Gabriela del Carmen Girón Valderrama

Characterizing the urban freight system
and the supporting infrastructure network

Gabriela del Carmen Girón Valderrama

A dissertation

submitted in partial fulfillment of the

requirements for the degree of:

Doctor of Philosophy

University of Washington
2022

Reading Committee:

Anne Goodchild, Chair

Don Mackenzie

Qing Shen

Program Authorized to Offer Degree:

Civil and Environmental Engineering

University of Washington

Abstract

Characterizing the urban freight system and the supporting infrastructure network

Gabriela del Carmen Girón Valderrama

Chair of the Supervisory Committee:
Professor Anne Goodchild
Department of Civil and Environmental Engineering

The urban freight system is essential to today's economy and cities' livability. In the last decade, densification, the growth of e-commerce, and the changing mobility ecosystem have amplified commercial vehicles' challenges in navigating the city streets or finding adequate parking space. Much of the current research and transportation planning efforts at the urban scale have focused on passenger mobility, giving little attention to commercial vehicle flows and their parking behavior. Furthermore, collecting freight data at the urban scale is challenging as this sector is fragmented, and its operations are complex, fast-changing, and heterogeneous. This results in local governments having limited insight into urban commercial operations patterns when developing appropriate and data-driven initiatives and policy measures. In response to this urban challenge, this research focuses on the need for cities and researchers to collect comprehensive and high-quality data; and develop evidence-based knowledge about urban commercial operations and their supporting infrastructure. This dissertation combines empirical case studies and analytical research focusing on two main aspects of urban commercial vehicle operations: on-street parking and traffic flow. First, it documents

and analyzes commercial vehicles' parking patterns around five prototype buildings in the Greater Downtown area. Second, it develops and implements a new comprehensive vehicle classification system for collecting urban traffic flow data, focusing on the urban commercial fleet's heterogeneity. Third, it examines the effectiveness of the clustering technique, i.e., K-Means and Hierarchical Clustering, for identifying subgroups of CV traffic daily profiles and helps evaluate what vehicle and direction features may influence the CV Daily Flow patterns. Time-related features have the largest influence on daily temporal variations. Then, the resulting clusters are displayed and evaluated spatially to perform a spatial interpretation of commercial vehicle traffic patterns to evaluate the underlying feature relations among road network attributes and typical traffic flow patterns. As local conditions can significantly affect traffic patterns, Seattle's specific cluster results are then translated to "Typical" CV traffic patterns that can help draw insights into the variations of urban CV traffic flows. This effort will pave the way to further studies in Seattle and other cities aiming to compare and validate the results.

TABLE OF CONTENTS

GLOSARY

Acronym	Definition
ADT	Average Daily Traffic
AMS	Analysis, Modeling and Simulation
AT	Articulated Truck
ATR	Automatic Traffic Recorder
ATRI	American Transportation Research Institute
AVC	Automated vehicle classification
BH	Business Hours [6 a.m. – 6 p.m.]
BINMIC	Ballard-Interbay Northend Manufacturing Industrial Center
BNSF Railway	Burlington Northern and Santa Fe Railway
CBD	Central Business District
CFS	Commodity Flow Survey
CVLZ	Commercial vehicle loading zone
CV	Commercial Vehicle
DB Index	Davies-Bouldin index
DFP	Daily Flow Profile
DTW	Dynamic time warping
FHWA	Federal Highway Administration
GDA	Greater Downtown Area
GIS	Geographic information system
GHG	Greenhouse Gas
GPS	Global Positioning System
GT	General Trend
HAC	Hierarchical Agglomerative Clustering algorithm
FMP	Freight Master Plan
FN	Freight Network
HHN	Heavy Haul Network
K	Number of clusters
MIC	Manufacturing Industrial Center
MPO	Metropolitan Planning Organization
Non-CV	Non-Commercial Vehicle
O	Number of outliers
OD	Origin-Destination
P	Daily Flow Patterns
PCA	Principal component analysis
PLZ	Passenger Load Zones
PSRC	Puget Sound Regional Council
RV	Recreational Vehicle
S	Silhouette Index
SDOT	Seattle Department of Transportation
SU	Single Unit Trucks

SUV	Sports Utility Vehicle
SV	Service Vehicle
TMG	Traffic Monitoring Guide
UFS	Urban Freight System
UPS	United Parcel Service
VIUS	Vehicle Inventory and Use Survey
WSDOT	Washington State Department of Transportation
WSS	Within-Cluster Sum of Squares

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This manuscript is the culmination of a long journey with lows and highs, full of gains, hardships, successes, frustrations, learning, challenges, and experiences that will last a lifetime. Now that my Ph.D. journey has come to an end, I cannot help myself from having a feeling of fulfillment; the job is finally done. However, although I am listed as the sole author of this document, this endeavor was only possible with the support and guidance of several influential individuals.

First and foremost, I am incredibly thankful for my advisor and committee chair, Dr. Anne Goodchild, who accepted me as her Ph.D. student and offered her mentorship. Her passion, strong will, and conviction have always inspired me, especially as a young professional woman in the transportation field. I am grateful for her guidance, assistance, and encouragement during my doctoral program. Under her guidance, I successfully overcame many difficulties and learned a lot.

I also want to thank Dr. Don MacKenzie, Dr. Qin Sheng, and Dr. Sarah Curran, outstanding educators, and innovative researchers, for all the knowledge shared, guidance, and support as part of my Ph.D. committee. I also want to acknowledge the assistance from the staff, faculty, and collaborators of CEE, especially Bryan Crocket, for providing his institutional knowledge and support throughout my program.

This work would only materialize with the financial support of the City of Seattle and Urban Freight Lab members. I want to recognize the crucial roles of Christopher Eaves, Jude Whillcher, and Venu Nemani, whose valuable advice, positive appreciation, and counsel throughout this research, helped me succeed.

I am thankful to the Portland Bureau of Transportation (PBOT), especially to my knowledgeable and experienced colleagues at the Planning Division, for giving me the opportunity to serve as the Urban Freight Coordinator for the last year and a half. My work at PBOT has been an incredible opportunity to expand the skills acquired during my academic program and has helped me attain a higher level of theoretical and practical understanding of the policymaking process, regulation of transportation systems, and multimodal planning. I would like to recognize Denver Igarta and Francesca Jones for sharing their knowledge, giving me new insights, and their support as I managed and balanced my academic responsibilities with my newly acquired job and parental duties.

I am also very thankful for my colleagues and friends at the CEE department and the Urban Freight Lab, a group of hard-working, intelligent, and balanced individuals. Special thanks to Dr. Klaas Fiete Krutein, Polina Butrina, Daniel Acosta, Ru'a Al-Abweh, Hiroyuki Hinohara Chelsea Greene, and Manali Seth. Thank you for your support through the challenging moments and the beautiful and happy memories we built together. Your friendship has been one of the most significant gains this journey has given me. Thanks should also go to my friends and colleagues Dr. Haena Kim, Panagiota Goulianou, Dr. Giacomo Dalla Chiara, Tom Maxer, Anna Alligood, Katie Ward, Jennifer Ross, Dr. Ed McCormack and Dr. Andisheh Ranjbari from the Supply Chain Transportation and Logistics Center. You provided me with a sense of community where we could exchange ideas, challenge each other, and support each other's work. I admire your motivation, innovative nature, and perseverance.

To my Seattle friends, a beautiful community of kind and inspiring individuals, thank you for becoming a family and source of strength one thousand miles away from my home country. Big thanks to Graciela and Dale Kingsman, who have been my parents, and Jessica Canet and Spencer Ogden, my siblings, for your kindness and affection and for lifting my spirit when needed.

It's my fortune to gratefully acknowledge my friends and family in Panama and Spain for their support and love, even if they did not understand the reasoning behind my endeavors. Thanks to my dearest friends, Yinlo Wong and Cristina Pan, for your sisterly love and always encouraging me through the high and lows. Special thanks to my mom, Gloria Valderrama-Barragan, for loving me unconditionally, teaching me that there is always an opportunity to grow, giving me the courage to push forward during hardships, and making me believe that no peak was out of reach.

Last but not least, I want to thank my husband, Dr. Jose Luis Machado, for his continuing love, insightful feedback, and support during my pursuit of a Ph.D. that made completing this dissertation possible. He was always present, especially in moments of doubt, helping me keep things in perspective and always believing in me. My deepest gratitude to my baby girl, who arrived in this world not knowing she had to share her mom during the early months of her life. Her arrival filled my life with a love that words cannot explain. I am incredibly thankful for her patience as I wrote my dissertation; her beautiful smile infused me with courage during such a pivotal moment, the homestretch of this journey.

DEDICATION

To my piece of the sun, my baby girl, Nicole Andrea. You provided the inspiration and the love necessary for me to complete this journey.

To my best friend and soulmate, my beloved husband Jose Luis Machado. The strength and support you have given me throughout this entire process goes beyond what words can adequately express.

My beautiful, loving and amazing mom, Gloria Valderrama-Barragan, who has never left my side encouraging me to her fullest and truest attention.

To my abuelo Augusto and my abuela Gloria, in your shoulders I stand. Thank you for your endless love, stories, and an example of a lifetime.

Chapter 1. Introduction and Research Questions

The urban freight system (UFS) is a core of today's economy and cities' livability. However, the limited insight and evidence-knowledge of the local governments about the commercial operations due to the lack of comprehensive data sources and consideration for freight in the planning process have created a gap between the infrastructure supply and the commercial operations' needs. Therefore, to address this research need, this dissertation attempts - through a set of empirical case studies, conceptual and analytical research- to:

a. build evidence knowledge of the urban logistics system,
b. assess and characterize the heterogeneity of the urban commercial operations, and
c. develop a framework for effectively understanding and classifying the urban freight infrastructure network, including its current usage and physical and temporal features.

This chapter serves a signpost for this document introducing the problem background, research motivation, research objective and the structure of the dissertation manuscript.

1.1. Background and Motivation

A constant goods movement and service activities is required for any functional city. However, densification, the growth of e-commerce, and the changing mobility ecosystem have amplified the challenges for commercial vehicles (CVs) navigating through the city streets and searching for a space to park and load/unload. With increasing road capacity unlikely to relieve traffic congestion, cities are under growing pressure to efficiently plan and manage both the infrastructure network supply and the transportation demand.

The movement of goods and services within the boundaries of cities, called the last mile, is the most expensive and time-consuming section of the supply chain. Moreover, in the last decade, the growth of online shopping and population increase have heightened commodity demand and freight flows, and shifted commercial vehicles to local street, collectors, and arterials in many cities around the world. These changes have resulted in an increasing concern over UFS negative externalities (i.e., congestion, pollution, noise, safety concerns).

The UFS relies on the design, provision, and usage of adequate infrastructure to supports its operations. Therefore, unlocking the potential of every public infrastructure element as density and demand for urban resources increase is a must for every city. However, the fragmentation of the UFS makes it challenging. In this system, the private stakeholders (i.e., shippers, carriers, providers, and receivers) are responsible for day-to-day operations. They operate in an infrastructure supplied, controlled, and managed by different public (i.e., municipalities) and private entities (i.e., developers, building administrators and receivers). For example, a carrier may optimize its route plan to avoid congestion, but it can't control the supply of adequate infrastructure at the delivery address.

Much of the current research and transportation planning efforts at the urban scale have focused on passenger mobility and parking operations, giving little attention to commercial vehicle flows and parking behavior and its supporting infrastructure. On the other hand, collecting freight data at the urban scale is challenging as this sector is fragmented, and its operations complex, changing and heterogenous. Moreover, the presence of interrupted traffic flow conditions and the relative density of access point to the urban street network create challenges for the analysis, characterization, and classification of commercial traffic flow patterns. This results in a lack of understanding of the logistics sector, as Heitz, Launay, and Beziat (2019) states: “it is difficult to guide public action in the absence of detailed and precise data.”.

Without having accurate, up-to-date information on the UFS, cities face challenges in devising effective strategies to minimize issues that hamper urban freight delivery efficiency, such as unauthorized parking and congestion. As Cherrett et al. (2012) states, “urban authorities having limited insight into urban freight operating patterns when attempting to develop suitable strategies and policy measures”. Providing cities with replicable data-collection methods that will support to build (and maintain) their own database of the freight infrastructure and CV operations, can bolster their in-house knowledge and planning capacity.

In response to this urban challenge, this research acknowledges the need for cities and researchers to collected comprehensive and high-quality data; and develop evidence-knowledge about the urban commercial operations and its supporting infrastructure. This research focuses on main public elements of the last-mile infrastructure network that provide access to land uses: the curb and the roadway. This PhD dissertation considers two independent research plans:

1. The on-street load/unload infrastructure and commercial vehicle parking operations, and
2. The roadway and the characterization of daily commercial vehicle traffic flow patterns.

1.2. On-street Commercial Vehicle Parking Operation

The competition for space among road users and lack of adequate infrastructure force delivery drivers either to search for vacant spaces or to park in unsuitable areas, which negatively impacts road capacity and causes inconvenience to other users of the road. The purpose of this chapter is to advance research by providing data-based insight into what is happening at the curb by answering the following questions:

- *Research Question 1: What are the characteristics of urban on-street commercial vehicle parking operations?*
- *Research Question 2: What is the current usage of the urban commercial load/unload infrastructure by commercial and non-commercial vehicles?*

To answer these questions, a data collection to quantify the usage of curb space and the diversity of the commercial operations was developed and implemented in the densest urban area of Seattle, the Greater Downtown. This study captures the parking behavior of commercial vehicles (CVs) everywhere along the block face as well as the parking activities of all vehicles (including passenger vehicles) in commercial vehicle loading zones (CVLZ). Based on the empirical findings, important characteristics of Seattle's urban freight parking operations are described, including a detailed classification of commercial vehicle types, dwell time distribution, usage of the CVLZ and choice of curb use for parking (e.g., authorized, and unauthorized spaces). The relationship between land use and commercial vehicle parking operations at the curb is discussed. The results from this research aimed to provide practical insight into the current CVs parking behavior at the urban scale.

1.3. Urban Commercial Traffic Flow

Understanding commercial vehicle (CV) activity on the roadway network is essential to informing, developing, and evaluating strategies that improve the efficiency and sustainability of the urban freight system (UFS). Although the significance of commercial vehicle flows, current data collection efforts often treat all vehicles alike without discriminating between the general traffic and CV or between different CV classes. These conditions create challenges for the analysis, characterization, and classification of commercial traffic flow patterns.

The Federal Highway Administration (FHWA) established a 13-category axle-based classification system. The FHWA's Traffic Monitoring Guide (TMG) points out that although some engineering and planning analysis may not require data in the detailed FHWA categories, all of them need information about truck volumes versus car volumes. Thus, many regionals and interstate studies use either a simple car/truck split or a very simplified classification system, commonly a 3- or 4- bin classification system based on vehicle length (FHWA, n.d.).

Understanding and evaluating commercial flows is challenging at the urban scale as this sector is fragmented, and its operations, complex, changing, and heterogeneous. Therefore, only applying a car/truck split or a simplified classification system does not comprehensively understand the commercial vehicle flows' complexity (size, activity, capacity, fleet characteristics, number of axles, etc.). Additionally, interrupted traffic flow conditions and the relative density of access points to the urban street network create challenges for the analysis, characterization, and classification of commercial traffic flow patterns. This results in many research and planning efforts not treating passenger and commercial flows separately. This information is required to obtain a more profound and accurate understanding of traffic patterns' variations and the attributes that influence them.

In response to this urban challenge, this chapter focuses on providing comprehensive and high-quality data of commercial flows; and the development of evidence-knowledge about characteristics of CV traffic flows and their temporal and spatial variations. Specifically, this effort aims to answer the following research questions in Chapter 3 of this dissertation:

- <u>*Research Question 3:*</u> *How heterogenous is the commercial vehicle fleet at the urban scale?*
- <u>*Research Question 4:*</u> *What are the characteristics of commercial traffic flows at the urban scale and how does it differ from the general traffic?*

To answer these questions, first, a robust, replicable, and comprehensive vehicle typology focused on freight vehicles was created. Next, a data collection based on manual traffic video processing was developed and implemented to capture spatial and temporal variations of urban traffic patterns with a higher level of granularity based on the established typology. Finally, an evidence-based understanding of commercial traffic is provided based on two Seattle case studies. One case study includes a first-of-its-kind cordon study in an urban area in the United States.

Building upon the research on CV traffic flows described in the last section of this research aims to answer the following questions in Chapter 4:

- <u>*Research Question 5:*</u> *What are the typical weekday commercial Daily Flow Patterns for commercial vehicles in urban street segments?*
    - *RQ5.a.: Can we use cluster analysis to identify CV Daily Flow Pattern?*
    - *RQ.5.b: How does body type, CV activity and directionality influence the CV Daily Flow patterns?*

- *Research Question 6: What roadway infrastructure and land use attributes influence the typical weekday commercial traffic pattern in urban street segments?*

For this section, an unsupervised data mining method is implemented to discover "typical" urban CV traffic patterns and identify what vehicle and time-related features have the largest influence on the identified patterns. Finally, a spatial exploratory analysis is executed to determine the subset of attributes (geographical, network topology, connectivity, urban form) that are related to the identified CV Daily Flow Patterns.

By responding the proposed research questions 3-6, we derive practical implications for travel demand management strategies by evaluating daily traffic temporal and spatial variations by the implementation of a mathematical approach and improving the representation of urban CV flows for future research and practice

Chapter 2. On-Street Commercial Vehicle Parking Operation

This chapter was previously published as Giron-Valderrama, G., Machado, J., Goodchild., A. (2019). Commercial Vehicle Parking in Downtown Seattle: Insights on the Battle for the Curb. Transportation Research Record, 2673(10), 770-780.

This chapter includes four sections in addition to this introduction. The second section discusses previous freight parking studies and the existing freight parking policies in cities and explores which of these approaches are being used in Seattle. The third section proposes a data collection method to document freight-related parking operations at the curb though direct observations. The fourth section provides empirical findings from data collection in Seattle. The fifth and last section includes a discussion of the findings and concluding remarks.

2.1. Introduction

Finding an adequate space to park can be a major challenge in urban areas. For commercial vehicles used for freight transportation and provision of services, the lack of parking spaces and parking policies that recognize those vehicles' unique needs can have negative impacts which affect all users of the road and particularly the drivers of these commercial vehicles (Butrina et al. 2017; Dablanc and Beziat 2015; Aiura and Taniguchi 2005; Jaller, Holguín-Veras, and Hodge 2013).

The curb is an important part of the public right-of-way. It provides a space for vehicles to park on-street; for delivery vehicles (i.e., cargo bikes, cargo vans, and trucks), in particular, it also gives a dedicated space for the loading and unloading of goods close to destinations. Hence it is a key asset for urban freight transportation planning which local governments can administer to support delivery and collection of goods.

According to Marcucci, Gatta, and Scaccia 2015, the development of sustainable management policies for urban logistics should be based on site-specific data given the heterogeneity and complexity of urban freight systems. Current loading/unloading parking policies include time restrictions, duration, pricing, space management, and enforcement (Young and Miles 2015; Nourinejad et al. 2014). However, as Marcucci, Gatta, and Scaccia (2015) pointed out after an extensive review of the literature on freight parking policy, the quantification of commercial vehicle operations on the curb to inform policy decision making is nonexistent. Therefore, local governments often lack data about the current usage of the curb and parking infrastructure, which is necessary to evaluate and establish these policies and therefore make well-informed decisions regarding freight planning, especially in dense, constrained urban areas.

Given the importance of the curb as an essential piece of the load/unload infrastructure, this research investigates what is actually happening at the curb, developing an evidence-based understanding of the current use of this infrastructure. A systematic data collection method was developed and applied, resulting in empirical findings about the usage of public parking for loading and unloading operations in the Seattle downtown area.

This research documents and analyzes the parking patterns of commercial vehicles (i.e., delivery, service, waste management, and construction vehicles) in the area around five prototype buildings located in the Greater Downtown area. The results of this research will help to develop and inform parking management initiatives.

2.2. Literature Review

On-street parking is a scarce resource in urban areas, with many competing demands for its use. Many studies describe how competition for space among road users and lack of adequate infrastructure force drivers either to search for vacant spaces (adding time to the delivery route) or to park in unsuitable areas (negatively affecting road capacity and causing inconvenience to other users of the road). Both behaviors lead to congestion, safety issues, and conflicts between modes (Butrina et al. 2017; Dablanc and Beziat 2015; Aiura and Taniguchi 2005; Holguín-Veras and Patil 2005). On-street parking is often the focus of parking policies where there is not ample supply to fulfill demand. Parking policy relates to the management of the price, supply, duration, and location of parking to enhance the urban environment (Young and Miles 2015). Specific to urban freight parking, (Nourinejad et al. 2014) categorize the main vehicle parking policies as follows:

- Time restrictions,
- Pricing strategies,
- Land use and space management,
- Parking enforcement.

Alternatively, off-street parking policies generally focus on setting a rate (parking spaces per activity level) at which parking should be provided (Young and Miles 2015). A surrogate measure of activity (e.g., floor area, type of commercial activity, number of employees, etc.), which is relatively easy to measure, is used to calculate the number of required parking spaces. However, this approach is limited for both on- and off-street load/unload infrastructure because, as research suggests, the relationship between these measures and the demand for parking is not constant. For example, Muñuzuri et al. (2010) discussed the relationship between the floor area of retailers and the quantity of freight traffic. Both found that larger retailers do not always generate the greatest quantity of freight traffic. More specifically, Muñuzuri et al. (2010) claimed that larger establishments receive more freight per delivery but not more deliveries per day.

Moreover, Pierce and Shoup (2013) estimated price elasticity of parking demand based on the results of the curb management system SFPark in San Francisco, CA—a demand-based pricing system which adjusts prices based on occupancy of curb meter parking without distinguishing between commercial and passenger vehicles. Treating these two users of the curb equally may not be the correct approach, however. As the San Francisco County Transportation Authority report indicates ‘‘while demand for parking is variable and drivers can switch travel patterns or modes if parking is not readily available, commercial loading demand is more likely to remain constant regardless of the supply of loading zones because few alternatives exist to truck or other deliveries’’ (SFCTA, 2015).

In an effort to overcome the lack of empirical evidence about commercial vehicle parking behavior, a few studies have documented unauthorized behavior. For example, Jaller, Holguín-Veras, and Hodge (2013) documented parking operations of 374 commercial vehicles in Midtown, New York City, and found that almost one quarter occurred in unauthorized parking areas, including not paid/expired parking meter, blocking a fire hydrant, and double parking. Richards (2017) described how the Washington DC Department of Transportation used data to support the implementation of a commercial vehicle loading zone (CVLZ) management program and a new regulation which required commercial vehicles to display annual or daily passes to park. By using data from pay-by-phone transactions of meter parking for trucks, this research documented the ratio of truck transactions versus unauthorized users' transactions. Additionally, the research team used parking citations to document aggregate trends of parking violations including double parking, overstays of parking stall time, and non-truck parking in load zones. They found that between Monday and Friday approximately half of the pay-by-phone transactions in loading zones were done by unauthorized users instead of trucks.

2.3. Seattle's context

Seattle's curb regulations consider ‘‘load zones’’ as the type of curb that provides areas solely for loading and unloading people and goods and should not be used for parking. ‘‘Passenger load zones’’

(PLZ) are allocated for quick passenger drop-off and pick-ups and the driver should remain in the vehicle. Load zones for commercial vehicles include two types of spaces (SDOT, 2018a).

- Truck-only load zone: Areas restricted to vehicles licensed as trucks for either delivery or pick-up of products, merchandise, or other commodities.
- CVLZ: Established in Seattle in 1989, their purpose is to provide space for service delivery vehicles with a 30-min limit. They are located in Seattle business districts with paid parking.

For CVLZs, permits are required for use. The Seattle Department of Transportation (SDOT) is the institution that manages, and issues permits for CVLZ use. According to SDOT, an average of 4,000 CVLZ permits are issued per year (SDOT, 2018b).

Off-street freight load/unload parking requirements consider three categories of loading demand based on land use: high, medium, and low demand, and have a different set of thresholds and requirements for the number of loading zones depending on the demand category (City of Seattle, 2018). Width requirements for parking spaces are segmented according to demand and the largest weekly delivery truck.

Regarding curb parking operations in Seattle, since 2010 SDOT has collected and reviewed occupancy data on all paid parking areas in the city, which is, to the extent of the authors' knowledge, the only quantitative initiative to measure parking operations at the curb in Seattle. The data is used to set and adjust on-street parking rates and hours through the Performance-Based Parking Pricing Program. This data-driven approach uses the principles of supply and demand to help ensure the city's goals of one to two spaces available per block (SDOT, 2018c); but has the limitation that it only applies to paid parking locations, and monitoring is applied to commercial vehicles and other curb users without distinction.

2.4. Data Collection

This research documents and analyzes the parking patterns of commercial vehicles (i.e., delivery, service, waste management, and construction vehicles) in the area around five prototype buildings. Data collectors recorded the type of vehicle and type of curb where the vehicles were parking with a reasonable level of accuracy and detail in a defined three-by-three city block grid. The data collected provided data-based evidence of commercial vehicle parking patterns anywhere along the curb (i.e., where, and how long they parked on the curb); and an understanding of how the CVLZs were used by any type of vehicle for five different three-by three city block grids.

Because of the challenges of street visibility and the complexity of vehicle and behavior studies, the curb observation study involved the use of human observers to collect data in the field. The researchers designed a ‘‘position’’ system for collecting data. Positions are fixed locations which provide the data collector with a clear view to record each parking operation of interest on his/her assigned area meeting the time precision defined in the study. From their positions, data collectors monitored several CVLZs, PLZs, hydrants, and other zones (e.g., travel lanes) where unauthorized commercial vehicle parking behavior might occur (such as double parking).

Data collectors recorded:

- the start and end parking time of commercial vehicles in each curb space or area;
- location where the driver parked; and
- the vehicle types.

An initial field assessment of the study area was necessary to define the configurations of the positions. Figure 2.1 shows the size of the study area surrounding one of the prototype buildings as an example. The number of positions in a study area will depend on the size of the area, the configuration of the urban environment, and the precision required for the study. Position maps and data collection forms were prepared for each position. The data collection forms are spreadsheets structured with the curb spaces and zones to be monitored. The curb spaces and zones in the spreadsheet are ordered to allow an easy scan of the area by the data collector and are color-coded to facilitate their localization in the position layout map.

During the implementation of the method, various position configurations were tested to determine which would enable collectors to collect the needed information reliably within a 1-min interval. Based on the field pilot results, up to four positions were established for each building for a total of 14 positions across all the study areas. Figure 2.2 shows one of the defined the study areas (i.e., Insignia Towers) curb space map with three positions.

2.5. Vehicle Typology

A detailed vehicle typology was designed to track specific vehicle categories consistently and accurately. The typology covers a wide range of vehicle types that can load/unload at the curb and is based on prior fieldwork and knowledge of curb and alley operations in the downtown Seattle area (see Table 2.1).

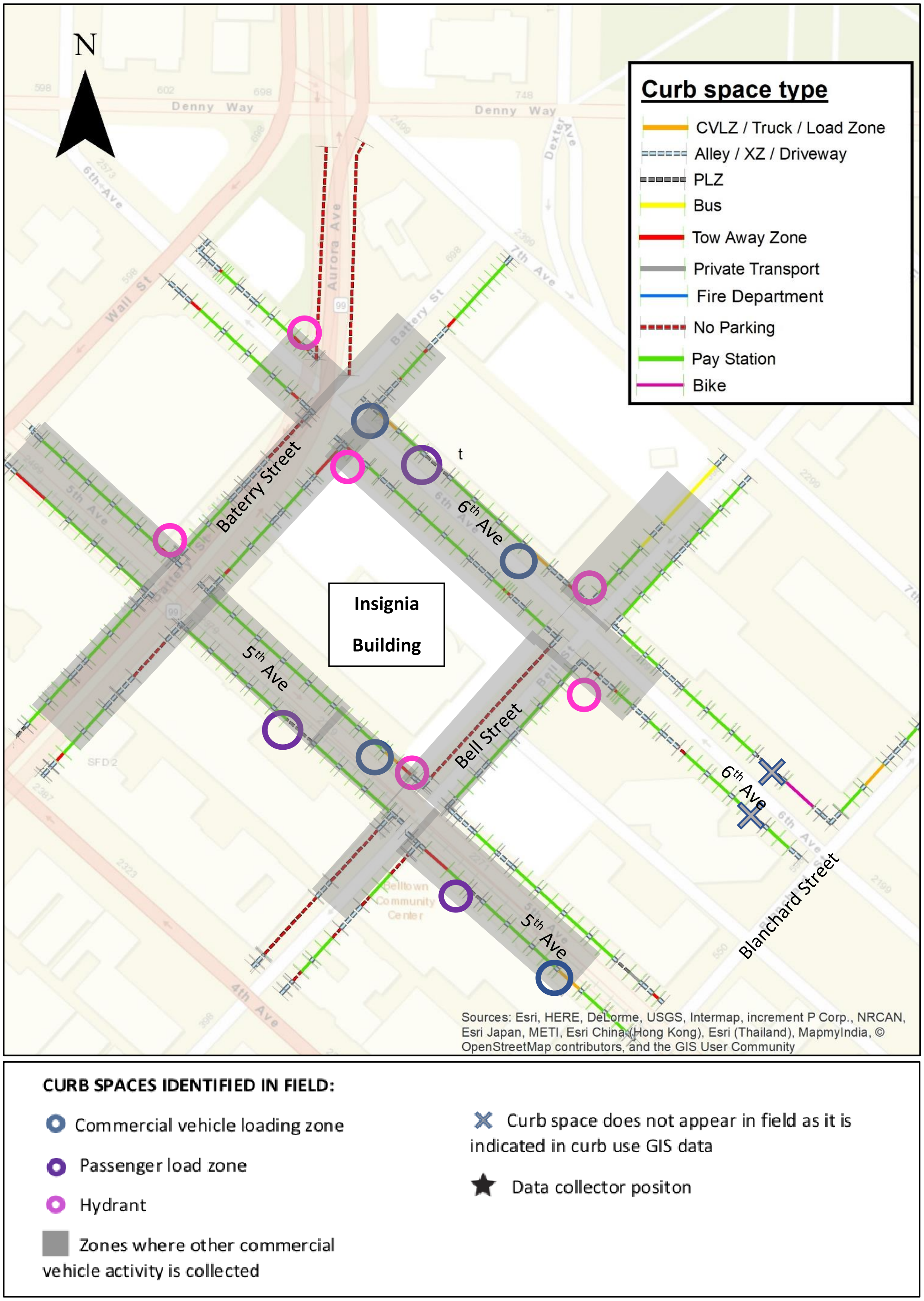


Figure 2-1. Insignia Towers building study area in downtown Seattle.

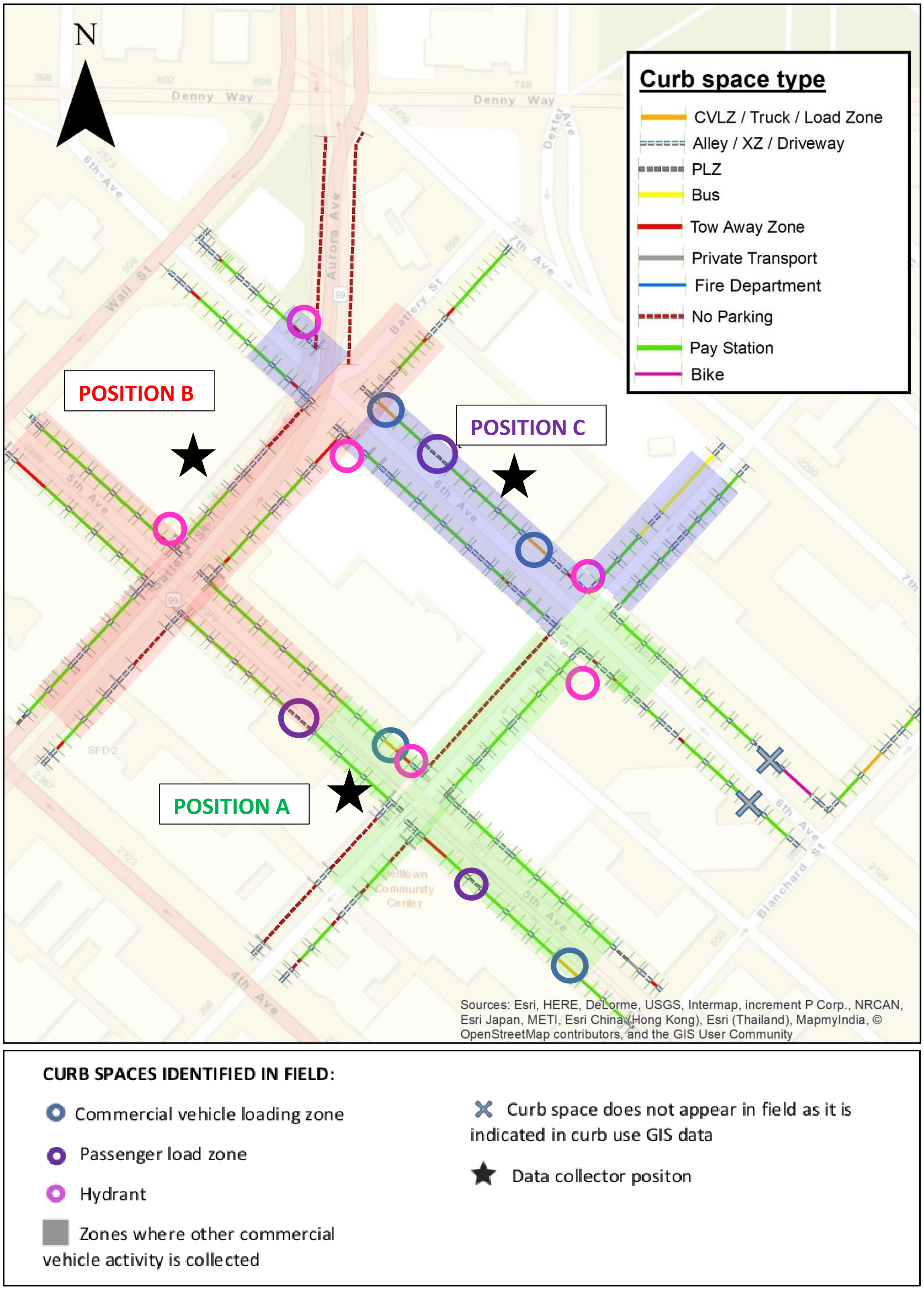
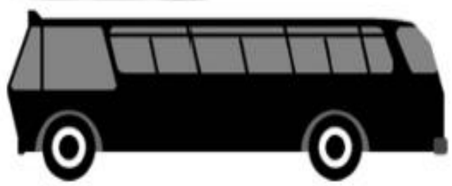


Figure 2-2. Positions of data collectors at Insignia Towers building study area.

Table 2-1- Types of Vehicles

Commercial vehicles	
Delivery vehicles	
Trailer truck	
Single unit truck—box truck	
Cargo van	 
Cargo bike	
Waste management trucks	 
Service vehicles[a]	  
General van[b]	
Construction vehicles	 

Other categories	
Passenger vehicles	
Taxi	
Motorcycle	
Buses	
Emergency vehicles	

[a]Service vehicles include vans and pick-up vehicles used for service operations.
[b]Cargo or service vans usually display a business logo. If there was not enough information visible, vehicle was marked as a general van.

For this research, the commercial vehicles of interest included trailers, box trucks, cargo vans, general vans, cargo bikes, service vehicles, waste management trucks, and construction vehicles. Passenger vehicles with commercial permits were not distinguished from those without a permit.

Additionally, this research uses the term ‘‘delivery vehicle’’ to group commercial vehicles used by carriers to transport and deliver different types of commodities (i.e., trailer trucks, box trucks, cargo vans, and cargo bikes). Although, passenger vehicles are also used for delivery of goods (e.g., Uber Eats, Amazon Prime Now, Amazon Fresh), these activities were not recorded as commercial vehicle activities.

2.6 Seattle's Case Study

A curb occupancy study was conducted in five different areas of downtown Seattle with different combinations of land uses, see Figure 2.3. The areas studied surround five prototype buildings preselected and studied in previous research on urban goods delivery The preselected buildings represent five archetypes: a hotel (Four Seasons Hotel), a high-rise office building (Seattle Municipal Tower), a historical building (Dexter Horton), a retail center (Westlake Center), and a residential building (Insignia Towers).

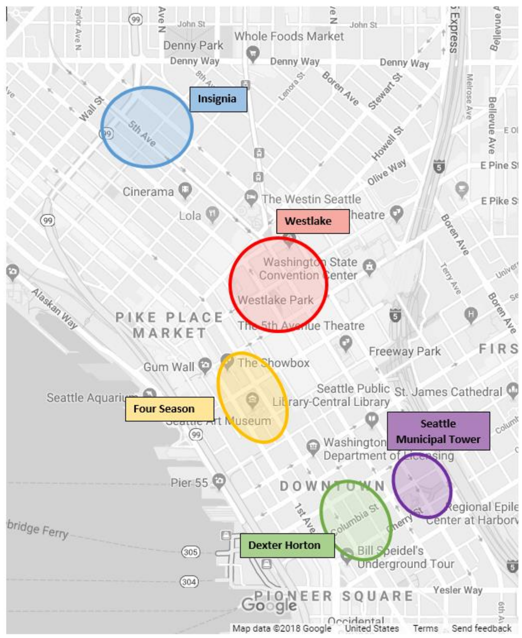


Figure 2-3. Study areas in Downtown Seattle.

2.6.1.Study parameters

Based on the project scope and in-field assessment of the areas surrounding the prototype buildings, a three-by-three city block grid was defined around each area because delivery vehicles need to park close to the delivery address. It was assumed that they would not park more than one block away from the delivery address. Block faces in downtown Seattle are typically between 300 and 400 feet long.

An inventory of the CVLZs and PLZs that serve each of the five prototype buildings was conducted. Additionally, since commercial vehicle parking operations could also take place outside of the CVLZs and PLZs, the database included, to the extent possible, areas where parking operations might occur, such as travel lanes, bus lanes, curb segments close to hydrants, tow-away-zones, and on-street meter parking. Table 2.2 shows the total length and number of CVLZs and PLZs in each building area during the data collection effort.

Table 2-2 Distribution of CVLZs and PLZs by building area.

Building area	Overall land use	Times of day	Total length of CVLZs (ft)	Count of CVLZs	Total length of PLZs (ft.)	Count of PLZs
Four Seasons Hotel and Harbor Steps Area	Hotel, retailers, art museum, restaurants and residential buildings	8:30 to 12:30 a.m.	197.3	6	424.8	13
Seattle Municipal Tower	Offices and government offices	9:00 a.m. to 1:00 p.m.	258.9	4	267.4	5
Dexter-Horton	Offices and hotels	One day from 8:00 a.m. to 12:00 p.m. Two days from 8:00 a.m. to 1:00 p.m.	643.4	17	525.3	9
Westlake Center	Retail center, hotel, commercial and office buildings	8:00 a.m. to 12:00 p.m.	90.7	3	370.7	6
Insignia Towers	Residential and university buildings	8:30 a.m. to 4:30 p.m.	117.3	4	88.8	3

Note: CVLZ = commercial vehicle loading zone; PLZ = passenger load zones.

