ABSTRACT

Title of Dissertation: ADVANCING TRANSPORTATION ANALYSIS WITH MOBILE DEVICE LOCATION DATA: INSIGHTS INTO BEHAVIOR, CONGESTION, AND ENVIRONMENTAL IMPACTS

Guangchen Zhao, Doctor of Philosophy, 2026

Dissertation directed by: Professor, Cinzia Cirillo, Department of Civil and Environmental Engineering

Emerging technologies, such as Mobile Device Location Data (MDLD), are reshaping the landscape of transportation analysis by offering continuous, large-scale observations of travel behavior. This dissertation explores the potential of MDLD in addressing critical challenges in transportation planning, focusing on data integration, travel behavior analysis, road congestion, and environmental impacts.

The research begins by addressing the limitations of traditional travel surveys, which often suffer from small sample sizes and short observation periods. To overcome these challenges, this study integrates MDLD with regional travel survey data through probabilistic record linkage algorithms, creating comprehensive datasets that combine detailed demographic information with extensive, multi-day travel records. These fused datasets capture underreported travel behaviors and provide a more holistic representation of population mobility, laying a strong foundation for advanced transportation research.

The second project validates MDLD's ability to capture personal travel behaviors by examining the relationship between vehicle miles traveled (VMT) and gasoline prices in the Washington, DC, metro area. Using longitudinal MDLD datasets, we uncover reasonable price elasticities of VMT and demonstrate how aggregate VMT informs local gasoline demand. This analysis establishes the reliability of MDLD for studying travel behavior and energy consumption trends.

The study then investigates changes in road congestion in the post-pandemic era, with a particular focus on the role of teleworking. Using MDLD-based measures of travel demand, trip distance, and congestion patterns in the Washington, D.C. region, the analysis reveals that while the frequency of commute trips has not fully recovered, average trip distances and non-work travel have increased, contributing to heightened congestion, especially during weekends. These findings highlight a fundamental shift in travel behavior rather than a simple rebound in traffic volume.

Finally, the research explores the environmental implications of transportation using MDLD to estimate greenhouse gas emissions. By integrating MDLD with vehicle fleet characteristics, fuel economy, and contextual factors, the study employs engineering models and Monte Carlo simulations to quantify emissions. This work highlights the impacts of shifting travel behaviors and fuel prices in the post-pandemic era, offering actionable insights for emissions reduction strategies.

Collectively, this dissertation showcases the transformative potential of MDLD in understanding mobility patterns and addressing energy and environmental challenges, providing a pathway for innovative transportation planning and policy development.

ADVANCING TRANSPORTATION ANALYSIS WITH MOBILE DEVICE LOCATION DATA: INSIGHTS INTO BEHAVIOR, CONGESTION, AND ENVIRONMENTAL IMPACTS

by

Guangchen Zhao

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2026

Advisory Committee:

Professor Cinzia Cirillo, Chair
Professor Anna Alberini, Dean’s Representative
Professor Paul Schonfeld
Professor Ali Haghani
Associate Professor Xianfeng Yang

Acknowledgements

I would like to express my sincere gratitude to my advisory committee for their guidance, support, and valuable feedback throughout my doctoral studies. Their expertise and encouragement have played an essential role in shaping this dissertation.

I am deeply grateful to Dr. Cinzia Cirillo, my advisor and committee chair, for her continuous mentorship, insightful advice, and unwavering support throughout my Ph.D. journey. I am truly thankful for her generosity with her time, patience, and for sharing her vast knowledge and experience with me, which have greatly advanced my research and professional development. It has been an honor and privilege to learn from her and to be her student. I would like to thank Dr. Anna Alberini, Dean's Representative, for her thoughtful comments and constructive feedback, as well as for the generous support she provided in helping me continue my academic career. I am especially thankful to Dr. Paul Schonfeld, who not only served on my doctoral advisory committee but also on my master's thesis committee. His consistent kindness, encouragement, and thoughtful guidance over the years have been deeply appreciated and have had a lasting impact on my academic development. I also thank Dr. Xianfeng Yang and Dr. Ali Haghani for their valuable suggestions, careful review, and insightful perspectives, which greatly improved the quality of this dissertation.

Finally, I would like to express my heartfelt gratitude to my family and friends for their unwavering support, patience, and encouragement throughout my doctoral journey. Their understanding and belief in me provided constant motivation during challenging times, and this achievement would not have been possible without them.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	**Description**
API	Application Programming Interface
CDR	Call Detail Record
CI	Congestion Index
COPERT	Computer Programme to Calculate Emissions from Road Transport
COVID	Coronavirus Disease
CV	Coefficient of Variation
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
EPA	Environmental Protection Agency
GHG	Greenhouse Gas
GIS	Geographic Information System
GPS	Global Positioning System
HBO	Home-Based Other
HBW	Home-Based Work
ID	Identifier
LBS	Location-Based Service
LRI	Location Recording Interval
MC	Monte Carlo
MDLD	Mobile Device Location Data
MOVES	Motor Vehicle Emission Simulator
MWCOG	Metropolitan Washington Council of Governments
NHTS	National Household Travel Survey
OBO	Other-Based Other
OD	Origin–Destination
PADD	Petroleum Administration for Defense Districts
POI	Point of Interest
RTS	Regional Travel Survey
SOC	State of the Commute
TIGER	Topologically Integrated Geographic Encoding and Referencing
TPB	Transportation Planning Board
VMT	Vehicle Miles Traveled
WBO	Work-Based Other
WFH	Work From Home

1. Chapter 1: Introduction

1.1 Background

Urban transportation systems play a critical role in shaping the quality of life in cities, and understanding travel behaviors is essential for effective transportation planning and policy development. Historically, personal transportation choices have been tracked nationwide by the National Household Travel Survey, and locally by similar studies conducted by state and local governments [1-3]. These studies typically collect data from a sample of households regarding their socio-demographic characteristics, including work status and work commute, and the vehicles available to them, and request the survey participants to maintain a diary of all the travel activities undertaken by each member of the household on a given day [4]. Initially conducted through in-person interviews, these surveys have transitioned to computer-assisted telephone interviews and, more recently, online formats [5, 6].

These travel surveys can in principle be customized to the specific needs of agencies, focusing, for example, on specific types of trips (e.g., long-distance trips) or vehicles (e.g., passenger cars or trucks) [7-9]. However, these surveys have several limitations that can impact the accuracy and comprehensiveness of the data collected, such as reporting bias, where people may inaccurately recall or report all of their travel activities, leading to underreporting or misrepresentation of certain trips [10]. Non-response bias is another concern, as certain groups of people may be less likely to participate in the survey, potentially skewing the representation of travel behaviors [11]. Moreover, travel surveys often have a limited timeframe, capturing data for only a specific day or short period [12], and they may rely on relatively small sample sizes of participants [3], limiting the representation of travel behaviors. Therefore, to overcome these limitations and obtain a more comprehensive and precise understanding of travel behaviors, there is a growing need to explore alternative data sources that can complement and enhance the insights derived from travel surveys.

