrates. Finally, while most trucking activities were affiliated with one of the above 11 industries some trip activities were non-classifiable. The latter trips seem to be highly clustered in the Toronto/Peel and Montreal regions, but the clustering tends to fade out with the distance from these regions (Figure 13).

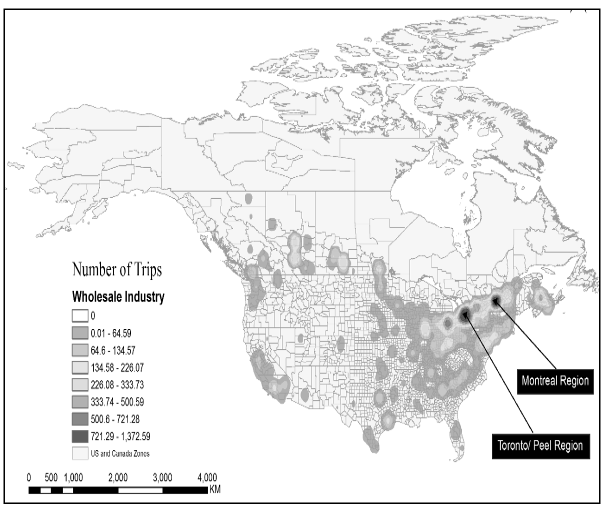

Figure 12: Trips associated with the Wholesale Industry

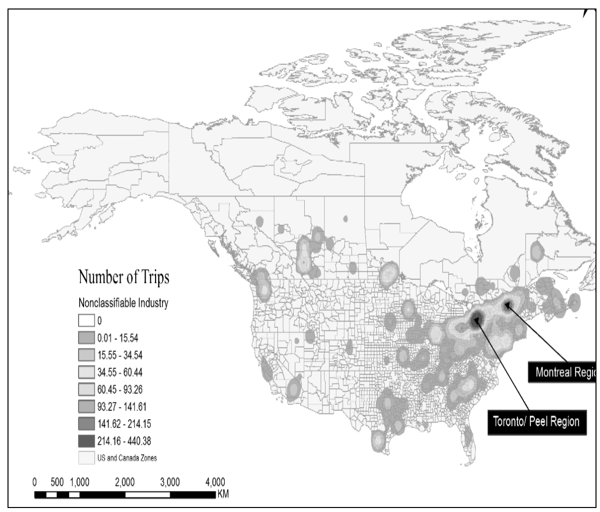

Figure 13: Trips in a non-classifiable industry

The above discussion explained the clustering of the 12 industry groups in term of their trip production patterns.The patterns pertaining to the trip attractions are very similar and will not be presented here for brevity. It is worth noting that when exploring the correlation matrix between the trip productions and attractions we also found the two patterns to be very closely related with a correlation coefficient of 0.90 and above for most industries. However, the non-classifiable trips seem to have lowest correlation coefficientof 0.82 which indicates a higher variety in its trip patterns. The same could be said about the Retailing industry trucking trips with a 0.85 correlation coefficient.

Conclusion

This study investigated the clustering of trucking trip ends around North America based on the 12 industries (Agriculture, construction, communication, finance, manufacturing, mining, public administration, services, retail, transportation, wholesale, and non-classifiable trips). Kernel estimation was employed using the ArcGIS Spatial Analyst extension to generate clustering maps of trucking trip ends of each industry.In general, the clustering patterns of each industry differsfrom one region to another. It was found that the Toronto/Peel Region had the most clusters in almost all industries, followed by Montreal in many cases. By comparison, the territories, as expected, have little to no truck trip activities.

Clustering map analysis based on trip starts and trip ends becomes very helpful for modeling the suitability of potential locations for logistic hubs and terminals. Also, mapping the transportation network flows can assist in identifying the segments of the transportation infrastructure with the highest levels of truck traffic. Future work spinning from the analysis presented in this paper will include combining network flows with the information generated in the cluster maps to run a multicriteria decision analysis (MCDA) in order to generate a suitability surface with the locations of potential logistic hubs. Clustering analysis and MCDA will be done with the help of the Spatial Analyst of ArcGIS. Further, employing network analysis techniques, namely, location-allocation modeling, will be utilized to determine the most optimal location of the potential hubs identified in the MCDA suitability map.