2.6.2.Study Sample

Data collectors were deployed to observe each study area for three days over roughly six weeks in October and December 2017. The five locations were monitored during three weekdays for between four and eight hours per day. For exact time of day when data was collected refer to Table 2-2. Between the five study locations, 1,816 parking operations by all vehicles parked in CVLZs and all commercial vehicles in the five study areas were observed. A total of 1,254 commercial vehicles were observed, 382 of which were parked in CVLZs and 872 were parked outside of CVLZs. An additional 562 non-commercial vehicles were parked in CVLZs, making a total of 948 parking operations observed in CVLZs.

2.7. Findings

Finding 1. Commercial Vehicles are Parking outside of CVLZs.

While commercial vehicles did park in CVLZs (35%), across all study areas an average of 40% of commercial vehicles (with delivery vehicles constituting the biggest share) parked in unauthorized locations. These results are detailed in Table 2.3. Observed unauthorized behavior included double parking, and commercial vehicles parked in PLZs, bus lanes, tow-away zones, and no-parking zones. Commercial vehicles parking in PLZs (26%) was the largest category of unauthorized commercial vehicle behavior. Delivery vehicles represent the largest share of these commercial vehicles (18%). Additionally, 22% of commercial vehicles (with service vehicles constituting most of those commercial vehicles) chose to park in metered parking spaces, which is considered an authorized space to park.

Table 2-3. Where are commercial vehicles parking across study areas?

Commercial Vehicles Type	Number of Vehicles Observed	CVLZ	PLZ	Meter Parking	Other Unauthorized Parking	Other	Temp. Construction Zone	Total share of parked vehicles.
Delivery vehicles *(trucks, cargo vans and cargo bikes)*	**694**	19.7%	17.8%	8.1%	9.0%	0.5%	0.3%	**55.3%**
Service vehicles	**456**	11.3%	7.3%	12.0%	3.5%	1.3%	0.9%	**36.4%**
General Van	**81**	3.5%	1.2%	1.0%	0.6%	0.2%	-	**6.5%**
Other CVs * *(include garbage trucks, construction vehicles)*	**23**	0.1%	-	0.3%	0.4%	0.2%	0.8%	**1.8%**
CM parked by type of curb use	**1254**	**34.6%**	**26.3%**	**21.4%**	**13.6%**	**2.2%**	**2.0%**	**100.0%**

Finding 2. Commercial and Passenger Vehicle Drivers use CVLZs and PLZs Fluidly.

Passenger vehicles made up more than half of all vehicles observed stopped in CVLZs (52%). Delivery vehicles made up just 26% of all vehicles parked in CVLZs; see pie chart in Figure 2.4. This finding suggests that commercial and passenger vehicles use marked load/ unload spaces fluidly. It is worth noting that Seattle parking policies allow passenger vehicles to hold commercial vehicle permits. This study does not distinguish between passenger vehicles with or without permits, however.

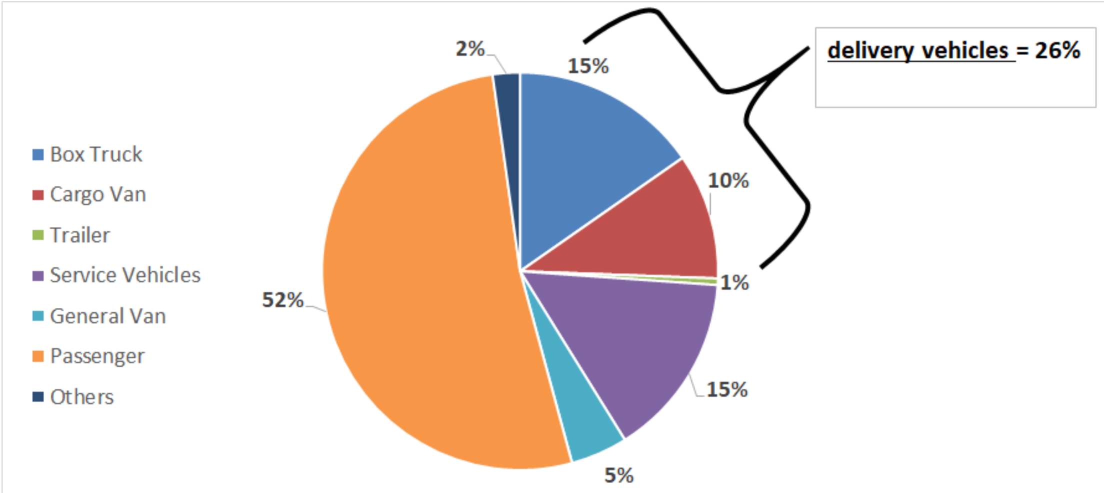


Figure 2-4. Distribution of vehicle types using CVLZs across all study areas.

Finding 3. Most Commercial Vehicle Demand is for Short-Term Operations but Some Commercial Vehicles Clearly Need Longer Parking at the Curb.

Across all study areas and curb uses, more than half (54%) of all commercial vehicles parked for 15 min or less. Furthermore, one third of all observed commercial vehicles were delivery vehicles parked for 15 min or less. Nearly one-third (28%) parked for 30 min or more, with service vehicles being the largest share of commercial vehicles parking at the curb for 30 min or longer (16%of all observed commercial vehicles). See Table 2.4.

When considering only CVLZ spaces, most vehicles parked for short-term operations. Across all vehicles, 63% parked for 15 min or less; 78% parked for 30 min or less; which is in compliance with the time restriction policy defined by the City of Seattle. When passenger vehicle drivers parked in CVLZs, they made very short-term use of them. Passenger vehicles made up the largest share of vehicles parking for 15 min or less (38.3%) in CVLZs. Delivery vehicles made up the second-largest share of vehicles parking for 15 min or less (14.1%) in CVLZs.

Table 2-4 How Long Did CVs Park in All Types of Curb Spaces in the Five Locations? % (#)

Commercial Vehicle Type	Total CVs by Vehicle Type	15 min or less	15 -30 min	30 – 60 min	>1 hr.
Delivery vehicles *(trucks, cargo vans and cargo bikes)*	**55.3% (694)**	33.7% (422)	11.2% (141)	6.8% (85)	3.7% (46)
Service CVs	**36.4% (456)**	15.1% (189)	5.4% (68)	6.1% (76)	9.8% (123)

General Van	**6.5% (81)**	3.9% (49)	1.6% (20)	0.8% (10)	0.2% (2)
Other CVs * *(include garbage trucks, construction vehicles)*	**1.8% (23)**	1.2% (15)	0.2% (2)	0.2% (3)	0.2% (3)
Total CVs by Time Parked	**100% (1,254)**	**53.8% (675)**	**18.4% (231)**	**13.9% (174)**	**13.9% (174)**

Finding 4. About One Third (36%) of all Commercial Vehicles Which Parked on the Curb Were Service Vehicles.

In contrast to delivery vehicles, which predominantly parked for 30 min or less, parking behavior of service vehicles was bifurcated. While 56% of them parked for 30 min or less; 44% parked for more than 30 min, and more than one quarter (27%) of the service vehicles parked for an hour or more. Because service vehicles make up such a big share of total commercial vehicles at the curb, this may have a disproportionate impact on parking space turn rates at the curb Urban towers require ongoing maintenance for heating, ventilation, and air conditioning; plumbing; electrical; and other systems.

Finding 5. Variation in the Distribution of Vehicle Types and Curb Uses Relates to the Spatial Distribution of the Current Infrastructure and Land Use.

The study areas showed significant differences regarding the most frequent locations for parking operations by commercial vehicles (see Table 2.5). For example, the Seattle Municipal Tower and Dexter Horton study areas had the most significant amount of curb length dedicated to CVLZs, with 259 ft and 643 ft, respectively. These buildings also showed the highest proportion of commercial vehicles in CVLZs. Conversely, the Four Seasons Hotel and Westlake Center areas had the largest share of commercial vehicles parked in PLZs (52% and 34%, respectively), both areas had the most curb length dedicated to PLZ (424.8 ft and 370.7 ft, respectively)., The Insignia study area showed the most significant proportions of commercial vehicles parked in meter parking spaces (57%). Perhaps unsurprisingly, this area has the longest share of meter parking along the curb.

Table 2-5 - Where Commercial Vehicles parked per study area?

Type of Curb	**Four Season**	**Seattle Municipal Tower**	**Dexter Horton**	**Westlak e**	**Insigni a**
CVLZ	19%	60%	58%	20%	16%

PLZ	52%	21%	18%	34%	10%
Meter Parking	15%	5%	13%	9%	57%
Other Unauthorized Parking	13%	9%	9%	21%	18%
Other	-	5%	1%	7%	
Construction Zone	2%	-	-	10%	-
Total of commercial vehicles observed	**256**	**152**	**359**	**215**	**272**
**Totals per column are 100%.*					

In addition to differences in where vehicles parked across the five study areas, this study also revealed significant differences in what kind of vehicles parked across the five study areas (see Table 2.6). Delivery vehicles were the largest share of vehicles parked along the curb for the Four Seasons and Westlake Center study areas (69% and 66%, respectively). Both areas have a dense concentration of commercial land use. The Four Seasons Hotel is surrounded by businesses such as Target, Pike Place Market, and several restaurants. Westlake Center is a four-story shopping center and 25-story office tower surrounded by a hotel and myriad of nearby retail shops and restaurants.

In contrast, the Seattle Municipal Tower and Insignia study areas showed the highest share of service vehicles of all observed commercial vehicle parking operations (53% and 45%, respectively). These two areas also showed the highest proportion of passenger vehicles with more than 50% in each area. This may be explained by the dense concentration of offices in the former area, and of residential and educational land use in the latter.

Table 2-6. Commercial vehicle type distribution by study area.

Type of Curb	**Four Season**	**Seattle Municipal Tower**	**Dexter Horton**	**Westlake**	**Insignia**
Delivery Vehicles	69%	33%	54%	66%	49%
Service Vehicles	28%	53%	35%	26%	45%
General Van	1%	14%	11%	2%	4%
Others	2%	-	-	6%	2%
Total of commercial vehicles observed at each location	**256**	**152**	**359**	**215**	**272**
**Totals per column are 100%.*					

2.8 Discussion and Conclusions

The Seattle-specific data collected provided a sample of 1,816 on-street parking operations with a granular vehicle typology.

Researchers found that the observed commercial vehicles and passenger cars were using the CVLZs and PLZs fluidly. High levels of unauthorized parking were found in all five study areas, ranging from 27% to 65%. Interestingly, in almost all the study areas, the most recurrent unauthorized behavior was parking in the PLZs space. Conversely, passenger vehicles made up more than half of all vehicles observed parking in CVLZs (52%).

Observed dwell times in CVLZs showed considerable variability between users. More than half of the delivery vehicles, but three quarters of observed passenger vehicles, stayed for up to 15 min. Approximately 20% of the parking operations lasted 30 min or more, with the largest share of these vehicles being service vehicles. When looking at all recorded commercial vehicle parking operations, this percentage is larger, with almost one third of vehicles parking for more than 30 min.

Buildings and equipment in the urban center in need of servicing and maintenance will often require providers to be on-site, as a van or other vehicle is generally required to carry parts and tools (TLF, 2018). Based on interviews with staff of service companies, (Allen et al. 2000) classified servicing activities in four categories: (i) quotation, (ii) installation, (iii) planned servicing/maintenance, and (iv) ad hoc servicing/emergency maintenance. Overall, servicing activities have received little research attention even though these operations are an important share of commercial operations. The Seattle data shows that they represent between 20% and 40% of parking operations across the five study areas. Furthermore, the Seattle data showed that servicing trips could skew the dwell time distribution of all commercial vehicles and tend to take over most commercial vehicle parking operations of 30 min or longer.

Finally, where commercial vehicles chose to park and the distribution of commercial vehicle types varied significantly from study area to study area, reflecting the fact that the service and freight demand is directly related to the land uses that generate them. Adequate supply of spaces, or the inability to meet demand, affects the levels of unauthorized behavior. The authors echo the popular opinion that, without an adequate and available supply of loading zones, on-street and off-street, drivers of commercial vehicles are forced either to spend more time looking for parking or to park in unauthorized spaces. These parking behaviors reduce the capacity of the roadways, causing inconvenience to pedestrians and conflicts with other modes, and ultimately lead to congestion and safety issues. This curb study provides a thorough evaluation of curb behavior in key Seattle locations and shows a diverse commercial vehicle demand for load/unload spaces. The insights drawn suggest a need to revise Seattle's existing parking policies, and a data-based foundation for doing so. While these insights are unique to a place, they likely reflect behaviors in other locations. However, because of the heterogeneity and complexity of the urban freight system, as Marcucci, Gatta, and Scaccia (2015) points out, approaches taken to develop policies and initiatives to improve curb management must be developed based on site-specific data.

This research aims to encourage data collection efforts, such as this one, to help reduce the gap in understanding commercial vehicles' use of the curb. The data collection approach developed and described in this chapter can and should be implemented in other cities, allowing for tailored solutions to improve curb operations and management.

Finally, further research is necessary to understand the nature of the activities which drivers of passenger vehicles are performing when they park in CVLZs. Moreover, with the increase of crowdsourcing of last-mile transportation services, future data collection methods should capture the magnitude and behavior of passenger vehicles used for delivery and pickup of goods (e.g., Uber Eats, Amazon Prime Now, Amazon Fresh).

Chapter 3. Urban Commercial Vehicle Data Collection for Flow Classification

This chapter is under review for publication for the Transportation Research Record as Giron-Valderrama, G., Goulianou, P., Goodchild., A. (2023). Characterizing Seattle's Commercial Traffic Flows Based on Vehicle Count Data.

In the succeeding sections of this paper, we first present a review of the existing literature related to the identification and classification of traffic flows, focusing on urban commercial vehicle flows. It also provides context to the study area where the empirical studies were conducted. Second, we describe the research method and its implementation in the city of Seattle. The research method in this section includes a robust and detailed vehicle classification system that focuses on the heterogeneity of the urban commercial fleet considering the vehicle body (e.g., van, single-unit truck, passenger vehicle), the activity performed (e.g., service, goods transport, construction), and the number of axles. Third, we present the main findings of the collected empirical data. The chapter concludes with a summary of findings and recommendations for policy makers.

3.1 Introduction

The collection of detailed traffic data is essential for the successful implementation of traffic management strategies. Several studies have studied the travel demand patterns and the potential factors that contribute to temporal, seasonal, weather-related, and spatial variations. By studying these systematic variations, this research has aimed to provide policymakers and traffic managers with the information to predict and forecast what would happen in the network in certain circumstances (Thomas, Weijermars, and van Berkum 2008).

CVs are critical and growing components of urban traffic. Accurate CV traffic monitoring, characterization, and analysis through the road segments can reveal the UFS's impacts on the transportation network and uncover trends for proper future planning and management. It helps inform and develop policies aiming to improve roadway performance, enhance sustainability and optimize infrastructure condition monitoring. However, current data collection efforts and research efforts aiming to evaluate and predict "typical" traffic patterns in the urban areas typically treat all vehicles alike, without discriminating between the general traffic and CV, or within the different CV categories. Little has been published about commercial vehicle (CV) variations in the urban space. Furthermore, limited information has been generated to support public agencies to understand typical CV traffic patterns.

Several efforts evaluating CV flows have focused on highway truck volume at the metropolitan, regional, and state levels. However, insights into highway/interstate CV traffic cannot directly translate to the urban context. Urban CV traffic is different and more complex than highway traffic for several reasons:

- Multiple traffic modes coexist, interact, and compete in the urban space.
- The urban network contains many intersections that result in traffic characterized by minor disturbances compared to highways that generally show fewer disturbances yet higher impact.
- The commercial travel demand characteristics (distance, motive, route length, number of stops, origins-destination patterns, and fleet configuration) are more diverse than traffic on highways.

Regarding distance the urban network serves long-distance traffic to and from the highways and a considerable amount of local or short-distance traffic (W. Weijermars and Berkum 2005). CV travel motives include various activities, such as delivery and pick-up of goods, service provision, construction traffic, waste management, and reverse logistics. Finally, CV traffic fleet configuration at the urban scale includes a diverse set of body types and a number of axles.

In response to this challenge, this research supports the need for cities and researchers to collect comprehensive and high-quality data and develop evidence-based knowledge about urban commercial operations and their supporting infrastructure. This paper through the development and implementation of an empirical approach creates foundational knowledge of CV traffic spatial and temporal variations, CV fleet heterogeneity, and variations between CV traffic flows and general traffic.

3.2. Literature Review

3.2.1. Commercial Vehicle Classification

The most widely used vehicle classification scheme is the 13-category system established by the Federal Highway Administration (FHWA) based on the number of axles and the number of units compromising the vehicle. The FHWA's Traffic Monitoring Guide (TMG) focuses on highway and bridge infrastructure. Hallenbeck et al. (1997) points out that although some engineering and planning analyses may not require data in the detailed categories, all of them need information about trucks volumes versus car volumes. The most common length classification systems based on the FHWA scheme essentially consist of either four generalized length bins (i.e., cars, small trucks, large trucks, and multi-trailer trucks) or a simple car/truck split. However, these simplifications may not be appropriate at the urban scale because they fail to capture the heterogeneity of the CV fleet and/or miss to capture the segment of the smaller CVs (e.g., light utility vehicles, delivery vans, service vans, or commercial pick-ups), a growing segment in the last decade.

Determining the body type or vehicle class of the CV is a key research problem often driven by the need to design and maintain the pavement infrastructure, or more often associated with understanding the commodity or activity for which the vehicle is being used (He et al. 2019; Allu, Sun, and Tok 2020). Including the body configuration attribute for vehicle classification, as He et al. (2020) states: "can be utilized to further distinguish vehicles within each axle-based category, connecting vehicle classification to freight planning, in which the operating characteristics…are present".

3.2.2. Current approaches to collecting empirical freight traffic data at the urban scale

Cities often rely on origin-destination (OD) surveys, participatory GPS data, or commodity flows to estimate, characterize, and classify traffic flows. These approaches require carriers, shippers, and/or receivers to fill in questionnaires or provide data about the commodity type, vehicle configuration, origin, destination, etc.

These survey techniques have common limitations: the sample size is too low; data reliability and the aggregation level of OD zones are too high to analyze travel demand patterns on a local scale (W. Weijermars and Berkum 2005). Therefore, these databases do not show the complete picture of CV flows as they are spatially and temporally limited (Yuksel et al. 2020). Participatory GPS data to estimate flow conditions, as an OD survey, can suffer from inaccuracy to often limited participation and the inherent imprecision of the current GPS (Banaei-Kashani, Shahabi, and Pan 2011b). The American Transportation Research Institute (ATRI) provides a GPS-based spatial and temporal database for a large sample of trucks with onboard, wireless communication systems in the U.S. Despite constituting the largest sample in the industry, it does not represent all truck flows as it is over-represented by medium- to large-fleets, the truckload sector and combination trucks (Class 7 and 8 of the Federal Highway Vehicle Classification System) (Short and Peters 2020).

In contrast, comprehensive road traffic counts efforts can provide a complete picture (spatially and temporally). Vehicle counts locations are distributed in the road network to form screen lines and/or cordons. These cordon surveys provide data about the CV share of total traffic, help monitor the changes in traffic volumes across time, evaluate the impacts of implemented policies, inform, and develop infrastructure and traffic management future policies.

Only the cities of London, UK, Toronto, Canada, and Dublin, Ireland, have conducted cordon counts to monitor trends across time and the impact of city-wide policies on traffic volumes. These cases have established a vehicle typology that considers the distinction between passenger, transit, and CVs. The focus of these programs has been to compile a comprehensive picture of modes share, classify road network elements based on daily volumes and traffic composition, evaluate traffic trends across years, and impact evaluation of implemented policies (Dublin City Council, 2019; MMM Group, n.d.; TFL, 2017). However, the level of aggregation of CVs in these approaches is still not enough to enhance the understanding of commercial activities by commodity type and vehicle body type.

3.2.3. Classification techniques using technology sensors

Most recent research on CV classification has been focused on truck classification using technologies that include both intrusive or in-pavement methods (e.g., inductive loops, weigh-in-motion (WIM), and pneumatic tubes) and non-intrusive or roadside sensors (e.g., radar, passive acoustic sensors, GPS, vision-based sensors) (Javadi et al. 2018). In the last two decades, substantial research efforts have aimed to prove more robust and detailed vehicle classification methods using various sensors and approaches such as artificial neural networks, deep learning, and inductive signature reading (Hernandez, Tok, and Ritchie 2016; Jeng and Ritchie 2008; Zhang, Wang, and Wei 2006). The studies mentioned above have focused on truck volumes on highways. However, to capture the spatial, temporal, and heterogeneity of CV flows in the urban transportation network, cities require a spread-out deployment of these technologies (i.e, GPS, sensor, loops, high-quality video cameras, etc.). In recent years, video cameras have been increasingly deployed on roadway networks providing necessary support for infrastructure management. The benefits of turning these cameras into data collection instruments are obvious, but storing, processing, and analyzing traffic data would generally require developing a cost-effective method to calibrate them efficiently and accurately (Nam and Nam 2018). Automated vehicle classification (AVC) equipped with vision-based sensors can support agencies to identify vehicle body configuration, manufacturers, and models. Several studies have focused on investigating video data feasibility as an important data source for vehicle detection, classification, tracking, and parking surveillance (Hsieh et al. 2006; Aqel et al. 2017). Significant, and successful efforts have been made to distinguish between buses and trucks, passenger and delivery vans, multi-trailers, trucks, and trailers. However, accurate vehicle classification in the urban context is challenging due to the heterogeneity of the CV fleet and the inter-class similarity between certain types of vehicles such as single-unit trucks and trailers (Taek Lee and Chung 2017). Therefore, for the time being, AVC methods cannot produce the detail or accuracy of human interpretation for the urban environment.

3.2.4. Knowledge of urban CV activity

With the growth of e-commerce and home delivery, there has been an increase of smaller CVs due to the nature of the last mile operations: smaller, fragmented, and more frequent deliveries. Additionally, commercial, and municipal service, maintenance, and construction logistics (discussed later by the authors) are often "hidden or forgotten freight transportation operations" (Ballantyne, Lindholm, and Whiteing 2013). Only a handful of studies have included service activities in the evaluation of the UFS. Two studies by Cherrett et al. (2012) and Visser, Nemoto, and Browne (2014) suggest service transport could represent as much as 43% of the total number of vehicles that arrive at a typical shop located in the main street during a typical week. In Seattle, Giron-Valderrama et al., (2019) found that about 36 % of all CVs using the curb space were service vehicles. That study highlights the importance of recognizing that service acts as a major contributor to urban CV volume and the higher curb utilization they require compared to delivery vehicles, as they use the space for longer periods of time.

Some studies show that the peak periods for general traffic do not match the goods delivery peak period in the urban context. Both Cherrett et al. (2012) and Browne et al. (2012) found that the 6:00 am to 12:00 pm period generates the most urban delivery activity. When it comes to UFS impacts on road capacity at a city scale, modeling typically adopts a classical four-step process to produce OD matrices (Gonzalez-Feliu et al, 2014). After a thorough literature review of both European and the USA approaches to estimating CVs volumes, Holguin-Veras and Jaller (2014) stated: "the freight traffic volume within a network is not covered by the sampling framework with sufficient integrity to support urban freight modeling". Moreover, Aditjandra et al. (2015) identified the following research gaps in data collection:

- inability of separating light goods vehicles activity from heavy goods vehicles.
- the infrastructure in which urban CVs operations takes place; and
- geographical detail about vehicle trips (Allen et al., 2014).

In summary, vehicle volume classification data allows the analysis of traffic flow patterns at a local level. However, the vehicle classification granularity that is achievable depends on the sensors employed, and the data reduction and classification approach. Although several methods have been proposed to classify vehicles based on their body frame (i.e van, trucks, passenger), CV classification at the urban scale is still largely unavailable. Ultimately, there is a lack of granular, large-scale, and high-resolution data (both spatially and temporally) for commercial traffic in the urban environment. CV flow estimations are too often based on sporadic surveys and/or simplistic simulation/models, which results in inaccurate and sparse traffic flow estimations.

3.3. Seattle's Context

3.3.1. Seattle's Freight Network

In 2005, the City of Seattle establish the Major Truck Street (MTS) network with the goal of identifying Seattle's main freight corridors. This MTS network focused on critical connections for freight movement through the city, seaport highway or intermodal connectors, and to/from major freight traffic generator, including Port of Seattle (Port) terminal, inter-modal rail facilities and the regional freeway network (SDOT, 2016).

In the 2016 Freight Master plan (FMP), additional analysis was conducted to identified roadway links about side the MTS network that may have higher significant truck volumes, and which in the MTS network may have lower. Additionally, an effort to include the local freight movements was made. Almost all freight network elements identified in the FMP 2016 are owned and operated by the City of Seattle. The regional highways (Interstate 5, Interstate 90, and State Route 99) are in the jurisdiction of the Washington State Department of Transportation (WSDOT). Most of the Seattle freight network is designated on arterial streets, with the non-arterial segments operating as first/last mile connectors within the MICs. This classification system does not necessarily change the elements' overall function, design, or character. But it does underscore the importance of ensuring that commercial flows can be safely accommodated on all relevant roads.

The City of Seattle's Freight Master Plan established four designations for Seattle's freight network (SDOT, 2016):

1. Limited Access Facility –Facilities like interstate and state highways (e.g. I-5 and SR 99) that support through movement and/or long-distance trips throughout the region and movements from MICs and from the urban center.
2. Major Truck Street – Arterial with a minimum threshold of 500+ trucks per day, serving connections to the regional network, between and through industrial land uses (MICs and intermodal terminals), commercial districts, and urban centers.
3. Minor Truck Street - Connections for goods delivery to urban villages and neighborhood commercial districts that also provide critical secondary connections to the significant

truck street network, creating system redundancy and resiliency. Minimum threshold of 500+ trucks per day.

4. First/Last Mile Connector – Locations within MICs where short truck movements are required to access key freight activity centers, such as Port facilities and intermodal terminals. These roads may not have high enough traffic volume to be classified as arterials, but the relatively high percentage of trucks they carry and their access to the Port and industrial lands make them essential to the overall freight network. Minimum threshold of 250+ trucks per day.

The City also has 2 regulatory networks related to freight – the Heavy Haul network (HHN) and the over-legal routes. Most of the HHN and most of the over-legal routes overlap with the freight network.

3.4. Methodology

3.4.1. Vehicle Typology

This research designs a detailed vehicle typology aiming to capture the complexity of the urban CV fleet. This list of vehicle types is grouped in categories based on vehicle trait similarities, and differences to capture heterogeneity in CV traffic flow operation. Leveraging the FHWA classification system and additional visual information contained in the CV bodies such as truck trailer type, carried equipment, rear window configuration, vehicle frame, and logo allows to capture a highly disaggregated CV volume data.

The proposed typology covers 65 separate vehicle categories, differentiating between delivery vans, service providers, construction vehicles, and delivery trucks, among others. An additional six "catch-all" categories are also included to account for cases of unknown vehicle types. The categories are based on three attributes body type; activity type; and the number of axles, see Figure 3-1.

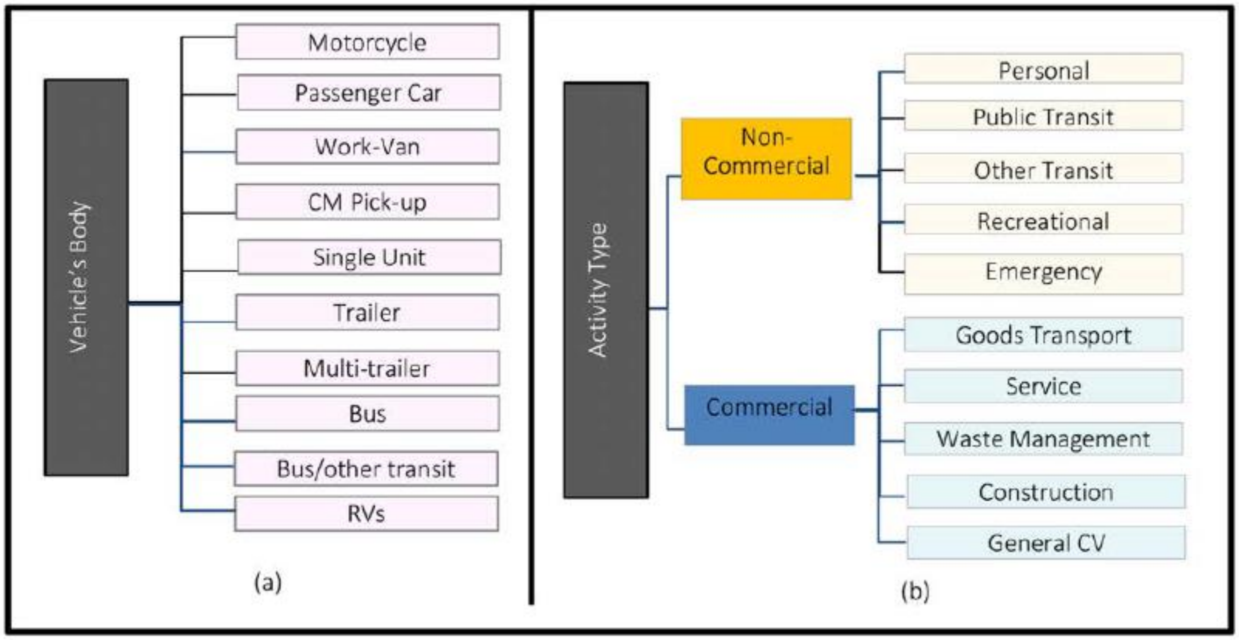


Figure 3-1. Vehicle typology: Attributes of vehicle body and activity type features.

The sections below describe the body type and activity type categories developed focusing only on the commercial vehicle categories (for a complete list of categories refer to Appendix A)

3.4.1.1. Body Type

This attribute indicates the type of body the vehicle has. Body types related to commercial activities include:

- **CM pick-up**: Pick-up used for commercial purposes. Figure 3-2 shows the features and frame configuration of the vehicles with this body type. This category is limited to pick-ups that meet at least one of the following conditions:

  a. Pick-up with two or more of the following features:

  (1) rails for mounting with or without ladders,

  (2) covered cargo area at roof's height,

  (3) roof clearance lights,

  (4) company logo,

  (5) truck toolboxes, and

  (6) sideboards.

  b. Pick-up with covered cargo area higher than the cabin roof, and

c. Pick-up carrying service equipment, barricades, and road signs

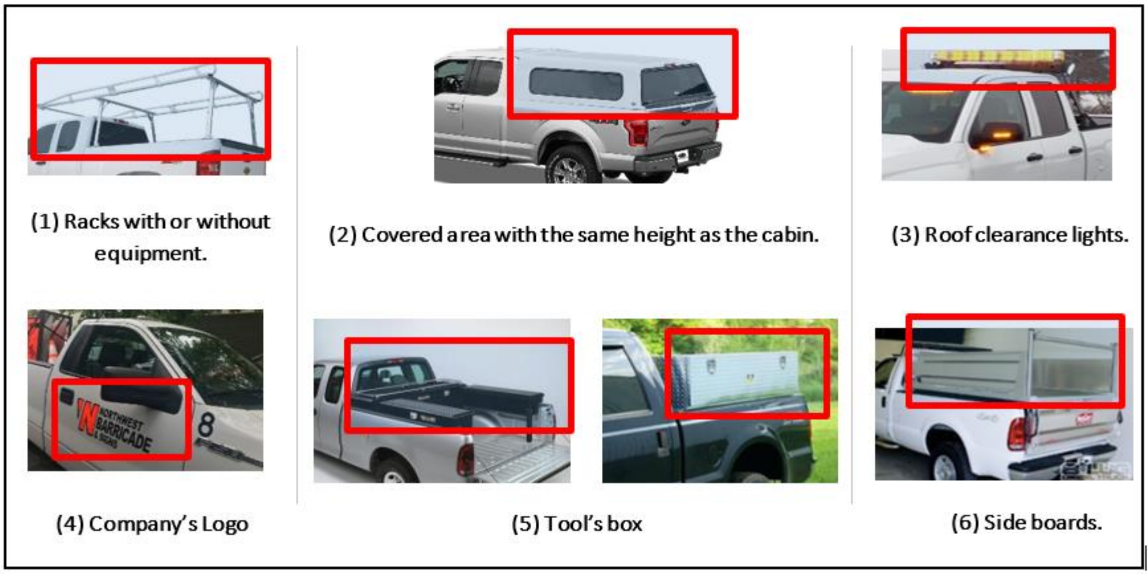


Figure 3-2. Features and frame configuration of the vehicles with the body type classified as CM pick-up.

- **Work van:** Unibody vehicle, which includes mini-vans, vans, and step-vans, with partial or no rear windows, manufactured primarily for commercial or emergency purposes (e.g., ambulances).
- **Single unit:** Truck on a single frame, including truck tractor units traveling without a trailer.
- **Trailer:** Truck consisting of two units in which the pulling unit is a tractor (semi-trailer unit trucks) or a single-unit truck (single trailer).
- **Multi-trailer:** Truck consisting of three or more units in which the pulling unit is a tractor or single-unit truck.

In this chapter, the term "small CV fleet" refers to both CM pick-up and Work van categories, see Table 3-1.

Table 3-1. Vehicles that followed the "small CV fleet" body type category.

SMALL CV FLEET		VISUAL DESCRIPTION
CM Pick-Up		 
Work van	a. Mini-Van	
	b. Van	
	c. Step-Van	 

3.4.1.2. Activity Type

This attribute indicates the primary purpose for which the vehicle was manufactured or its primary usage purpose, including commercial and non-commercial categories. Although private vehicles are also used to deliver goods or provide services, this study does not record these vehicles as CVs as they have no discernable identifiers. All passenger vehicles, regardless of potential commercial purpose, are counted as private (non-commercial) vehicles. Commercial purposes include:

- **Goods Transport:** CVs designed or used for carrying commodities (e.g., carrier and shipper work van; auto-transporters; tankers; box trucks; and containers).
- **Service:** CV is designed or used typically by maintenance or service providers (e.g., electricians or plumbers, internet service providers, caterers, gardeners, etc.); including food trucks, bucket trucks, service provider pick-ups, and utility trucks. Figure 3-4 shows examples of single-unit utility trucks used as service vehicles.

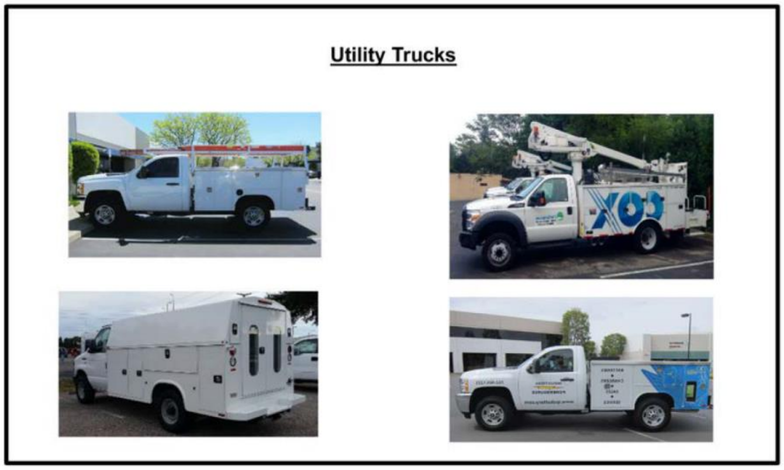


Figure 3-3. Examples of service's single unit utility truck.

- **Waste Management (WM):** Trucks that collect and transport waste, including street sweepers and sewage waste trucks.
- **•Construction:** trucks sold by manufacturers primarily for building, civil engineering, or engineering work. (e.g., rack trucks; stake trucks; concrete mixers; dumpers; empty flatbeds; and flatbeds carrying construction materials or equipment).
- **General CV:** When it was not possible to assign a CV in any of the commercial categories described above, it followed the "General CV" category.

3.4.1.3. Number of Axles

This attribute represents the vehicle's number of axles. Certain truck configurations utilize axles that can be lifted when the vehicle is empty or lightly loaded. The position of these axles (touching the ground) affects the vehicle classification. For example, a data collection location may exhibit directional differences in vehicle classification even though the same trucks may be traveling one direction loaded (with axles down) and the other direction empty (with axles lifted). Following the FHWA classification system, the axles of recreational or other light trailers attached to vehicles are not considered.

3.4.2. Instruments

Using traffic video data and human observers, we captured road-based freight data in real-world urban environment with a higher level of granularity and scale that previous efforts.

The VLC software (an open-source multimedia player) allowed data collectors to modify the speed in which the video footage was played to capture and classify vehicle volume; and reduce data classification errors. Video-image detection technology is explored as an option to process the video footage (e.g., detection and classification of the vehicles). However, existing algorithms cannot distinguish at the needed level of granularity. For example, service van vs. delivery van, pick-up vs. van, or the number of axles per vehicle.

3.4.3. Quality Control

The data collection method for this study ensured that the results are robust by completing a quality control process. Data-agreement checks were completed during data collectors' training to establish a standardized vehicle classification processing system, using the detailed vehicle typology created for this project. Three data collectors reduced one-hour videos and confirmed the vehicles' number and classification. Data flagged during the check were reviewed in more depth to evaluate the reason behind the differences and to improve the final vehicle classification system. For cases where data collectors were not able to categorize the vehicle type of the passing vehicle, this attribute was defined as unknown.

By establishing this process and using manual video processing, the research team was able to address and overcome several of the challenges related to video processing including:

- **Illumination variations:** An excess or lack of light blurs the scene, making the vehicles harder to track and classify.
- **Occlusion:** Other vehicles or stationary objects may partially or fully obstruct a view of the vehicles passing through the video frame.
- **Camera configuration:** A low-viewing camera angle will increase the level of occlusion. Low-resolution videos will increase the rate of vehicle misclassification and miscounting.
- **Weather conditions:** Weather conditions such as fog, haze, heavy rain, or snow will reduce visibility, making the vehicles harder to track and classify.
- **Variety of movement:** In an intersection, vehicles may stop and turn in several directions, making vehicle trajectories more unrestricted and unpredictable than highway traffic.

3.5. Seattle's Case Study

For the last decade, Seattle has been—and continues to be—one of the fastest-growing cities in the United States. Seattle's unprecedented growth and geographic constraints (wedged between water and mountains) create significant challenges in effectively managing the movement of people, services, and goods.

Currently, urban traffic volume counts by regional agencies are limited in spatial and vehicular detail. SDOT is responsible for recording traffic counts through the year on selected arterial streets in Seattle, providing a seasonally adjusted average weekday total vehicle traffic for all lanes at all locations (i.e., locations where data was collected for this study) (SDOT, 2018). WSDOT provides annual average daily traffic volumes in select locations of their jurisdiction, including the major interstates and state highways in the Seattle area. The data include truck volumes separated into three types: single, double, and triple units (WSDOT, 2016). Puget Sound Regional Council (PSRC) regional truck model has three levels of vehicle classification: light commercial, medium trucks, and heavy trucks. They base their model on WSDOT Annual Traffic Flow's locations and additional manual counts for model validation through the Puget Sound Region (Childress, 2015). But none of these existing efforts produce enough detail to effectively inform policy and understand Seattle's commercial vehicle movements.