In recent years, the landscape of travel data collection and analysis has been dramatically transformed by the rise of mobile devices and cutting-edge data collection technologies. With the widespread adoption of GPS-enabled smartphones and other mobile devices, an invaluable resource known as Mobile Device Location Data (MDLD) has become available [13]. This data source captures the movements and locations of individuals with unprecedented precision and continuity throughout their daily routines [14]. In contrast to conventional travel surveys, which rely on sporadic and self-reported data, MDLD distinguishes itself by accurately documenting the precise locations and movement trajectories of individuals throughout their daily routines [15]. This continuous influx of high-resolution location data provides an intricate portrayal of travel behaviors, showcasing origin and destination points, travel routes, and durations spent at specific sites. Through extended data collection periods, MDLD unveils the temporal dynamics of travel patterns, revealing differences between weekdays and weekends, peak and off-peak hours, and even seasonal variations in travel habits [16]. Additionally, MDLD illuminates recurring travel behaviors, exposing habitual routes and preferred modes of transportation among individuals, along with their frequently visited destinations [17].

This research aims to explore how MDLD can enhance our understanding of key issues such as data fusion, travel behavior analysis, road congestion, and the environmental impacts of mobility, contributing to more sustainable, efficient, and data-driven transportation planning and policy-making.

1.2 Objectives

This dissertation aims to explore and utilize the potential of MDLD to advance sustainable and efficient travel analysis. Specifically, the objectives of this research are to:

1) Develop and implement probabilistic record-linkage techniques to merge MDLD with traditional travel survey data, creating comprehensive datasets that combine detailed trip characteristics with socio-demographic information.
2) Utilize longitudinal MDLD datasets to examine travel behavior responses to economic factors, with a focus on quantifying the relationship between vehicle miles traveled (VMT) and gasoline prices, providing insights into travel behavioral responses to economic factors.
3) Investigate changes in road congestion and driving behavior in the post-pandemic era by examining how teleworking-related shifts in trip frequency, trip distance, spatial travel patterns, and travel timing contribute to congestion dynamics across weekdays and weekends.
4) Develop a trip-based framework for estimating transportation-related greenhouse gas (GHG) emissions using MDLD, integrating trip-level distance and timing information with vehicle fleet characteristics and emission factors, and incorporating uncertainty in unobserved vehicle attributes to support behaviorally informed environmental policy and emissions reduction strategies.

Through these objectives, the research seeks to address critical knowledge gaps and provide actionable insights for policymakers and planners in the era of big data and digital mobility solutions.

1.3 Contributions

This dissertation contributes to the growing body of knowledge on leveraging Mobile Device Location Data (MDLD) for advanced transportation analysis and planning in several key ways:

1) This study introduces a novel approach to integrating MDLD with traditional travel survey data, exemplified by the Regional Travel Survey (RTS), using probabilistic record linkage methods. The integration addresses the limitations of traditional travel surveys—such as small sample sizes and short observation periods—by linking the geospatial precision and continuous temporal coverage of MDLD with the demographic richness of survey data. The resulting data panels capture detailed travel patterns, including peak-hour commutes, off-peak travel, and non-home-related trips, while providing insights into daily, weekly, and seasonal variations. Additionally, the combined dataset offers a longitudinal perspective on travel habits, enabling an unprecedented level of analysis that is both comprehensive and temporally robust, enhancing our understanding of individual and population-level travel behaviors.
2) This research demonstrates the utility of MDLD in constructing high-resolution longitudinal datasets to analyze travel behavior and its sensitivity to external factors such as socioeconomic conditions, land use, and gasoline prices. Using MDLD from the Washington, DC metro area, both individual-level and geographic pseudo-panels (at block-group and census-tract levels) were developed, enabling robust and granular analysis. By incorporating fine-grained gasoline price data at the station level and leveraging advanced panel data modeling techniques, the study evaluates how daily vehicle miles traveled (VMT) responds to changes in gasoline prices and other variables. The analysis confirms that VMT from MDLD exhibits expected relationships with community characteristics, infrastructure, and land use, validating its quality

and applicability for behavioral research. The study also explores how VMT elasticity varies with respect to gasoline prices, finding patterns consistent with economic theory and prior research. Additionally, it distinguishes the impact of gasoline prices on commute versus non-commute travel, highlighting differential effects. These findings underscore the potential of MDLD as an independent and complementary data source to traditional surveys, offering unparalleled opportunities to explore travel patterns and their economic and environmental determinants with exceptional temporal and spatial precision.

3) This study provides new empirical evidence on post-pandemic road congestion dynamics using high-resolution mobile device location data. By leveraging longitudinal MDLD and consistent origin–destination (OD) pairs across multiple time periods, the research disentangles changes in travel demand, trip distance, and spatial travel patterns before, during, and after the COVID-19 pandemic. The findings demonstrate that increased post-pandemic congestion is driven less by a recovery in trip frequency and more by shifts toward longer-distance, more spatially dispersed trips and changes in travel timing associated with teleworking. This contribution advances the understanding of how evolving work arrangements reshape congestion patterns and highlights the limitations of relying solely on aggregate traffic volumes to assess system performance in the post-pandemic era.

4) This study develops a trip-based framework for estimating transportation-related greenhouse gas emissions that enables behaviorally grounded and spatially resolved emission analysis. By integrating MDLD-derived trip-level distance and timing information with spatially resolved vehicle fleet characteristics and emission factors, the framework estimates resident-based CO_2 emissions while explicitly accounting for uncertainty in unobserved vehicle attributes. The proposed approach allows emissions to be attributed to travel behavior, trip purpose, and

origin–destination flows, providing insights into who generates emissions and why. Comparisons with traditional roadway-based inventories demonstrate that MDLD-based estimates are consistent at aggregate levels while offering additional behavioral and spatial detail, supporting more targeted and efficient transportation climate policy analysis.

5) This study utilizes high-quality MDLD with extensive spatial and temporal coverage, making it a unique and valuable resource for transportation analysis. The MDLD is sourced from multiple leading providers of passively collected passenger travel data, capturing travel patterns from over 15% of the U.S. population. With a temporal coverage spanning the entire years of 2019 through 2022, this dataset provides a rich opportunity for longitudinal analysis, enabling the examination of travel behavior changes across four years. This extended time frame offers valuable insights into evolving transportation trends, particularly in the post-pandemic era.

1.4 Organization

This dissertation is organized into seven chapters. Chapter 2 provides a comprehensive review of the literature on the use of MDLD in transportation analysis, including its advantages over traditional data sources, challenges, and the potential for integrating MDLD with other datasets. Chapter 3 explores data fusion techniques, focusing on how MDLD can be combined with traditional travel survey data, such as the Regional Travel Survey (RTS), to improve travel behavior analysis. Chapter 4 examines the impact of gasoline prices on VMT using MDLD, analyzing how fluctuations in fuel prices influence travel patterns and exploring factors like socioeconomic characteristics and land use. In Chapter 5, we investigate changes in road congestion and driving behavior in the Washington, D.C. region before, during, and after the

COVID-19 pandemic. Using longitudinal MDLD and consistent OD pairs, this chapter examines how shifts in trip frequency, trip distance, spatial travel patterns, and travel timing, particularly those associated with teleworking, contribute to post-pandemic congestion dynamics. Chapter 6 presents a trip-based framework for estimating transportation-related GHG emissions by integrating MDLD-derived trip information with vehicle fleet characteristics and external datasets. This chapter evaluates emission patterns across space, time, and trip purpose, compares MDLD-based estimates with traditional roadway-based inventories, and highlights the behavioral drivers of emissions. Finally, Chapter 7 concludes the dissertation by summarizing the key findings, discussing the contributions of the study, and suggesting directions for future research.