References

Aaron, F., Sai, K., James, C., &Zilora, S. (2007). Development of an Intermodal Network for Freight Transportation Analysis. ESRI 2007 User Conference , 1-14.

Adam, t. (2017). Strategies for Reducing Logistics and Supply Chain Costs. A&A.

Adrienko, N.; Adrienko, G. Spatial Generalization and Aggregation of Massive Movement Data. IEEE Trans. Vis. Comput. Graph. 2011, 17, 205–219.

Bernardin, V. L., Jr., S. Trevino, and J. Short. Expanding Truck GPSBased Passive Origin–Destination Data in Iowa and Tennessee. Presented at 94th Annual Meeting of the Transportation Research Board, Washington, D.C., 2015.

Crainic, T. G., & Laporte, G. (2011, January 13). Planning models for freight transportation. Retrieved February 2, 2019, from https://www.sciencedirect.com/science/article/pii/S0377221796002986

Ruriani, D. C. (2014). Reducing Freight Cost . *Inbound Logistics* , 13-21 .

Gingerich, K., Maoh, H., & Anderson, W. (2016). Characterization of International Origin–Destination Truck Movements Across Two Major U.S.–Canadian Border Crossings. Transportation Research Record: Journal of the Transportation Research Board, 2547(1), 1-10. doi:10.3141/2547-01

Guo, D.S.; Zhu, X.; Jin, H.; Gao, P.; Andris, C. Discovering spatial patterns in origin-destination mobility data. Trans. GIS 2012, 16, 411–429.

He, B., Zhang, Y., Chen, Y., & Gu, Z. (2018). A Simple Line Clustering Method for Spatial Analysis with Origin-Destination Data and Its Application to Bike-Sharing Movement Data. ISPRS International Journal of Geo-Information, 7(6), 203. doi:10.3390/ijgi7060203Mao, F.; Ji, M.; Liu, T. Mining spatiotemporal patterns of urban dwellers from taxi trajectory data. Front. Earth Sci. 2016, 10, 205–221.

Johnson, E. M., & Prentice, C. (2018, February 28). Corporate America's new dilemma: Raising prices to cover higher... Retrieved from https://www.reuters.com/article/us-usa-freight-transportation-insight/corporate-americas-new-dilemma-raising-prices-to-cover-higher-transport-costs-idUSKCN1GA0DS

Boliang, L., Liu, S., Lin, R., Wang, J., Sun, M., Wang, X., Liu, C., Wu, J., Xiao, J. (2019) The location-allocation model for multi-classification yard location problem. Transportation Research Part E, 122, 283-308.

Mathew, T. V. (2009, July 28). *Role of transportation in society.* Retrieved from Lecture notes in Transportation Systems Engineering: https://www.civil.iitb.ac.in/tvm/1100_LnTse/102_lnTse/plain/plain.html

Ma, X., E. D. McCormack, and Y. Wang. Processing Commercial Global Positioning System Data to Develop a Web-Based Truck Performance 10 Transportation Research Record 2547 Measures Program. In Transportation Research Record: Journal of the Transportation Research Board, No. 2246, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 92–100.

Sönke, B., Maria, L., & Johan, W. (2007). THE IMPACT OF URBAN FREIGHT TRANSPORT. NOFOMA Annual conference (pp. 1-15). Reykjavik, Iceland: Chalmers University of Technology.

Stopher, P. R., and S. P. Greaves. Household Travel Surveys: Where Are We Going? Transportation Research Part A: Policy and Practice, Vol. 41, No. 5, 2007, pp. 367–381.

Sandag, J. (n.d.). Understanding Clustering Analysis. Retrieved February 28, 2019, from https://www.sandag.org/rta/transfer/cluster_analysis.pdf