3.5.1. Study area

Greater Downtown Area

The Greater Downtown Area (GDA) consists of five different neighborhoods including Downtown Seattle and is the densest urban center in Seattle. This area consists of a mix of commercial, residential, office, retail, industrial, and harbor front land uses. To the west boundary of the GDA, there is the Downtown waterfront where several Port Facilities are located including the cruise terminal and terminal 30, which offers access to the rail yards located nearby. Additionally, two key interstates run throughout the GDA: I-5 (north-south movement) and I-90 (east-west movement).

Ballard/Interbay

The Ballard-Interbay area consists of maritime, industrial, commercial, and residential land uses with critical local and regional freight routes, including one of the two most important manufacturing industrial center (i.e., BINMIC) in the region and two HUB urban villages. The BINMIC consists of a mix of diverse uses including, light manufacturing, rail, warehousing, and several Port of Seattle facilities.

3.5.2. Data Collection

Traffic cameras collected traffic footage at 71 locations across the two study areas (Figure 3-5). Each location represents a roadway leg for which a screen-line was performed. Therefore, capturing the directionality of each of the vehicles crossing the screen-line (e.g., North, South, East, or West). Two weekdays of video footage were recorded (the two weekdays were chosen between Tuesday, Wednesday, and Thursday) for 24 hours per day for each location in both study areas. Additionally, seven days of data were collected for three GDA gateways to capture variations in the volume of CV traffic throughout the week.

The 71 locations were selected with the Seattle Department of Transportation (SDOT) based on:

- The average annual daily traffic for roadways and waterways obtained from the Washington State Freight and Goods Transportation System framework (WSDOT, 2018); ensuring all streets with high truck volume where captured (+500 daily truck traffic).
- Seattle's Freight Network (FN) classification (SDOT, 2016); prioritizing all major FN designations (e.g., Major, Minor and Last/First Mile Connectors).
- The location of existing traffic cameras in the GDA and along I-5; helping to reduce data collection costs when possible.

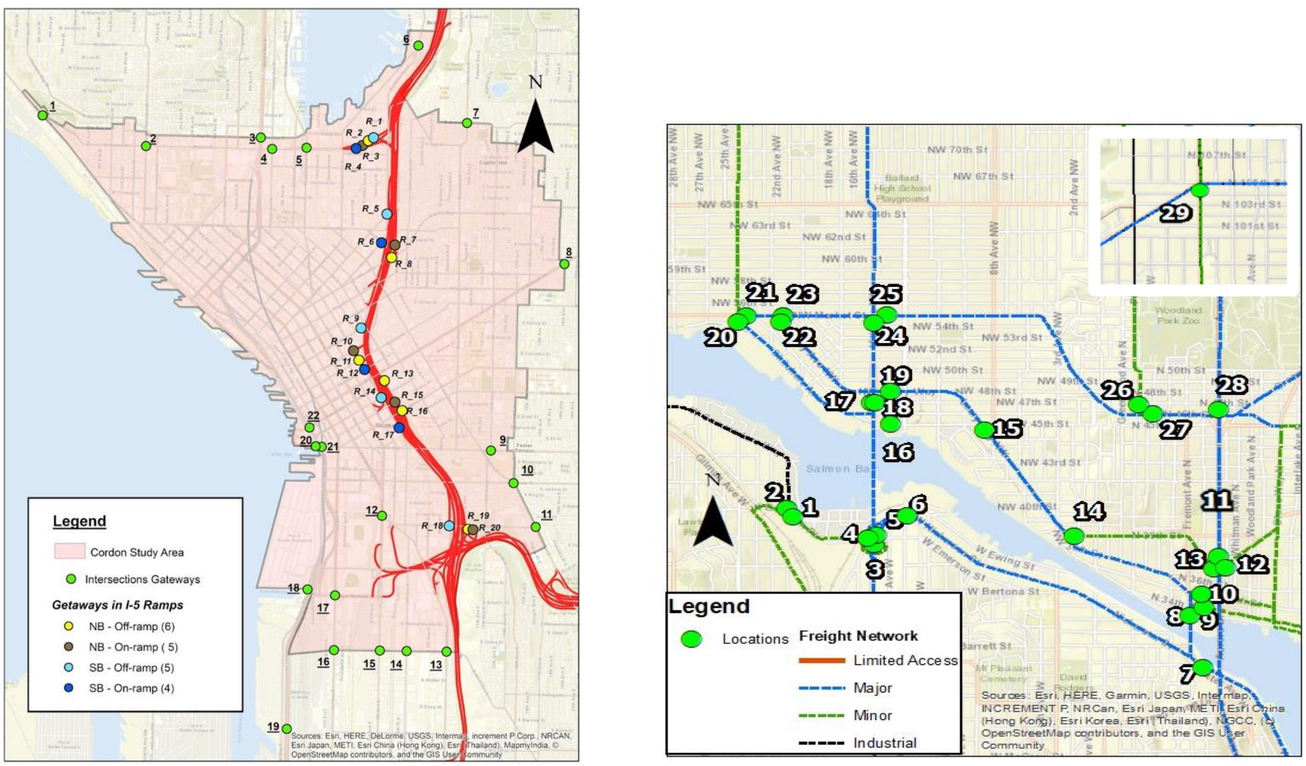


Figure 3-4. GDA and Ballad/Interbay count locations

Also, a cordon was defined based on the GDA's locations (i.e., gateways), and the GDA's formal boundaries that enclosed the study area to gather inbound and outbound CV volumes. The objective behind this effort was to capture a baseline cordon for Seattle's most constrained and densest neighborhoods. All vehicles crossing the cordon line were counted. As such, the data here refer to vehicle movement in two directions across all gateways 1) inbound (into the urban center) and 2) outbound (out of the urban center). However, the GDA vehicle count cordon study does not capture vehicle movement inside the cordon. Moreover, while it provides a comprehensive picture of center city traffic composition, this effort cannot be considered a comprehensive count of all of Seattle's urban center traffic. Rather, the study uses the data to identify traffic patterns and vehicle types in established gateways.

Data collectors watched the videos and produced manual counts of all vehicles crossing predefined screen lines in a 15-minute interval aggregated by day of the week, time of day, vehicle body type, vehicle use, number of axles, and directionality. The gateways include arterials, collectors, interstate on/off ramps, and ferry terminal entries/exits, capturing major commercial traffic routes in both study areas, see Table 3-2. for further detail.

Table 3-2. GDA cordon and Ballard count locations.

Time frame of the Data Collection Effort	***Study Area***	***Location Type***	***No. of locations***	***No. of days surveyed***	***Hours Collected***
Sept. & Nov.2018	**Greater Downtown area (GDA)**	Reference Intersections	3	7	168 hrs. [120 hrs *(only CVs counted + 48 hrs (all veh. counted)]*
		Intersections	16	2	48 hrs. (all vehicles counted)
		Additional locations * *(non-gateways outside or inside the cordon perimeter)*	3	2	48 hrs. (all vehicles counted)
		I-5 On and Off-Ramps	20	2	48 hrs. (all vehicles counted)
Dec. 2019 & Jan. 2020	**Ballard/Interbay Northend (BI)**	Roadway legs	23	2	48 hrs. (all vehicles counted)
		N-B On and Off-Ramps	6	2	48 hrs. (all vehicles counted)

The GDA gateways include arterials, collectors, I-5 on/off ramps and ferry terminal entries/exits, comprehensively capturing all major commercial traffic routes. For the Ballard area, the locations comprehensively capture all major commercial traffic routes including arterials, collector arterials and the ramps for the Ballard, Fremont, and Aurora bridges. The selected roadways also capture CV volumes along the study area's major corridors:

- <u>North-South:</u>
    - **15th Ave W/NW,**
    - **Fremont Ave N, and**
    - **Aurora Avenue N.**

- <u>East-West between the N-S corridors:</u>
    - **NW Market St, and**
    - **NW Leary Way.**

3.5.3. Data Sample

A total of 146,541 CVs were observed, 87,718 in the GDA locations and the rest, in the Ballard/Interbay. Table 1 shows all observed vehicles in a 48-hour period.

Table 8 – Data sample.

Table 3-3. Data Sample

Activity Type	GDA (veh)	BI (veh)
Private	1,141,711	907,393
CVs	87,718	58,823
Public Transit	13,611	11,733
Other Transit	10,188	4,737
Emergency	1,381	483
RVs	194	116
Unknown	1,684	-
TOTAL	**1,256,487**	**983,285**

CV Share of Total Traffic

The average CV share of total traffic volume is 7%. The CV share per location varies between 3.6% and 10%, with just eight locations above 10% (all located in industrial zoning areas). It is worth noting that, in the GDA, this percentage drops to 5% when only considering the ramps and rises to 9%, when considering only intersections. This result may be explained by geometric limitations that drive away heavier CV traffic reducing the total share of CV that uses this infrastructure.

3.6. Discussion

The following section describes the insights derived from two Seattle case studies to bridge the data gap of CV traffic and, thus, build the foundation knowledge needed to evaluate and develop suitable strategies to improve the UFS movement and its supporting infrastructure.

Weekends and weekdays showed statistical differences in volume and daily profiles

For three GDA gateway seven-day worth of data was collected to evaluate day-of-week traffic variations, refer to Figure 3-5. For all locations, weekends showed the lowest volume of CVs, with Sunday having the lowest share of all locations (with day-of-week adjustment factors varying between 2.7 and 5.6). The day-of week-adjustment is used to represent the difference between days of the week with a typical average day volume based on the collected data.

Also, each of the three gateways showed slightly different behavior when evaluating the day of the week with the highest volume of CV. However, daily CV volume showed no statistically significant differences between weekdays within locations (at a 5% level based on a chi-squared test.)

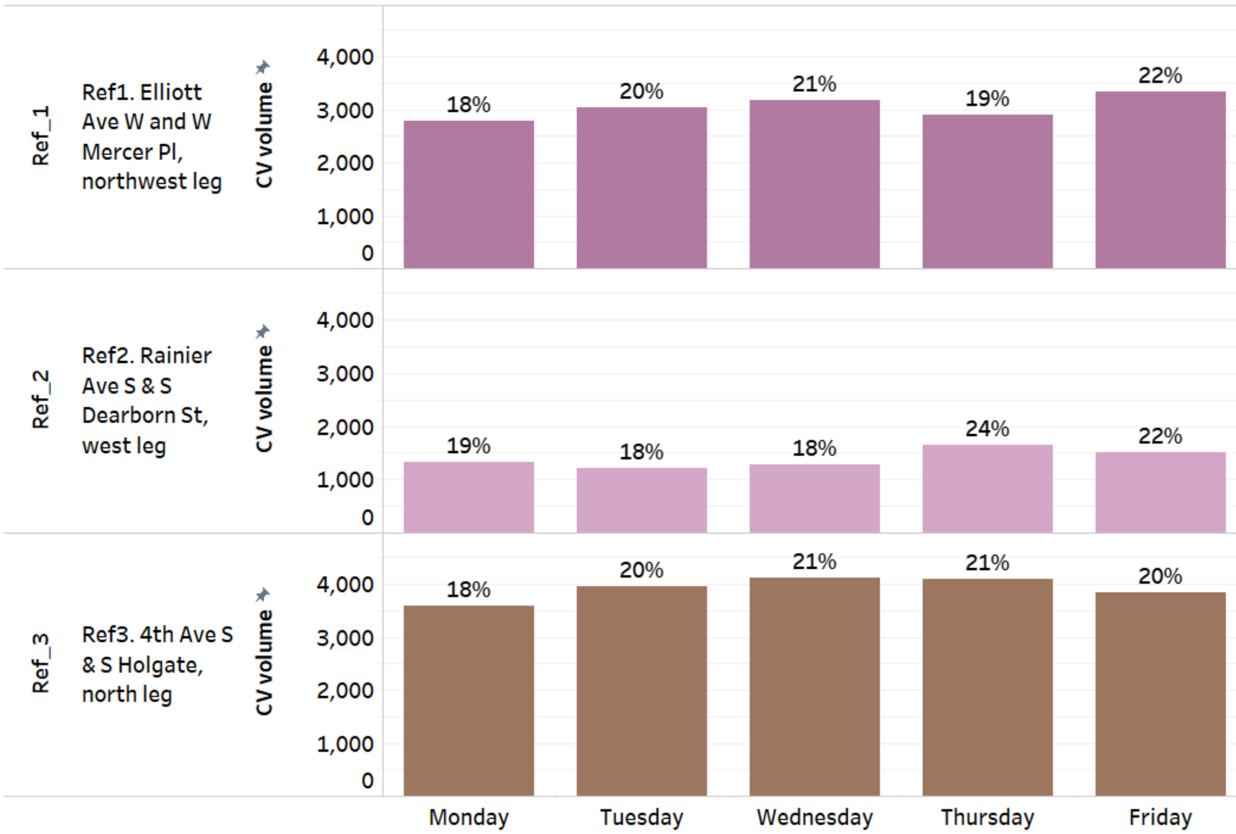


Figure 3-5. Share of CV volume per day of the week.

Smaller CV vehicles make up the largest share of all CV traffic.

In general, the most common CV body types in the traffic volume were small CV fleet - 54% of CVs in GDA and 60% in BI -. The second-highest proportion of CVs was single-unit 2-axle vehicles, at nearly 30% of all recorded CVs. When considering only the I-5 ramps and minor truck routes, the share of the smaller CV fleet rises to 60% and the volume for all truck categories drops to 5% (except for 2-axle single-unit trucks).

Service vehicles are a significant share of CV traffic.

Goods transport and service vehicles (SV) constituted the highest share, each accounting for 30% of all CV traffic in the GDA. For the Ballard-Interbay area, SV had far higher volumes than any other

CV category, accounting for 41% of all CV traffic. This share is significantly larger (10% larger) than what was observed in the GDA, highlighting, even more, the importance of service movements in this area.

Among the myriad ramifications of this finding, one is parking planning. As Giron-Valderrama et al. (2020) found service vehicles tend to have longer dwell times, with 44% of all observed service vehicles parked for more than 30 minutes and 27% parked for an hour or more. Given this study's finding of service vehicles representing a significant share of commercial traffic volume, these vehicles may have a disproportionate impact on parking space utilization on the curb.

Fleet size variations are influenced by CV activity type.

Different variations of the fleet configuration can be observed when evaluating body and activity types for all CVs (Figure 3-6). Service vehicles had the highest share (90%) of pick-ups and vans (smaller CVs). In contrast, single-unit trucks represented 75% or more of the total percentage of goods transport, construction, and waste management fleet. Additionally, trailers were between 11% to 15% of the total share of Construction and Goods Transport vehicles. This finding highlights the importance of accounting for the different operational and design requirements needed for mobility and access (e.g., turning radius, pavement considerations, and allocated space for load/unload activities).

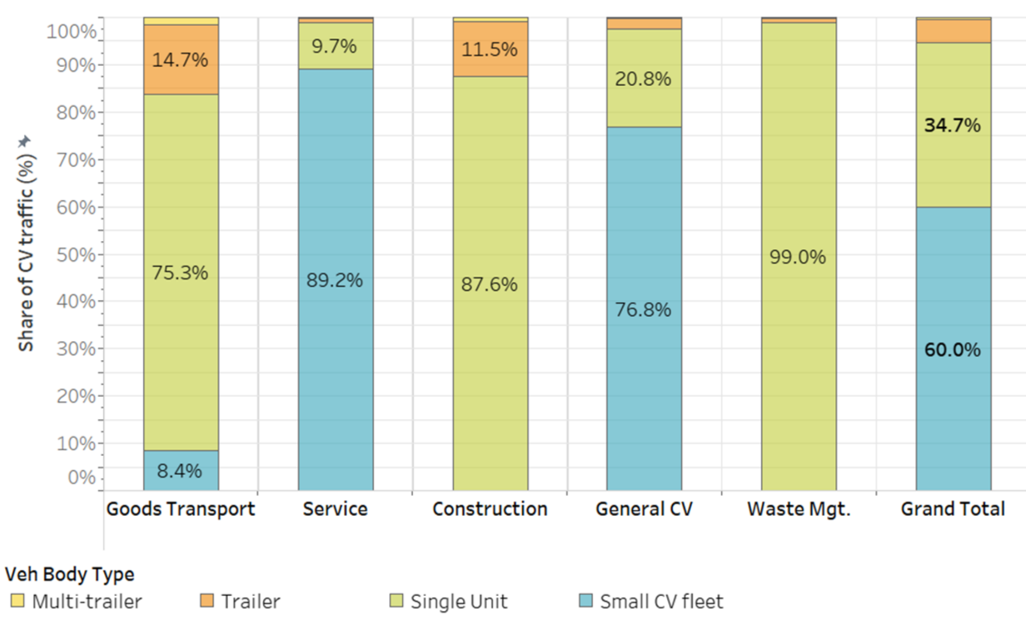


Figure 3-6. Share of CV traffic by activity type.

On average, daily share of CV volume is low in the evening and night and high during peak volume hours of 9:00 AM-12:00 PM

Since the counts were collected in a 24-hour period, it was possible to examine and compare traffic between 'daytime' (defined as 06:00-17:59) and 'nighttime' (defined as 18:00-23:59 and 00:00-05:59). Over the course of 24 hours, just 15% of CV daily traffic on average flows from 6 PM – 6 AM in the GDA and just 9% in Ballard-Interbay. Conversely, 27% of CV traffic occurs in the GDA and 31% in Ballard-Interbay during the three hours of maximum volume (9:00 AM -12:00 PM). This finding suggests no considerable travel demand in off-peak hours for CVs, while twice as much private traffic volume was at nighttime for both locations (28% for the GDA and 27% for the Ballard-Interbay).

Most CVs were observed during regular business hours, 6 AM – 6 PM, and had only one bump throughout the day.

The vehicle counts were accumulated for each hour, with each hour averaged over the 48-hour period to obtain the study area time profile. Figure 3.7. shows that CV volumes throughout the day are significantly smaller than private vehicle volumes. Private vehicle traffic has very distinct peaks during the commuter AM/PM peak periods, reflecting passenger vehicles' general profile across Seattle.

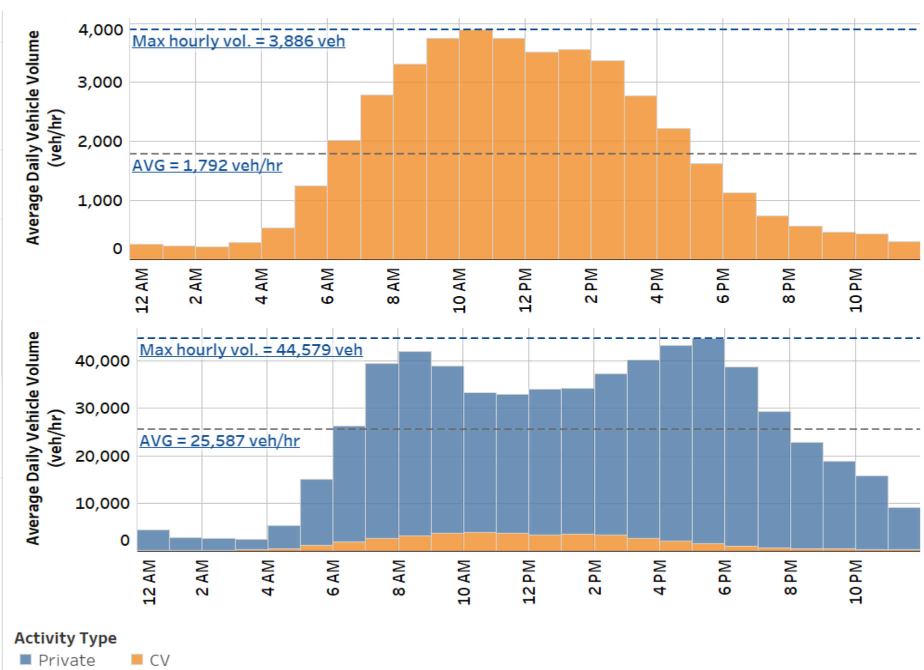


Figure 3-7. Average daily CV and private vehicle volume (veh/hr).

Most CVs were observed during business hours (6:00 AM-6:00 PM). In contrast to the private vehicle pattern, CV's time-of-day pattern has only one "hump," peaking during the morning and early afternoon, and steadily declining over the day. The 'spare' capacity freed up by the decline in private vehicles between commuter peaks was utilized by the number of CVs observed in the selected gateways. This is consistent with the findings in other research and data collection efforts [15].

Figure 3-8 shows the vehicle volume share during the three hours with the highest volume for CV, the private vehicle AM peak and the private vehicle PM peak. The three most intense hours for CV traffic (9:00 AM – 12:00 PM) hold 27% of the CV traffic. In contrast, 19% of private vehicles were observed during the AM peak period (7:00 – 10:00 AM) and 21% during the PM peak period [4:00 PM – 7:00 PM]. This finding suggests that CVs have a more intense peak than private vehicles.

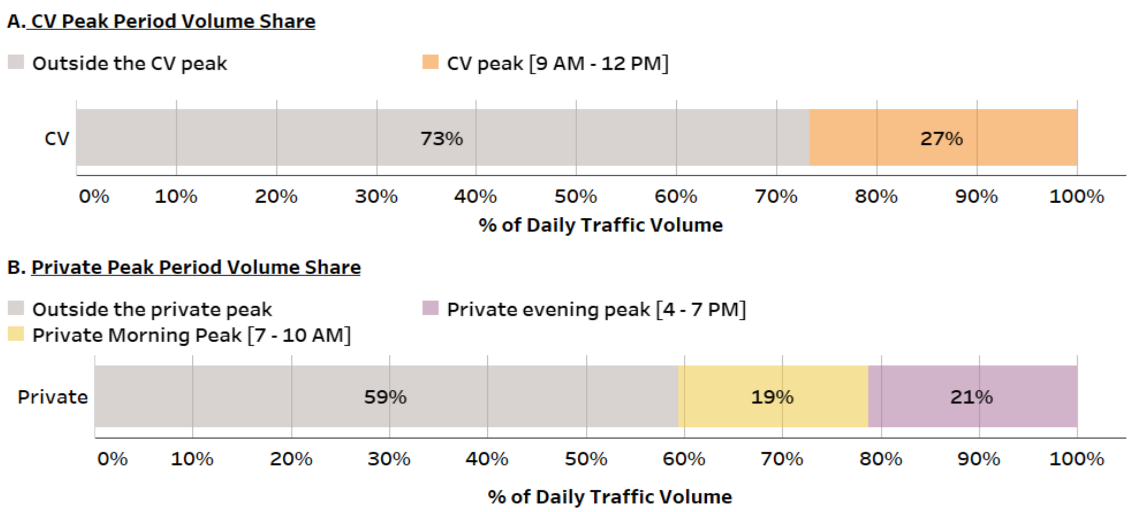


Figure 3-8. Comparison of peak vs non-peak traffic counts for commercial and private.

Overall, CV traffic peaked in the late morning time with observed variations between locations inside each study area.

Overall, GDA CV traffic peaked at 9:45 AM, two hours after the private vehicle AM peak (7:45 AM), and 7 hours before the private vehicle PM peak (5:00 PM). Very similarly, Ballard/Interbay CV traffic peaked at 10:15 AM, roughly 2 hours after the private vehicle 8 AM peak and 6 hours before the private vehicle 4:15 PM peak. Looking across gateways, CV peak took place between 7:00 AM and 1:45 PM for 90% of all locations.

Regarding directionality in the GDA, overall inbound and outbound CV volume both peaked in the morning: inbound at 8:30 AM and outbound at 11:30 AM (Figure 3-9). Additionally, the aggregated daily average inbound and outbound CV volumes in the GDA were balanced with a daily hourly volume of 459 CVs/hr. and 463 CVs/hr., respectively.

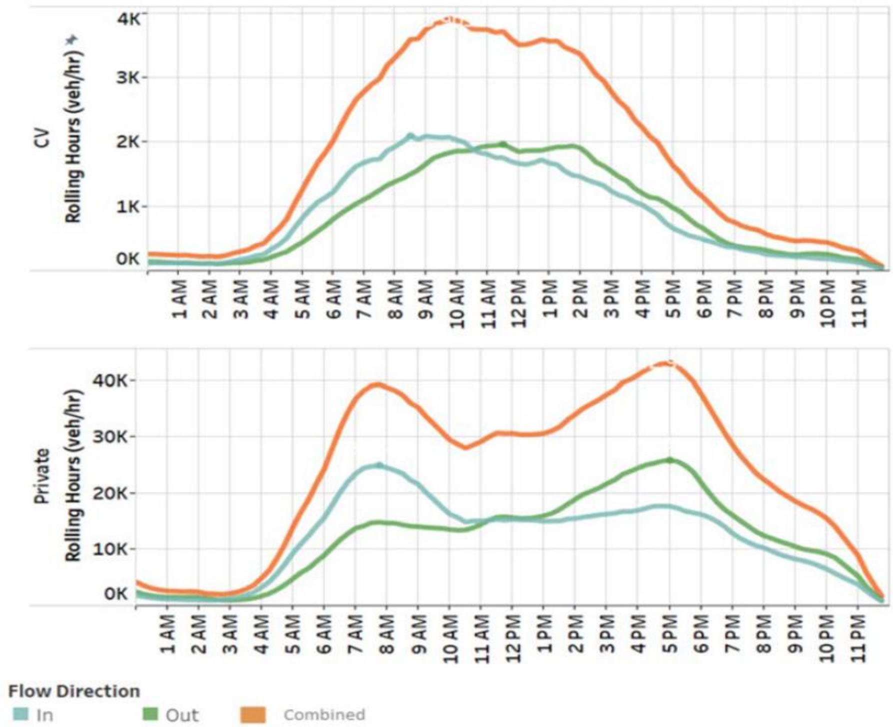


Figure 3-9. Private and commercial vehicle rolling hour by the time of day (different scales used).

When evaluating CV activity in the GDA, each category peaked at different times in the GDA:

1. Goods Transport peaked at 9:00 AM.
2. Construction and General CM were consistent with overall trends, peaking at 10:00 AM and 11:15 AM, respectively.
3. Service peaked in the early afternoon, at 2:00 PM.
4. Waste Management (WM) peaked at 6:15 AM. Considering that the WM vehicle stations are located outside of the study areas (City of Seattle, n.d.), it is logical for the peak hour to be before the start of waste collection (7:00 AM) and after street sweeping (6:00 AM) is completely based on services provided by the City of Seattle.

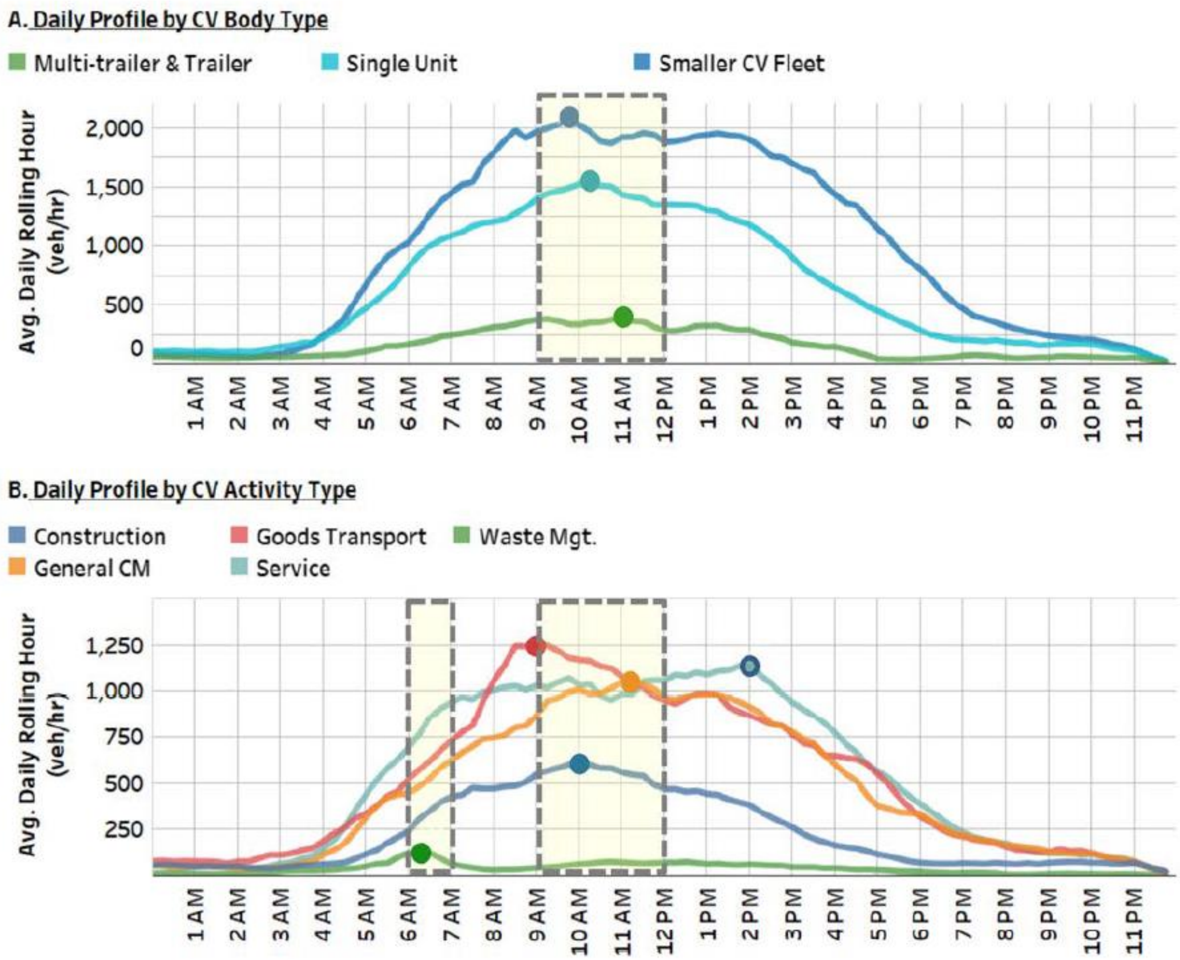


Figure 3-10. Daily traffic profiles by (a) CV body type, and (b) activity type.

While there are peak differences in each CV activity category, overall CVs across body types peaked between 9:45 AM – 11:00 AM, as indicated in Figure 3-10 and consistent with the overall trend findings described earlier. In contrast, in the BI, each CV activity type and CV body type showed similar peak periods between 10 AM – 12 noon. The only exception was the Waste Management (WM) category, which peaked at 7 AM.

Freight Network designation should account for CV fleet heterogeneity and changes in CV flows

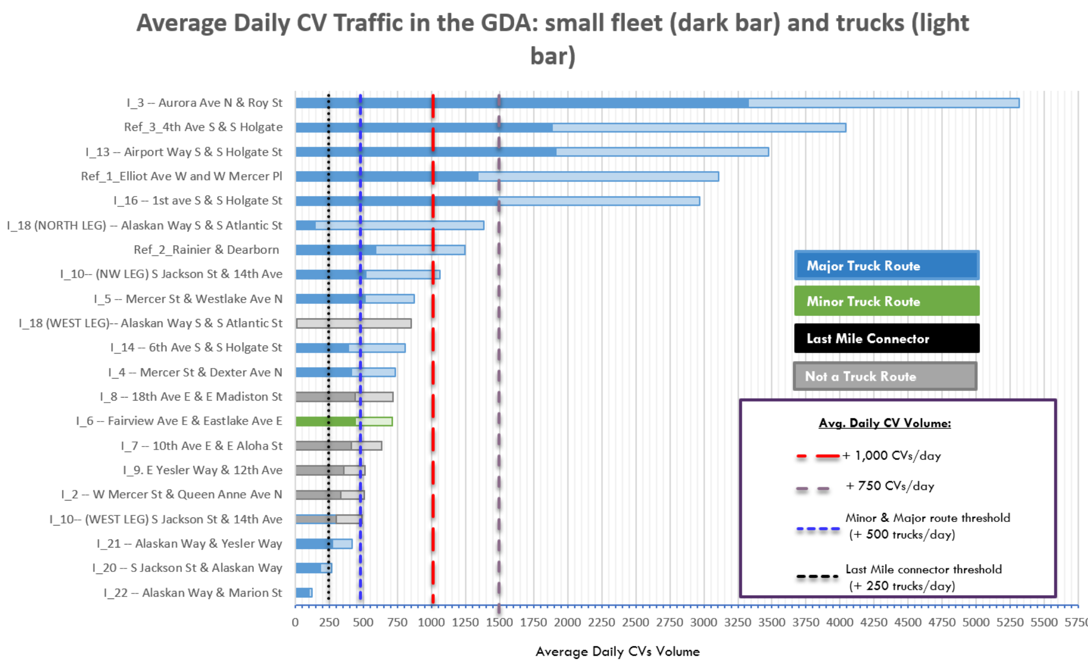
Several cities in the U.S. have established a Freight Network (FN) to identify the key roadway segments that carry the highest CV volumes of freight and provide connectivity to key freight facilities. This classification system does not necessarily change the elements' overall function, design, or character. But it does underscore the importance of ensuring that commercial flows can be safely accommodated. Municipalities often identify the FN elements based on a set of attributes or criteria which may include average daily truck volumes, connectivity, design vehicle, adjacent land use, and resilience (SDOT, 2016). However, this categorization is mostly qualitative and often relies on often spatial and temporal limited vehicle counts; and misses capturing the smaller commercial fleet. Moreover, the current FN conceptualization focuses mainly on the mobility of the big and heavy CV. It neglects the significant share of smaller CV vehicles navigating the city streets and the infrastructure's essential access function for freight movement, particularly for the last-mile connection to the commercial district and neighborhoods. For example, when considering the smaller commercial fleet, 40% of locations reported more than double the volume threshold defined by Seattle’s FN system.

Without capturing the share of smaller CV vehicles in city streets, the current FN designation will:

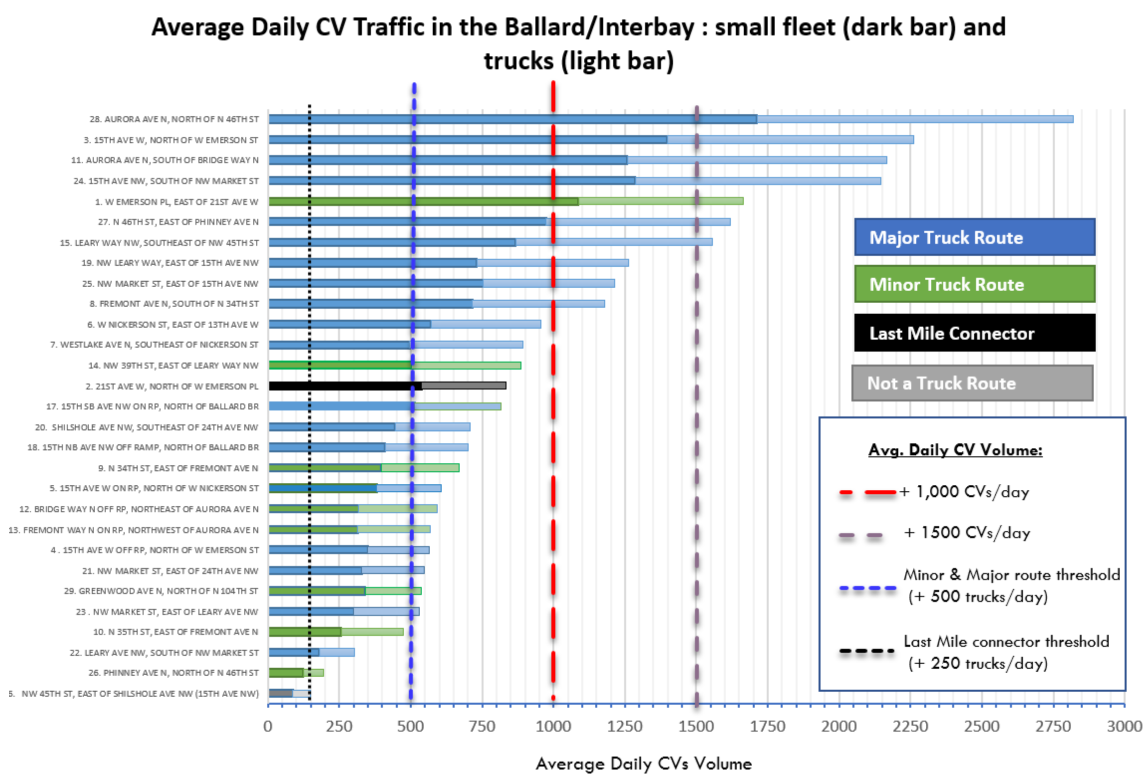
1. Underestimate the total CV traffic flow load in the network elements, because of not capturing the smaller CV fleet.
2. Misidentify minor arterials, neighborhood collectors, and local streets that carry a high share of CV movement (because of the increase in last-mile deliveries) that may not have been designated as part of the Freight Network before.
3. Be unable to distinguish the roadway segments that are providing a key access function for the freight movement, information necessary for load/unload infrastructure planning.

In 2016 Seattle’s Freight Master Plan (FMP) (SDOT, 2016). the most recent city’s FN was established. The network designation included a four-tier categorization, including First/Last Mile Connector, Major, and Minor Truck Streets.

For the Ballard/Interbay study area, all data collection locations are designated as part of the FN. On the other hand, only 70% (15 out of 21) of the studied intersection and none of the I-5 ramps are elements of Seattle's FN in the GDA. When evaluating the data collection locations defined as elements of the SDOT's FN, the observed truck volumes differed from the daily volume threshold for their assigned designations (more than 60% did not meet the threshold). However, when considering all CV volume (i.e., small fleet and trucks), almost 90% of locations meet the Major Truck Route threshold (500+CVs/day), including more than 40% of locations not designated as FN's elements (Figure 3-11). Even more, 40% of all locations (excluding the I-5 ramps) reported more than double, and 25% exceeded by nearly three times the daily volume threshold for the Minor and Major Truck Routes defined in the SDOT's Freight Master Plan (SDOT, 2016). This finding highlights the importance of including smaller vehicles in the freight planning process to identify and characterize the roadway network elements supporting the urban CVs flows.



(a)



(b)

Figure 3-11. Evaluation of average daily CV volume based on Freight Network threshold in the (a) GDA (excluding I-5 ramps), and (b) Ballard/Interbay area.

Significant variation in traffic patterns due to local conditions relates to local infrastructure and land use attributes.

Although all locations showed the typical CV daily pattern for the aggregated volume, some locations in both study areas had directionality-oriented traffic patterns, with CV movement in one direction during some parts of the day and in the opposite direction during others. These patterns may be explained by the temporal distribution of activities in the area, for example, where specific fleets may be leaving or returning to a warehouse, commercial area, or the port terminal.