2. Chapter 2: Literature Review

2.1 Mobile Device Location Data in Transportation Research

2.1.1 Travel Surveys and Traffic Monitor Data

Two foundational data sources in transportation research are travel surveys and traffic monitoring systems, each offering distinct capabilities and limitations. Travel surveys, typically conducted by local and national governments, gather information on trips taken by travelers over a specific travel day. These studies typically collect data from a sample of households regarding their socio-demographic characteristics, including work status and work commute, and the vehicles available to them, and request the survey participants to maintain a diary of all the travel activities undertaken by each member of the household on a given day [4]. Initially conducted through in-person interviews, these surveys transitioned to computer-assisted telephone interviews and, more recently, online formats [5, 6]. Traffic monitors, such as inductive loop detectors, video detection systems, and microwave radar sensors, are commonly installed along freeways or at intersections to collect data on traffic speed, volume, density, and vehicle classification [18].

Both data sources have their strengths. Travel surveys provide self-reported travel data with detailed background information, ensuring accurate locations of trip ends, modes, and purposes, and reducing the need for imputation of key trip attributes [19, 20]. Traffic monitors, being mature technologies, can continuously monitor traffic conditions 24 hours a day, offering precise and real-time information [21]. However, they also have limitations in analyzing travel behaviors. Travel surveys are prone to reporting and nonresponse biases [10, 11], are limited to specific timeframes (usually only a day or short period) [12], and often rely on relatively small sample sizes [3]. In addition, national or local personal travel surveys are conducted infrequently (every 5-8 years), and each wave covers different households than the previous ones, making it impossible to observe how travel habits change with someone's economic and personal circumstances and stock of vehicles [1, 22]. Traffic monitors, on the other hand, always face the challenge of high installation and maintenance costs. The systems often have limited spatial coverage, focusing mainly on major roads while neglecting smaller streets and rural areas [23]. Additionally, traffic monitors provide only a snapshot of traffic conditions without detailed origin-destination (OD) information, making it difficult to fully understand travel behaviors and patterns for comprehensive transportation planning [21, 24]. These limitations have motivated growing interest in alternative data sources, particularly Mobile Device Location Data, as discussed in the following section.

2.1.2 Mobile Device Location Data (MDLD)

Table 2-1 Mobile Device Location Data for Mainstream Use

MDLD	**Example**	**Spatial Coverage**	**Temporal coverage**	**Frequency of collection**	**Accuracy**
GPS data	GPS-based travel survey; GPS Data without User Recall	Low	Low	High	High
Cellular Data	Call Detail Record (CDR); sightings	High	High	Low	Low
LBS data	Smartphone activities	High	High	Low	High

For transportation research purposes, MDLD can be broadly classified into three categories: Global Positioning System (GPS) data, Cellular Data, and Location-Based Service (LBS) data. **Error! Reference source not found.** compares these data types across key dimensions including spatial coverage, temporal coverage, and collection frequency.

GPS data has been integrated into travel surveys since the late 1990s [25, 26]. Early implementations relied on in-vehicle GPS loggers that recorded location at approximately one-second intervals during vehicle movement [27]. While efficient, this approach was limited to motorized trips, leaving walking, cycling, and other non-motorized activities unrecorded. Subsequent surveys addressed this gap by introducing wearable GPS loggers capable of capturing the full range of travel activities [28, 29]. More recently, smartphone applications running in the background have largely replaced dedicated hardware, offering greater portability and ease of use [30]. Beyond survey contexts, GPS data without user recall is also widely deployed for real-time traffic speed estimation and travel time monitoring [31, 32].

Cellular data encompasses two primary forms: Call Detail Records (CDR) and sightings data. CDR, the more widely used of the two, logs the location of the nearest cell tower at the time of a call or text [33-35]. Sightings data, though less common, offers finer spatial resolution by estimating user location through triangulation across multiple towers [33]. Both forms have been applied extensively in studies of individual human mobility patterns [34-36].

LBS data has gained increasing prominence alongside the proliferation of smartphones. Unlike GPS or cellular data, LBS data draws on multiple positioning technologies simultaneously — including Bluetooth, Wi-Fi, GPS, and cellular networks — selecting the most accurate source available at any given moment [33]. This results in high locational accuracy and low location recording intervals (LRI), making LBS data the dominant source for both academic mobility research and commercial data applications [37-39].

Each data type carries distinct trade-offs. GPS-based travel surveys offer the highest collection frequency and precision, but are constrained by limited spatial and temporal coverage due to their reliance on recruited participant samples. Cellular and LBS data, by contrast, achieve broad geographic and temporal coverage since the vast majority of mobile users engage with calls, messaging, or location-based applications on a near-daily basis. Together, these complementary MDLD sources offer a rich, multi-perspective lens on human mobility and destination choice.

Despite these strengths, MDLD presents notable limitations. Most critically, it lacks explicit demographic information, which constrains its standalone utility for comprehensive travel behavior analysis [40]. Demographic attributes, such as age, employment status, household size, and income, are well-established determinants of travel patterns, influencing trip frequency, purpose, mode choice, and distance [41, 42]. Integrating MDLD with traditional travel surveys, such as the Regional Travel Survey (RTS), offers a pathway to address this gap. Current approaches to such integration fall into two broad directions. The first involves GPS-based travel surveys in which participants use wearable devices [43] or smartphone applications [30] to record their activities alongside self-reported demographic information, yielding reliable and detailed data, though subject to the same volunteer sampling biases as conventional surveys [44]. The second approach applies population synthesis techniques to infer socio-demographic attributes from aggregated census data [45], though this method depends on the availability of a demographically known sub-sample of location data, which is often difficult to obtain[46].

2.2.2 Imputation of Missing Information Based on MDLD

While MDLD offers more frequent and precise location records than traditional travel surveys, it typically lacks the ground truth information that surveys provide — such as travel mode, trip purpose, and users' home and work locations. These attributes are essential inputs for core transportation applications including travel demand modeling, traffic assignment, and accessibility analysis. A growing body of methodological research has emerged to address these gaps, broadly organized around four imputation tasks.

1) Home and Work Location Identification

Identifying home and work locations serves as a foundational step in analyzing human activity patterns [47]. The process typically begins by grouping geographic coordinates into candidate activity centers using either rule-based [48, 49] or clustering-based methods [50, 51]. These candidate locations are then classified by type using one of two general approaches: context-based or behavior-based methods [52].

Context-based methods infer the nature of an activity center from its surrounding environment, drawing on land-use characteristics and nearby points of interest (POI) [53, 54]. Behavior-based methods, which are more widely applied, classify activity centers according to observed patterns such as arrival time, dwell time, and visit frequency [55]. Home locations are generally identified as places where individuals spend the greatest total time, particularly during nighttime hours, while work locations are associated with extended daytime stays and frequent visits [52, 56-58]. Pan (2023) [59] refined this framework by incorporating both overall and nighttime hours for home identification, and introducing a temporal similarity ratio for work location detection — specifically to avoid misclassifying frequently visited locations near home as workplaces. Given that MDLD rarely contains personal identifiers, behavior-based methods remain the most practical approach for this task.

2) Trip End Identification

Accurately detecting trip ends from location data is a prerequisite for analyzing travel behavior, route choice, and transportation demand [60]. he conventional approach relies on rule-based methods, in which domain experts define parameters — such as dwell time thresholds and speed cutoffs — that are applied sequentially to location points to determine where trips begin and end [61-63]. More recently, supervised machine learning methods have been introduced as a complement to rule-based approaches, classifying individual location points as either stationary or in motion [64-66]. Clustering algorithms offer another avenue, inferring trip ends by first identifying activity locations within the data [50, 51, 67, 68]. One recent study combined spatiotemporal clustering with three optimization models to improve trip end detection [68]. For LBS data specifically, a "Divide, Conquer and Integrate" (DCI) framework has been proposed to extract mobility patterns from the bimodally distributed data structure of the Puget Sound region, merging rule-based and incremental clustering methods and validating results against household travel surveys at the census tract level [69].