Additionally, we observed significant variations in traffic patterns related to peak hours, CV share of total traffic, CV volume, directionality, CV fleet composition, and peak/average CV flow ratio. For example, in the GDA area, the average daily traffic flow aggregate showed a lot of variations when aggregated by gateway location (Figure 3-12). However, nearby intersection gateways showed similar commercial traffic patterns. The finding illustrates how the adjacent land, proximity to freight facilities, street type, and network connectivity impact network elements' use patterns (Figure 3-13).

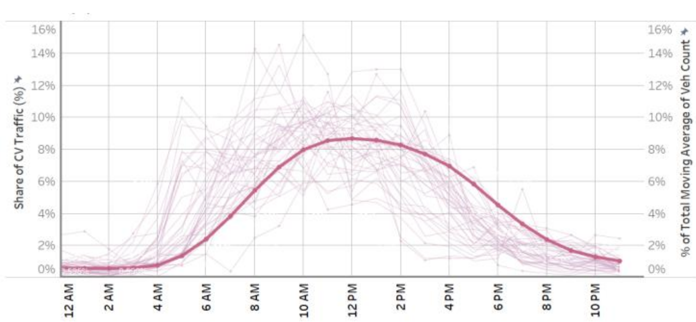
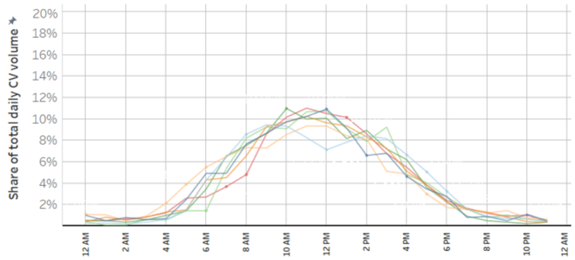
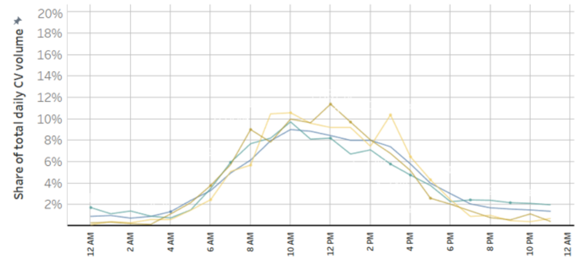


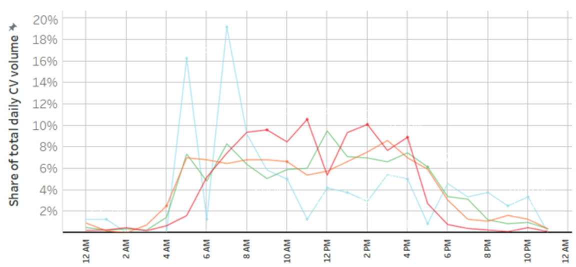
Figure 3-12. Average CV Daily Profile for the GDA locations aggregated by vehicle body, activity type and direction.



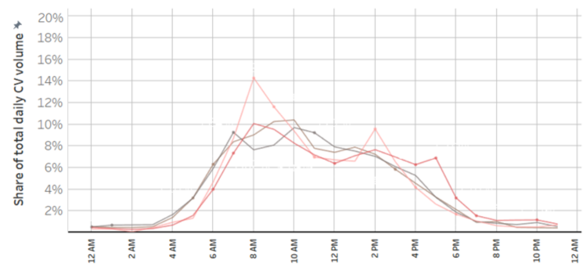
(a) CV share of total traffic for northern gateways.



(b) CV share of total traffic for eastern gateways.



(c) CV share of total traffic for western gateways.



(d) CV share of total traffic for southern gateways.

Figure 3-13. Average CV Daily Profile for the GDA locations aggregated by vehicle body, activity type and direction and displayed by proximity along the GDA cordon.

3.7. Conclusions

Classification of vehicles using traffic data is one of the most important parameters for traffic management. However, in the urban environment, most of the efforts have been focused on the movement of passenger vehicles, and little attention has been paid to CV classification. Therefore, there is a need for methods that produce empirical and comprehensive databases that allow the detection and classification of disaggregated urban CVs flow lay the foundation for freight analysis and transportation planning at the city level.

This study proposes a new vehicle typology and shows the CV flows through the application of cordon counts in two case studies in Seattle. The detailed vehicle typology was created to capture the complexity of urban traffic using video data and human observers and covers 65 different vehicle categories based on body type and the number of axles. This detailed vehicle typology was used to identify the vehicle types during the data collection in Seattle's case studies, GDA, and Ballard/Interbay areas. Traffic data were collected in 71 points in the two study areas, while a cordon was defined in the GDA area to gather inbound and outbound volumes.

This work's contributions include the development of a replicable and comprehensive typology to record granular CV and a demonstration of the methodology with two Seattle case studies. It also paves the way for providing tailored solutions that consider the spectrum of commercial vehicle fleets and their different needs for the operations; and builds a broad understanding of CV activity surely lacking from transport planners.

The results of this research produced several key insights:

1. There are fundamental operational differences between passenger and CV traffic
2. Passenger traffic patterns are influenced by commuting patterns and the profile of the city residents.
3. Service vehicles are a significant share of the CV fleet, which require attention in the transportation planning process.
4. CV movements depend on the operational hours of commercial businesses to complete deliveries and provide services.
5. CV traffic patterns have one "hump" during the day as opposed to passenger traffic, which has two.
6. It is important to account for CV fleet heterogeneity and changes in CV flows.

While grounded in specific locations, the findings contribute to a broader urban theoretical and conceptual understanding of CV urban configuration fleets and flow patterns. It also supports the development of traffic monitoring systems and CV classification. Finally, this study provides the possibility of detecting daily CV traffic patterns and developing applications that respond to these patterns as a basis for the freight planning policies at the municipality level.

Chapter 4. Characterizing Commercial Vehicle Within Day Traffic Variations

This chapter is related to the CV within-day traffic temporal and spatial variations, in which we examine the related work and the proposed research method for the dissertation work ahead. Then, a methodology for the evaluation of temporal and spatial traffic flow pattern variations is proposed and implemented. The proposed method includes the implementation of an unsupervised classification approach (i.e., clustering) and an exploratory spatial analysis. Finally, a summary of the findings and a discussing on the policy-related implications are presented.

4.1. Introduction

Commercial vehicles (CVs) are critical and growing components of urban traffic. Accurate CV traffic characterization, monitoring, and analysis can reveal the Urban Freight System (UFS) system's impacts on the transportation network and uncover trends for proper future planning, investment allocation, maintenance, forecasting, and infrastructure management. However, most of the traffic analysis research of CVs focuses on the highway system, partly due to the absence of traffic data at the urban scale. This leaves the local governments with little insight into the temporal and spatial patterns of the CVs traffic flows to inform the decision-making process adequately.

At the national and state scale, CV's origin-destination (OD) level flows are reasonably well captured with several freight models and statistical data that provide insight into spatial, temporal, and activity variations. Specifically, the Freight Analysis Framework complies the nation’s most significant freight dataset, such as the Commodity Flow Survey (CFS), which includes 153 zones that mostly show ingress and egress at the state level. Most recently, the 2021 Vehicle Inventory and Use Survey (VIUS), which was reinstated in 2021 after being discontinued for nearly 20 years, provides data on the physical and operational characteristics of the nation's truck population; it breaks down the truck fleet into light (pickups, SUV and minivans) and heavy vehicle categories.

A gap exists in the datasets described above between the regional and local goods movement as the CFS’s primary goal is to produce national and state-level estimates of the total number of trucks and truck miles and does not show how vehicles move around within cities and towns. Similarly, VIUS does not include spatial granularity at the city level. This leaves municipalities with no nationally consistent sources, models, or guideline for collecting, evaluating, and estimating freight activity data.

Furthermore, highway/interstate traffic cannot directly translate to the urban landscape, as urban CV flows differ and are more complex than highway traffic. Reasons explaining this difference include:

1. Multiple traffic modes coexist, interact, and compete in the urban space.
2. The urban network contains many intersections, which results in traffic characterized by many minor disturbances, compared to highways that generally show fewer disturbances yet with higher impact.
3. The travel demand characteristics (distance, motive, and fleet configuration) are more diverse than traffic on highways.
4. The more fragmented nature of freight flows and increased delivery challenges such as higher cost and timeliness.

Evaluating, modeling, and estimating the urban CV flows requires high-detailed data often unavailable to the local agencies. In return, agencies have no option but to rely on intuitive assumptions to account for commercial vehicle flows and their impacts on congestion, infrastructure conditions, safety, and emissions. As discussed in Chapter 3, only applying a car/truck split, a simplified classification system or just considering general traffic under the assumptions of passenger traffic. However, all these options do not adequately account for the complexity in commercial vehicle flows.

For example, regional and urban traffic estimations models often focus on passenger travel and dismiss the particularities of urban CV flows by simplifying the problem on the premises that urban CV flows are a small share (5-15%) of Average Annual Daily Traffic (ADT) compared to Passenger Vehicle (PV) traffic. However, as PV flows differ from CV flows. For example, modeling PV flows often relies on the assumption of a relatively common set of trip purposes, including to/from work, shopping, recreational and purpose trips. This assumption explains a high percentage of the observed PV flows, which means it is sufficient to use the resulting model for traffic planning purposes, perhaps after additional calibration using traffic counts. On the contrary, the urban network serves freight trips with a different set of distance travelled (i.e., long-haul trips as well as a considerable amount of medium and local trips), travel motives (e.g., reverse logistics, drayage, fulfillment, home delivery, service provision, construction), fleet configuration and route configurations (e.g., number of stops; OD; time windows; route frequency and length; shipment size).

Traffic analysis process evaluate annual AADT and the design hour volume (peak hour). The first metric is primarily used for network and maintenance planning and evaluation. In contrast, the peak hour is used for design work. However, the daily variability is of crucial importance. The information on the state of the traffic system at different locations and on different time periods provides temporal and spatial insight of congestion, reliability, and spare capacity. Finally, the information provided in the current average traffic conditions does not provide enough detail to understand the fleet configuration or connect these movements with economic activity.

In response to this urban challenge, our research leverages the comprehensive and high-quality data set of the urban CV flows in Seattle, Washington presented in Chapter 3. A clustering approach is implemented to discover "typical" urban CV traffic patterns and identify vehicle and time-related features that best describe the identified traffic patterns. Clustering is a popular unsupervised method used in data mining to discover potential patterns, used specially for geographical data that visually may be tough to identify (Zhao et al., 2004). Lastly, an exploratory spatial analysis is executed to determine the subset of attributes (infrastructure geometrical characteristics, network topology and urban form) that are related to these CV Daily Flow Patterns.

Finally, the objective of this chapter is to improve insight and build foundational knowledge into urban CV Flow patterns by developing and applying a framework to analyze temporal and spatial variations in an urban center. It aims to motivate transportation agencies and researchers to leverage on the insights from this research to 1) formulate and prioritize CV traffic data collections needs, 2) develop city-level freight networks and guide infrastructure investment, 3) characterizing CV traffic demand at the urban scale and 4) support the development and evaluation of data-driven policy.

4.2. Literature Review

Most of the current research focused only on the accuracy of estimation or prediction results and not on the interpretation of spatiotemporal traffic patterns, particular in the urban context (Ma, Li, and Chen, 2020). However, in the last two decades, there has been an increased interest in choosing clustering analysis for the identification of “typical” traffic patterns that are representative of traffic conditions to assist the evaluation, management, improvement, and maintenance of the transportation network. In fact, the FHWA updated their guidance for utilizing analysis, modeling, and simulation (AMS) to incorporate clustering analysis to detect operation scenarios as an important component of AMS (Saha et al. 2019).

Saha et al. (2019) points out that while clustering analyses have been used extensively in other disciplines, their use has been limited in the transportation engineering field. However, there has been a growing interest in its use in recent years due to the increasing availability of detailed transportation data and the identified needs for scenario identification for AMS and decision support. Transportation agencies and researchers have devoted significant attention to clustering approaches aiming to identify traffic patterns (i.e., travel times, volume variation or speeds) for large datasets in transportation networks (Salamanis et al. 2017; Almannaa, Elhenawy, and Rakha 2020). Table 4-1. summarizes nine of these studies, including the traffic metric studied, data inputs for clustering analysis, temporal and spatial features analyzed and the clustering approach.

Our literature review found that different types of data have been investigated in cluster analysis, including traffic density and speed data (Xia and Chen 2007), traffic volume (Park 2002; Banaei-Kashani, Shahabi, and Pan 2011a; Chen, Yang, and Xu 2017), AADT (Gecchele et al. 2011), and freeway traffic flow conditions (congested and level of service) (Azimi and Zhang, 2010). Data source for these evaluations include manual traffic counts, GPS traces (Necula, 2015), loop detector occupancy (Saha et al. 2019; Banaei-Kashani, Shahabi, and Pan 2011a; Weijermars 2007) and LIDAR (Tok et al., 2017).

Based on induction loop data (continuous observations of vehicular traffic for a selection of major road segments), a very detailed analysis of traffic distributions has been conducted by Weijermars (2007). This paper evaluated weekly patterns, seasonal and weather variations. Traffic flow profiles were identified using Ward's hierarchical clustering algorithm and subsequently what factors were responsible for the resulting clusters were investigated.

Banaei-Kashani et al., (2011a) applied unsupervised clustering to evaluate and identify a set of distinct "signature" speed traffic patterns aiming to accurately represent the observed traffic flows on all segments of the Los Angeles County Road network. Based on their correlation analysis, they found that the combination of direction, connectivity and locality of a road segment can best predicted the defined "signature" patterns.

Gecchele et al. (2011) performed a comparative performance analysis of different clustering techniques to identify traffic patterns variations and similarities based on traffic volume data in the Province of Venice, Italy. Differently from most of the studies, traffic adjustment factors were calculated separately for passenger vehicles and truck. Kim & Chang, (2012) also considered passenger and trucks separately while using a hybrid clustering technique to calculate peak-hour ratios for road traffic volumes based on for 24-hours traffic data in Korea.

Salamanis et al., (2017) implemented a Density-Based spatial clustering to create cluster of partial time series with the objective to develop a method for accurate traffic prediction under both normal and abnormal conditions. The experiment took as input a 4-month traffic dataset of 120 locations with counts collected in 5-minute intervals. They pointed out that future work should include the examination of the impact that specific clustering algorithm have on the performance of the mode. Necula (2015) tested K-means, hierarchical clustering, and density-based clustering DBSCAN to find groups of street segments with similar course of traffic flow over time (volume hourly variation). This analysis found K-means to be the best model (in terms of volume, shape, and orientation for all clusters). The dataset consisted of 10000 vehicle GPS traces from around 3600 drivers in New Haven, Connecticut, USA.

Yang et al., (2017) utilize clustering to analyze the traffic state variation based on quantitative speed data for three sections of the Beijing network. This study concluded that the frequency distribution of the section for each cluster were related to road hierarchies, location, and function. However, when it comes to road function, as Weijermars (2007) argue: "the use of [road] function classes in a [predictive] model fails to capture the underlying causes of varying traffic volumes". For CV flows, besides road function class, other non-temporal features should be consider including number of lanes, connectivity, adjacent land, commercial fleet configuration.

In summary, the literature review shows that while applying cluster analysis to analyze traffic patterns is still a very fluid area of research, there is a substantial body of work supporting the application of this approach. Multiple cluster analysis algorithms have been considered (including K-means, Ward's, DBSCAN, and most recently spectral clustering) with some studies doing a comparative analysis to find the superior model. Also, multiple types of traffic data have been considered including speed, volume and trajectory and its relationship with non-temporal attributes, such as roadway function and infrastructure geometric characteristics. The motivation of these studies generally includes identifying traffic patterns for improved forecasting by including a range of variation sources such as seasonal and temporal variations, compositions of flows (passenger and trucks), normal and abnormal conditions and spatial-temporal variations. However, and most importantly, when it comes to urban CV travel, there is not well documented understanding about its daily variations disaggregated from the general traffic.

Table 4-1. Summary of Traffic Flow Clustering Studies

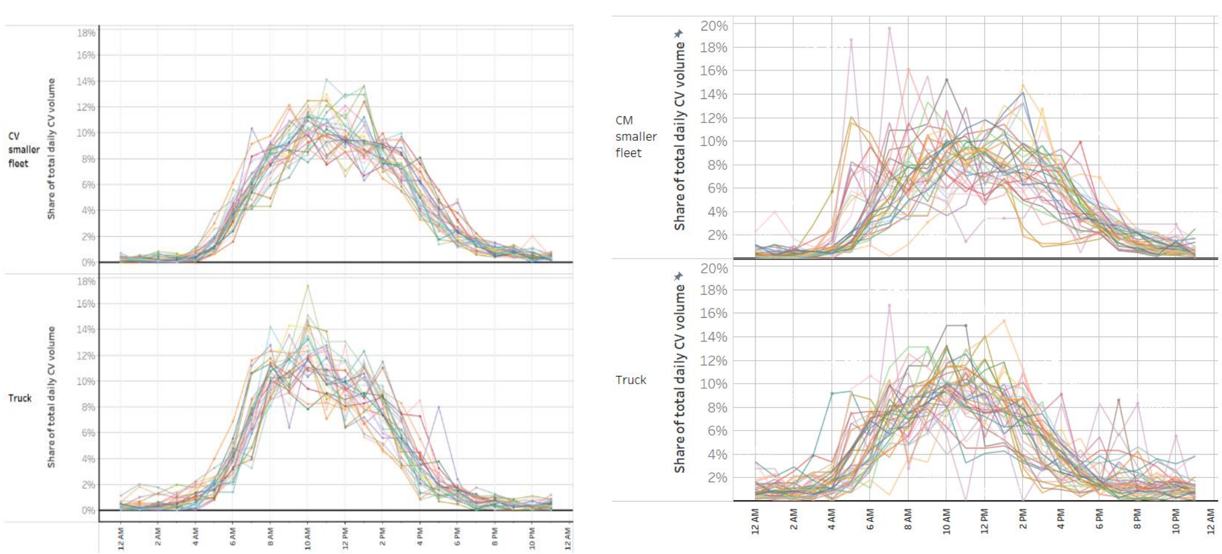
Source	Data	Cluster Object	Temporal Features	Spatial Features	Distance Metric	Clustering Method
Salamanis et al., (2017)	Traffic counts from one freeway vehicle detection station in Oakland, CA, during 4 months. The data is aggregated in 5-min. intervals by traffic direction.	Daily flow profile	Traffic incidents lasting over 3 minutes, weekdays, and weekends	None	NA	PCA and Density-Based Spatial Clustering of Applications with Noise
(Park, 2002)	836 traffic loop data 18 days freeways and ramps in San Antonio, Texas, in 5-min. intervals	Daily flow profile	None	None	Euclidean distance	fuzzy C-means
(Weijermars and Berkum, 2005) (Thomas & Weijermars, 2008)	Traffic counts form loop detectors in highway A50 and road segments in Almelo, Netherlands.	Daily flow profile	Day of the week, month, and weather	Location	Euclidean distance	K-means and Ward algorithm
(Banaei-Kashani et al., 2011)	Los Angeles traffic loop-sensor data: 10 month / 1,592 sensors on freeways in 15-min. intervals	Daily flow profile	None	Road segment length, number of lanes, direction, network topology (i.e., connectivity and density), locality	-	X-means
(Gecchele et al, 2011)*	24 hour Traffic counts from 54 automatic traffic recorder (ATR) sites for a 6-month period. Province of Venice, Italy.	AADT estimates per location	3-day types (Weekdays, Saturdays and Sundays) and 2-month periods	None	Euclidean distance	K-means, PAM, K-medoids, Ward, X-Means, several Model-based algorithms.
(Necula, 2015)	10000 vehicle GPS traces, from around 3600 drivers for 8200 roadway segments of an urban area in onnecticut, USA	Daily flow profile	3-day types (Weekdays, Saturdays and Sundays)	Land use and road functional classification	-	Hierarchical clustering, DBSCAN, K-means.
(Calafate et al., 2015)	Traffic data for 421 urban segments based on loop detectors	Daily flow profile	Day of the week	None	Correlation distance	PAM algorithm (K-medoids, partitioning methods)
(Yang et al., 2017)	55 road segments, 28 days of consecutive data for three section of the Beijing urban road network.	Daily speed profile	Holidays, weekday	Road functional classification	Spectral Embedding	Ng-Jordan-Weiss NJW algorithm, (better performance a lower computational complexity for points in a high-dimensional space)
Saha et al, 2019	Microwave detectors in a freeway in Fort Lauderdale, Florida.	Volume, speed, occupancy, travel lane blockage, incident severity, weather.	None	None	Euclidean distance	K-prototype, K-medoids, a set of hierarchical methods, K-means on PCA-mix

**Only one out of 9 references consider commercial vehicle volumes in the cluster analysis*

4.3 Data

For this study, the traffic state variation is described by the average flow (also known as flow rate, intensity, or volume), which is defined as the number of vehicles through a road link by a unit of time. This dataset is considered a geo-referenced time series database as its storages the history of the evolving object (i.e., time period) while providing geo-reference (i.e., road segment) to the measure variables (i.e., flow) (Maimon and Rokach 2010).

Average weekday CV flow profiles at the individual link (i.e., roadway segments and ramps) were based on two days of collected traffic data for each location. Figure 4-1. describes the average daily flow profile per each count location by plotting the average 15-min share of total daily CV volume (%) by study area and by CV vehicle size (i.e., small CV fleet vs. Truck). Trucks showed an earlier peak for both study areas than the smaller CV fleet. Count locations in Ballard/Interbay showed more consistent patterns between each other than the ones in the GDA.



a. (b)

Figure 4-1 Small fleet vs trucks: average daily flow profiles per location expressed as share of traffic volume per CV body type aggregated by direction. (a) Ballard/Interbay and (b) GDA.

4.3.1. General Trend

To identify and characterize the Daily Flow Patterns, the resulting cluster schemes are evaluated against the CV Traffic General Trend of the data collected. This comparison helps identify variables that explain the observed temporal variations. As discussed in the previous section, most CVs are observed during business hours (6:00 AM-6:00 PM). In contrast to the private vehicle pattern, CV's time-of-day pattern has only one "hump," peaking during the morning and early afternoon, and steadily declining over the day. The ‘spare’ capacity freed up by the decline in private vehicles between commuter peaks is utilized by the number of CVs observed in the selected gateways, see Figure 4-2. This is consistent with the findings in other research and data collection efforts.

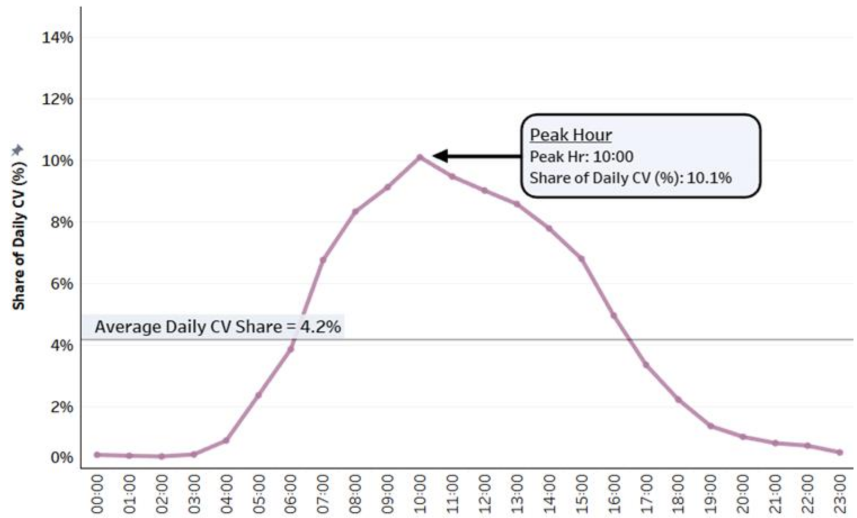


Figure 4-2 – Commercial Vehicle Flow General Trend for both study areas.

4.4 Methodology

This research evaluates within-day urban CV traffic temporal and spatial variations and groups the observed daily flow profiles to identify "typical" CV traffic patterns. It leverages Seattle's traffic count data under the assumption of three popular instinctive understandings about traffic flows on road segments described by Banaei-Kashani, Shahabi, and Pan (2011a):

1. Each road segment has a typical traffic flow,
2. Segments can be categorized and grouped into a set of distinct clusters based on the similarity of their traffic volume variations,
3. Within each category or cluster, road segments have not only similar traffic flows but also similar other characteristics (geographical, infrastructure-related, connectivity).

In other words, there is a limited number of "typical" traffic patterns (i.e., labels), each of which can accurately represent all daily CV volume variations observed in its subset. If such a typical pattern exists, it can be exploited to label each element of the road network (daily flow profile) that accurately characterize the typical flow of that segment.

The focus of this research is not to distinguish between locations with high CV traffic loads and locations with lower CV traffic loads, yet in differences in distribution of traffic over the day (i.e., daily temporal variations). Therefore, due to the considerable variation in the volume magnitude between locations, using proportions of total daily traffic volume (i.e., share of daily CV vol per time period) are a more appropriate representation of a daily flow profile than actual commercial traffic volumes.

No pre-existing categories (i.e., labels) can be used to validate the pattern groups. Therefore, an unsupervised machine learning approach is proposed for this analysis. Clustering is the most common expression of problems aiming to group unlabeled training data based on similarity and exploring attribute variations within and between the groups. By implementing a clustering approach, this research aims to detect correlations and group similar flow patterns (hourly volume variations) while trying to account for the effects of spatial correlation (i.e., the mutual interference between objects due to their spatial proximity) (Maimon and Rokach 2010).

This research methodology includes four main phases (see Figure 4-3). Phase 1 entails the design of the clustering procedure. Phase 2 deals applying the clustering procedure to identify a limited number of clusters that group CV Traffic daily flow profiles. Phase 3 involves characterizing and translating Seattle's CV traffic flow clusters into generalized "Typical" urban CV Daily Flow Patterns. Phase 4 entails an exploratory spatial analysis to determine the subset of attributes that relates to the identified patterns.

The remainder of this section will describe each of the steps of the implemented methodology.

4.4.1. Phase 1: Design of Clustering Procedure

4.4.1.1 Level of Aggregation

Based on the database aggregation levels should be chosen to calculate the daily flow profiles for the following features: time, CV body and CV activity. A pre-definition of the bounds needs to be set for each of these features before this analysis, as they constitute the input of the cluster algorithm.

Time Aggregation

The optimal number of time intervals describing daily traffic loads by time of day for clustering analysis (shaped based) depends on the amount and frequency of short-term variations (Weijermars 2007). When the aggregation level is too low, differences between days can be due to deference in the number of green periods or other random short-term variations in traffic volumes. On the other hand, when the aggregation level is too high, temporal variations (peak periods, activity dips, peak volumes) might be missed. For our analysis, aggregation levels if 30 mins, 1 hour, 2 hour and 3 hours are evaluated to determine the optimal level for the daily flow profiles input in the clustering algorithm. Our analysis showed that 1-hour intervals performed best as shorter period produced clusters very difficult to represent due to the excessive variability. Longer intervals resulted in large clusters with high within-cluster variance, and thus, poor clustering.

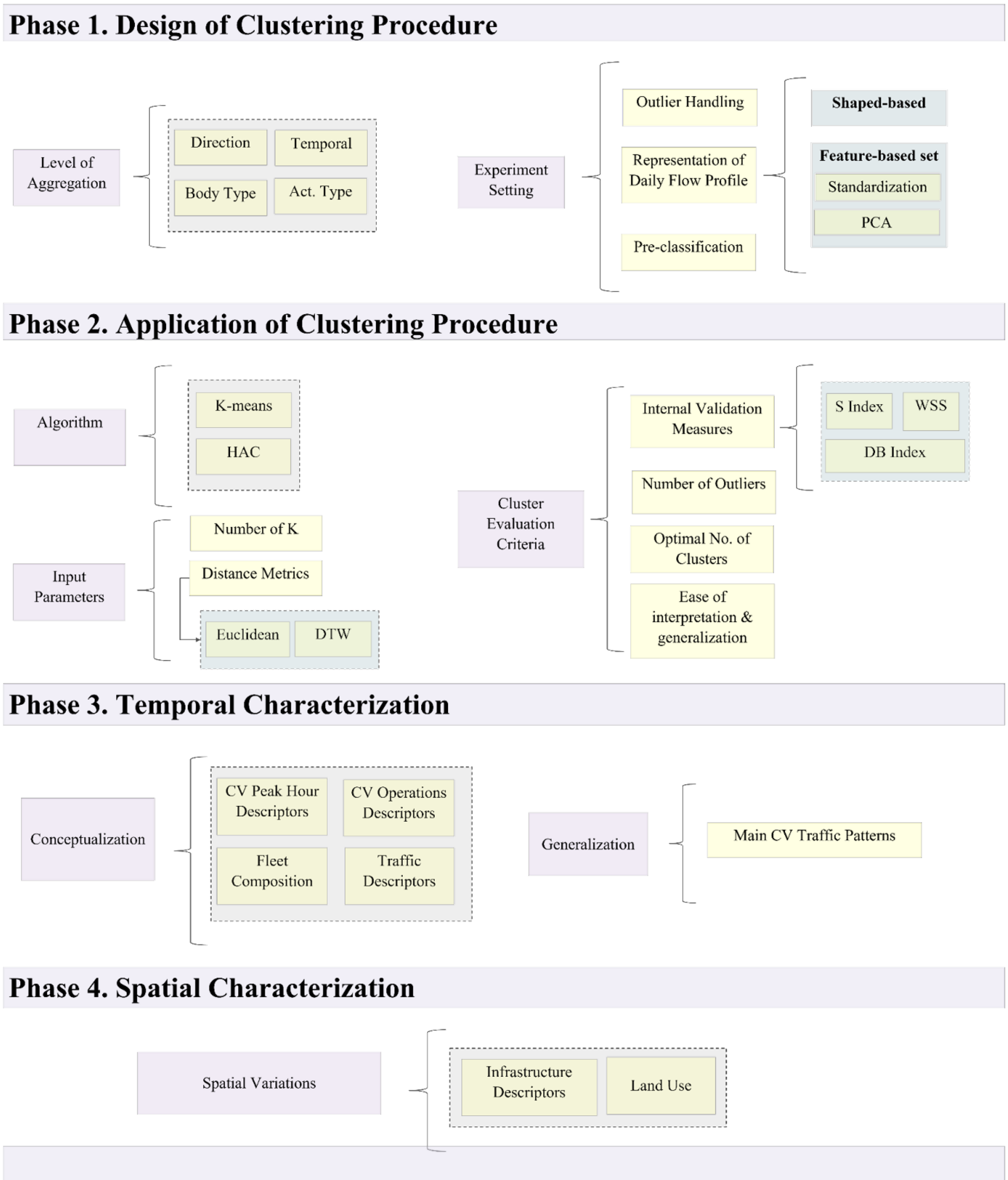


Figure 4-3 Methodology for CV Traffic Characterization.

Vehicle Body Type Aggregation

Low volume within vehicle classes can be problematic as small changes in volume within a vehicle class lead to large changes in percentage of volume (Hallenbeck et al. 1997). For example, for a daily flow profile with an average ADDT of only 30 trucks, an increase of 5 vehicles per hour represents a 15% increase in the hourly CV share. Thus, daily flow profile with low volume show excessive variability impacting the stability of the resulting clusters and their interpretation. To address this issue, when testing the impact of vehicle body on the daily temporal variations, the level of aggregation is as follows:

- **CV small fleet**, Class 3 and Class 5 vehicles that are categorized as urban delivery vans, services vans, and service pick-up trucks based on the typology described on the previous chapter;
- **Single Unit Trucks (SU),** Class 5 (not included in the previous category) through Class 7; and
- **Articulated Trucks (AT),** Class 8 till Class 13, thus, including both trailer trucks and multi-trailer trucks.

Directionality Aggregation

The actual pattern on a street segment may be directionally oriented, with CV movements in one direction during some parts of the day and in the opposite direction during other parts. The Seattle data shows that, aggregating by direction and not considering the bidirectionality of flow on the street segments, CV directional patterns at specific locations are hidden. Thus, our methodology considers a level of traffic flow aggregation by direction.

4.4.1.2 Experiment Settings

Different clustering results can be obtained by varying experiment settings related to pre-classification, daily flow profile representation, and handling outliers.

Infrastructure Type Pre-classification

The conditions of traffic are very different between ramps and street intersection legs. Traffic flow in ramps has only one direction - coming to/from - a major interstate corridor. At street segments adjacent to an intersection, the traffic can be bidirectional and affected using a more diverse mix of users that cross the intersection. Pre-classification can help the clustering algorithm to form tighter groups by filtering out the difference between these two types of infrastructure before applying cluster analysis. Previous studies of highway flow patterns have executed pre-classification using calendar or site-specific data (Weijermars and Van Berkum, 2005b; Weijermars and Van Berkum, 2005a; Wild, 1994; Chung, 2003). Once multiple clustering with pre-classification are obtained, a smaller set of clusters can be obtained by collapsing similar clusters Weijermars and Van Berkum, 2005b)

Outlier Handling

To address outliers in the clustering results, all small clusters that are hard to profile by means of the cluster-dependent distributions are categorized as outlier. A significant share of outliers discover during the cluster procedure were related to low ADDT daily flow profiles. A sensitivity analysis of minimum AADT for 50 and 100 is conducted to find the optimal threshold that minimizes the presence of outlier while preserving the identification of the CV Daily Flow Patterns in our data.

Representation of Daily Flow Profile

To run the cluster analysis, the Daily CV flow profiles (DFP) need to be defined mathematically to describe their height and shape (Weijermars and Berkum 2005). Two options are evaluated to represent each of the profiles:

- *Shaped based.* As a univariate time-series, the daily flow profile can be treated as one observation consisting of a series of traffic counts as a function of the time of the day.
- *Feature based.* To obtain more insight into the type of temporal differences, a set of flow-related attributes describe the profile. Since describing the daily flow profiles as a time series does not explicitly provide information about the type and magnitude of flow patterns between locations, this approach can help overcome this drawback. For clustering of the feature-based representation of the Daily flow profile, standardization is necessary for proper interpretation of the cluster results, as these features are not all measured on the same scale and, consequently, cannot be combined directly into one vector of features. Additionally, PCA (principal component analysis) is utilized as a tool for dimensionality reduction and feature selection since it aims to find the intrinsic dimensionality of the data (Ding & He, 2004; Chavent et al., 2017;Salamanis et al., 2017; Saha et al., 2019).

After carefully evaluating the shaped-based and feature-based approaches, a shaped-based approach is a preferred alternative, and the basis of our cluster analysis results. Theoretical and empirical reasons lead to this choice of the clustering method.

An approach that considers a feature-based description of the Daily CV flow profile requires the definition of a set of time series features. In common practice, the traffic analysis process deals with the traffic situation described with features representing an average or critical basis, including the average weekday volume and design hour volume (primarily described as the nth highest hourly volume). However, besides the average condition of a transportation network element, the variability at different times and locations is also crucial.

Hourly variations of urban CV traffic throughout the day are an under researched phenomenon for which no labeled data is available, and preconceptions of non-CV patterns may not suit well. Therefore, as this research focuses on analyzing the variation in CV daily flow profiles to identify patterns, the feature-based approach is not optimal as it:

- Loses on the variability at different locations and times and by vehicle type, which is also vital to understand the spare capacity and the spread of demand peaks to assist traffic demand strategies.
- May incorrectly assume what temporal variables best describe the variations that may inform the grouping of entities and pattern discovery.
- A shaped-based approach with richer information is hard to beat if a suitable cluster algorithm is chosen. This is in line with previous studies applying cluster analysis to understand traffic flows, as the shaped-based fits are the most frequently used, refer to table 4.1 in the literature review section above.

4.4.2. Phase 2: Application of Clustering Procedure

In its most common form, cluster analysis is based on heuristics that try to maximize the similarity between in-cluster elements and the dissimilarity between inter-cluster elements. Among the similarity-based techniques, there are two major approaches: partitional approach (e.g., k-means) and hierarchical clustering (Depaire et al., 2008). A brief description of these algorithms follows. The reader is referred to some classical text for a more complete discussion and review of existing clustering techniques (Gordon, 1999; Kaufman and Rousseau, 2005).

4.4.2.1. Clustering Algorithm

K-means clustering

K-Means clustering algorithm is a well-known unsupervised classification method and by far the most common partitional algorithm. This category of clustering approaches forms clusters by the closeness of data points to the C centroid of cluster.

The K-means algorithm partitions the dataset into a predefined number (k groups) of clusters by minimizing the distance of observations within clusters and maximizing the distance of observations in different clusters. An optimal solution to this problem minimizes the total distance between points and the center of their respective cluster. In other words, this solution finds k centers C that minimizes the potential function for any given set of training data X.

Hierarchical Clustering

The Hierarchical Clustering algorithm seeks to build a hierarchy of clusters by combining observations (agglomerative approach) or disaggregating observations (divisive approach). In contrast to K-Means, Hierarchical Clustering does not require the specification of the number of clusters initially. A required input for Hierarchical Clustering is the dissimilarity measure between groups of observations or "linkage". Typical linkage methods include Ward's method, single, complete and average linkage.

4.4.2.2. Distance Metric

We consider two different distance metrics in this research, including Euclidean distance and Dynamic Time Wrapping (DTW). Euclidean distance is a metric invariant to time shifts. Since the approach to use shape-based inputs in the cluster analysis clearly holds such variants, DTW is the chosen metric due to its ability to find the optimal non-linear alignment between *tw* time periods, in other words, ability to deal with time shifts.

4.4.2.2. Cluster Number Selection and Cluster Validation

Clustering validation evaluates the goodness of clustering results and includes two main categories: internal and external cluster validation (Liu et al., 2010). Since our application of cluster analysis is an unsupervised technique and there is no ground truth data to compare, an assessment of quantitative external validation is impossible. Still, the data distribution among all the clusters is evaluated through visualization (Weijermars, 2007).

Internal validation measures can be used to choose the optimal clustering algorithm on a dataset and the optimal cluster number without additional information. Internal validation measures are typically based on two criteria:

- Compactness: describes how closely related the objects in a cluster are
- Separation: relates to how distinct a cluster is from other clusters.

There are several widely used internal validation measures in the literature. We select three metrics that represent a mix of compactness and separation evaluation criteria:

1. Within-cluster sum of squares (WSS): The sum of distances between the observations and the corresponding centroids are evaluated. The methods help determine the optimal value of K, typically obtained as the shift point of the curve (Elbow method) and utilized in this research only for the k-mean algorithm.
2. Silhouette Index (S): considers the pairwise difference between- and within-cluster distances. Maximizing the value of S leads to optimal clustering.
3. Davies-Bouldin Index (DB): measures the similarity between clusters. The smaller the index is, the better the clustering result is.

Additionally, the following aspects are documented for each cluster model run and are part of our model selection and validation process.

- The number of outliers (O): defined as the number of small clusters that we defined as clusters with three or fewer daily flow profiles. Clusters typically indicate outliers and are hard to profile using cluster-dependent distributions.
- The optimal number of clusters (K): a parsimonious model achieves the desired level of explanation with as few clusters as possible.
- Ease of interpretation and generalization of the resulting clusters.