3) Travel Mode Imputation

Methods for imputing travel mode from MDLD generally fall into two categories: trip-based and segment/point-based approaches. Trip-based methods operate on already-identified trips, assigning a single mode to each trip [70]. Segment-based methods divide location data into fixed-length intervals by time or distance, impute a mode for each segment, and then merge consecutive same-mode segments into unified trips [71]. Speed and acceleration are among the most consistently useful features across both approaches [70-75]. When the Location Recording Interval (LRI) is below approximately 10 seconds, these kinematic features alone are often sufficient for accurate mode classification. At higher LRIs — around 30 seconds — supplementary features such as real-time transit information [71], multimodal network data [70, 71], or sociodemographic attributes [74] are typically incorporated to maintain accuracy. Yang et al. developed a hybrid model combining heuristic rules with a random forest classifier to distinguish among six modes: driving, bus, rail, walking, cycling, and air travel [76].

4) Trip Purpose Imputation

Trip purpose imputation from MDLD is framed as a classification problem, with methods falling into three broad categories [48, 77]:

- **Rule-based models** apply predefined heuristic rules to selected features to assign trip purposes [78, 79]. Valued for their simplicity and interpretability, these models were prominent in early research, though their limited transferability across contexts has constrained their broader adoption [77].

- **Probabilistic models** estimate the likelihood of each possible trip purpose rather than assigning a deterministic label. For example, rather than selecting the nearest POI as a trip destination, these models calculate destination probabilities as a function of distance [80].

- **Machine learning models** have increasingly outperformed both rule-based and probabilistic approaches over the past two decades. Decision trees, random forests, artificial neural networks, and support vector machines have all been applied as primary classifiers [81-84]. A key consideration in these models is the construction of balanced training and test sets, with equal representation of each trip purpose being the most common strategy [83].

In this study, an efficient algorithm that accounts for device residency and worker status is applied to impute work and non-work trip purposes from MDLD [85].

2.2 Vehicle Miles Traveled (VMT) and Economic Sensitivity

2.2.1 Tracking VMT

Tracking mobility and miles traveled — whether by private car or other means of transport — is a major focus of transportation agencies and in transportation research, and serves as a key input for origin-destination modeling [86, 87], infrastructure planning [88, 89], accident prevention [90, 91], fuel efficiency regulations [92], and energy forecasts and emissions calculations [93, 94].

As discussed in Section 2.1.1, travel surveys have historically been the primary tool for measuring personal vehicle miles traveled (VMT) at both the national and local levels [95]. Previous studies of personal, household, and neighborhood-level VMT have largely relied on these surveys [96-99]. In recent years, trip distances have been measured more precisely using mapping software based on respondent-reported origins and destinations — as in the 2017 NHTS, which uses the Google Maps API — or through GPS-enabled devices installed in vehicles, as adopted in VMT-tax programs in states such as Oregon and Virginia.

However, as noted in Section 2.1.1, these surveys are conducted infrequently (every 5–8 years) and cover different households across waves, making it impossible to track how individual travel habits evolve alongside changes in personal circumstances, income, or vehicle ownership. Single-day travel diaries have also been criticized as inadequate for constructing reliable annual VMT estimates [12, 100], and reporting bias may further compromise data quality [10].

The emergence of MDLD over the past decade has opened new possibilities for VMT tracking, as outlined in Section 2.1.2. A key advantage of MDLD is its ability to follow the same device continuously over time, enabling near-continuous monitoring across many days and trips. Given near-universal mobile phone ownership, these datasets often encompass thousands of devices tracked over extended periods [14] , offering a level of temporal depth and scale that traditional surveys cannot match.

Researchers have begun leveraging these advantages for VMT estimation. Zhang et al. [24] examined the feasibility of estimating local road VMT from GPS trajectories collected from a small device sample. Calabrese et al. [36] used large-scale mobile tracking data to measure individual VMT, though their triangulation-based positioning introduced relatively large spatial errors. More recent work has scaled these efforts to the state and national level [101-103], and some studies have constructed nationwide annual origin-destination matrices [104].

Despite this progress, research on using large-scale MDLD to estimate personal or neighborhood-level VMT remains limited. This gap likely reflects the anonymous nature of MDLD, which arrives without personal identifiers and requires heuristic rules and assumptions to infer home and work locations, trip origins and destinations, travel modes, and trip purposes. The computational demands of processing raw location data at scale present an additional barrier.

2.2.2 Elasticity studies on VMT

The relationship between socioeconomic factors, fuel prices, and vehicle miles traveled (VMT) is complex and multifaceted. Fuel prices themselves play a significant role in shaping VMT. Research has shown that a 10% increase in average gasoline prices can lead to reductions in annual VMT of 0.86% to 2.65% [105]. This is further supported by a study finding that a 10% increase in fuel prices would reduce driving by approximately 2% in the long term [106]. The short-run fuel price elasticity of VMT, estimated at -1.71, underscores a strong immediate response to fuel price changes, with households quickly adjusting their travel patterns in reaction to rising costs [107].

Several studies have examined how socioeconomic factors influence VMT. It is consistently demonstrated that higher-income households exhibit greater elasticity in response to fuel price changes compared to lower-income households. For example, one analysis found that the highest income groups had fuel price elasticities ranging from -0.41 to -0.35, while the lowest income group displayed an elasticity of -0.24 [108]. This greater sensitivity among wealthier households may stem from their discretionary income and flexibility in travel choices, making them more responsive to fluctuations in fuel prices. Similarly, wealthier households have been found to be more elastic with respect to time costs than fuel costs, further influencing their VMT behavior [109].

Another critical aspect influencing VMT is the rebound effect, wherein improvements in fuel efficiency lead to increased travel as a behavioral response to lower per-mile costs. This occurs because higher fuel economy reduces the per-mile fuel cost, incentivizing additional travel. The effect is particularly pronounced among lower-income households, with estimates indicating a rebound effect of 0.7 for the lowest income group [108]. Broader studies have found short- and long-term rebound effects of approximately 8% and 20%, respectively, demonstrating the persistent challenge of offsetting the benefits of fuel efficiency improvements [110].

Urban form also interacts with fuel prices to influence VMT. Households in low-density neighborhoods, often characterized by higher incomes and more vehicles per driver, show varying responses to gasoline price changes. While these households are less responsive to price changes for work trips, they exhibit some sensitivity for non-work trips; for instance, a 1% increase in gas prices results in a 0.171% reduction in non-work trip VMT. This suggests that urban form moderates the relationship between fuel prices and travel behavior, with differences in density and land use shaping household responses [111].

While a substantial body of literature has examined the impact of socioeconomics and fuel prices on vehicle miles traveled, these studies often rely on conventional travel surveys to produce personal or household-level VMT [108, 111-115], and for these reasons their sample sizes have been small and/or they have lacked a longitudinal component. Another important difference between this work and earlier studies is that we have the weekly price of motor fuels from each gas station in our study area. By contrast, the majority of the earlier studies have deployed state-level [116] or Petroleum Administration for Defense Districts (PADDs)-level prices [117], which assign the same price figure to locales in different states—a situation with a clear measurement error in one key regressor.