4.4.2.3. Programing Implementation

Python 3.7 is used for data wrangling and machine learning implementation for this research. Scikit-learn provides a set of python modules for machine learning and data mining. Specifically, we use the following modules: PCA, AgglomerativeClustering (Hierarchical Clustering), KMeans, davies_bouldin_score, silhouette_samples, silhouette_score, StamdardScaler (Standarization). We also use modules in the Scipy library, including bartlett (Bartlett's Test of Sphericity), dendogram, linkage. amd the dtw module in the dtaidistance package to compute DTW for shaped-based hierarchical clustering. Appendix C includes the Python script for this research.

4.4.3. Phase 3: Cluster Characterization

Following the work of Depaire et al. (2008), we label each cluster based on the distribution of variables conditional on each cluster. For categorical variables we select the highest conditional probability obtained for a determined category of a variable given its membership to a specific cluster. For continuous variables, we use the median value. Different variables were considered for characterization in relation to:

- CV Peak Hour Descriptors
- CV Operations Descriptors
- Fleet Composition
- Traffic Descriptors

A complete list of the features of the variable used for the differentiation of clusters is listed on Table 4-2 in the Results section. After characterizing the clusters resulting from the model, this analysis assesses the minimum number of patterns required to adequately describe the daily variation of CV Traffic Flows for the city of Seattle and propose a number of Main CV Daily Flow Patterns.

4.4.4. Phase 4: Spatial Characterization

Analyzing spatial variations of traffic volumes without taking the temporal component into account results in a characterization of location only by the magnitude of traffic volume at a certain time interval (high, medium vs low volume locations). The results of these analyses are location specific and do not provide more insight in urban traffic in general. In contrast, the analysis spatial variations in combination with temporal variations can provide additional insights into the interaction of CV flows and the infrastructure characteristics.

Once typical CV Daily Flow Patterns are identified and characterized based on temporal features, they are evaluated against spatial properties to label the clusters following a similar process than described for the temporal characterization to:

1. Analyze the underlying feature relations among road network features and typical traffic flow patterns
2. Draw location-dependent conclusions about the usage and CV Daily Flow Patterns.

Different spatial variables were captured through ArcGIS software and considered for characterization in relation to:

- Infrastructure Design including infrastructure type (ramps vs street segments) and number of lanes.
- Infrastructure Designation, including Heavy-Haul network, Arterial Designation, Street Designation and Freight Network Designation.
- Topological Features, including Urban Village designation, key freight generators, adjacent land use.

4.5 Results and Discussion

4.5.1 Selected Clustering Model

We follow the process described in Figure 4.3 of the methodology section to find the best model. We calibrate a total of 17 models considering variations of the clustering algorithm (i.e., K-Means and Hierarchal), pre-classification (i.e., infrastructure type and body type), and minimum CV ADT to handle outliers (i.e., 50, 100). Appendix B shows the internal validation result comparisons between the models. The final model considers a total of **17 clusters** and considers the following model input and parameters:

a. Disaggregation by location, flow direction, and body type (i.e., smaller fleet, articulated trucks, and single-unit trucks_.
b. Pre-classification of the flow daily profiles by infrastructure type (ramps vs. street segments);
c. Pre-classification of the street segments' daily flow profiles based on the vehicle body, meaning articulated trucks and non-articulated trucks (i.e., single-unit trucks and smaller fleet);
d. A minimum of 50 veh./day to account for the high variation related to low CV ADT volumes;
e. Hierarchical clustering algorithm.
f. Euclidean distance metric; and
g. Ward linkage method.

After carefully considering the multiple competing models, shaped-based hierarchical clustering with ward linkage is the preferred model for all the CV daily flow profile partitions resulting from pre-classification. The selected cluster is based on the proposed multi-criteria clustering evaluation approach described in section 4.4.2.2. A few general conclusions result from the comparison between cluster algorithms.

Hierarchical clustering with complete linkage tends to show better internal validation indicators (i.e., smaller DB and greater S). However, the result is challenging to characterize because it includes one or two large clusters and many smaller ones (outliers). For example, the result for the ramp pre-classification with hierarchical clustering and complete linkage method suggests 8 clusters with a local minimum of DB = 0.786, smaller than HR Ward (1.189) and K-Means DTW (1.822), but 6 out of the 8 clusters show three or fewer DFPs (see Figure 1-3 in appendix C). Most of the time, K-Means showed the poorest internal validation scores for DB and S (highest and lowest, respectively) regardless of the number of clusters among the three clustering approaches.

All cluster algorithms showed an increasing number of outliers (small clusters) with an increasing number of clusters (K). As described in the methodology section, the presence of these outliers can be partly explained by the excessive variability of some DFPs due to the low CV ADT. To test for the impact of low volumes for the Hierarchical Ward Linkage cluster model, we run an additional test with a threshold of a minimum of 50 and 100 CV ADT per profile for each of the three pre-classifications (i.e., ramps, street intersections - articulated trucks and street segments small fleet / single-unit trucks). This analysis generally shows a reduction in the number of outliers (clusters of three DFPs or less) and a better representation of the resulting clusters for all pre-classification groups (see Figures 4-6 in the appendix C). A 50 veh/hr and 100 veh/hr minimum cut-off thresholds for CV ADT result in dropping 89 DPFs (23.5% of all DFPs) and 135 DPFs (35.7% of all DFPs), respectively, most of which are articulated truck DFPs. To put this number into context, articulated trucks represent 9.9% of total CV volume between all locations in the study. A value of a minimum of 50 for the CV ADT of the DPFs is selected as the best cut-off because it improves the clustering result while reducing the number of DFPs lost.

Figure 4-4 below shows internal validation metrics, Davies-Boulding and Average Silhouette index, and the number of outliers for the final clustering result. As described in the methodology section, the smaller the Davies-Boulding and the greater the Average Silhouette index are, the better the clustering result.

Selecting the best cluster result is based on a multi-criteria and may differ from what a specific index indicates for an optimum result for the same experiment. For example, a Hierarchical ward linkage cluster of street segment articulated trucks DPFs showed a local minimum DB value at 6 clusters (0.985) (see Figure 4-4), and a local Average Silhouette Index maximum at 7 clusters. Our analysis shows that increasing the number of clusters to 7 for the street segment articulated trucks pre-classification improves the representativeness of the patterns. It is also worth noting that for the three pre-classifications, the number of outliers drastically increases after about 10 clusters (see Figure 4-4).

The selection process of the number of clusters based on internal validation indices resulted in 7 clusters for ramps, 8 for articulated truck DPFs on street segments, and 9 for smaller fleet / single-unit trucks on street segments.

Following Saha (2019) and Weijermars (2007), after selecting the best number of K for each pre-classification cluster model, we further examine the cluster results to determine if further clustering and/or merging of clusters is needed. Also, some clusters with similar temporal characteristics and from different pre-classifications are merged. The resulting number of clusters is 17.

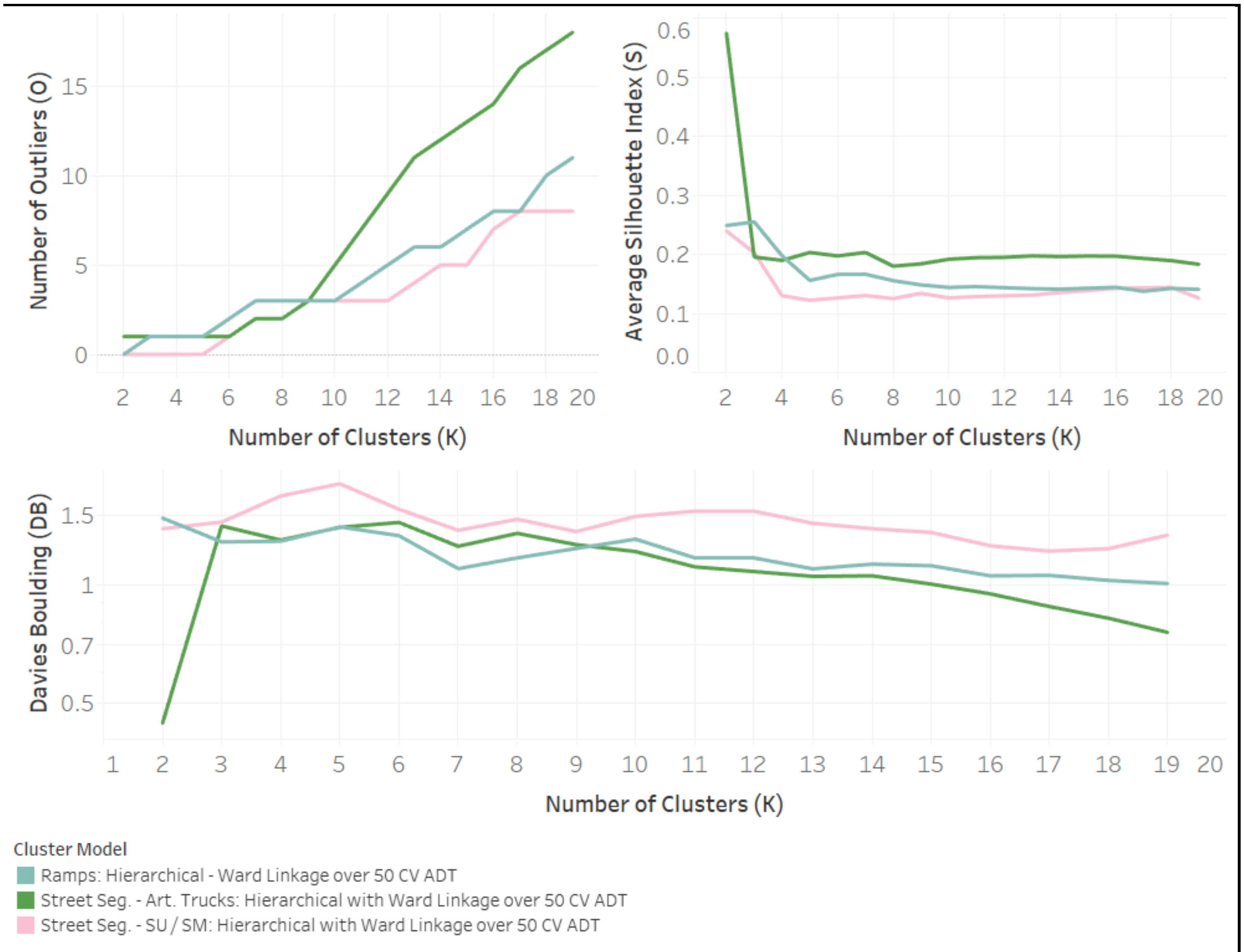


Figure 4-4. Internal Validation Metrics of Final Cluster Models with Pre-classification by Infrastructure and Vehicle Body Type

4.5.1 Conceptualization of CV Daily Flow Patterns

The final set of features used for profiling the clusters is listed in Table 4-2 related to:

- *CV Peak Hour Descriptors:* peak hour, Number of humps and traffic intensity
- *CV Operations Descriptors:* Interaction with General traffic, Activity Dip and Off-hours activity

- *Fleet Composition:* Body Type, Activity Type, FHWA vehicle classification
- *Traffic Descriptors:* Volume description, hump width, skewness and hump standard deviation

Table 4-2. Attributes for feature-based Daily CV Flow Profile description.

Name	Temporal-related Features	Definition	Characterization Definition	Characterization
Hump Description	Hump Width (hr)	Continuous Time interval with 7% CV hr. share during the Peak Hour	-	Narrow (<= 4 hrs.) Wider (> 7 hrs.)
	Hump Std (%)	Standard deviation of the Hump during the Peak	Characterization is based on the metric Distribution based on Quartiles	Smooth (< =1.5) Intermediate (1.6 - 2.9) Uneven (> 3.0)
	Skewness	Center of Mass [measure of asymmetry of the distribution]	-	AM Skew (< 11:00 AM) No Skew (11:00 AM - 12:00 PM) PM Skew (> 12:00 PM)
Volume Description	CV ADT (CV/day)	Daily flow profile ADT	Characterization is based on the metric Distribution based on Quartiles & ADT Threshold defined by Seattle's Freight Network	High (500) Medium High (250 -500) Medium Low (150 – 250) Low (150)
	Daily Variation	Standard deviation of Daily Flow Profile	Characterization is based on the metric Distribution based on Quartiles	High (> 4.5) Medium (3.8 - 4.4) Low (< 3.7)
Interaction with General Traffic	Difference in CV and non-CV Peak times (hr)	The minimum distance to a non-CV Peak (i.e., peaks higher than 6.5% share of non-CV ADT)	-	During (< ± 1 hr) Shoulder (±1 - ±2 hr) Spare Capacity (± 2.1 - ± 4 hr) Different Period (> ± 4hr)
	Non-CV Peak Time (hr.)	Peak time of non-CV		
Activity Dip	Time of occurrence (hr.)	Drop of CV hr. share of at least 7% between a double peak daily flow profile	-	Lunch Break (Time of occurrence = noon) Am PV Peak (Time of occurrence = non-CV AM Peak)
Off-hour Activity	Off-hours CV Vol Share (%)	Share of CV ADT during 12:00 AM – 5:45 AM & 6:25 PM – 11:45 PM	-	High (=> 50 %) Medium (50 -20%) Low (<= 20%)

Peak Composition	Number of Humps	Number of time intervals with CV hourly share more than 7%	-	Multiple / Double / Single
	Peak Hour Time	Peak Hour based on rolling hour distribution	-	Early AM Peak (6:00 AM – 8:45 AM) Midday Peak (9:00 AM – 12:45 PM) Early PM Peak (1:00 PM – 3:00 PM) PM Peak (3:15 PM– 6:00 PM) Off-hours Peak (12:00 AM – 5:45 AM & 6:25 PM – 11:45 PM)
	Traffic Intensity	Peak Hour Volume (%) based on rolling hour distribution	Characterization is based on the metric Distribution based on Quartiles	High (> 14.8 %) Medium (14.7 - 11.9%) Low (< 11.8%)
Fleet Composition	Share by CV Body Type (%)	Small CV Fleet, Single Unit, Art. Truck	Relative Frequency of the categories within the cluster	-
	Share of FHWA Vehicle Category (%)	Class 4 through Class 13	Relative Frequency of the categories within the cluster	-
	Share of Activity Type	Waste Management, Construction, good transport, service	Relative Frequency of the categories within the cluster	-

The values for each of these variables for each of the 17 clusters can be found in Table 4-3. Additionally, the definition of these metrics supports the generalization of the resulting clusters on seven Main CV Daily Flow Patterns in urban areas. These patterns provide useful and generalizable insights for urban freight transportation. Each of the patterns and the main insights of their temporal and spatial classification are described in the following sections below:

- Main CV Daily Flow Patterns
- Effect of Vehicle Body Type and Activity Type
- Geospatial and Infrastructure Differences
- Directional Differences

Table 4-3. Attributes for feature-based Daily CV Flow Profile description.

Pattern Number	Pattern Name	Cluster Number	Peak Configuration			Traffic Descriptors			CV Op. Descriptors			Fleet Composition		
			Median Max. PKH Vol (%)	**Median PKH (hr)**	**Median No. of Humps**	**Median Hump Width (hr)**	**Median CV ADT (veh/day)**	**Median Hourly Std. (%)**	**Median Off-hour Volume Share (%)**	**Dip median (%)**	**Dip median Std. (%)**	**AT (%)**	**SM (%)**	**SU (%)**
1	**Business Hours (BH)**	**1**	12.4 (1.3)	10.1 (1.3)	1.0 (0.2)	7.0 (0.7)	300 (246)	4.1 (0.4)	7.9 (2.8)	2.4	(1.5)	0	**59**	41
		2	11.5 (1.4)	12.4 (1.7)	1.0 (0.8)	7 (2.0)	430 (248)	3.8 (0.2)	13.1 (3.1)	2.8	(2.2)	0	**62**	39
		3	12.2 (0.2)	11 (0.8)	1.0 (0)	7 (1.1)	172 (57.)	3.5 (0.3)	22.8 (2.6)	0.9	(0)	**67**	0	33
2	**Partial Business Hours**	**4**	13.9 (1.7)	11.1 (1.2)	1.0 (0.4)	6 (1.5)	174 (182)	4.3 (0.4)	9.8 (3.2)	2.1	(2)	**87**	0	13
		5	11.9 (1.5)	10.4 (1.7)	1.5 (0.6)	5 (0.9)	283 (363)	3.8 (0.4)	13.3 (4.5)	2.6	(1.1)	0	14	**86**
		6	15.7 (1.6)	9.6 (0.6)	1.0 (0.4)	6 (1.0)	189 (82.)	4.6 (0.3)	10.3 (3.9)	2.6	(1.1)	0	40	**60**
		7	14.5 (2.1)	9.8 (1.6)	**2.0 (1.0)**	4 (2.1)	**65 (49)**	4.3 (0.6)	13.0 (7.8)	**4.5**	**(1.3)**	3	**50**	**47**
3	**AM Distribution Peak**	**8**	14.0 (4.1)	8.4 (1.1)	1.0 (0.6)	6 (1.4)	393 (339)	4.3 (0.5)	10.6 (4.8)	2.2	(1.9)	17	30	**53**
		9	11.2 (1.0)	7.6 (2.2)	1.0 (0.7)	4 (2.1)	317 (134)	3.0 (0.5)	27.3 (7.4)	2.2	(1.3)	0	33	**67**
4	**PM Distribution Peak**	**10**	11.5 (2.8)	15.3 (1.0)	1.5 (0.6)	6 (2.0)	586 (505)	3.7 (0.4)	13.9 (6.4)	1.6	(3.1)	46	**54**	0
		11	12.5 (1.7)	13.0 (0.7)	1.0 (0.6)	7 (0.6)	311 (227)	3.8 (0.3)	15.1 (4.2)	1.8	(0)	0	**87**	13
5	**Activity Dip**	**12**	12.7 (1.7)	6.5 (2.8)	**2.0 (0.4)**	3 (1.7)	234 (152)	3.4 (0.3)	25.5 (3.3)	**4.5**	**(0.9)**	**100**	0	0
		13	14.1 (1.5)	10.0 (1.4)	**2.0 (0.7)**	5 (2.7)	152 (134)	4.3 (0.4)	7.8 (5.2)	**3.7**	**(1.8)**	0	33	**67**
		14	14.7 (1.5)	11.0 (1.9)	**2.0 (0.5)**	4 (1.2)	**613 (287)**	5.1 (0.6)	3.5 (5.6)	**6.5**	**(4)**	0	**73**	27
6	**Through Truck-Traffic**	**15**	12.1 (1.0)	10.5 (1.8)	1.5 (0.5)	3 (1.1)	**100 (33.)**	2.6 (0.3)	32.6 (2.5)	**4.2**	**(1.7)**	**100**	0	0
7	**Night-Time Operations**	**16**	11.9 (0.1)	21.3 (0)	1.0 (0)	7 (1.4)	**95 (1.4)**	4.1 (0.1)	84.4 (2.3)	-	-	**100**	0	0
8	**Intermittent Operations**	**17**	17.8 (3.2)	8.8 (3.7)	**3.0 (0.6)**	3 (1.3)	**83 (19.)**	4.5 (0.6)	19.8 (11.)	**6.7**	**(5.6)**	0	**50**	**50**

4.5.3 Main CV Daily Flow Patterns

The descriptive cluster analysis in this study characterizes Seattle's CV flow and provides site-based insights into its temporal variations and the nature of body type, directionality, and infrastructure type. In order to leverage this information to extract generalizable insights for other urban areas, the set of 17 clusters can be generalized into a smaller set of eight Main CV Daily Flow Patterns that explain the temporal and spatial variability of Seattle's case. The specifics of the curve's skewness, the peak's sharpness, and the exact location of the peak travel periods vary somewhat from cluster to cluster. Table 4-4 provides an overview of the main patterns, which clusters fall within them, their aggregated size, and the application of the definitions introduced in Table 4-2 above to facilitate the interpretation. Additionally, Figure 4.5. shows the resulting clusters aggregated by the proposed Main CV Daily Flow Patterns; and Figure 4.6 shows the distribution of DFPs in each cluster by CV vehicle type.

The eight main patterns can be described as follows:

- **Pattern 1**. **Business Hour (BH) Hump.** The clusters that follow this pattern showed a characteristic broad hump (i.e., 7 hrs. or more) and a lower level of variability during the hump hours (i.e., smooth, or intermediate) and the 24-hr. period (i.e., medium or low). Most CV traffic starts and ends during the regular extended business hours and peaks at a period of spare road capacity outside the non-commercial vehicle traffic peak hours. The clusters that followed this pattern and their share of CV daily flow profiles (DFP) include Cluster 1 (17%), Cluster 2 (9%), and Cluster 3 (1%), which follow variations that make them distinct. Cluster 1 shows no skewness and earlier peak during the late AM period compared to Clusters 2 and 3, which peak midday. Cluster 2, on the other hand, shows a low traffic intensity (compared to the medium intensity of the other two clusters) and is skewed towards the PM period. Last, Cluster 3 shows an AM skewness due to a higher share of volume during the early morning hours and a higher hump variability (i.e., intermediate). Concerning the fleet composition, Cluster 1 and 2 mainly include smaller fleets (~60%) and some single-unit trucks (~40%) DFPs, and Cluster 3 showed mostly articulated trucks (~67%) and some smaller fleets (~33%) DFPs.
- **Pattern 2. Partial Business Hour Hump**. This pattern is like Pattern 1 but with a narrower hump (i.e., less than seven hrs.). Four clusters follow this cluster which represent 35% of the CV DFPs, including Cluster 4 (12%), Cluster 5 (10%), Cluster 6 (8%) and Cluster 7 (5%). Clusters 4 and 5 show peak CV activity outside and between the non-CV peak hours (CV peaks during midday and late AM, respectively). Cluster 6 shows the highest traffic intensity

and variability throughout the day among the four clusters. Cluster 6 is also the only cluster skewed (towards the AM) and shows the largest separation from the non-CV peak hour. Cluster 7 shows the narrowest hump width and the highest within-hump hourly variation. CV DFPs in Cluster 7 tend to peak closer to the non-CV peak and during its shoulder hours. CV fleet configuration differs among these groups, with Cluster 4 showing a mix of smaller fleet and single-unit truck profiles and Cluster 5 including mostly single-unit trucks and some smaller fleet profiles. Clusters 6 and 7 mostly include profiles of one vehicle type; articulated trucks and single-unit truck profiles, respectively.

- **Pattern 3. Distribution AM Peak.** Pattern 3. Distribution AM Peak. Includes CV DFPs with a single AM skewed peak that reaches maximum hourly volume between 6 and 9 am and simultaneously with the non-CV peak. Cluster 8 (10% of CV DFPs) and Cluster 9 (3% of CV DFPs) follow this pattern. Cluster 9 shows a narrower hump (vs. medium width), lower level variability during the day, and the hump than Cluster 8. Cluster 9 also shows a higher share of off-hours activity, mostly during the early AM period. Both clusters mostly show single-unit trucks (~53%-67%) and some smaller fleet profiles (30-33%), but Cluster 8 also shows a few articulated truck profiles (~17%).
- **Pattern 4. Distribution PM Peak.** Describes a PM skewed single-peak pattern with the latest CV peak hours among all clusters (2 - 6 pm). Two main variations are included in this pattern. Cluster 10 (5% of CV daily DFPs) shows a lower hourly volume variability during the day and hump and tends to occur on the shoulder hours of the non-CV PM peak. Cluster 11 (4% of CV DFPs) shows a wide hump width and peaks during the period of spare capacity between the non-CV peaks. Smaller fleet DFPs are the most frequent in Cluster 10 (~87%), and Cluster 11 shows a mix of the smaller fleet (~54%) and single unit trucks (~46%).
- **Pattern 5. Activity Dip.** Consist of two peaks throughout the day in the same direction. The valley between the peaks is observed in either of these periods: a) 7- 8 am, associated with a decline in operations trying to avoid the commuter AM peak, and b) noon, associated with the lunchtime break. Three clusters follow the activity dip pattern, including Cluster 12 (4% of DFPs), Cluster 13 (4% of DFPs), and Cluster 14 (2% of DFPs). Clusters 13 and 14 show the noon lunch break. However, Cluster 14 shows a narrower hump width and a higher level of hourly variability during the hump (i.e., uneven vs. intermediate) and throughout the day (i.e., high vs. medium). On the other hand, Cluster 12 shows an activity dip during the non-CV AM peak and a higher share of off-hours activity, mainly during the early AM. Concerning CV

fleet composition, most DFPs in Cluster 12 are smaller fleets (~73%) and some single units (~27%), which are inverted in Cluster 13 (~67% smaller fleet and ~33% single unit). Cluster 14 is entirely articulated truck DFPs.

- **Pattern 6**. **Through Truck-Traffic**. This pattern reflects the fact that long-distance trucks are not constrained to the regular business day. Many travel at night to avoid traffic congestion; others drive whenever possible, as most drivers are paid by the mile rather than by the hour. Thus, the percentage of daily travel at night by these vehicles is much higher than for other vehicle classes. Cluster 15 (1% of DFPs) represents this pattern, which has a fairly flat volume distribution (i.e., low hourly variability during the day). This pattern shows a midday peak of medium intensity, probably due to a mix of through traffic with other CV travel demands during business hours.
- **Pattern 7. Nighttime Operations.** This cluster shows the highest off-hours activity (mainly during the late PM) of all clusters and a wide hump during the night hours. This pattern includes a small but distinct set of DFPs in Cluster 16 (1% of DFPs), representing nighttime demand for CV activity. Only articulated truck DFPs were grouped in this cluster and associated with construction activities.
- **Pattern 8. Intermittent Operations**. This pattern refers to narrow multi-humps DFPs with high-intensity CV peaks resembling a "saw-tooth" pattern. This cluster showed high variability in hourly volumes throughout the day and simultaneous CV and non-CV peaks. This cyclical short-frequency pattern is related to intermittent operations such as ferry unloading. This cluster formed a mix of single-unit trucks and smaller fleet DFPs (50% split).

It is worth noting that 4 DFPs (<1.4%) are considered outliers and not included in any of the abovementioned patterns. The cluster models group these DFPs into three clusters and are very difficult to interpret.

Table 4-4 Main CV Daily Flow Patterns Descriptors

Pattern Number	Pattern Name	Cluster Number	Size of Clusters		Peak Configuration		Traffic Descriptors				CV Operations Descriptors			Fleet Composition
			Count of DPFs	Share	Peak Period	Traffic Intensity	Peak Hump Descriptor	Hump Std	Skewness	Daily Variations	Off-Hours Activity	CV vs. non-Cv PKH (hr)	Activity Dip	Share by CV Body Type
1	**Business Hours (BH)**	**1**	49	43%	Late AM Peak	Medium	**Wide**	**Smooth**	No Skew	Medium	Low	Spare btw peaks	-	Mosly SM and some SU
		2	26	23%	Midday Peak	**Low**	**Wide**	**Smooth**	**PM Skew**	Medium	Low	Spare btw peaks	-	Mosly SM and some SU
		3	3	3%	Midday Peak	Medium	**Wide**	Inter.	**AM Skew**	**Low**	**Medium (most of during Early AM)**	Spare btw peaks	-	Mostly AT and some SM
2	**Partial Business Hours**	**4**	34	30%	Midday Peak	Medium	Medium	Inter.	No Skew	Medium	Low	Spare btw peaks	-	AT
		5	30	27%	Late AM Peak	Medium	Medium	Inter.	No Skew	Medium	Low	Spare btw peaks	-	SU
		6	22	19%	Late AM Peak	**High**	Medium	Inter.	**AM Skew**	**High**	Low	**Different Period**	-	Mosly SU and some SM
		7	15	13%	Late AM Peak	Medium	**Narrow**	**Uneven**	No Skew	Medium	Low	**Shoulder (+)**	-	Mix of SM and SU
3	**AM Distribution Peak**	**8**	30	27%	**AM Peak**	**Medium with high Variation**	Medium	Inter.	**AM Skew**	Medium	Low	**During**	-	Mostly SU and some SM and AT
		9	9	8%	**AM Peak**	**Low**	**Narrow**	**Smooth**	**AM Skew**	**Low**	**Medium (most of during Early AM)**	**During**	-	Mostly SU and some SM

4	**PM Distribution Peak**	**10**	15	13%	**PM Peak**	**Low with high Variation**	Medium	**Smooth**	**PM Skew**	**Low**	Low	**Shoulder (+)**	-	Mix of SM and AT
		11	13	12%	**PM Peak**	Medium	**Wide**	Inter.	**PM Skew**	Medium	Low	Spare btw peaks	-	SM
5	**Activity Dip**	**12**	11	10%	**AM Double Peak**	Medium	**Narrow**	Inter.	No Skew	**Low**	**Medium (most of during Early AM)**	**Shoulder**	**AM PV Peak Hr.**	AT
		13	12	11%	Late AM Peak	Medium	Medium	Inter.	No Skew	Medium	Low	**Different Period**	**Lunch Break**	Mostly SU and some SM
		14	7	6%	Late AM and PM Peak	Medium	**Narrow**	**Uneven**	No Skew	**High**	Low	**Spare & Shoulder**	**Lunch Break**	Mostly Smaller Fleet and some Single Unit Trucks
6	**Through Truck-Traffic**	**15**	4	4%	Late AM Peak	Medium	**Narrow**	**Smooth**	No Skew	**Low**	**Medium**	**Different Period**	-	AT
7	**Night-Time Operations**	**16**	2	2%	**Off-hours Peak**	Medium	**Wide**	Inter.	**PM Skew**	Medium	**High (mostly during PM)**	**Different Period**	-	AT
8	**Intermittent Operations**	**17**	6	5%	**Multiple Hump**	**High with High variation**	**Narrow**	**Uneven**	No Skew	**High**	Low	**During**	**High Variation**	Mix of SM and SU

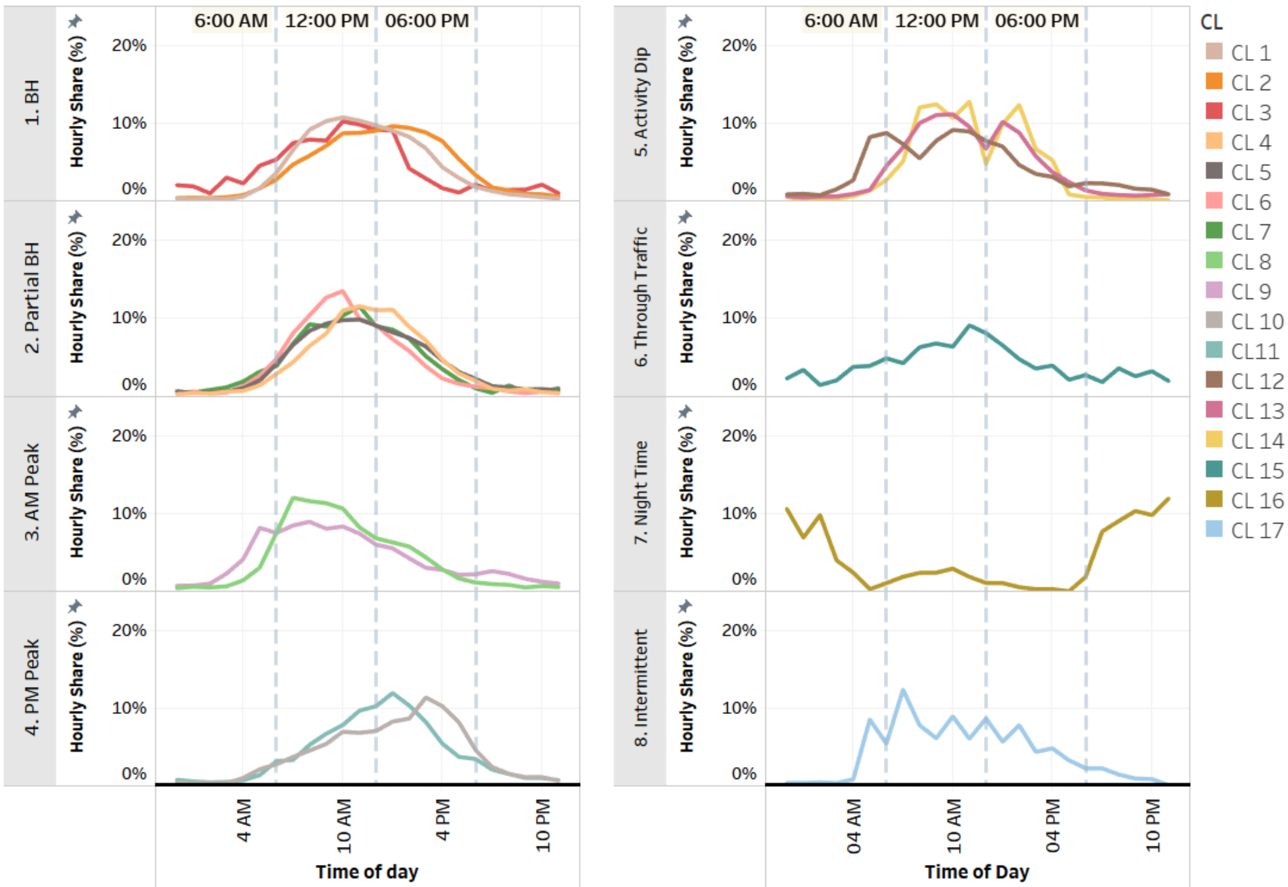


Figure 4-5. Clusters center lines grouped by Main CV Daily Flow Patterns.

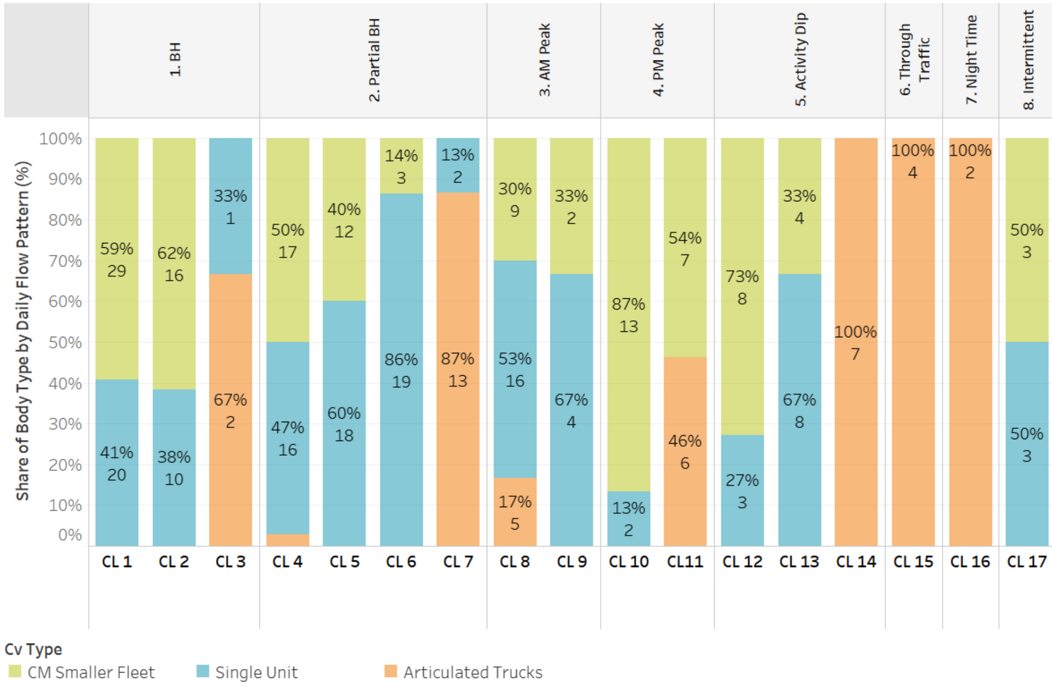


Figure 4-6. Distribution of DFPs in each cluster by CV vehicle type

4.5.4 Effect of Vehicle Body Type

The clustering analysis helps to uncover the influence of CV body type in the daily temporal variations of commercial flows in the urban area. The variations are mainly related to the dominant traffic pattern, traffic intensity, and peak-hour and off-hour CV volume.

As indicated in Figure 4-7, there are dominant patterns for each vehicle's body type:

- Small Fleet dominant pattern is the Business Hours Hump Pattern (37% of small fleet DFPs) but also shows a significant share of the Partial-Business Hour pattern (26%).
- 45% of single unit trucks DFPs are grouped in the Partial-Business Hour pattern and followed by Business Hour (25% of SU DFPs).
- The most frequent pattern among articulated trucks PDF is the Partial-Business Hour pattern (35% of articulated trucks DFPs) and followed by Activity Dip (18%).

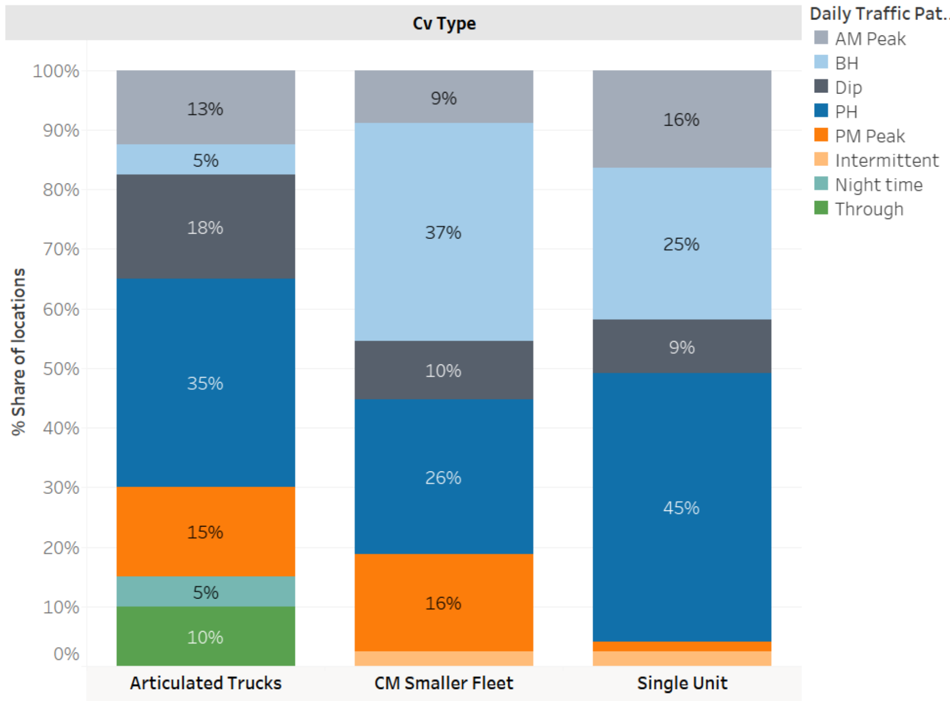


Figure 4-7. Distribution of Daily Flow Patterns by vehicle type and location.

Additionally, as described in the section above, several Main CV Traffic Patterns show a dominant CV body type. For instance, single-unit trucks represent 55% of all the DFPs falling in the Partial-Business hump pattern variations (Clusters 4-7) (see Figure 4-7 above). At the same time, vehicle body types can follow different patterns that may be characteristic of another vehicle type. These variations can be observed spatially and are described in the sections below.