2.3 Teleworking and Post-Pandemic Travel Behavior

2.3.1 Teleworking

The COVID-19 pandemic has fundamentally altered travel behavior worldwide, with teleworking emerging as one of the most significant and lasting changes [118]. During the pandemic, work-from-home arrangements were widely adopted to reduce physical contact, and in many regions teleworking has persisted well beyond the acute public health phase [119, 120]. Evidence from regional and national surveys consistently shows both a substantial increase in the share of workers who telework and a marked rise in telework frequency relative to pre-pandemic levels. According to the 2022 State of the Commute (SOC) Survey Report from the Metropolitan Washington Region, 2.14 million regional commuters—66% of the total—were teleworking at least occasionally in 2022, nearly double the 1.07 million (35%) recorded in 2019 before the pandemic, and the average telework frequency rose, nearly tripling from 1.2 telework days in 2019 to 3.37 telework days in 2022 [2, 121]. These shifts reflect not only temporary responses to health risks but also longer-term changes in work culture, supported by perceived benefits for both employees and employers.

From the employee perspective, teleworking reduces commuting time and costs, offers greater flexibility in residential location choice, and improves work–life balance [122]. Employers similarly benefit from reduced office space and operational costs, improved employee satisfaction, and access to a broader labor pool [123]. As a result, teleworking has become embedded in many labor markets, raising important questions about its long-term implications for travel demand and transportation system performance.

A critical dimension of teleworking's transportation implications is the uneven distribution of telework-compatible jobs across occupations and industries. Knowledge-intensive sectors such as finance, technology, and professional services have the highest shares of teleworkable positions, while industries such as construction, manufacturing, and retail remain largely place-dependent [124].

The rise of hybrid work arrangements — in which employees split their time between home and the office — has emerged as the dominant post-pandemic work model and carries distinct transportation implications compared to full remote work. Unlike fully remote workers who eliminate commute trips entirely, hybrid workers retain regular commuting patterns but shift them temporally, contributing to the redistribution of peak-period demand rather than its elimination [125]. Evidence suggests that hybrid workers tend to concentrate office visits on midweek days, compressing demand into Tuesday–Thursday peaks while leaving Monday and Friday relatively quiet [126].

Teleworking has also been associated with residential relocation, as reduced commuting frequency loosens the geographic constraints on residential location choice. Some workers have taken advantage of remote work flexibility to relocate farther from employment centers, trading shorter commutes for larger homes or lower costs of living [127]. While this may reduce individual commute frequency, it increases average trip distances when workers do commute, with implications for VMT and emissions that partially offset the benefits of fewer commute days.

2.3.2 Post-COVID Congestion

Early expectations in the literature suggested that widespread teleworking would lead to sustained reductions in traffic congestion by eliminating or reducing daily commute trips. Several studies hypothesized that fewer peak-period commute trips would ease pressure on roadway networks, particularly in large metropolitan regions [128, 129]. However, emerging post-pandemic evidence has complicated this narrative. Despite persistent reductions in traditional commuting, many regions have experienced a rebound, or even an increase, in congestion relative to pre-pandemic conditions [130, 131]. Reporting from the Washington Post (September 2023) indicated that traffic volumes along major highways in the Washington region are rising, with many drivers experiencing increased difficulty navigating the area due to more crowded roads than before the pandemic [132]. The article suggests that a key reason for this trend is the significant reduction in public transit ridership, contributing to greater reliance on private vehicles. Transit ridership in major U.S. cities has recovered only partially from pandemic-era lows, with the overall ridership in September 2023 at 74 percent of September 2019 levels [133]. This trend is further corroborated by comparisons of the 2019 and 2022 SOC Survey Reports from the Metropolitan Washington Region, as well as between the 2017 and 2022 National Household Travel Survey (NHTS) at the national level [2, 22, 121, 134].

Academic studies from diverse international contexts further reinforce this complexity. For instance, Loo et al. analyzed the congestion index (CI) during peak hours in Hong Kong and found that the afternoon peak was not significantly affected by work-from-home policies in many areas [135]. Similarly, Xu et al. studied hourly average speeds in traffic analysis zones in Shanghai and found that congestion had either recovered or even exceeded pre-pandemic levels in central areas [136]. U.S.-based evidence corroborates this pattern of congestion recovery and intensification. Studies analyzing traffic volume data from major metropolitan areas have documented that while peak-period volumes remain below pre-pandemic levels, consistent with reduced commuting, midday and weekend traffic has recovered more rapidly and in some cases surpassed 2019 benchmarks [137-139]. The growth of weekend congestion has received particular attention as a distinguishing feature of post-pandemic travel patterns. Prior to the pandemic, weekend traffic was generally lower and more diffuse than weekday traffic in most metropolitan regions. Post-pandemic evidence indicates that this gap has narrowed or reversed in some areas, as the redistribution of discretionary travel and the decoupling of work and commute schedules have elevated weekend travel demand [137-139].

These findings challenge the assumption that reduced commuting alone is sufficient to alleviate congestion and point instead to broader structural changes in travel behavior, which this dissertation examines in the context of the Washington, D.C. metropolitan region.

2.4 Transportation-Related Greenhouse Gas Emissions

Transportation is the largest source of greenhouse gas (GHG) emissions in the United States, accounting for nearly one-third of total GHG, predominantly in the form of carbon dioxide (CO_2) [140]. Quantifying these emissions accurately and continuously is critical for evaluating policy effectiveness, supporting regional climate action plans, and monitoring progress toward decarbonization goals. Effective mitigation policies require not only accurate measurement of total emissions, but also a clear understanding of the sources and spatial distribution of emissions, the populations responsible for producing them, and the travel behaviors that drive them [141-143].

However, existing emissions inventory frameworks remain limited in their ability to answer these policy-relevant questions.

The U.S. Environmental Protection Agency's *MOtor Vehicle Emission Simulator* (EPA's MOVES) is the primary tool used for on-road emissions estimation in the United States [144]. MOVES calculates emissions based on vehicle activity occurring within a specified geographic area, incorporating inputs such as vehicle miles traveled (VMT), fleet composition, road types, speeds, and environmental conditions [145]. This framework provides robust and widely accepted estimates of emissions at multiple spatial and temporal scales and is extensively used for regulatory compliance, transportation conformity analysis, and regional planning.

Despite these strengths, MOVES and similar roadway-based inventories are fundamentally location-based and operating-oriented. Emissions are attributed to where vehicles operate on the roadway network rather than to the origins, destinations, or purposes of trips [144, 146]. As a result, emissions are accurately quantified by where they occur on the roadway network but remain behaviorally anonymous, limiting the ability to attribute emissions to specific travel patterns, populations, or underlying socioeconomic drivers [147, 148]. This limitation constrains the design and evaluation of targeted mitigation strategies, including land-use planning, demand management, transit investment, and equity-focused climate policies [149].

Beyond roadway-based inventories, a growing body of research has sought to link transportation emissions directly to travel behavior at the trip, household, or individual level. These approaches leverage activity-based models, GPS trajectory data, and household travel surveys to attribute emissions to specific trip purposes, origins, destinations, and traveler characteristics [150, 151]**.** Such behavioral emissions frameworks offer a richer picture of the demand-side drivers of transportation GHGs, enabling more targeted policy interventions such as congestion pricing, land-use intensification near transit, and mode-shift incentives [152, 153]. However, these approaches have historically been constrained by the same data limitations that affect travel demand modeling more broadly, namely, small sample sizes, limited temporal coverage, and the absence of continuous longitudinal tracking [154, 155].