The effect of the vehicle body type can also be observed on the traffic intensity during the peak hour, as the bigger the vehicle size, the higher probability of high traffic intensity during the peak hour (see Figure 4-8). 37% of small fleet DFPs have a low traffic intensity during peak hours (below 11.8%). In contrast, only 8% of articulated trucks fall into that category.

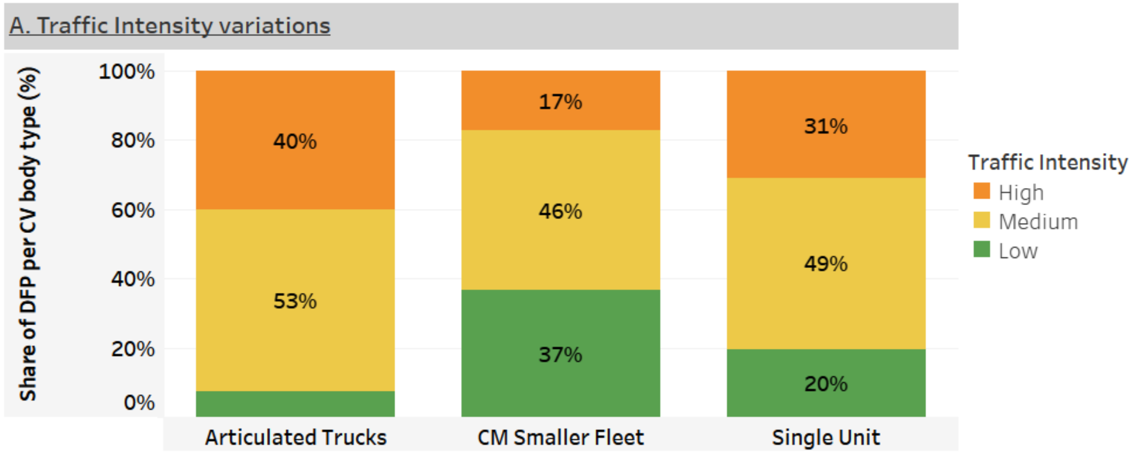


Figure 4-8. Traffic Intensity by Vehicle Body Type.

Temporal variations are also observed during the non-CV peak hour. Only 12% of all evaluated DFPs have a peak hour between 1:00 PM and 2:45 PM. The percentage is even lower between 3:00 PM and 6:00 PM (5%). This is in line with the (Weisbord and Fitzroy, 2010) depiction of the PM non-CV traffic congestion impacts on commercial hours of operation. They indicated that "the growth of PM peak traffic congestion has reduced, and in many cases even eliminated, late afternoon stock/merchandise deliveries in some larger urban areas." This has pushed transportation and distribution businesses to define starting times for their operation into the early morning hours for some. Even more so, large cities are implementing restrictions after 3 PM that impact business operations (irrespective of the sector). For example, in Seattle, CVLZs at the curb are restricted during peak hours for both AM and PM periods.

Based on the data set, it is concluded that DPFs with the latest peak primarily correspond to small fleet, representing 75% of the ones peaking between 3 PM and 6 PM and 50% of the ones peaking between 1 PM - 3 PM (50%). This finding may be explained by the extended delivery routes related to the increase in urban deliveries. Classes 3 and 5 typically perform urban goods delivery and service operations.

Finally, only 14% of all DPFs show medium or high volumes (> 20% of CV ADT) during off-hours, that is, 6 p.m. to 6 a.m. Among those, two main patterns are observed:

- Only articulated trucks DFP are associated with high and medium share off-hours traffic in street segments.

- Inbound GDA small fleet and single-unit trucks in the I-5 ramps show medium off-hours activity during early AM hours between 3-6 AM before the CV peak occurs.

4.5.5 Geospatial and Infrastructure Differences

4.5.5.1. SoDo in Greater Duwamish MIC

All street segments located in the southern boundary of the GDA cordon are located within or in the boundary of SoDo, which is one of the four neighborhood that makes up part of the Greater Duwamish MIC. Much of this land is designated as industrial (80.4%), including warehousing, distribution and intermodal. Notable uses include container port terminals, logistics and distribution, BNSF and UP railyard that support intermodal operations, UPS and Fedex distribution centers.

In this area, the highest share of waste management vehicles is observed, which may be explained by the Republic Services recycling facility at S Lander St, west of 3rd Ave S.

One of the main patterns observed in this area was related to intermodal flows. Intermodal containers that are not loaded on trains in a terminal are drayed to near-dock intermodal yards. Containers may also be trucked to a local warehouse or distribution center, repackage from an ocean-going 20 or 40-foot container to a 53-foot domestic container. The most concentrated Port truck trip volumes are between the container terminals and BNSF's SIG and UP's Argo intermodal rail terminals (SDOT, 2015). Data was collected in the access of Port of Seattle's marine container Terminal 46, which is a containership terminal operated by Pacific Terminal Service Company (PTSC) during the time of data collection. Additionally, this data also reflects the movements related to the intermodal yards, particularly for SIG yard, which is in proximity to the marine terminal and a few facilities other intermodal facilities along Holgate St.

Seaport Connectors. A Multi-peak pattern with most of the daily activity concentrated between 8:00 am and 4:00 pm describe the heavy truck traffic along the SoDo's marine terminals and nearby rail facilities (Activity Dip pattern). This pattern averages two and three peaks occurring at 9 am, 10 am, and 2 pm, with an activity dip at noon. The dip point can be explained by break times for lunch or the driver's work shift (Tok et al., 2017). This cluster includes locations that serve outbound and inbound flows from the marine terminals and nearby rail facilities, with several street segments with an SDOT designation of *seaport connectors*. The trajectories captured along these street segments may include regional movement, drayage, and trans-loading activities in the city. The time distribution of these locations can be explained by the marine terminal and railyard hours of operation. Thus, Port and the municipality can evaluate and implement city programs that would encourage CV to use off-hours peaks as a demand management strategy to reduce heavy truck traffic in the urban area during congested times.

4.5.5.1. BINMIC

The 932 acres, the Ballard-Interbay-Northend Manufacturing and Industrial Center (BINMIC), as identified in Seattle's Comprehensive Plan, is the region's smallest MIC, with around 60% of the land designated as industrial. Compared to other MICs, it has a smaller parcel size with a mix of diverse uses, spanning light manufacturing, maritime, food processing, warehousing, a rail yard, and several Port of Seattle facilities. Port facilities in the BINMIC include the Fishermen's Terminal, a terminal that provides anchorage to fishing and commercial workboats, and numerous businesses serving the fishing industry. Other key industrial facilities in this area include a brewery district, Salmon Bay shipyard, and BNSF railyard. The BNSF railyard is primarily used for maintenance with little truck activity. Finally, Seattle's Public Utility (SPU) operates a municipal waste transfer station near the BINMIC, east of Aurora Ave.

Hub and Spoke Operation. Within the BINMIC industrial districts, a balanced ingress/egress throughout the day was observed. The wide hump shape of single unit and CM smaller fleet combined with the directional balanced flow and fishing industry land use resembles hub-and-spoke operations, where commercial vehicles have a preferred base in the center of their operations, and travel between the hub and other destinations back and forth. This pattern may be related to wholesale distribution to supermarkets or direct distribution to other retailers.

4.5.5.2. Ferry Terminals

Ferry operations captured in our study area are related to the loading/unloading of the Colman Dock. This Seattle ferry terminal provides access to Washington State Ferries to Bainbridge Island and Bremerton. Two distinct traffic patterns describe these operations. First, the Intermittent traffic pattern (Cluster 8) is associated with most street segments providing access to the ferry terminals for both the small CV and SU fleet. These saw-tooth profiles can be explained by the intermittent nature of the ferry schedule. The ferries arrive at the dock in their defined schedules, and the unloading happens in a short period. Along the same lines, ferry users arrive at the dock anticipating the vessel's arrival and wait to load. The second pattern is a Distribution PM Peak, which describes the flow of the small CV fleet arriving at the ferry terminal, a smooth curve with an afternoon peak, similar to the outbound movement from the CBD to the industrial area south of GDA.

Additionally, all small CV fleet and non-CV DFPs follow the same patterns. In particular, both vehicle body types peak during the early morning (5 and 7 AM) for vehicles coming into the GDA; and during the afternoon (3 PM) for those leaving the GDA. On the other hand, the SU DFPs peak during the morning for vehicles taking the ferry towards Bainbridge Island and at noon for the ones arriving on the ferry at Seattle. This characteristic may be indicative of service and delivery routes serving Baimbridge Island coming from the industrial and distribution facilities.

4.5.6 Directional Differences

Time-of-day analyses can be (but are not always) affected by directional distributions. The size of the directional differences varies from site to site and depends significantly on local conditions rather than on functional class or other routinely stored parameters (Hallenbeck et al. 1997). For Seattle's case study, several locations follow different patterns based on the traffic direction. Two main categories of locations present directional differences in this study:

4.5.6.1. Proximity to significant changes in land use (SODO)

Significant changes in land use designation occur between the GDA urban center, the industrial area in SoDo, and the industrial pocket in the southeast boundary of GDA. When combining inbound and outbound flows for the street segments in these areas, the classic two peaks associated with commuter trips are followed by the smaller CV fleet. SU also has a Distribution AM peak pattern for the flow going into GDA. However, the single-unit flows going out of GDA show a business hump pattern (a wider hump earlier in the day). Although not following a Distribution PM Peaks as the smaller fleet, the single-unit DFPs follow the latest peaking variation among the Business Hour Patterns, see Figure 4-9.

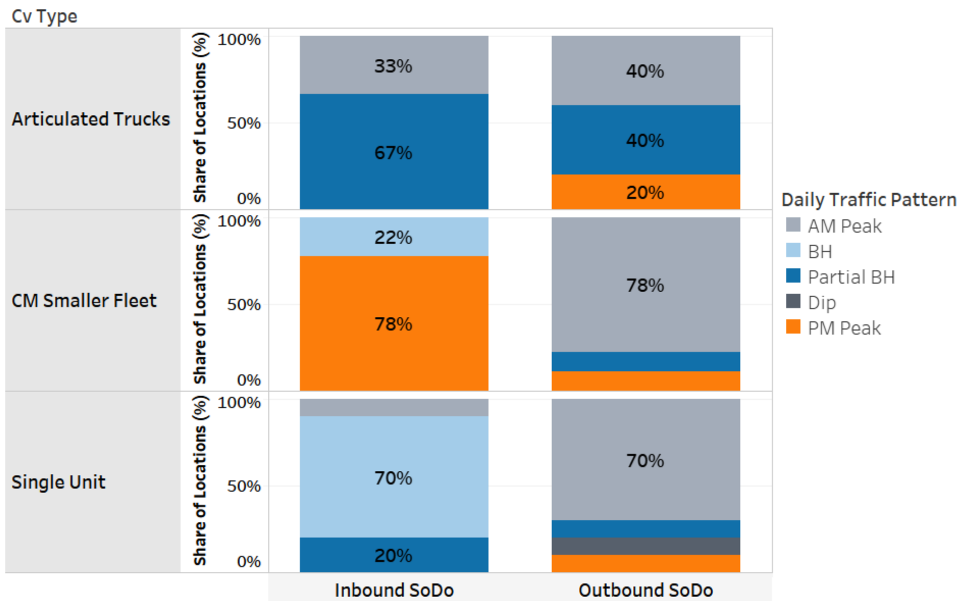


Figure 4-9. Distribution of Daily Flow Patterns by vehicle type and location for the industrial area south of GDA.

For both vehicle types, the morning peak occurs only in one direction (toward the urban center), and the afternoon peak occurs in the opposite direction (inbound to the MIC), see figure 4-10. These peaks are still prominent when DFPs for both directions are combined. One potential cause of the similarity of directional movement between non-CV and CV traffic is that delivery and pick-up of goods operations and service activities occur at commercial activity centers, the same locations that generate the influx of commuter hours. Furthermore, this can be related to movements leaving and returning to the distribution and industrial facilities after the last service or delivery.

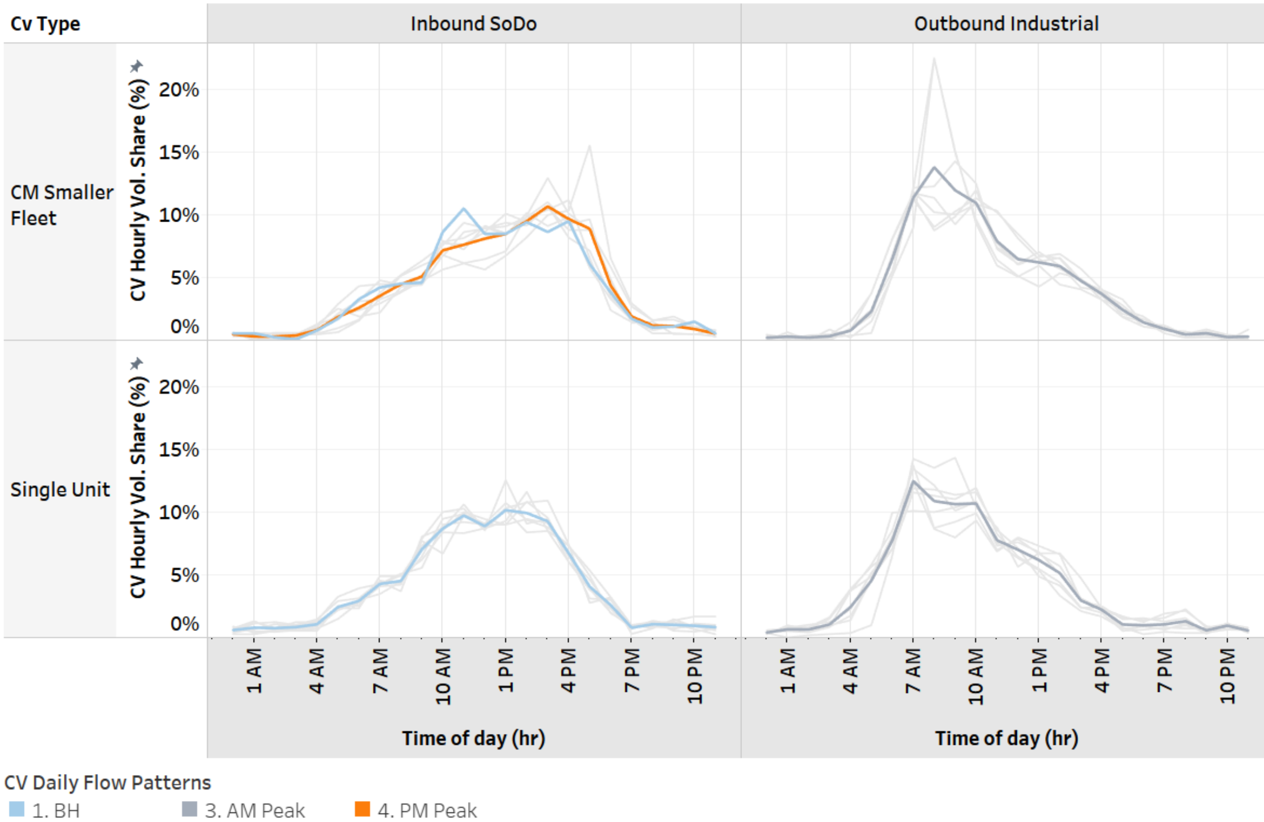


Figure 4-10. Dominant Daily Flow Pattern by vehicle type by direction in the industrial area south of GDA.

4.5.6.2. Connector to the final delivery

Interstate/freeway ramps serve a critical access role, connecting high-speed limited-access infrastructure to surface streets such as arterial and collector. They provide connection between area of service and the industrial/distribution facilities for the CV transportation demand for vehicle traveling from middle or longer distances. We can assume that the CV trajectories consolidated in the ramp capture the section of the route that has either just left home-based or has stop serving the area.

As ramps have no bidirectional traffic, the CV ADT collected for each is unidirectional. The ramps considered in this study include:

- four ramps along the two of Seattle's N-S main corridors (north and south of the Interbay),
- 18 ramps of I-5 providing access to the GDA; and
- and two ramps providing access to I-84 (W-E corridor).

Inbound GDA. Smaller fleet and single unit vehicles arriving to the GDA mainly show two patterns, including *Pattern 3 Distribution AM Peak and Pattern 5 - Activity Dip*, respectively (see Figure 4-11). All small fleet on I-5 off-ramps follow the Cluster 14 variation, and most of single unit trucks (80%) on NB on-ramps follow Cluster 9. These pattern variations represent the earliest AM Peak (6.5 AM and 7.5 AM, respectively) of all locations in the study. Additionally, both present a flow increase in the AM off-hour period compared to all patterns describing smaller and single unit trucks flows. Smaller fleet shows an 8 a.m. activity dip pattern, corresponding to the passenger peak at the I-5 off-ramps. This pattern also shows a double AM peak (6 am and 10 am peak) with an average of 240 CV/day, avoiding the PV commuter peak for both directions (I-5 SB and NB). On the other hand, single-unit trucks on NB on-ramps show an average of 320 CV/day and a flat hump during the non-CV peak hour (5 -11 a.m.), which indicates that freight operations utilizing SU trucks coming from locations south of the GDA must occur despite the PV AM commuter peak.

Outbound GDA. For outbound volumes, smaller fleet and single-unit trucks show distinct patterns, including a *Distribution PM Peak* and *Partial Business Hour* pattern, respectively (see Figure 4-11). A *Cluster 11* smoother pattern peaking at 2 p.m. describes the flow of the small CV fleet getting out of the GDA urban centers using I-5 regardless of direction (SB or NB). This outbound traffic average 600 CV ADT. This peak represents the latest CV traffic peak between all the ramps in our study, which happens before non-CV traffic starts increasing at these locations and four hours before the commuter PM peak leaving GDA. In contrast, single-unit outbound flow shows a partial business hour hump (*Cluster 5*) peaking earlier than the smaller fleet pattern (between 10-11 a.m.) with medium traffic intensity (above 12% during the peak hour) and average of 300 CV/day.

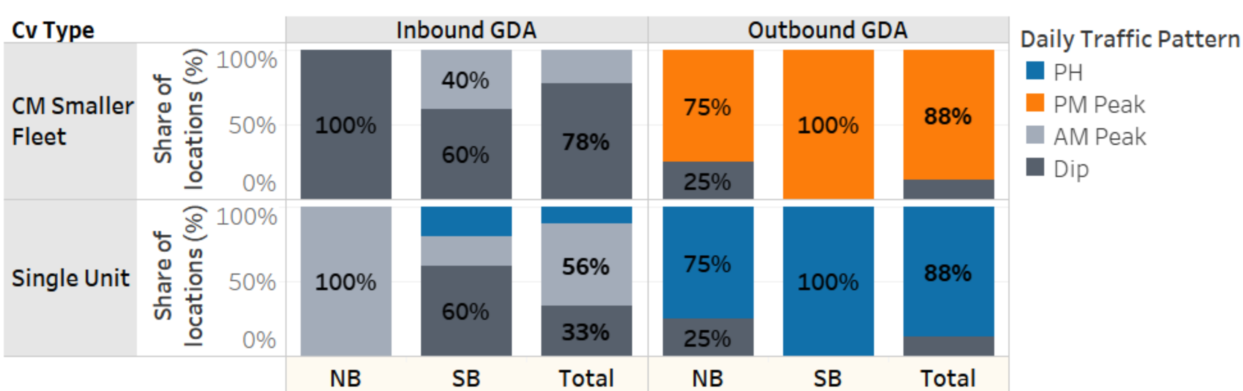


Figure 4-11. Frequency of CV Daily Traffic Flow Patterns by Vehicle Type and Direction in GDA I-5 Ramps

Ballard-Interbay connection. In contrast, the majority of DFPs associated to the ramps located in Ballard (along 15th NW and Aurora Ave N), see Figure 4-12. are grouped in Cluster 5 for non-articulated truck body types. This pattern peaks around 9:30 a.m. with an peak hour volume share above 14% and a CV ADT close to 250. For this pattern, directionality has no influence over temporal variations of CV DFPs.

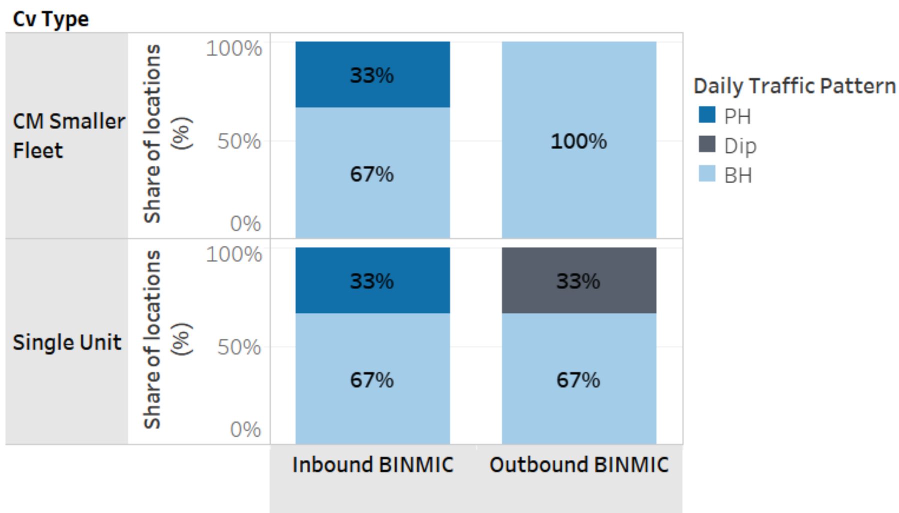


Figure 4-12. Frequency of CV Daily Traffic Flow Patterns by Vehicle Type and Direction in Ballard I-5 Ramps

4.6 Discussion and Conclusions

Having a realistic description of the traffic patterns in a specific city is a requirement to obtain adequate understanding and meaningful freight demand forecasting that supports the development, prioritization, and assessment of infrastructure and policy scenarios. To develop an accurate depiction of the urban CV flows, a level of granularity is necessary to capture the heterogeneous nature of these operations and to define tailored solutions. In addition, if aimed at developing advanced traffic management solutions, it becomes further necessary to have a more in-depth understanding of how traffic is distributed in a particular city, which basically requires performing a correct analysis and classification of such traffic.

To date, efforts to understand the UGS have been hindered by the lack of relevant and available data. To address this critical gap, we leverage on our CV typology and collected data introduced in Chapter 3 to derive insight about the UGS. This chapter addresses this challenge by developing and evaluating a method to extract representative and unique CV flow patterns from traffic count data from70 roadway segments in Seattle.

A 4-phase methodology is developed in which daily traffic flow profiles, fleet composition, and locations' spatial location characteristics are extracted from traffic count and GIS data and then used to find common but unique CV daily flow patterns. The daily flow profiles are defined by the directional volume share by the time of day, vehicle body type, and location and feed a clustering unsupervised learning procedure in the second phase, which evaluates multiple algorithms, including K-Means with DTW distance and Hierarchical clustering with Ward and complete linkage. The clustering procedure considers a pre-selection before clustering by infrastructure type and body vehicle type into three groups: ramps, street segments - articulated trucks, and street segments - small fleet and single unit trucks. Based on traffic count data for the Seattle case study, we identified 17 CV daily flow clusters among 289 CV daily flow profiles from 70 locations. A hierarchical clustering Ward linkage approach produces the best clustering results and internal validation indices. Phase 3 of our methodology develops a generalization of the cluster results and application to gain insights into CV urban traffic flows. To conduct a temporal characterization of the daily flow patterns, we define 15 metrics in four categories: CV Peak Hour Descriptors, CV Operations Descriptors, Fleet Composition, and Traffic Descriptors. The 17 clusters are be generalized into a smaller set of eight Main CV Daily Flow Patterns that explain the temporal and spatial variability of Seattle's case.

The results of our cluster analysis contain information that can provide useful and generalizable insights for urban freight transportation. With regards to the effect of vehicle body type and activity type, some patterns showed a characteristic vehicle body type such as business hour hump and partial-businesses hump that most frequently showed small fleet and single-unit truck DPFs, respectively. On the other hand, vehicle body types followed more than one flow pattern depending on the land use, directionality, infrastructure type and nature of the CV operations. For example, smaller fleet tended to show a business hour hump pattern in a commercial area such as Ballard, an Activity Dip Pattern in interstate ramps and a Distribution AM and PM peaks in the industrial area of downtown Seattle. Time-of-day analyses can be (but are not always) affected by directional distributions. The size of the directional differences varies from site to site and depends significantly on local conditions rather than on functional class or other routinely stored parameters (Hallenbeck et al., 1997). The effect of directionality is indicated by the different patterns followed by single-unit trucks going into GDA (AM Distribution Peak and Activity Dip patterns) and going out of GDA (Partial Business Hours Pattern). By not capturing the heterogeneity in CV flows, key operational aspects may be missed. By looking at the aggregated General trend of CVs, one could incorrectly assume that freight firms mostly can minimize congestion and delay adjusting their schedules (e.g., deliveries) to avoid the peak traffic periods, (i.e., morning and evening PV commuting times). However, as the data shows, significant volumes and sometimes, even peak volume of the overall CV traffic occurred during the PV peak hour. This finding highlights how some supply chain operations or business needs can't be adjusted. This agrees with prior findings of freight and package pick-up ending before commute rush hours in the urban core (Pivo et al., 2002) start affecting traffic flow on it.

The obtained insight into existing CV flow patterns can be used for traffic monitoring, forecasting, traffic management and transport modeling scenarios. One important example application is emissions modeling of the CV fleet in urban areas as tracking activities in urban logistics networks represents a major source of greenhouse gas (GHG) emissions. Different trucking activities, defined by trip purpose as well as other logistics characteristics such as vehicle type and route characteristics, may impact the associated GHG emissions and designing management measures should take those purposes into consideration (Gan et al., 2018). Therefore, understanding the heterogeneity of the CV vehicle activities in an urban area, the share of those activities of the total CV traffic, temporal and spatial variations and traffic intensity must be considered for CV emissions modeling.

Typical CV flow patterns can also be used for traffic forecasting and traffic management scenarios. For these scenarios, it is important that the cluster average is representative for the CV flows within the clusters. Our 17 clusters capture the variability of overall CV flows into a representative limited set of groups. Among the three vehicle body types, including SM, SU and AT, AT showed the highest level of variability and is the vehicle type most difficult to cluster, therefore, a limited number of AT specific clusters were identified. On the other hand, an activity dip pattern was discovered, which can be associated with AT marine container and port intermodal operations representing an important localized high-intensity traffic pattern in port cities and cities with critical intermodal facilities. Analysis of traffic management scenarios would benefit of the understanding of typical CV flow patterns to develop applications that enhance CV travel time reliability and speed such as Freight and Transit (FAT) Lanes and Freight Signal Priority (FSP).

On a planning level, setting priorities for improving road networks has become a high-demand activity for agencies financing infrastructure development. With limited funding resources, developing and implementing infrastructure projects that improve transportation efficiency and safety, ensure the resiliency of the freight network, and improve the economic competitiveness is a difficult task. Our research defines 15 metrics to characterize CV flow patterns and generalizes 17 clusters into a smaller set of eight Main CV Daily Flow Patterns that explain the temporal and spatial variability of Seattle's case. This enhances agencies' data-driven ability to classify elements of the street network that provide a critical support function to different CV flows. Since we explained CV flows based on general travel demand factors, our results can, at a minimum, produce insight into regular variations in CV traffic volumes in other urban areas, if not be extended to other cities with similar characteristics to Seattle.

Chapter 5. CONCLUSION

Understanding commercial vehicle activities throughout the transportation network are essential to strategic freight planning, infrastructure maintenance, funding allocations, and traffic operation management. The research on CV demand is minimal, as most of the work has focused on passenger activity. As a result, practitioners are frequently unable to estimate the magnitude of the CV demand and the nature of its operations in their jurisdictions, making the development of fitted solutions challenging at best. Moreover, recent studies have revealed a significant void in the availability of detailed CV activity data in urban areas.

The case studies presented in this dissertation and the clustering exercise demonstrate the potential of the data obtained in analyzing and understanding current and historical CV flows and parking operation in a dense urban area.

The study on CV parking operations provides insights on dwell time, fleet configuration, parking choice and the current use of the infrastructure, an essential piece to evidence-support the development and implementation of effective curb management strategies.

On the other hand, Chapter 2 and 3 provides a robust and replicable framework to categorize and record granular CV flows; and a demonstration of the methodology with two Seattle case studies. The typology presented in this dissertation has already been applied to other areas of Seattle to continue capturing site-based and detailed information about the temporal and spatial patterns of CV traffic.

The dissertation paves the way for providing tailored solutions that consider the spectrum of CV fleets and their different operational needs; and builds a broad understanding of CV activity surely lacking among transport planners. As an example, this research provides key insights into service trips, a typical disregarded section of the freight operations both in the literature reviewed and municipalities efforts in the US and a significant share of the overall CV flows.

To summarize, our research includes the following contributions:

- Develops a replicable and transferable data collection method to record commercial vehicle parking behavior.
- Gives the opportunity to revise existing parking policies, regarding usage restriction, time restriction, and space management; to rethink where should different vehicles park, for how long, and how many spaces should be provided to support an efficient urban goods distribution system.

- Creates knowledge into the temporal variations and typical within-day traffic patterns for commercial vehicles in the urban area, considering the urban commercial fleet's heterogeneity.
- Identifies temporal and spatial similarities/dissimilarities between commercial vehicle flows that can be summarized in an actionable way with pattern groups for urban freight modeling and planning.
- Compares of commercial vehicle group patterns with general traffic that highlights the importance of considering commercial vehicle flows.
- Displays the importance of providing tailored solutions that consider the spectrum of load/unload operations. For example, servicing trips could skew the dwell time distribution of all CVs, as they tend to take over most of the longer CV parking operations.
- Develops a replicable and comprehensive commercial vehicle typology to record granular commercial vehicle volumes and provides a demonstration of the methodology with two Seattle case studies.
- The 3,740 hours of video coded data provide a helpful asset in future video detection and image classification efforts
- Develops a method to describe and classify urban road segments based on their CV load throughout the day.
- Identify and characterize the commercial demand daily variation in the Seattle area that will assist road maintenance, infrastructure and demand management policies.
- Provides a demonstration of methodology with a Seattle Case Study. Urban logistics policy recommendations are provided based on a review of Seattle's Freight Network designations.
- Recommends level of granularity that municipalities should consider in future data collection efforts and analysis; and what are the viable technologies to detect typical CV patterns.

Researchers and practitioners can use the Seattle case studies conducted in this dissertation as a measurable, in-depth investigation of curb and roadway network that suggest possible outcomes of vehicle operations in these spaces with practical implications for curb management and traffic demand management policies. With all the variations on freight operational needs and fleet configurations, our research represents a steppingstone necessary to provide a comprehensive view and understanding to support policies/strategies that efficiently manage this public asset.

BIBLIOGRAPHY

1. Allen, J., M. Piecyk, M. Piotrowska, F. McLeod, T. Cherrett, K. Ghali, T. Nguyen, T. Bektas, O. Bates, A. Friday, S. Wise, and M. Austwick. Understanding the Impact of E-Commerce on Last-Mile Light Goods Vehicle Activity in Urban Areas: The Case of London. Transportation Research Part D: Transport and Environment. 2018; Vol. 61, pp. 325–338. https://doi.org/10.1016/j.trd.2017.07.020.
2. Aditjandra, P. T., F. Galatioto, M. C. Bell, and T. H. Zunder. Evaluating the Impacts of Urban Freight Traffic: Application of Micro-Simulation at a Large Establishment. European Journal of Transport and Infrastructure Research. 2016; 16, pp. 4–22. https://doi.org/10.18757/ejtir.2016.16.1.3110.
3. Aiura N., Taniguchi E. Planning On-Street Loading-Unloading Spaces Considering the Behaviour of Pickup-Delivery Vehicles. Journal of the Eastern Asia Society for Transportation Studies, Vol. 6, 2005, pp. 2963–2974.
4. Akgün, J. Monios, T. Rye, A. Fonzone. Influences on urban freight transport policy choice by local authorities, Transport Policy. 2019; 75, pp. 88-98
5. Allen J., Anderson S., Browne M., Jones P. A Framework for Considering Policies to Encourage Sustainable Urban Freight Traffic and Goods/Service Flows. Transport Studies Group, University of Westminster, London, UK, 2000.
6. Allen, J., Ambrosini, C., Browne, M., Patier, D., Routhier, J. L., and Woodburn, A. Data collection for understanding urban goods movement. In Sustainable Urban Logistics: Concepts, Methods and Information Systems. 2014; pp. 71-89. Springer, Berlin, Heidelberg.
7. Allu, K. R., Sun, Z., & Tok, A. Lidar Based Reconstruction framework for Truck Surveillance. UC Irvine: Institute of Transportation Studies. Available from: https://escholarship.org/uc/item/68b3g3362020
8. Almannaa, M. H., M. Elhenawy, and H. A. Rakha. 2020. "A Novel Supervised Clustering Algorithm for Transportation System Applications." IEEE Transactions on Intelligent Transportation Systems 21 (1): 222–32. https://doi.org/10.1109/TITS.2018.2890588.
9. Aqel, S., A. Hmimid, M. A. Sabri, and A. Aarab. Road traffic: Vehicle detection and classification. In 2017 Intelligent Systems and Computer Vision (ISCV). 2017; pp. 1-5.
10. ATRI. A Financial Analysis of Toll System Revenue: Who Pays & Who Benefits. January, 2020

11. Ballantyne, M. Lindholm, A. Whiteing. A comparative study of urban freight transport planning: addressing stakeholder needs. Journal of Transport Geography. 2013; 32, pp. 93-101
12. Banaei-Kashani, F., C. Shahabi, B. Pan. "Discovering Patterns in Traffic Sensor Data." In Proceedings of the 2nd ACM SIGSPATIAL International Workshop on GeoStreaming - IWGS. 2011; pp. 10–16. Chicago, Illinois: ACM Press. https://doi.org/10.1145/2064959.2064963.
13. Banaei-Kashani, Farnoush, Cyrus Shahabi, and Bei Pan. 2011. "Discovering Patterns in Traffic Sensor Data." In Proceedings of the 2nd ACM SIGSPATIAL International Workshop on GeoStreaming - IWGS '11, 10–16. Chicago, Illinois: ACM Press. https://doi.org/10.1145/2064959.2064963.
14. Browne, M., J. Allen, T. Nemoto, D. Patier, and J. Visser. Reducing Social and Environmental Impacts of Urban Freight Transport: A Review of Some Major Cities. Procedia-Social and Behavioral Sciences. 2012; 39, pp. 19–33. https://doi.org/10.1016/j.sbspro.2012.03.088ï.
15. Butrina P., Giron G., Machado J., Goodchild A., Ayyalasomayajula P. From the Last Mile to the Last 800 Feet: Key Factors in Urban Pick-up and Delivery of Goods. Transportation Research Record: Journal of the Transportation Research Board, 2017. 2609: 85–92.
16. Calafate, C. T., Soler, D., Cano, J. C., & Manzoni, P. (2015). Traffic management as a service: The traffic flow pattern classification problem. Mathematical Problems in Engineering, 2015.
17. Cherrett T., Allen J., McLeod F., Maynard S., Hickford A., Browne M. Understanding Urban Freight Activity–Key Issues for Freight Planning. Journal of Transport Geography, Vol. 24, 2012, pp. 22–32.
18. City of Seattle Land Use Code, 2018. City of Seattle, Seattle, WA, 2018.
19. Dablanc L., Beziat A. Parking for Freight Vehicles in Dense Urban Centers—The Issue of Delivery Areas in Paris. MetroFreight Center of Excellence, Marne la Vallée, France, 2015.
20. Delivery and Servicing Plans Making Freight Work for You. Transport for London, London, U.K., 2018. https://tfl.gov.uk/info-for/urban-planning-and-construction/transport-assessment-guide/guidance-by-transport-type/freight. Accessed July 28, 2018
21. Depaire, Wets and Vanhoof (2008). Traffic accident segmentation by means of latent class clustering. Accident Analysis and Prevention 40, 1257-1266
22. Dublin City Council, National Transport Authority, Canal Cordon Report 2019; Report on trends in mode share of vehicles and people crossing the Canal Cordon. May, 2019.

23. FHWA. Verification, Refinement, and Applicability of Long-Term Pavement Performance Vehicle Classification Rules. November, 2014. Available from: https://www.fhwa.dot.gov/publications/research/infrastructure/pavements/ltpp/13091/002.cfm
24. Fossheim, J. Andersen. Plan for sustainable urban logistics–comparing between Scandinavian and UK practices. European Transportation Research Review. 2017; 9, pp. 52-65
25. Gecchele, Gregorio, Riccardo Rossi, Massimiliano Gastaldi, and Armando Caprini. 2011. "Data Mining Methods for Traffic Monitoring Data Analysis: A Case Study." Procedia - Social and Behavioral Sciences, The State of the Art in the European Quantitative Oriented Transportation and Logistics Research – 14th Euro Working Group on Transportation & 26th Mini Euro Conference & 1st European Scientific Conference on Air Transport, 20 (January): 455–64.
26. Giron-Valderama, G. Goodchild, A. Characterization of Seattle's commercial traffic patterns: A Greater Downtown Area and Ballard/Interbay vehicle count and evaluation. 2020.
27. Girón-Valderrama, G. del C., J. L. Machado-León, and A. Goodchild. Commercial Vehicle Parking in Downtown Seattle: Insights on the Battle for the Curb. Transportation Research Record. 2019. https://doi.org/10.1177/0361198119849062.
28. González-Feliu, J., Cedillo-Campo, M. G., and García-Alcaraz, J. L. An emission model as an alternative to OD matrix in urban goods transport modelling. Dyna. 2014; 81(187), pp. 249-256.
29. Gordon, A. D. (1999). Classification. Chapman and Hall. New York.
30. Hallenbeck, M., B. Smith, C. Cornell-Martinez, and J. Wilkinson. VEHICLE VOLUME DISTRIBUTIONS BY CLASSIFICATION BY CLASSIFICATION. Washington State Transportation Center; 1997.
31. He, A. Wu, X. Huang, A. Rangarajan, S. Ranka. Video-Based Machine Learning System for Commodity Classification. Proceedings. VEHITS 2020 - 6th International Conference on Vehicle Technology and Intelligent Transport Systems. 2020.
32. He, P., Wu, A., Huang, X., Scott, J., Rangarajan, A., & Ranka, S. Deep learning based geometric features for effective truck selection and classification from highway videos. In 2019 IEEE Intelligent Transportation Systems Conference (ITSC). 2019; pp. 824-830.
33. Hernandez, S. V., Tok, A., and Ritchie, S. G. Integration of Weigh-in-Motion (WIM) and inductive signature data for truck body classification. Transportation Research Part C: Emerging Technologies. 2016; 68, pp. 1-21.

34. Holguin-Veras J., Patil G. Observed Trip Chain Behavior of Commercial Vehicles. Transportation Research Record: Journal of the Transportation Research Board, 2005. 1906: 74–80.
35. Holguín-Veras, J., & Jaller, M. Comprehensive freight demand data collection framework for large urban areas. In Sustainable urban logistics: Concepts, methods and information systems. 2014; pp. 91-112. Springer, Berlin, Heidelberg.
36. Hsieh, J.-W., Yu, S.-H. Chen, Y.-S. Hu, W.-F. Automatic traffic surveillance system for vehicle tracking and classification. IEEE Trans. Intell. Transp. Syst. 2006; 7, pp. 175–187.
37. Jaller M., Holguín-Veras J., Hodge S. Parking in the City: Challenges for Freight Traffic. Transportation Research Record: Journal of the Transportation Research Board, 2013. 2379: 46–56.
38. Javadi, S., M. Rameez, M. Dahl, and M. I. Pettersson. "Vehicle classification based on multiple fuzzy c-means clustering using dimensions and speed features." Procedia Computer Science. 126 (2018): 1344-1350.
39. Jeng, S. T., and S. G. Ritchie. Real-Time Vehicle Classification Using Inductive Loop Signature Data. Transportation Research Record. 2008; 2086, pp. 8–22. https://doi.org/10.3141/2086-02.
40. Kaufman, L., & Rousseeuw, P. J. (2009). Finding groups in data: an introduction to cluster analysis. John Wiley & Sons.
41. Kim, Hyung-Joo, and Justin S. Chang. 2012. “Calculation of the Peak-hour Ratio for Road Traffic Volumes using a Hybrid Clustering Technique.” Journal of Korean Society of Transportation 30 (1): 19–30.
42. Lee, J. T., and Y. Chung. Deep Learning-Based Vehicle Classification Using an Ensemble of Local Expert and Global Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2017; pp. 920-925. https://doi.ieeecomputersociety.org/10.1109/CVPRW.2017.127
43. Lindholm, M. Blinge. Assessing knowledge and awareness of the sustainable urban freight transport among Swedish local authority policy planners. Transport Policy. 2014; 32, pp. 124-131
44. Liu, Y., Li, Z., Xiong, H., Gao, X., & Wu, J. (2010, December). Understanding of internal clustering validation measures. In 2010 IEEE international conference on data mining (pp. 911-916). IEEE.

45. Ma, Xiaolei, Yi Li, and Peng Chen. 2020. “Identifying Spatiotemporal Traffic Patterns in Large-Scale Urban Road Networks Using a Modified Nonnegative Matrix Factorization Algorithm.” Journal of Traffic and Transportation Engineering (English Edition) 7 (4): 529–39. https://doi.org/10.1016/j.jtte.2018.12.002.
46. Maimon, Oded, and Lior Rokach, eds. 2010. Data Mining and Knowledge Discovery Handbook. Boston, MA: Springer US. https://doi.org/10.1007/978-0-387-09823-4.
47. Marcucci E., Gatta V., Scaccia L. Urban Freight, Parking and Pricing Policies: An Evaluation from a Transport Provider’s Perspective. Transportation Research Part A: Policy and Practice, Vol. 74, 2015, pp. 239–249.
48. Marsden, K.T. Frick, A.D. May. How do cities approach policy innovation and policy learning? A study of 30 policies in Northern Europe and North America. Transport Policy. 2011; 18, pp. 501-512
49. McDonald, N., Yuan, Q., & Naumann, R. Urban freight and road safety in the era of e-commerce. Traffic Injury Prevention.2019; 20 (7), 764-770.
50. MMM Group. Greater Toronto Area Cordon Count Program Transportation Trends Technical Report.
51. Muñuzuri J., Cortés P., Onieva L., Guadix J. Modelling Peak-Hour Urban Freight Movements with Limited Data Availability. Computers and Industrial Engineering, Vol. 59, No. 1, 2010, pp. 34–44.
52. Nam, Y., and Y. C. Nam. Vehicle Classification Based on Images from Visible Light and Thermal Cameras. Eurasip Journal on Image and Video Processing. 2018; 2018, No. 1, https://doi.org/10.1186/s13640-018-0245-2.
53. Necula, Emilian. "Analyzing traffic patterns on street segments based on GPS data using R." Transportation Research Procedia 10 (2015): 276-285.
54. Nourinejad M., Wenneman A., Habib K. N., Roorda M. J. Truck Parking in Urban Areas: Application of Choice Modelling within Traffic Microsimulation. Transportation Research Part A: Policy and Practice, Vol. 64, 2014, pp. 54–64.
55. Ogden. A framework for urban freight policy analysis. Transportation Planning Technology. 1984; 8, pp. 253-265
56. Park, Byungkyu “Brian.” 2002. “Hybrid Neuro-Fuzzy Application in Short-Term Freeway Traffic Volume Forecasting.” Transportation Research Record 1802 (1): 190–96
57. Performance-Based Parking Pricing Program. Seattle Department of Transportation, 2018. https://www-seattle-gov.offcampus.lib.washington.edu/transportation/projects-and-

programs/programs/parking-program/performance-based-parking-pricing-program. Accessed July 26, 2018.

58. Pierce G., Shoup D. Getting the Prices Right: An Evaluation of Pricing Parking by Demand in San Francisco. Journal of the American Planning Association, Vol. 79, No. 1, 2013, pp. 67–81.
59. Richards L. Urban Goods Movement Sustainability in Washington, D.C. Presented at 96th Annual Meeting of the Transportation Research Board, Washington, D.C., 2017.
60. Saha, R., Tariq, M. T., Hadi, M., & Xiao, Y. (2019). Pattern Recognition using Clustering Analysis to support transportation system management, Operations, and Modeling. Journal of Advanced Transportation, 2019.
61. Salamanis, Athanasios, Giorgos Margaritis, Dionysios D. Kehagias, Georgios Matzoulas, and Dimitrios Tzovaras. 2017. “Identifying Patterns under Both Normal and Abnormal Traffic Conditions for Short-Term Traffic Prediction.” Transportation Research Procedia, 19th EURO Working Group on Transportation Meeting, EWGT2016, 5-7 September 2016, Istanbul, Turkey, 22 (January): 665–74
62. San Francisco County Transportation Authority. Geary Corridor Bus Rapid Transit. Project EIS/EIR. San Francisco County Transportation Authority, San Francisco, CA, 2015.
63. Seattle Department of Transportation, Load Zones. City of Seattle. https://www-seattle-gov.offcampus.lib.washington.edu/transportation/projects-and-programs/programs/parking-program/parking-regulations/load-zones. Accessed November 8, 2018.
64. Seattle Department of Transportation. CITY OF SEATTLE FREIGHT MASTER PLAN. 2016.
65. Seattle Department of Transportation. Commercial Delivery Load Zone Program / Urban Goods Delivery Strategy. City of Seattle. https://www-seattle-gov.offcampus.lib.washington.edu/transportation/projects-and-programs/programs/parking-program/urban-goods-delivery-strategy. Accessed November 8, 2018.
66. Sheth, M., P. Butrina, A. Goodchild, and E. McCormack. Measuring Delivery Route Cost Trade-Offs between Electric-Assist Cargo Bicycles and Delivery Trucks in Dense Urban Areas. European Transport Research Review. 2019; 11, No. 1. https://doi.org/10.1186/s12544-019-0349-5.
67. The Final 50 Feet Urban Goods Delivery System. Final Report. Supply Chain Transportation and Logistics Center, Seattle, WA, 2018.

68. Thomas, T., and W. Weijermars. Variations in Urban Traffic Volumes. EJTIR Issue. 2008; 8, pp. 251–263.
69. Tok, A., Hyun, K., Hernandez, S., Jeong, K., Sun, Y., Rindt, C., & Ritchie, S. G. (2017). Truck activity monitoring system for freight transportation analysis. Transportation Research Record, 2610(1), 97-107.
70. Transport for London. Travel in London. 2017.
71. Visser, J., T. Nemoto, and M. Browne. Home Delivery and the Impacts on Urban Freight Transport: A Review. Procedia - Social and Behavioral Sciences. 2014; 125, pp. 15–27. https://doi.org/10.1016/j.sbspro.2014.01.1452.
72. Weijermars and E. van Berkum, Analyzing highway flow patterns using cluster analysis, Proceedings. 2005 IEEE Intelligent Transportation Systems. 2005; pp. 308-313, doi: 10.1109/ITSC.2005.1520157.
73. Weijermars, W. A. M. (2007). Analysis of urban traffic patterns using clustering.
74. Weisbrod, G., & Fitzroy, S. (2011). Traffic congestion effects on supply chains: Accounting for behavioral elements in planning and economic impact models. In Supply Chain Management-New Perspectives. IntechOpen.
75. Xia, Jingxin, and Mei Chen. 2007. “Defining Traffic Flow Phases Using Intelligent Transportation Systems-Generated Data.” Journal of Intelligent Transportation Systems 11 (1): 15–24.
76. Yang, Senyan, Jianping Wu, Geqi Qi, and Kun Tian. 2017. “Analysis of Traffic State Variation Patterns for Urban Road Network Based on Spectral Clustering.” Advances in Mechanical Engineering 9
77. Young W., Miles C. F. A Spatial Study of Parking Policy and Usage in Melbourne, Australia. Case Studies on Transport Policy, Vol. 3, No. 1, 2015, pp. 23–32.
78. Yuksel, E., R. L. Bertini, N. Menon, S. Ozkul, B. Staes. A Contemporary Approach for Visualizing Temporal and Spatial Urban Freight Movement by Leveraging Mobility Portal Data. 2020 Forum on Integrated and Sustainable Transportation Systems (FISTS). 2020; pp. 272-279, doi: 10.1109/FISTS46898.2020.9264848.
79. Zhang, G., Y. Wang, and H. Wei. Artificial neural network method for length-based vehicle classification using single-loop outputs. Transportation research record. 2006; 1, pp. 100-108.

APPENDIX

Appendix A - Vehicle Typology

Table A-1 Vehicle Typology for both Seattle's Greater Downtown cordon and Ballard/Interbay studies.

No	CV	General Body Type	Activity Type	Vehicle Body	No. of axles	FHW Class
1	Non-commercial	Motorcycle	Private	Motorcycles	2 axles	Class 1
2	Non-commercial	Car	Private	Passenger vehicle	2 axles	Class 2
3	Non-commercial	Car	Emergency	Passenger vehicle	2 axles	Class 2
4	Non-commercial	Transit	Public Transit	Bus	2 axles +	Class 4
5	Non-commercial	Transit	Other Transit	Bus/other transit	2 axles +	Class 3 & Class 4
6	Non-commercial	Recreational	Recreational	RVs	Unknown	Unknown
7	Non-commercial	Emergency van	Emergency	Work Van	2 axles	Class 3 & Class 5
8	Commercial	CM smaller fleet	Goods Transport	Work Van	2 axles	Class 3 & Class 5
9	Commercial	CM smaller fleet	Service	Work Van	2 axles	Class 3 & Class 5
10	Commercial	CM smaller fleet	Service	CM Pick-Up	2 axles	Class 3 & Class 5
11	Commercial	CM smaller fleet	General CM	Work Van	2 axles	Class 3 & Class 5
12	Commercial	Truck	Goods Transport	Single Unit	2 axles	Class 3 & Class 5
13	Commercial	Truck	Goods Transport	Single Unit	3 axles	Class 6
14	Commercial	Truck	Goods Transport	Single Unit	4 axles +	Class 7
15	Commercial	Truck	Goods Transport	Trailer	3 or 4 axles	Class 8
16	Commercial	Truck	Goods Transport	Trailer	5 axles	Class 9
17	Commercial	Truck	Goods Transport	Trailer	6 axles +	Class 10
18	Commercial	Truck	Goods Transport	Multi-trailer	5 or less axles	Class 11
19	Commercial	Truck	Goods Transport	Multi-trailer	6 axles	Class 12
20	Commercial	Truck	Goods Transport	Multi-trailer	7 axles +	Class 13
21	Commercial	Truck	Goods Transport	Unknown	Unknown	Unknown
22	Commercial	Truck	Service	Single Unit	2 axles	Class 3 & Class 5
23	Commercial	Truck	Service	Single Unit	3 axles	Class 6
24	Commercial	Truck	Service	Single Unit	4 axles +	Class 7
25	Commercial	Truck	Service	Trailer	3 or 4 axles	Class 8
26	Commercial	Truck	Service	Trailer	5 axles	Class 9
27	Commercial	Truck	Service	Trailer	6 axles +	Class 10
28	Commercial	Truck	Service	Multi-trailer	5 or less axles	Class 11
29	Commercial	Truck	Service	Multi-trailer	6 axles	Class 12
30	Commercial	Truck	Service	Multi-trailer	7 axles +	Class 13
31	Commercial	Truck	Service	Unknown	Unknown	Unknown
32	Commercial	Truck	Waste Management	Single Unit	2 axles	Class 3 & Class 5
33	Commercial	Truck	Waste Management	Single Unit	3 axles	Class 6
34	Commercial	Truck	Waste Management	Single Unit	4 axles +	Class 7
35	Commercial	Truck	Waste Management	Trailer	3 or 4 axles	Class 8
36	Commercial	Truck	Waste Management	Trailer	5 axles	Class 9

37	Commercial	Truck	Waste Management	Trailer	6 axles +	Class 10
38	Commercial	Truck	Waste Management	Multi-trailer	5 or less axles	Class 11
39	Commercial	Truck	Waste Management	Multi-trailer	6 axles	Class 12
40	Commercial	Truck	Waste Management	Multi-trailer	7 axles +	Class 13
41	Commercial	Truck	Waste Management	Unknown	Unknown	Unknown
42	Commercial	Truck	Construction	Single Unit	2 axles	Class 3 & Class 5
43	Commercial	Truck	Construction	Single Unit	3 axles	Class 6
44	Commercial	Truck	Construction	Single Unit	4 axles +	Class 7
45	Commercial	Truck	Construction	Trailer	3 or 4 axles	Class 8
46	Commercial	Truck	Construction	Trailer	5 axles	Class 9
47	Commercial	Truck	Construction	Trailer	6 axles +	Class 10
48	Commercial	Truck	Construction	Multi-trailer	5 or less axles	Class 11
49	Commercial	Truck	Construction	Multi-trailer	6 axles	Class 12
50	Commercial	Truck	Construction	Multi-trailer	7 axles +	Class 13
51	Commercial	Truck	Construction	Unknown	Unknown	Unknown
52	Non-commercial	Truck	Emergency	Single Unit	2 axles	Class 3 & Class 5
53	Non-commercial	Truck	Emergency	Single Unit	3 axles	Class 6
54	Non-commercial	Truck	Emergency	Single Unit	4 axles +	Class 7
55	Non-commercial	Truck	Emergency	Trailer	3 or 4 axles	Class 8
56	Non-commercial	Truck	Emergency	Trailer	5 axles	Class 9
57	Non-commercial	Truck	Emergency	Trailer	6 axles +	Class 10
58	Non-commercial	Truck	Emergency	Multi-trailer	5 or less axles	Class 11
59	Non-commercial	Truck	Emergency	Multi-trailer	6 axles	Class 12
60	Non-commercial	Truck	Emergency	Multi-trailer	7 axles +	Class 13
61	Non-commercial	Truck	Emergency	Unknown	Unknown	Unknown
62	Commercial	Truck	General CM	Single Unit	2 axles	Class 3 & Class 5
63	Commercial	Truck	General CM	Single Unit	3 axles	Class 6
64	Commercial	Truck	General CM	Single Unit	4 axles +	Class 7
65	Commercial	Truck	General CM	Trailer	3 or 4 axles	Class 8
66	Commercial	Truck	General CM	Trailer	5 axles	Class 9
67	Commercial	Truck	General CM	Trailer	6 axles +	Class 10
68	Commercial	Truck	General CM	Multi-trailer	5 or less axles	Class 11
69	Commercial	Truck	General CM	Multi-trailer	6 axles	Class 12
70	Commercial	Truck	General CM	Multi-trailer	7 axles +	Class 13
71	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown

Table A-2. Vehicle Typology metadata.

Attribute	Code domain	Description
VEH_ID	Text	Vehicle's category unique identifier.
CV_TYP	Non-commercial, Commercial	Indicates if the vehicle is used for commercial or non-commercial activity.
ACT_TYP	Private, Public Transit, Other Transit, Recreational, Emergency, Freight, Service, Construction, Waste Management	Indicate the primarily purpose for which the vehicle was manufactured or is primarily usage purpose. . **If CV_TYP = 'Non-Commercial',** **Private:** Vehicles manufactured primarily for the purpose of carrying passengers for private or business needs with maximum seating capacity of eight people. **Emergency:** Vehicles used by emergency responses teams (e.g., fire trucks, ambulances and police cars). **Recreational:** Vehicles designed or modified for recreation or camping. **Public Transit:** Vehicles manufactured primarily for the purpose of carrying passenger with a minimum seating capacity of 10 people. Only included those vehicles owned by local agencies to provide transit services by public conveyance with established routes, schedules and transit fee. **Other Transit:** School; public; private; or commercial passenger-carrying buses and vans excluding public transit. **If CV_TYP = 'Commercial',** **Goods Transport:** Vehicles design for carrying commodities (e.g.., carrier and shipper's work van; autotransporters, cargo tanks, box trucks, containers, and tankers). Only work-vans with either readable and/or recognizable carrier or shipper logo are considered in this category. **Service:** Vehicles designed typically to be use by maintenance or service providers (e.g., electricity, plumbing, internet telecommunication, catering, public utilities, pest control); including food trucks, buckets trucks, service providers' pick-ups or work vans, and any other service-body truck. Only work-vans working as food-trucks or with either readable and/or recognizable service providers; racks and/or service equipment are considered in this category. **Waste Management:** Vehicles manufactured primarily for the purpose of collection, transportation, disposal or recycling, and

		monitoring of waste; including street sweepers and sewage waste trucks. Only trucks are considered in this category. **Construction:** Vehicles primarily sold by manufactures for building, civil engineering or engineering work. (e.g.., rack trucks; stake trucks; concrete mixers; dumpers; empty flatbeds; and flatbeds carrying construction materials or equipment). Only trucks are considered in this category. **General CV:** When a commercial vehicle can't be classified in any of the commercial **ACT_TYP** described below it will follow this category. This can be attributed to low resolution; occlusion; or absence of logo; and/or lack of equipment. Otherwise, **"Unknown".**
BODY_TYP	Motorcycle, Passenger vehicle, Bus, Bus/other transit, RVs, Work Van, CM Pick-up, Bus, Single Unit, Trailer, Multi-trailer.	Indicate the vehicle's body type. The relationship between this attribute and **ACT_TYP** is also described below**.** **If ACTY_TYP = "Private",** **Motorcycle:** All two-or three wheeled motorized vehicles. Typical vehicles in this category have saddle type seats and ae steered by handlebars rather than steering wheels. Includes motorcycles, motor scooters, and 3-wheel motorcycles. Due to low video resolution is not possible to accurately distinguish between private and emergency motorcycles. Therefore, for this project, <u>all motorcycles</u> are classified as "**Private**". **Passenger vehicle:** Sedan, coupes, SUVs, mini-van, passenger-vans and pick-ups manufactured primarily for the purpose of carrying passengers with maximum seating capacity of 8 people. It includes those pulling recreational or other light trailers. Please see **CM Pick-up** and **Bus/Van**'s description for <u>exceptions to this category</u>. This category can also be classified as **ACT_TYP = "Emergency"** (e.g., police cars)**.** **If ACT_TYP = "Recreational",** **RVs:** Vehicle designed or modified for recreation or camping, including campervans, motor-house, campervans and truck campers. **If ACT_TYP = "Public Transit",** **Bus**: Vehicles manufactured as traditional bus passenger-carrying buses with two axles and six tires; or three or more axles.

		If ACT_TYP = "Other Transit", **Bus/vans:** Vehicles manufactured as traditional bus passenger-carrying buses (e.g., chatter bus, coach bus, school bus, short bus) with a minimum seating capacity of ten people; including passenger vans (FHWY - Class 3). Pick-ups used for commercial purposes follow the category **"Service"** of the **ACT_TYP** attribute. For this project **ACT_TYP = "Construction"** was not considered due to low resolution of the video footage., **CM Pick-up:** This category is limited to pick-ups that meet at least one of the following conditions: a. Pick-up with covered cargo area higher than the cabin roof; b. Pick-up carrying service equipment, barricades and road signs; c. Pick-up with two or more of the following features: i. rails for mounting with or without ladders, ii. covered cargo area with same height as the cabin, iii. roof clearance lights, iv. Company logo, v. truck tool boxes, and vi. side Boards. **Work-Van:** Unibody vehicle which includes mini-vans, vans and step-vans, with partial or without windows in the rear, manufactured primarily for the purpose of commercial or emergency (e.g., ambulances). Some are similar in size and design as passenger vans and passenger mini-vans; others are much larger, often having roll-up rear doors. Depending on the company or/and presence of service equipment follow one of the following **ACT_TYP** categories: **"Freight", "Service" or "General CM".** Truck categories depending on their configuration follow on of the following **ACT_TYP** categories: **"Freight", "Service", "Waste Management", "Construction" or "Emergency":** **Single unit:** Truck on a single frame, including truck tractor units traveling without a trailer. **Trailer:** Truck consisting on two units in which the pulling unit is tractor (i.e., semi-trailer unit trucks) or single unit truck (i.e., single trailer). **Multi-trailer:** Truck consisting on three or more units in which the pulling unit is tractor or single unit truck. Otherwise, **"Unknown"**
FHWA_CLASS	Class 1, Class 2, Class 3 & Class 4, Class 4, Class 3 & Class 5, Class 6,	Indicate vehicle class according to the Federal Highway (FHWA) classification's system. The relationship between this attribute (**FHWA_CLASS**) and **BODY_TYP** is also described below**.**

	Class 7, Class 8, Class 9, Class 10, Class 11, Class 12, Class 13.	**If BODY_TYP = "Motorcycle",** **Class 1:** All two- or three-wheeled motorized vehicles. **If BODY_TYP = "Passenger vehicles",** **Class 2:** Two-axle and four-tire vehicle. Due to low visibility and occlusion, Class 3 and Class 2 are combined for passenger vehicles recommended by the FHWA. **If BODY_TYP = "Bus",** **Class 4:** All vehicles manufactured as traditional passenger-carrying buses with two axles and six tires or three or more axles. **If BODY_TYP = "Bus/other transit",** **Class 3 & Class 4:** Traditional passenger-carrying buses and two-axles with four-tire vehicles (Class 3). **If BODY_TYP = "CM Pick-up" or "Work-Van",** **Class 3 & Class 5:** Two-axle single unit vehicles with four (Class 3) or six -tires (Class 5). **If BODY_TYP = "Single Unit",** **Class 5:** Truck on a single unit frame with two axles and dual rear wheels. **Class 6:** Truck on a single unit frame with three axles and dual rear wheels. **Class 7:** All trucks on a single frame with four or more axles. **If BODY_TYP = "Trailer",** **Class 8:** All vehicles with four or fewer axles consisting of two units, one of which is a tractor or a straight truck power unit. **Class 9:** All vehicles with five axles consisting of two units, one of which is a tractor or a straight truck power unit. **Class 10:** All vehicles with six or more axles consisting of two units, one of which is a tractor or a straight truck power unit. **If BODY_TYP = "Multi-trailer",**

		Class 11: All vehicles with five or fewer axles consisting of three or more units, one of which is a tractor or a straight truck power unit. **Class 12:** All vehicles with six axles consisting of three or more units, one of which is a tractor or a straight truck power unit. **Class 13:** All vehicles with seven or more axles consisting of three or more units, one of which is a tractor or a straight truck power unit. Otherwise, **“Unknown”**
NO_AXLES	2 axles, 2 axles +, 3 axles, 4 axles +, 3 or 4 axles, 5 axles, 6 axles +, 5 axles or less, 6 axles, 7 axles +	Finally, data collectors will classify each vehicle by the number of axles touching the ground, without considering recreational or other light trailers, based on the Federal Highway Administration (FHWA) vehicle’s classification system.

Appendix B – Cluster Evaluation

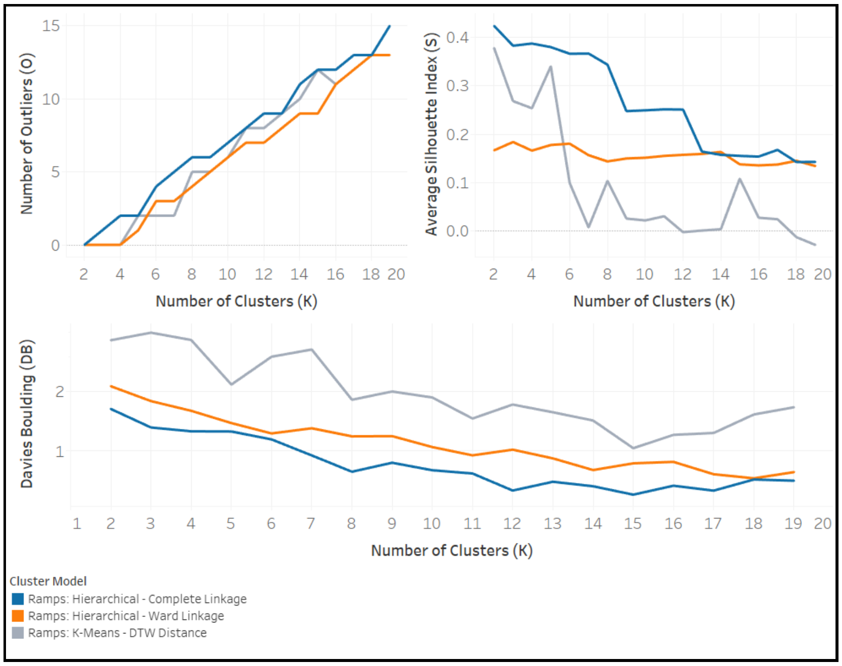


Figure B-1. Internal Validation Metrics of K-Means, Hierarchical Ward and Complete Linkage for Ramps Pre-classification

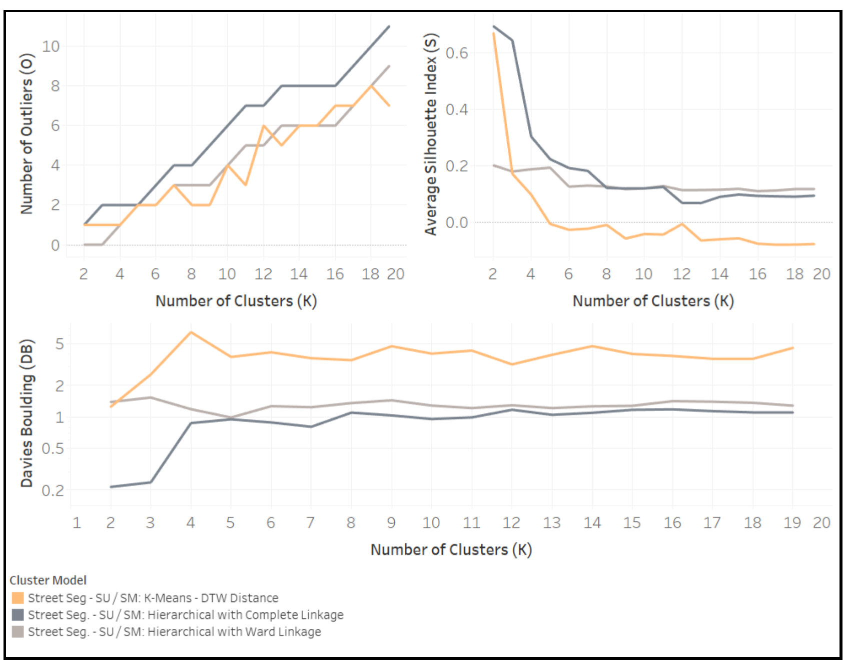


Figure B-2. Internal Validation Metrics of K-Means, Hierarchical Ward and Complete Linkage for Street Segment - Single-Unit Trucks / Small Fleet Pre-classification

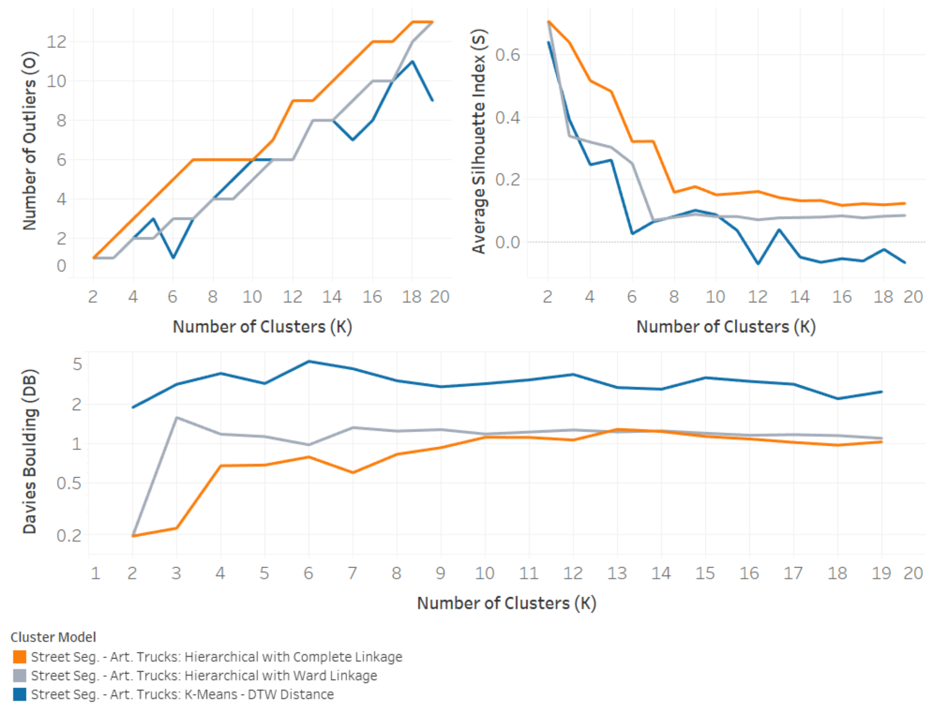


Figure B-3. Internal Validation Metrics of K-Means, Hierarchical Ward and Complete Linkage for Street Segment - Articulated Trucks Pre-classification

Figure B-4. Internal Validation Metrics of Hierarchical Ward Linkage for Ramps Pre-classification and 0, 50 and 100 CV ADT Minimum Thresholds

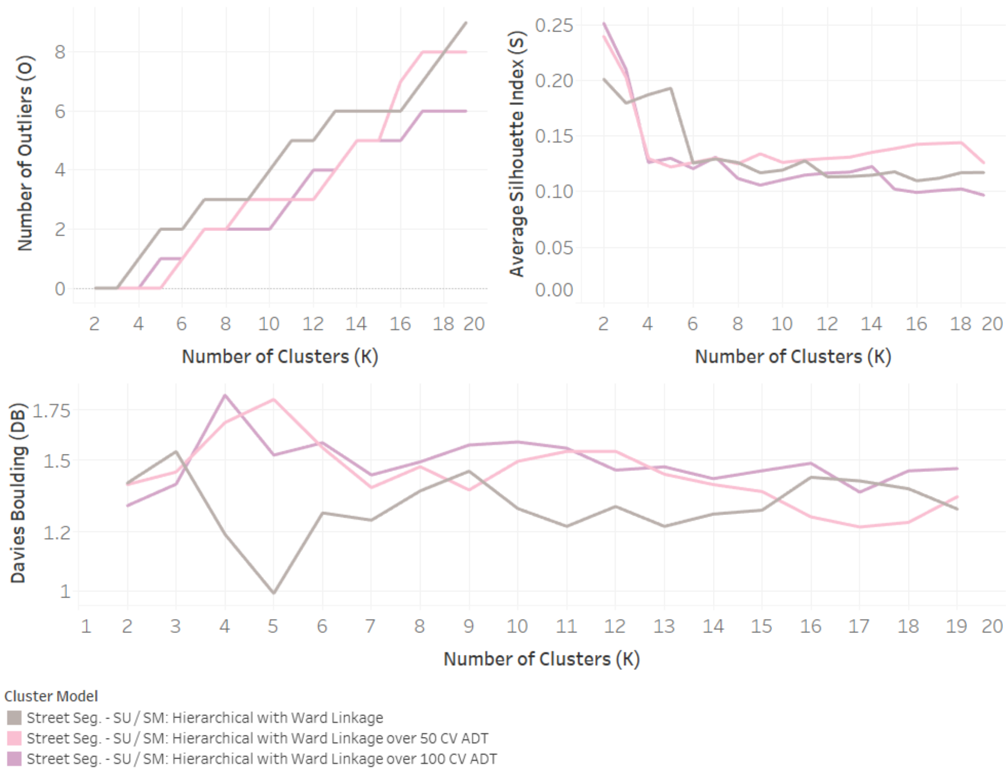


Figure B-5. Internal Validation Metrics of Hierarchical Ward Linkage for Street Segment - Small Fleet / Single Unit Pre-classification and 0, 50 and 100 CV ADT Minimum Thresholds

Figure B-6. Internal Validation Metrics of Hierarchical Ward Linkage for the pre-classification of AT in Street Segment and 0, 50 and 100 CV ADT Minimum Thresholds