Vehicle fleet composition and fuel economy are critical moderating factors in the relationship between travel behavior and emissions. The ongoing electrification of the passenger vehicle fleet introduces significant heterogeneity in per-mile emissions across vehicles, as electric vehicles (EVs) produce zero tailpipe emissions while their lifecycle emissions depend on the electricity generation mix [156, 157]. Conventional fleet fuel economy has also improved steadily over the past decade, driven by federal CAFE standards and market trends toward more efficient powertrains [158, 159]. Accurately capturing this fleet heterogeneity is essential for producing reliable emissions estimates, and is a methodological consideration addressed explicitly in this dissertation through the integration of MDLD with vehicle fleet and fuel economy data.

3. Chapter 3: Data Fusion for Travel Analysis: Linking Travel Survey and Mobile Device Location Data

This study proposes a novel approach to enhance travel behavior analysis through the integration of MDLD and RTS datasets based on record linkage algorithms. Two distinct data linkage approaches are utilized to connect individuals with similar travel behavior across datasets. The resulting data panels offer comprehensive and accurate representations of travel behaviors over multiple days, capturing not only peak-hour commutes but also off-peak travel and non-home-related trips. Additionally, they include representative population demographics, enhancing the overall richness and reliability of the dataset.

The primary objectives of this study are as follows:

1) To integrate MDLD and RTS datasets using two data linkage approaches: one based on a probabilistic method and the other on similarity-based techniques.
2) To construct data panels that offer a longitudinal perspective on individuals' travel behaviors, overcoming the limitations of traditional survey methods.
3) To evaluate the effectiveness of each data linkage approach in capturing accurate travel patterns and compare the characteristics of the integrated data panels with the original RTS and MDLD datasets.
4) To explore the representativeness of the data panels and discuss the implications of the findings for transportation planning and policy development.

In the following sections, we present the data used for data linkage, the methodology employed for panel construction, the results of the linking process, and a comprehensive analysis of the data panels, discussing the implications and potential applications of our findings for transportation planning and policy development.

3.1 Data Description and Preprocessing

Figure 3.1 presents the methodological framework of this study. There are three major components: Data Preprocessing of RTS and MDLD, Identification of individuals with similar travel behaviors by two distinct approaches, and Construction of the data panel.

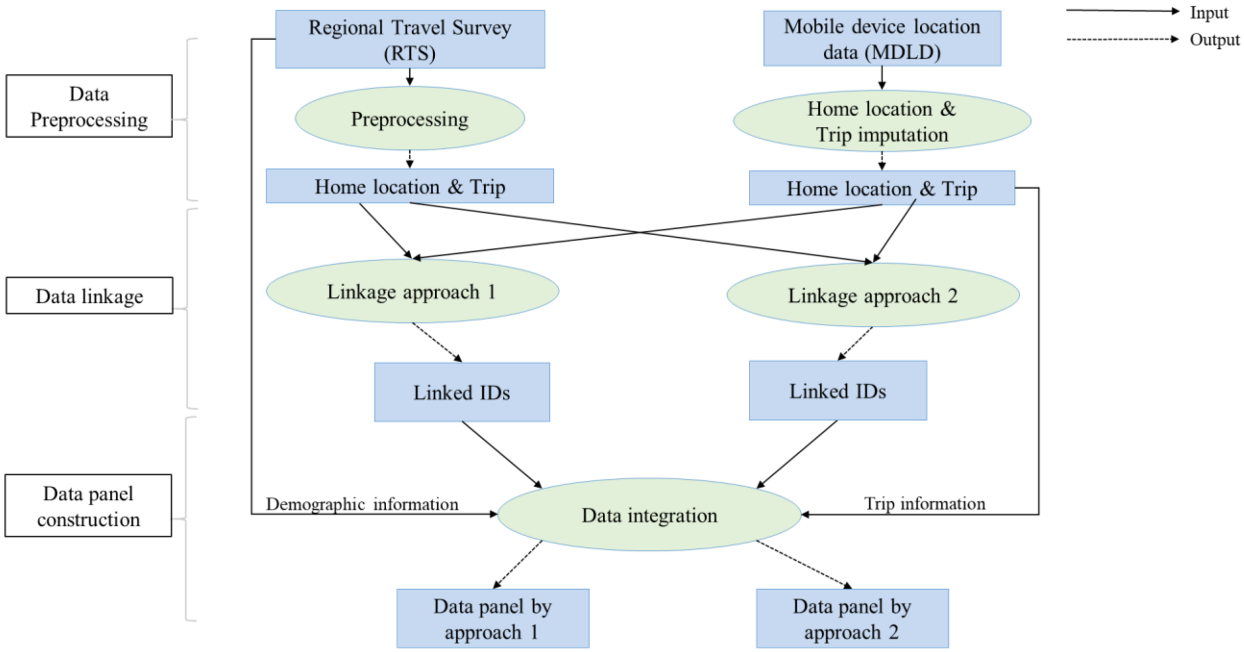


Figure 3.1 Methodological Framework of Data Fusion

Two primary datasets are utilized to identify individuals with similar travel behaviors, and construct the data panel: trips collected in the Regional Travel Survey (RTS), and trips imputed from MDLD.

3.1.1 Regional Travel Survey (RTS)

The Regional Travel Survey (RTS) conducted by the Metropolitan Washington Council of Government (MWCOG) from October 2017 through December 2018 collected demographic and travel information from a randomly selected sample of households in the metropolitan Washington region. Key components in the RTS Dataset include household, person, trip, and vehicle files. Trip details, including origin/destination, start/end times, mode of travel, and trip purpose are recorded. Demographic information of individuals, such as age, gender, household size, income level, and employment status are also included.

3.1.2 Mobile Device Location Data (MDLD)

The Mobile Device Location Data utilized in this study comes from a consistent national mobility data panel produced by the University of Maryland (UMD) team from multiple data vendors. While the MDLD provides valuable insights into travel patterns, the absence of demographic data limits the ability to analyze the relationship between travel behavior and specific demographic factors.

By combining the RTS's detailed demographic information with MDLD's high-precision, long-term trip records, the research aims to create a unique dataset that can provide insights into travel behavior trends and their connection to demographics at a broader scale. Anna Arundel County in the State of Maryland is chosen as the case study in this paper. To ensure data accuracy and relevance, both datasets were filtered to include only individuals and devices residing within Anna Arundel County. Additionally, the analysis was limited to trips occurring solely within the county, excluding any trips departing from or entering the county from different regions. It should be noted that the methodology is scalable to regional, state or national level.

3.1.3 Data Preprocessing

The data record linkage requires that both datasets have the same structure and comparable attributes. In the case of this study, the RTS already includes person information such as the census tract of the home location, as well as trip details like census tract of origin/destination, start/end times, and more. However, the raw MDLD only contains location sightings with anonymized device identifiers (ID), latitude and longitude coordinates, time stamps, and positioning accuracies. To bridge this gap, a series of imputation algorithms are applied to derive the home location and trip information for each individual in the MDLD dataset [85]. A behavior-driven approach is utilized to identify devices' home locations, which are imputed as the census tract where the device is observed with the highest count of nights, hours, and sightings during nighttime [59].

For extracting trip information from the sightings, a tour-based method is employed. The algorithm recognizes essential tours based on the daily life centers—such as a typical home-destination-home tour composed of multiple trips between the daily life centers and other stops—and further examines location observations within each tour to construct the complete chain of trips. Between the major activity locations, the point-to-point travel time, distance, and speed from the current location observation to the previous sighting and the next sighting are examined through a recursive algorithm to identify the stops and trips. Finally, a daily trip roster is generated, where for each trip, a unique trip identifier, start and end time, and coordinates of the origin and destination are included.

Additionally, the MDLD dataset may contain data from one or multiple recorded days for each device, while the RTS dataset records only one day for each person. To account for this difference, one day is carefully selected for each device in the MDLD data. Specifically, the focus is on including trips from the day with the highest number of recorded trips, as this day is deemed to best represent the person and allows for meaningful comparisons between the datasets.

3.2 Data linkage

In this research study, we present our methodology that builds upon the existing conventional approach for data record linkage. Figure 3.2 visually represents the data record linkage framework:

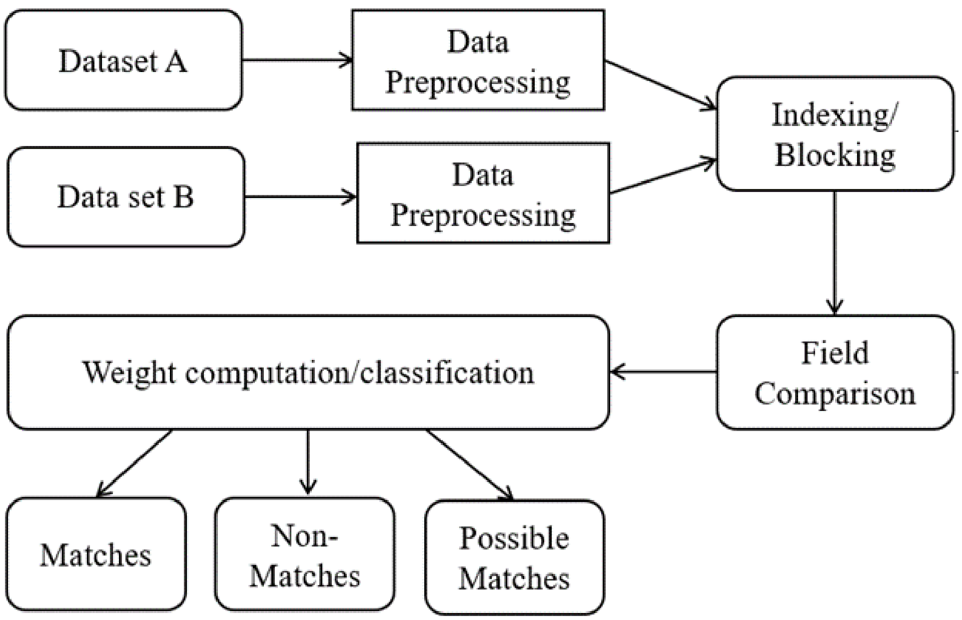


Figure 3.2 Data Record Linkage Framework

The framework presented above encompasses the fundamental steps involved in any data record linkage model, whether it is exact matching or supervised/unsupervised probabilistic matching. These steps include indexing/blocking, similarity measurement, and linking decision. However, each step may vary in its implementation across different record linkage models. The selection of a specific model depends on various factors, such as the specific needs of the research, the characteristics of the data, the availability of previous knowledge, and whether partially true data is accessible for model training or not [160].

After data processing step described in the previous section, these datasets are structured at the trip-unit level, where each device/person may have one or more records representing the total number of recorded trips. However, this data structure presents a unique challenge for the matching process, given our aim to link persons/devices. Most matching techniques are commonly applied to datasets involving entities with single records, such as health data records and other administrative data sources [161-163]. To address the challenges posed by the trip-unit level data structure in the matching process, we have developed two distinct approaches, both based on probabilistic record linkage principles. In the following sections, before presenting the two approaches, we provide a more detailed explanation of each of the record linkage steps that were incorporated into both approaches.

3.2.1 Indexing/Blocking

The first step of data linkage is called indexing or blocking. This involves creating a list of potential candidate pairs between the two datasets for the comparison in the subsequent step. Initially, all possible pairs are formed into a matrix including all possible cases by exhaustively trying every entity from dataset A with every entity from dataset B. However, this exhaustive approach can result in significantly high computational time and reduce the algorithm's efficiency in finding matches. To address this issue, a more efficient approach is employed by eliminating many pairs through a process known as "blocking." Blocking involves grouping entities based on certain criteria, such as shared attributes or properties [164]. For example, candidate pairs may only be considered if they share the same zip code area. This way, the algorithm can reduce unnecessary comparisons and focus only on potentially relevant matches, thereby improving its efficiency. Blocking techniques are diverse and not constrained to merely grouping entities with identical attributes (Standard Blocking). In our study, we used Standard Blocking and we restricted our comparisons to individuals/devices residing in the same home tract area.

3.2.2 Field Comparison

Following the indexing/blocking process, potential candidate pairs are compared for similarity using selected attributes called "matching variables." These variables are common to both datasets, and the comparison is done individually for each one. The results range from 0 to 1, with 0 indicating no match, 1 indicating a perfect match, and any value in between representing the degree of similarity. The similarity depends on the type of variables compared, such as string values, numeric, date, time, or location. This study focuses on time and geographic variables. Distances between departure and arrival times are computed in minutes, while distances between departure and arrival locations are computed in kilometers for each candidate pair. These values represent distances and are not similarity scores yet. To convert distances into similarity scores ranging from 0 to 1, a linear decay function is used [165]. The process involves two tuning parameters: "offset" and "scale." If a distance falls within the specified range, it gets a similarity score of 1. However, if the distance value exceeds the offset, the score starts to linearly decay from 1 to 0 at a rate of 1/scale (Figure 3.3).

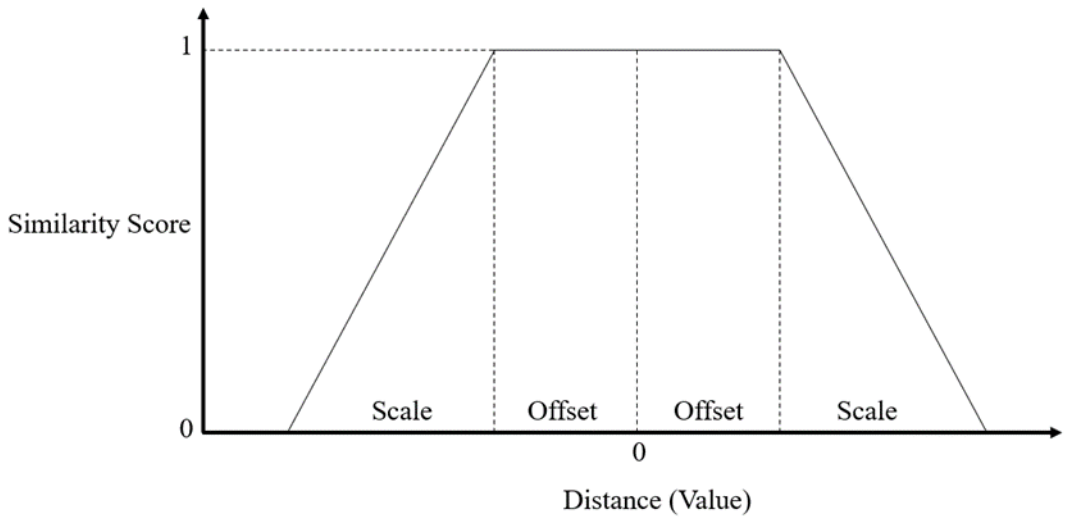


Figure 3.3 Conversion of Distances to Similarity Scores using Linear Decay Function

The selection of tuning parameters in this study is driven by the desired accuracy level and the allowable error defined by the user. For this investigation, the tuning parameters "offset" and "scale" were set to 15 minutes and 25 minutes, respectively, for departure and arrival times. Additionally, for departure and arrival locations, the tuning parameters "offset" and "scale" were configured to 1 km and 3 km, respectively. We made these choices to the best of our ability. We believe that the choices meet the specific accuracy requirements and permissible errors outlined for the study.