Appendix C – Python Script

```
#Date 10/26/22
#Title: Script for cluster analysis and characterization

#import packages
import pandas as pd
import os
import numpy as np
import matplotlib.pyplot as plt
from tslearn.clustering import TimeSeriesKMeans
from test import gdb_sample
from datetime import datetime
from scipy.ndimage import gaussian_filter1d
from sklearn.decomposition import PCA
from sklearn.preprocessing  import StandardScaler
from sklearn.cluster import AgglomerativeClustering
from scipy.stats import bartlett
from sklearn.cluster import KMeans
from sklearn.metrics import davies_bouldin_score
from sklearn.metrics import silhouette_samples, silhouette_score
import matplotlib.cm as cm
from statsmodels.sandbox.archive.linalg_decomp_1 import CholArray
from scipy.cluster.hierarchy import dendrogram, linkage
from dtaidistance import dtw
import scipy.spatial.distance as ssd
from dtaidistance.clustering import hierarchical
from varname import nameof

#general preferences
pd.set_option('display.max_columns', 1000)
pd.set_option('display.max_rows', 1000)

# Define globals

# database import parameters
dtypes_ = {'GATE_ID': 'str', 'FLOW_DIR': 'str', 'LEG': 'str',
           'DAY_WEEK': 'str', 'GEN_BODY': 'str', 'ACT_TYPE': 'str',
           'BODY_TYPE': 'str', 'NO_AXLES': 'str','ramp_name': 'str',
           'batch': 'str', 'Veh_Count': 'float64', 'NEW_FLOW_DIR': 'str',
           'Cor_Dir': 'str', 'Group_Dir': 'str', 'CV_TYPE':'str'}
parse_dates_ = ['time']

# Directory paths
# inputs

working_dir_path = r'G:\Shared drives\Dissertation\Cordon\Analysis'
file_path = working_dir_path + '\db_all_long.csv'
overall_features_path = working_dir_path + '\Cluster\Summary_Statistics - per
leg_dir.xlsx'
analysis_out_path = working_dir_path + r'\Final Analysis'

# cluster parameters
experiment_num = 'HR_Sh_Wrd_Lg_Art_100' # use this value to create outputs for each
experiment

timeseries_cl = True # Run clusters with time series (True) or features (False)
final_ft = True
PCA_override = False

hierarchical = True
DTW_hier = False
linkage_ = 'ward'

pre_select = True # if TRUE, apply pre-classification n by infra type
```

```
onlyramps = False # if TRUE, only ramps are considered. If false, only legs are
considered
pre_select_body = True # if TRUE, apply pre-classification by body type
art_trucks = True  # if TRUE, only art trucks are considered. If false, only Single Unit
Small Fleet are considered
pre_select_act = False # if TRUE, apply pre-classification by activity type (construction
and waste management)
trim_curves = False # if TRUE, trim ends of the curve between trim_st and trim_end hours
trim_st = 6
trim_end = 18
filter_curves = False # if TRUE, select curves from external list
AM_PM = True # if TRUE, select curves from AM_PM list, otherwise Business_hr list is used

# 2-step clustering inputs
if AM_PM:
    AM_PM_path = working_dir_path + '\Cluster\AM_PM_labels.csv'
else:
    AM_PM_path = working_dir_path + '\Cluster\Business_Hr.csv'

ACT_Filters = ['construction', 'waste management']
if pre_select_act:
    Veh_var = ['ACT_TYPE']
else:
    Veh_var = ['CV_TYPE']

min_sample_size = 100 # threshold to drop curves with low daily volume
outlier_min = 4 # threshold to define outliers (number of DPFs in cluster)
seed = 0 # random seed to obtain same results across cluster runs
cluster_number = 12 # number of clusters
cluster_range = 8 # +- number of clusters for model runs
minutes_agg = 60 # temporal aggregation of time series for clustering
max_plots_per_row = 5 # for time series clustering, number of figures per row in output
agg_grouper = ['ID', 'LEG', 'NEW_FLOW_DIR'] + Veh_var # feature grouping for experiment
gof_ncluster = cluster_number+cluster_range
y_axis = 30 # height of time series clustering output plots
smooth_par = 3 # for peak hour curve metrics, smoothing approximation parameter
concave_curve_len_over = 3 # minimum 15 min time period
PCA_var_cutoff = 0.85 # variance threshold to select # of PCA components
overall_feature_keep = ['ID1', 'LEG1']
PK_metric = 'Perc' # Calculate Peak hour metrics of curves ("Perc": Percentage,
"Veh_Count": count of vehicles)
min_PK_val = 6 # (PERCENT) WARNING: ONLY WORKS IF PK_metric = 'Perc'
activity_dip_thresh = 7

curve_metrics = ['MaxPKVol', 'MaxPKtime', 'PKspread',
                 'PKIncrRate', 'PKDecRate', 'NumPKs', 'CenterMass',
                 'PK_Start', 'PK_End', 'tail_length', 'hump_incr_t',
                 'hump_width', 'quartile_PK_ev', 'PK_time_2nd', 'hump_STD',
                 'sharpness', 'mean_Vol', 'median_Vol', 'hourly_STD',
                 'hump_num', 'hump_all_width']

noncv_metrics = ['MaxPKVol_noncv', 'MaxPKtime_noncv', 'non_cv_ampkvol', 'non_cv_ampkh',
'non_cv_pmpkvol', 'non_cv_pmpkh']
script_metrics = ['NonCV_vs_CV_PK_time', 'PT_Share', 'Daily_All', 'Daily_PT',
'Daily_CV']# This metrics are added to data frame later in the script

# Features to include in feature-based cluster
features = ['hump_width',
                       'hump_STD',
                       'CenterMass',
                       'Daily_CV',
                       'hourly_STD',
                       'MaxPKtime_noncv',
                       'non_cv_overlap',
                       '18_6_CV_Perc',
                       'hump_num',
                       'MaxPKVol',
```

```
                        'MaxPKtime',
                        '12_CV_Perc',
                        '8_CV_Perc']


export_peak_figs = False
hump_stats_agg_min = 60
hump_width_threshold = 7 #percent
hump_tail_threshold = 2 #percent
quantile_PK_spread = 0.75

# Create Parameter Log Export for Model Run
par_values = [timeseries_cl, final_ft, PCA_override,
    hierarchical, DTW_hier, linkage_, pre_select,
    onlyramps, pre_select_body, art_trucks,
    pre_select_act, ACT_Filters, outlier_min,
    seed, cluster_number, cluster_range,
    minutes_agg, max_plots_per_row, agg_grouper, min_sample_size,
    gof_ncluster, y_axis, smooth_par, concave_curve_len_over,
    PCA_var_cutoff, overall_feature_keep, PK_metric,
    min_PK_val, curve_metrics, experiment_num, noncv_metrics, script_metrics,
    original_features, final_features, export_peak_figs, hump_stats_agg_min,
    hump_width_threshold, hump_tail_threshold, quantile_PK_spread]
par_keys = ['timeseries_cl', 'final_ft', 'PCA_override',
    'hierarchical', 'DTW_hier', 'linkage_', 'pre_select',
    'onlyramps', 'pre_select_body', 'art_trucks',
    'pre_select_act', 'ACT_Filters', 'outlier_min',
    'seed', 'cluster_number', 'cluster_range',
    'minutes_agg', 'max_plots_per_row', 'agg_grouper', 'min_sample_size',
    'gof_ncluster', 'y_axis', 'smooth_par', 'concave_curve_len_over',
    'PCA_var_cutoff', 'overall_feature_keep', 'PK_metric',
    'min_PK_val', 'curve_metrics', 'experiment_num', 'noncv_metrics', 'script_metrics',
    'original_features', 'final_features', 'export_peak_figs', 'hump_stats_agg_min',
    'hump_width_threshold', 'hump_tail_threshold', 'quantile_PK_spread']

par_dict = {}
for key, val in zip(par_keys, par_values):
    par_dict[key] = val

par = pd.DataFrame(par_dict.items(), columns = ['Parameter', 'Value'])

# Classes
class curve(object):

    global smooth_par, analysis_out_path, min_concave_curve_len, PK_metric

    def __init__(self, id, time_series, experiment_num):

        # Store id
        self.id = id

        # store time series
        self.time_series = time_series

        # store roll hour
        self.rollhr = time_series.rolling(4).sum().reset_index()
        self.rollhr[PK_metric] = self.rollhr[PK_metric].shift(-3)
        self.rollhr = self.rollhr.dropna()

        self.experiment_num = experiment_num

        self.label = "{}_{}".format(self.experiment_num, self.id)

        self.smooth = []

        self.smooth_d2 = []
```

```
        self.infls =[]

        self.ConcaveInfl = []

        self.peaks = []

    def meanVol(self, period_min):

        ts_agg = self.time_series.reset_index().groupby(pd.Grouper(key = 'time', freq =
str(period_min) + 'min')).sum().reset_index()
        mean_vol = ts_agg[PK_metric].mean()

        return mean_vol

    def quantileVol(self, period_min, q = 0.5): #default set to median

        ts_agg = self.time_series.reset_index().groupby(pd.Grouper(key = 'time', freq =
str(period_min) + 'min')).sum().reset_index()
        q_vol = ts_agg[PK_metric].quantile(q)

        return q_vol

    def getConcaveInfl(self):
        # Smooth Curves

        self.smooth = gaussian_filter1d(self.rollhr[PK_metric], smooth_par)

        # compute second derivative
        self.smooth_d2 = np.gradient(np.gradient(self.smooth))
        concave_indx = [indx for indx, val in enumerate(self.smooth_d2) if val < 0] #
locations with concave curve

        # find inflexion points
        self.infls = np.where(np.diff(np.sign(self.smooth_d2)))[0]
        infls_shift = self.infls.tolist().copy()

        # add start/end curve  points
        infls_shift.insert(0, -1)
        infls_shift.append(len(self.smooth_d2)-1)

        curve_pairs = [[infls_shift[i]+1, infls_shift[i+1]] for i in
range(0,len(infls_shift)-1)]
        concave_curves = [curve for curve in curve_pairs if curve[0] in concave_indx]
        self.ConcaveInfl = [curve for curve in concave_curves if curve[1]-curve[0] >
concave_curve_len_over]

    def getPeaks(self):

        peak_list = []
        peak_vol = []

        for peak in self.ConcaveInfl:

            peak_i = peakhr(peak, self.rollhr.loc[peak[0]:peak[1]])
            peak_i_vol = peak_i.PKVol

            peak_list.append(peak_i)
            peak_vol.append(peak_i.PKVol)

        self.peaks = peak_list

        return self.peaks, peak_vol

    def getCenterMass(self): #get center of mass of time series
        ts = pd.DataFrame(self.time_series).reset_index()
        ts['time_num'] = ts.time.dt.hour + ts.time.dt.minute/60
        x = np.average(ts.time_num, weights = ts[PK_metric])
```

```
        y = np.average(ts[PK_metric], weights = ts.time_num)

        return x, y

    def getlentail(self, period_min, tail_threshold):

        ts_hourly = self.time_series.reset_index().groupby(pd.Grouper(key = 'time', freq
= str(period_min) + 'min')).sum().reset_index()
        tail_len = ts_hourly[ts_hourly[PK_metric] <= tail_threshold].shape[0]

        return tail_len

    def gethumpshape(self, period_min, hump_threshold):

        ts_hourly = self.time_series.reset_index().groupby(pd.Grouper(key = 'time', freq
= str(period_min) + 'min')).sum().reset_index()
        hump = ts_hourly[ts_hourly[PK_metric] >= hump_threshold]

        lists_ = {}
        len_i = 0

        #activity dip descriptor
        max_i = 0
        min_i = 100
        dip_time = np.nan

        for index, row in ts_hourly.iterrows():

            if row[PK_metric] > hump_width_threshold:
                len_i += 1
            else:
                if len_i > 0:
                    lists_[index] = len_i
                    len_i = 0

        if len_i > 0: #store if hump ends at the end of the curve
            lists_[index] = len_i
            len_i = 0

        hump_width = hump.shape[0]
        PK_indx = hump[PK_metric].idxmax()

        try:
            hump_ends_after_PK = [t for t in lists_.keys() if t >= PK_indx]
            hump_width = lists_[hump_ends_after_PK[0]]
            hump_all_width = sum(lists_.values())
        except:
            print('ERROR: hump calcuation failed')
            print(ts_hourly)
            print(lists_)
            print(PK_indx)
            hump_width = np.nan
            hump_all_width = np.nan

        PK_st = hump['time'].loc[PK_indx].hour + hump['time'].loc[PK_indx].minute/60
        hump_st = hump['time'].iloc[0].hour + hump['time'].iloc[0].minute/60
        hump_end = hump['time'].iloc[-1].hour + hump['time'].iloc[-1].minute/60
        incr_time = PK_st - hump_st

        hump_indx = hump.iloc[[0,-1]].index.values.tolist()
        hump_indx = range(hump_indx[0], hump_indx[-1]+1)
        hump_STD = ts_hourly.loc[hump_indx, PK_metric].std()
        hourly_STD = ts_hourly.loc[:, PK_metric].std()
        hump_num = len(lists_.keys())
        return incr_time, hump_width, hump_STD, hourly_STD, hump_num, hump_all_width

    def __str__(self):   # used in print statement
```

```
        return self.label

    def getQroll(self, q):
        return self.rollhr[PK_metric].quantile(q)

    def savefig(self):

        # plot results
        plt.figure()
        plt.plot(self.rollhr[PK_metric], label='Noisy Data')
        plt.plot(self.smooth, label='Smoothed Data')
        plt.plot(self.smooth_d2 / np.max(self.smooth_d2), label='Second Derivative
(scaled)')
        for i, infl in enumerate(self.ConcaveInfl, 1):
            plt.axvline(x=infl[0], color='k', label=f'Inflection Point {i}')
            plt.axvline(x=infl[1], color='k')
        plt.legend(bbox_to_anchor=(1.55, 1.0))
        plt.title(self.label)
        plt.savefig(r'{}\fig\Peaks\{}.png'.format(analysis_out_path, self.label),
                    bbox_inches='tight')
        plt.close()

class peakhr(object):

    def __init__(self, infls_pair, rollhr):

        self.start = infls_pair[0]
        self.end = infls_pair[1]
        self.times = rollhr.time
        self.vols = rollhr[PK_metric]
        self.PKVol = max(self.vols)

    def getPKspread(self):

        return (self.end - self.start)*15/60 # hours

    def getPKtime(self):
        time = self.times.iloc[list(self.vols).index(self.PKVol)]
        return time.hour + time.minute/60 # hours

    def getIncreaseRate(self):
        peak_interval = self.getPKtime()*60/15
        hight = self.PKVol - self.vols.loc[self.start]
        base = (peak_interval - self.start)*15/60 #veh/hr
        try:
            result = hight/base
        except:
            result = np.nan
            print('getIncreaseRate calculation failed')

        return result

    def getDecreaseRate(self):
        peak_interval = self.getPKtime()*60/15
        hight = self.PKVol - self.vols.loc[self.end]
        base = (peak_interval - self.end)*15/60 #veh/hr

        try:
            result = hight/base
        except:
            result = np.nan
            print('getDecreaseRate calculation failed')

        return result

# import database
```

```python
db_raw = pd.read_csv(file_path,
                        dtype = dtypes_,
                        header = 0,
                        parse_dates = parse_dates_)
db1 = db_raw.copy()

if len(overall_feature_keep) > 2:
    db_ovll_ft_raw = pd.read_excel(overall_features_path,
                                        header = 0,
                                        sheet_name = 'Sum_Stadistics',
                                        parse_dates = ['D_PK_HR', 'D_PK_HR_1', 'D_PK_HR_2'])
    db_ovll_ft = db_ovll_ft_raw[overall_feature_keep].copy()

am_pm_labels = pd.read_csv(AM_PM_path)

print('Database import sucessful')

#Merge final database

################### FILTERS ####################################################
# Filter curves for AM PM test
if filter_curves:
    db1['label'] = db1[agg_grouper].agg('_'.join, axis = 1) # create unique ID for curves

    db1 = db1[db1.label.isin(am_pm_labels.label.values.tolist())]
    db1.drop(columns = 'label', inplace = True)
    print('filter_curves')
    print(len(db1.groupby(agg_grouper).size().reset_index().index))

    cl_label_dict =  {}
    pattern_label_dict = {}
    for index, row in am_pm_labels.iterrows():

        cl_label_dict[row.label] = row.CL_Label
        pattern_label_dict[row.label] = row.Pattern_Label

#Filter for ramps /legs
if pre_select:
    if onlyramps:
        # Only ramps
        db1 = db1[(db1.LEG == 'Off') | (db1.LEG == 'On')]
    else:
        #Only legs
        db1 = db1[(db1.LEG != 'Off')]
        db1 = db1[(db1.LEG != 'On')]

#Filters for Activity and Body Type
#db1 = db1[db1.ACT_TYPE == 'service']
if pre_select_body:
    if art_trucks:
        db1 = db1[db1.CV_TYPE == 'Articulated Trucks']
    else:
        db1 = db1[db1.CV_TYPE != 'Articulated Trucks']

elif pre_select_act:

    db1 = db1[db1.ACT_TYPE.isin(ACT_Filters)]

# filter out locations
db1 = db1[~db1.batch.str.contains('Batch')] # Filter locations of Chris new

# filter out non-CVs, Public Transit (PT)
db_noncv = db_raw[db_raw.CV_TYPE == 'Non-CV'].copy()

db_PT = db_noncv[db_noncv.ACT_TYPE == 'public transit']

db1 = db1[db1.CV_TYPE != 'Non-CV'].copy()
```

```
# Filter CV curves with little volume
daily_CV_vol = db1.groupby(agg_grouper)['Veh_Count'].sum().reset_index()
curves_keep = daily_CV_vol[daily_CV_vol.Veh_Count > min_sample_size] # curves that min
criteria
curves_keep = list(curves_keep[agg_grouper].itertuples(index = False, name = None)) #
transfor to tuples
db = db1[[x in curves_keep for x in list(zip(*[db1[col] for col in agg_grouper]))]] #
filtered database

# Calculate daily volumes for overall stats

# Daily CV volume
daily_CV_vol.rename(columns = {'Veh_Count': 'Daily_CV'}, inplace = True)

# Daily Public Transit (PT) ##Volume Activity_type public transit
# grouper is edited to ensure one curve per location or location/direction is computed
for non_cv flows
agg_gruper_noncv_all = ['ID', 'LEG', 'NEW_FLOW_DIR']
agg_grouper_noncv = [col for col in agg_grouper if col in agg_gruper_noncv_all]
daily_PT_vol = db_PT.groupby(agg_grouper_noncv)['Veh_Count'].sum().reset_index()
daily_PT_vol.rename(columns = {'Veh_Count': 'Daily_PT'}, inplace = True)

# Daily total volume
daily_all_vol = db_raw.groupby(agg_grouper_noncv)['Veh_Count'].sum().reset_index()
daily_all_vol.rename(columns = {'Veh_Count': 'Daily_All'}, inplace = True)

# filter by time of day
if trim_curves:
    db1 = db1[(db1.time.dt.hour >= trim_st) & (db1.time.dt.hour < trim_end)]

############################## PERCENTAGE CALCULATION ###################
# Calculate percentage of total
db['Perc'] = 100*db['Veh_Count'] / db.groupby(agg_grouper)['Veh_Count'].transform('sum')
db_noncv['Perc'] = 100*db_noncv['Veh_Count'] /
db_noncv.groupby(agg_grouper_noncv)['Veh_Count'].transform('sum')

############################## NONCV CURVE METRICS ######################################

# Aggregate by grouper
db_noncv_transf = db_noncv.groupby(agg_grouper_noncv+['time']).sum().reset_index()
ids = db_noncv_transf[agg_grouper_noncv].drop_duplicates() # Unique combinations of
curves
df_noncv_curve_fts = pd.DataFrame(columns = agg_grouper_noncv + noncv_metrics)

for i in range(len(ids.index)):

    # select data for each curve in ids
    id = ids.iloc[i].tolist()
    curves_keep = [tuple(id)]

    curve_data = db_noncv_transf[[x in curves_keep for x in
list(zip(*[db_noncv_transf[col] for col in agg_grouper_noncv]))]] # filtered database
    curve_data = curve_data.set_index('time')[PK_metric]

    label = ('_'.join(curves_keep[0]).replace(' ',''))

    if curve_data.sum() == 0:
        print('Curve ', label, ' shows no vehicles')

        continue

    # Calculate curve metrics
    try:

        curve_i = curve(label, curve_data, experiment_num)
        curve_i.getConcaveInfl()
```

```
        if export_peak_figs:
            curve_i.savefig()
        peaks, vols = curve_i.getPeaks()

        non_cv_ampkvol = 0
        non_cv_pmpkvol = 0
        non_cv_ampkh = np.nan
        non_cv_pmpkh = np.nan

        for indx, pk in enumerate(peaks):

            pkh_i = pk.getPKtime()
            pkvol_i = vols[indx]

            if pkh_i < 12 and pkvol_i > non_cv_ampkvol:
                non_cv_ampkh = pkh_i
                non_cv_ampkvol = pkvol_i
            elif pkh_i >= 12 and pkvol_i > non_cv_pmpkvol:
                non_cv_pmpkh = pkh_i
                non_cv_pmpkvol = pkvol_i


        overall_PK = peaks[vols.index(max(vols))]

        list_ = [max(vols), overall_PK.getPKtime(), non_cv_ampkvol, non_cv_ampkh,
non_cv_pmpkvol, non_cv_pmpkh]

        list_ = [item if item > 0 else np.nan for item in list_ ]

        # store curve metric results for noncv curve_i
        df_noncv_curve_fts.loc[i] = id+list_

    except:
        print('ERROR: could not obtain curve charactersitcs for ', label)

print('Non_CV Curve Metrics Succesful!')

non_cv_ts =
db_noncv.groupby(agg_grouper_noncv+['time'])['Perc'].sum().unstack('time').reset_index()

############################# CV CURVE METRICS ####################################
# Aggregate by grouper

db_transf = db.groupby(agg_grouper+['time']).sum().reset_index()

ids = db_transf[agg_grouper].drop_duplicates() # Unique combinations of curves

df_curve_fts = pd.DataFrame(columns = agg_grouper + curve_metrics)

for i in range(len(ids.index)):

    # select data for each curve in ids
    id = ids.iloc[i].tolist()
    curves_keep = [tuple(id)]

    curve_data = db_transf[[x in curves_keep for x in list(zip(*[db_transf[col] for col
in agg_grouper]))]] # filtered database
    curve_data = curve_data.set_index('time')[PK_metric]

    label = ('_'.join(curves_keep[0]).replace(' ',''))

    if curve_data.sum() == 0:
        print('Curve ', label, ' shows no vehicles')

        continue

    # Calculate curve metrics
```

```
    try:

        curve_i = curve(label, curve_data, experiment_num)
        curve_i.getConcaveInfl()
        if export_peak_figs:
            curve_i.savefig()
        hump_incr_time, hump_width, hump_STD, hourly_STD, hump_num, hump_all_width =
curve_i.gethumpshape(hump_stats_agg_min, hump_width_threshold)
        peaks_raw, vols_raw = curve_i.getPeaks()

        # filter out small peaks
        real_pks_indx = [indx for indx, vol in enumerate(vols_raw) if vol >= min_PK_val]
        peaks = [peaks_raw[i] for i in real_pks_indx]
        vols = [vols_raw[i] for i in real_pks_indx]

        overall_PK = peaks[vols.index(max(vols))]
        center_mass_x, center_mass_y = curve_i.getCenterMass()
        num_peaks = len(peaks)
        overall_PK_vol = max(vols)
        sharpness = overall_PK_vol / curve_i.meanVol(hump_stats_agg_min)
#curve_i.getQroll(quantile_PK_spread)

        if num_peaks > 1:
            pk_vols_order = vols.copy()
            pk_vols_order.remove(max(vols))
            pk_vols_order = list(set(pk_vols_order))
            pk_vols_order.sort()
            PKH_2nd_indx = vols.index(pk_vols_order[-1])
            PKH_2nd_time = peaks[PKH_2nd_indx].getPKtime()
        else:
            PKH_2nd_time = 24 #next day same peak

        list_ = [overall_PK_vol, overall_PK.getPKtime(),
                 overall_PK.getPKspread(),
                 overall_PK.getIncreaseRate(),
                 overall_PK.getDecreaseRate(),
                 num_peaks, center_mass_x, overall_PK.start, overall_PK.end,
                 curve_i.getlentail(hump_stats_agg_min, hump_tail_threshold),
                 hump_incr_time, hump_width,
                 curve_i.getQroll(quantile_PK_spread),
                 PKH_2nd_time, hump_STD, sharpness,
                 curve_i.meanVol(hump_stats_agg_min),
                 curve_i.quantileVol(hump_stats_agg_min),
                 hourly_STD, hump_num, hump_all_width]

        #sharpness
        # store curve metric results for curve_i
        df_curve_fts.loc[i] = id+list_

    except:
        print('ERROR: could not obtain curve characteristics for ', label)

if not timeseries_cl:
    experiment_num = experiment_num + '_PCA'
print('Curve - Peak Hour statistics successful!')

df_all_ftrs = df_curve_fts.copy()

# Add daily CV volume variable to features master dataframe
df_all_ftrs = df_all_ftrs.merge(daily_CV_vol, on = agg_grouper, how = 'left')
df_all_ftrs = df_all_ftrs.merge(daily_PT_vol, on = agg_grouper_noncv, how = 'left')
df_all_ftrs = df_all_ftrs.merge(df_noncv_curve_fts, on = agg_grouper_noncv, how = 'left')
df_all_ftrs = df_all_ftrs.merge(daily_all_vol, on = agg_grouper_noncv, how = 'left')

# CV Share in time periods 2 -5 am, 9 am - 12 pm and 7 - 10 pm.
# Share of volumes per period calculated throughout the day but only use 9am-12pm for now
period_min = 60
```

```
db_cv_period = db.groupby(agg_grouper+[pd.Grouper(key = 'time', freq = str(period_min) +
'min')]).sum().reset_index()
db_cv_period.time = db_cv_period['time'].dt.strftime('%H:%M')
db_cv_period =
db_cv_period.groupby(agg_grouper+['time'])['Perc'].sum().unstack('time').reset_index()

#rename col names
#new_vol_cols =
[str(int(col.split(':')[0]))+'_'+str(int(col.split(':')[0])+int(period_min/60))+"_CV_Perc
" for col in db_cv_period.columns.values.tolist()[len(agg_grouper):]]
new_vol_cols = [str(int(col.split(':')[0])) for col in
db_cv_period.columns.values.tolist()[len(agg_grouper):]]
db_cv_period.columns = agg_grouper + new_vol_cols

off_hours = list(range(0,5)) + list(range(18,24))
off_hours = [str(h) for h in off_hours]

####Variables Time
db_cv_period['3_6_CV_Perc'] = db_cv_period[['3', '4', '5']].agg(['sum'], axis ="columns")
db_cv_period['7_9_CV_Perc'] = db_cv_period[['7', '8']].agg(['sum'], axis ="columns")
db_cv_period['12_CV_Perc'] = db_cv_period[['12']].agg(['sum'], axis ="columns")
db_cv_period['10_CV_Perc'] = db_cv_period[['10']].agg(['sum'], axis ="columns")
db_cv_period['8_CV_Perc'] = db_cv_period[['8']].agg(['sum'], axis ="columns")
db_cv_period['14_CV_Perc'] = db_cv_period[['14']].agg(['sum'], axis ="columns")
db_cv_period['6_CV_Perc'] = db_cv_period[['6']].agg(['sum'], axis ="columns")
db_cv_period['17_19_CV_Perc'] = db_cv_period[['17', '18']].agg(['sum'], axis ="columns")
db_cv_period['18_21_CV_Perc'] = db_cv_period[['18', '19', '20']].agg(['sum'], axis
="columns")
db_cv_period['18_6_CV_Perc'] = db_cv_period[off_hours].agg(['sum'], axis ="columns")

intervals = ['3_6_CV_Perc', '7_9_CV_Perc', '12_CV_Perc', '10_CV_Perc', '8_CV_Perc',
'14_CV_Perc', '6_CV_Perc', '17_19_CV_Perc', '18_21_CV_Perc', '18_6_CV_Perc']
all_features = curve_metrics + noncv_metrics + script_metrics + intervals
display_only_cols = [f for f in all_features if not f in features]

# Add CV Perc variables to features master dataframe
df_all_ftrs = df_all_ftrs.merge(db_cv_period[agg_grouper+intervals], on = agg_grouper,
how = 'left')

df_all_ftrs['NonCV_vs_CV_PK_time'] = df_all_ftrs.MaxPKtime_noncv - df_all_ftrs.MaxPKtime
df_all_ftrs['PT_Share'] = df_all_ftrs.Daily_PT/df_all_ftrs.Daily_All

#calculate non_cv_peak_overlap
pass_threshold = 6.5
df_all_ftrs.loc[df_all_ftrs['non_cv_ampkvol'] < pass_threshold, ['non_cv_ampkh',
'non_cv_ampkvol']] = [np.nan, np.nan]
df_all_ftrs.loc[df_all_ftrs['non_cv_pmpkvol'] < pass_threshold, ['non_cv_pmpkh',
'non_cv_pmpkvol']] = [np.nan, np.nan]
df_all_ftrs['non_cv_AMoverlap'] = df_all_ftrs.non_cv_ampkh-df_all_ftrs.MaxPKtime
df_all_ftrs['non_cv_PMoverlap'] = df_all_ftrs.non_cv_pmpkh-df_all_ftrs.MaxPKtime
df_all_ftrs['non_cv_overlap'] = df_all_ftrs[['non_cv_AMoverlap',
'non_cv_PMoverlap']].min(axis = 1)

# Select features to include in analysis and only display
if display_only_cols:
    cols_remain = [col for col in df_all_ftrs if not col in display_only_cols]
    df_display_ftrs = df_all_ftrs[agg_grouper + display_only_cols].copy()
    display_col_labes = [col + '_displayonly' if col in display_only_cols else col for
col in df_display_ftrs.columns]
    df_display_ftrs.columns = display_col_labes
    df_all_ftrs = df_all_ftrs[cols_remain].copy()
print(df_all_ftrs.columns)

# Filter curves for AM_PM cluster check
print('Volume characteritsics successful!')

######################### Run PCA for dimension reduction #########################
```

```
if not timeseries_cl:
    # Standarized variables for PCA analysis

    # drop nans for PCA
    len_withnas = len(df_all_ftrs)
    df_all_ftrs.dropna(how = 'any', inplace = True, axis = 1)
    len_withoutnas = len(df_all_ftrs)

    # extract and standarize explanatory variables
    x_df = df_all_ftrs.drop(columns = agg_grouper)
    var_x = [np.var(x_df[col], ddof = 1) for col in x_df.columns] # variance of variables
(see Barlett test)
    print(x_df.columns)
    x = x_df.values
    x = StandardScaler().fit_transform(x) # standarized values for PCA

    # Barlett test, check appropiateness of PCA
    stats, p = bartlett(df_all_ftrs.iloc[:,7], df_all_ftrs.iloc[:,8])
    pca_nas_num = len_withnas - len_withoutnas

    if p > 0.05 or PCA_override:
        PCA_invalid = True
    else:
        PCA_invalid = False

    if PCA_invalid:
        print('Sample does not meet Barlett test')
        print('could not reject Ho that the population correlation matrix is identity
matrix)')
        print('PCA solution is not unique and PCA cannot help to reduce the data')
        test_eval = 'Barlett test does not reject Ho - PCA inappropiate; '
        barlett_result_string = test_eval + 'test value = '+ str(stats) + ' ; p = ' +
str(p)
        print(barlett_result_string)
        experiment_num = experiment_num+ 'skip'

    else:
        test_eval = 'Barlett test rejects Ho - PCA appropiate; '
        barlett_result_string = test_eval + 'test value = '+ str(stats) + ' ; p = ' +
str(p)
        print(barlett_result_string)

        # Plot PCA variance chart
        plt.figure()
        pca_chart = PCA().fit(x)
        plt.rcParams["figure.figsize"] = (12,6)

        fig, ax = plt.subplots()
        y = np.cumsum(pca_chart.explained_variance_ratio_)
        components_chart = len(y)
        xi = np.arange(1, components_chart+1, step=1)

        plt.ylim(0.0,1.1)
        plt.plot(xi, y, marker='o', linestyle='--', color='b')

        plt.xlabel('Number of Components')
        plt.xticks(np.arange(0, components_chart+1, step=1)) #change from 0-based array
index to 1-based human-readable label
        plt.ylabel('Cumulative variance (%)')
        plt.title('The number of components needed to explain variance')

        plt.axhline(y=PCA_var_cutoff, color='r', linestyle='-')
        plt.text(0.5, 0.85, '{} cut-off threshold'.format(PCA_var_cutoff), color = 'red',
fontsize=16)

        ax.grid(axis='x')
        plt.savefig(r'{}\fig\FitStats\{}.png'.format(analysis_out_path, experiment_num),
```

```
                        bbox_inches='tight')
        plt.close()

        # Select number of PCA components based on explained variance ratio:
        try:
            PCA_comp_num = [indx for indx, var_i in enumerate(y) if var_i >
PCA_var_cutoff][0]

        except:
            PCA_comp_num = 2
            print('PCA component calcuation based on ', PCA_var_cutoff, ' explained
variance cut-off failed')
            print('PCA components set to default of ', PCA_comp_num, ' components')

        #Run PCA for selected number of components
        pca = PCA(n_components = PCA_comp_num)
        scores_pca = pca.fit_transform(x)

        #PCA components column names
        PCA_cols = ['PC'+ str(n) for n in range(1,PCA_comp_num+1)]

        # scores for output workbook
        scores_pca_df = pd.DataFrame(data = scores_pca, columns = PCA_cols)

        # calculate loadings for interpretation
        pca_loadings_df = pd.DataFrame(pca.components_.T, columns = PCA_cols, index =
x_df.columns.values.tolist())

        pca_notes = ['Number of PCA components = '+str(PCA_comp_num),
                     'Variance explained = '+str(pca.explained_variance_ratio_),
                     'Number of features dropped due to nas = ', pca_nas_num]

        for note in pca_notes:
            print(note)

        pca_notes_df = pd.DataFrame(pca_notes)

    print('PCA successful!')

######################## Format Time Series ###########
# Aggregate by time increment
ts = db.groupby(agg_grouper+[pd.Grouper(key = 'time', freq = str(minutes_agg) +
'min')]).sum().reset_index()
ts.time = ts['time'].dt.strftime('%H:%M')
ts = ts.groupby(agg_grouper+['time'])['Perc'].sum().unstack('time')

# Define values for clustering

if timeseries_cl:
    # shape for clustering with time series
    X_volumes = ts.values
    sz = X_volumes.shape[1] # parameter for plotting clustering with time series
    df = ts.reset_index()[agg_grouper].copy()

    if hierarchical & DTW_hier:
        X_volumes = dtw.distance_matrix_fast(X_volumes)
else:
    df = df_all_ftrs[agg_grouper].copy()

    if PCA_invalid:
        X_volumes = x
    else:
        X_volumes = scores_pca


ts = ts.reset_index().copy()
############################## RUN CLUSTERS ######################################
```

```
# Estimate number of clusers for each iteration
cl_per_iter = [*range(cluster_number-cluster_range,cluster_number+cluster_range+1,1)]

# Internal validation indeces
results = {}
wcss = []
silhouette = {}
cl_range = range(2,gof_ncluster)
CL_outliers = {}

for n_clusters in cl_range:
    if hierarchical:

        if timeseries_cl & DTW_hier:
            dba_km = AgglomerativeClustering(n_clusters = n_clusters,
                                             affinity = 'precomputed',
                                             linkage = linkage_)
        else:
            dba_km = AgglomerativeClustering(n_clusters = n_clusters,
                                             affinity = 'euclidean',
                                             linkage = linkage_)

    elif timeseries_cl:
        dba_km = TimeSeriesKMeans(n_clusters = n_clusters,
                                  n_init=2,
                                  metric="dtw",
                                  verbose=True,
                                  max_iter_barycenter=10,
                                  random_state=seed)

    else:
        dba_km = KMeans(n_clusters = n_clusters,
                        n_init=2,
                        random_state=seed)

    dba_km.fit(X_volumes)

    if not hierarchical:
        #Elbow
        wcss.append(dba_km.inertia_)

    #Davies-Boulding Score
    cl_labels = dba_km.fit_predict(X_volumes)
    db_index = davies_bouldin_score(X_volumes, cl_labels)
    results.update({n_clusters: db_index})

    # Count of outliers
    counts = {x: list(cl_labels).count(x) for x in list(cl_labels)}
    counts = [1 for value in counts.values() if value < outlier_min]
    CL_outliers.update({n_clusters: len(counts)})

    #Silhouette
    silhouette_avg = silhouette_score(X_volumes, cl_labels)
    sample_silhouette_values = silhouette_samples(X_volumes, cl_labels)

    silhouette.update({n_clusters: silhouette_avg})

    fig = plt.figure()
    ax1 = fig.gca()
    ax1.set_xlim([-0.1, 1])
    ax1.set_ylim([0, len(X_volumes)+(n_clusters+1)*10])
    y_lower = 10

    for i in range(n_clusters):
        # Aggregate the silhouette scores for samples belonging to
        # cluster i, and sort them
```

```
        ith_cluster_silhouette_values = sample_silhouette_values[cl_labels == i]

        ith_cluster_silhouette_values.sort()

        size_cluster_i = ith_cluster_silhouette_values.shape[0]
        y_upper = y_lower + size_cluster_i

        color = cm.nipy_spectral(float(i) / n_clusters)
        ax1.fill_betweenx(
            np.arange(y_lower, y_upper),
            0,
            ith_cluster_silhouette_values,
            facecolor=color,
            edgecolor=color,
            alpha=0.7,
        )

        # Label the silhouette plots with their cluster numbers at the middle
        ax1.text(-0.05, y_lower + 0.5 * size_cluster_i, str(i))

        # Compute the new y_lower for next plot
        y_lower = y_upper + 10  # 10 for the 0 samples

    ax1.set_title("The silhouette plot for the various clusters.")
    ax1.set_xlabel("The silhouette coefficient values")
    ax1.set_ylabel("Cluster label")

    # The vertical line for average silhouette score of all the values
    ax1.axvline(x=silhouette_avg, color="red", linestyle="--")

    ax1.set_yticks([])  # Clear the yaxis labels / ticks
    ax1.set_xticks([-0.1, 0, 0.2, 0.4, 0.6, 0.8, 1])

    fig.savefig(r'{}\fig\FitStats\{}_silhouette_{}CL.png'.format(analysis_out_path,
experiment_num, n_clusters))
    plt.close()

# Plot elbow method
if not hierarchical:

    plt.figure()
    plt.plot(cl_range, wcss, marker = 'o', linestyle = '--')
    plt.xlabel('Number of Clusters')
    plt.ylabel('WCSS')
    plt.title('{} - Kmeans with PCA Clustering'.format(experiment_num))
    plt.savefig(r'{}\fig\FitStats\{}_Elbow.png'.format(analysis_out_path,
experiment_num))
    plt.close()

# Plot Davies-Boulding Score
plt.figure()
plt.plot(list(results.keys()), list(results.values()))
plt.xlabel("Number of clusters")
plt.ylabel("Davies-Boulding Index")
plt.savefig(r'{}\fig\FitStats\{}_DaviesBoulding.png'.format(analysis_out_path,
experiment_num))
plt.close()

# Plot Silhouette Scores
plt.figure()
plt.plot(list(silhouette.keys()), list(silhouette.values()))
plt.xlabel("Number of clusters")
plt.ylabel("Average Silhouette Scores")
plt.savefig(r'{}\fig\FitStats\{}_Silhouette.png'.format(analysis_out_path,
experiment_num))
plt.close()
```

```
DB_results = pd.DataFrame({'CL_Num':list(results.keys()), 'Davies_Boulding':
list(results.values())})
silhouete_results = pd.DataFrame({'CL_Num':list(silhouette.keys()), 'Silhouette_avg':
list(silhouette.values())})
CL_outliers_results = pd.DataFrame({'CL_Num': list (CL_outliers.keys()), 'CL_Outliers':
list(CL_outliers.values())})

stats_df = pd.merge(DB_results, silhouete_results, on = 'CL_Num')
stats_df = pd.merge(stats_df, CL_outliers_results, on = 'CL_Num')


if hierarchical:
    # Dendogram
    #####pd.DataFrame(X_volumes).to_csv(analysis_out_path+'\\dtw_matrix.csv')
    plt.figure()
    if timeseries_cl & DTW_hier:
        distArray = ssd.squareform(X_volumes)
    else:
        distArray = X_volumes
    #print(distArray)
    linkage_data = linkage(distArray, method = linkage_)
    dendro = dendrogram(linkage_data)
    plt.savefig(r'{}\fig\FitStats\{}_dendro.png'.format(analysis_out_path,
experiment_num), bbox_inches = 'tight')
    plt.close()

else:
    stats_df['WCSS'] = wcss

print('Cluster stats successful!')

sub_melt = []
# Run models, export workbook and figures
for cl_num in cl_per_iter: #cl_num is number of cluster in each iteration

    # Define any model types to be run in this section (i.e., DBA, Soft DTW, Euclidean,
etc.)
    # DBA-k-means and K-means
    if hierarchical:
        if timeseries_cl & DTW_hier:
            dba_km = AgglomerativeClustering(n_clusters = cl_num,
                                  affinity = 'precomputed',
                                  linkage = linkage_)
            model_names = ['HR_dtw']#+['_']+[linkage_[0]]

        else:
            dba_km = AgglomerativeClustering(n_clusters = cl_num,
                                  affinity = 'euclidean',
                                  linkage = linkage_)
            model_names = ['HR_e'] #+ ['_'] + [linkage_[0]]

    elif timeseries_cl:
        dba_km = TimeSeriesKMeans(n_clusters = cl_num,
                                  n_init=2,
                                  metric="dtw",
                                  verbose=True,
                                  max_iter_barycenter=10,
                                  random_state=seed)

        model_names = ['DTW'] # DTW k-means


    else:
        dba_km = KMeans(n_clusters = cl_num,
                              n_init=2,
                              random_state=seed)
        model_names = ['Km'] # k-means
```

```
    model_types = [dba_km]

    # Create labels for each model run
    subtitles = []

    for model in model_names:
        sub = 'Ex{}_{}_{}CL'.format(experiment_num,model,cl_num)
        subtitles.append(sub)
        sub_melt.append(sub)
    for run, subtitle in zip(model_types, subtitles):
        plt.figure()
        y_pred = run.fit_predict(X_volumes)

        if timeseries_cl: # if time series clustering, plot curves
            if not hierarchical:
                # Get number of rows/columns of plots
                rows = cl_num // max_plots_per_row
                if cl_num % max_plots_per_row > 0:
                    rows += 1
                cols = min(cl_num, max_plots_per_row)

                for yi in range(cl_num):
                    plt.subplot(rows, cols, yi+ 1)

                    for xx in X_volumes[y_pred == yi]:
                        plt.plot(xx.ravel(), "k-", alpha=.2)

                    plt.plot(run.cluster_centers_[yi].ravel(), "r-")
                    plt.xlim(0, sz)
                    plt.ylim(0, y_axis)
                    plt.text(0.55, 0.85,'CL %d' % (yi + 1),
                             transform=plt.gca().transAxes)
                    if yi == 0:
                        plt.title(subtitle)

                plt.savefig(r'{}\fig\TimeSeriesCL\{}.png'.format(analysis_out_path,
subtitle))
                plt.close()
        #plt.show()

        df[subtitle] = y_pred + 1

        print('##################################################')
        print('Model run ', subtitle, ' succesfully completed!')

# Merge clustering results with features df
df = df.merge(df_all_ftrs, on = agg_grouper, how = 'left')

# Merge clustering results with timeseries df
df = df.merge(ts, on = agg_grouper, how = 'left')
id_vars_ = [col for col in df.columns.tolist() if not col in sub_melt]

# Reshape clusters to long from wide format
df = df.melt(id_vars = id_vars_, value_vars = sub_melt, var_name = 'Model_Run',
value_name = 'CL_Label')
df['label'] = df[agg_grouper].agg('_'.join, axis = 1) # create unique ID for curves


if display_only_cols:
    df = df.merge(df_display_ftrs, on = agg_grouper, how = 'left')

Excelwriter =
pd.ExcelWriter(r'{}\Results_Workbooks\Ex{}_output.xlsx'.format(analysis_out_path,experime
nt_num),engine="xlsxwriter")
df.to_excel(Excelwriter, index = False, sheet_name = 'ClusterAnalysis')
```

```python
# If PCA clustering, add PCA results to dataframe
if not timeseries_cl:

    if not PCA_invalid:
        pca_loadings_df.to_excel(Excelwriter, index = True, sheet_name = 'PCA_loadings')
        pca_notes_df.to_excel(Excelwriter, index = False, sheet_name = 'PCA_notes')

non_cv_ts.to_excel(Excelwriter, index = False, sheet_name = 'Non_CV_DailyShare')
stats_df.to_excel(Excelwriter, index = False, sheet_name = 'CL_stats')
par.to_excel(Excelwriter, index = False, sheet_name = 'Model_Parameters')

Excelwriter.save()

print('files successfully exported!')
```