3.2.3 Approach 1

The first approach uses a probabilistic method based on the Fellegi and Sunter theorem [166] for record linkage. We adapt this method to the trip-unit level, which helps identify matching trips among the datasets. All the previous steps were applied at the trip-unit level to compute similarity scores across potential candidates. However, it is essential to note that these are similarity scores, not probabilities. This is why we have chosen to use the Fellegi and Sunter method to convert these similarity scores into probabilities of matching. To convert the similarity score into a match probability, Bayes' Theorem is applied. To do this, we require the following information:

- P-Probability (Overall): This is the prior probability, which represents the expected proportion of duplicates in a dataset before any comparisons are made between records.
- M-Probability (for each variable): P(Variable Matches | Record is a Match) is the probability that two records with the same entity agree on the linkage variable.
- U-Probability (for each variable): P(Variable does not match | Record is not a match) is the probability that two records with different entities agree on the linkage variable.

Unfortunately, none of this information is available in our case, as we lack ground truth data or training data. Therefore, in approach 1, we estimated these parameters for each matching variable using an unsupervised algorithm facilitated by the expectation/maximization algorithm [167]. In the decision step, at the trip-unit level, matched trips are accepted if their similarity score exceeds the specified threshold value of 0.5.

To find matches/links between devices and persons, we developed an algorithm that leverages unique matches to establish connections between individuals or devices. Since the matching at the trip level can result in multiple possible matches for each device/person, the algorithm searches for the most unique connections, assuming that the most unique links are the more reasonable matches.

3.2.4 Approach 2

The second approach is based on a similarity-based method. In this approach, we restructured the dataset, transforming it from the trip-unit level to the person/device level. This involved aggregating all trip information pertaining to each individual or device into a single trip itinerary. Each trip itinerary is represented as a vector of one or more trips, with each trip including selected matching variables as departure and arrival times, as well as departure and arrival locations in longitudinal and latitudinal format. By restructuring the data in this way, we create a more manageable dataset that facilitates the application of similarity measures. To measure the similarity between any candidate pairs, we have chosen to use the dynamic time warping algorithm [168]. This algorithm calculates the distance between the two candidate pairs, taking into account variations in the time and location axes. The resulting distance is then converted into a similarity score using a linear decay function, similar to the tuning parameter used in approach 1. The dynamic time warping technique is particularly effective in capturing patterns and similarities in time-series data, making it well-suited for our trip-based datasets. This method has been widely used in various data trajectory similarity studies [169-171]. In the decision step, we determine the acceptance of matched devices based on a similarity score. Devices are linked if their similarity score is equal to the maximum value and greater than or equal to a specified threshold value. For this study, the threshold value was set to 0.5, meaning that candidate pairs with a similarity score of 0.5 or higher are accepted as matches. By leveraging trip itineraries, we achieved direct linking of devices to individuals without the need for the uniqueness algorithm utilized in approach 1.

For indexing/blocking and estimating probabilities (P, M, and U) in both approaches, we utilized the 'recordlinkage' package in Python [165]. By incorporating these two complementary approaches, the challenges posed by the original trip-unit level data structure can be effectively addressed.

3.3 Data panel construction

Data panels are constructed for travelers linked by approach 1 and approach 2, respectively. The trip records in the data panel are sourced from MDLD, which continuously captures trip information over multiple days without reporting bias, providing a longitudinal perspective on individuals' travel behaviors. The demographic information, including age, gender, household size, income level, and employment status, is obtained from RTS, where detailed and reliable demographic data is reported. By integrating these two data sources, we combine geospatial accuracy with detailed demographic information, addressing the limitations of both datasets and enabling a more nuanced analysis.

In the next section, we will present and compare the different characteristics of the four datasets: the original RTS, MDLD, and the panel data constructed with travelers linked by approach 1 and approach 2. We will also discuss and analyze the effectiveness of the two linking algorithms and the representativeness of the panel data.

3.4 Results

3.4.1 Sample size, timespan, and trip rate of the data panel

Table 3-1 presents a summary of key statistics, including sample size, timespan, and trip rate, for different datasets: RTS, MDLD, travelers linked by approach 1, and travelers linked by approach 2. Note that the linkage process involved utilizing RTS and a representative day of MDLD to identify linked travelers. Subsequently, the travel patterns of these linked travelers were analyzed for the entire month to which the observed week belonged. This approach simulates the construction of such data panel, allowing for longer-term tracking of travel behaviors using MDLD data.

As shown in Table 3-1, MDLD originally covers a significantly larger sample size compared to RTS, with 23,224 travelers observed within just one week. This number is over 16 times greater than the total number of travelers in RTS, which includes 1,383 individuals. The substantial sample size in MDLD enhances the likelihood of finding travelers with similar travel behaviors in both datasets. Consequently, approach 1 successfully matched 617 travelers from MDLD with RTS, while approach 2 achieved 977 matches. These numbers represent approximately 44.6% and 70.6% of the travelers in RTS, respectively. This suggests that incorporating trip records from more days in MDLD could further increase the number of travelers matched with RTS.

Table 3-1 Sample size, timespan, and trip rate comparison

	RTS	MDLD	Linked_approach1	Linked_approach2
Number of travelers	1,383	23,224	617	977
Median number of days in a month	1	10	10	13
Mean number of days in a month	1	11.4	10.89	13.14
Median number of longest consecutive days in a month	1	4	3	5
Mean number of longest	1	5.413	4.843	6.273

consecutive days in a month				
Median trips per day	2	2	2	2.5
Mean trips per day	2.837	2.772	2.215	3.140

One notable distinction is that RTS collects trip records for each responder on a single travel day, while MDLD provides trip records for an average of 11.4 days per traveler in a month. On average, the longest consecutive travel days observed for each MDLD traveler is 5.413 days. This highlights one of the advantages of MDLD, as it captures daily, weekly, and seasonal variations in travel behaviors due to its multiple-day coverage. For travelers linked by approach 1, the average number of days is 10.89, and the average longest consecutive days is 4.843, both of which are less than those in the entire MDLD dataset. Conversely, travelers linked by approach 2 show averages of 13.14 days and 6.273 longest consecutive days, both exceeding the entire MDLD dataset. This indicates that approach 1 tends to link travelers with shorter timespans in MDLD to RTS, while approach 2 is more capable of identifying travelers appearing over more days.

Regarding the trip rate, travelers in RTS are recorded to have an average of 2.837 trips per day, which closely aligns with the average of 2.772 trips per day made by travelers in MDLD. It should be noted that these numbers are slightly lower than the typical trip rate found in other datasets. This discrepancy can be attributed to our exclusion, as mentioned in Section 2, of any trips departing from or entering the county from different regions. This filtering step was applied to ensure the accuracy and relevance of the data. Upon examining the travelers linked by approach 1, an average of 2.215 trips per day was observed. Conversely, travelers linked by approach 2 demonstrated an average of 3.140 trips per day. This suggests that approach 1 is more inclined to identify individuals who undertake fewer trips, whereas approach 2 is better suited for detecting more active travelers. Considering that approach 2 can also identify travelers appearing in more days, it is reasonable to assume that the data panel constructed based on the linking results from approach 2 will include a higher proportion of active travelers.

3.4.2 Trip characteristics

1) Trip duration

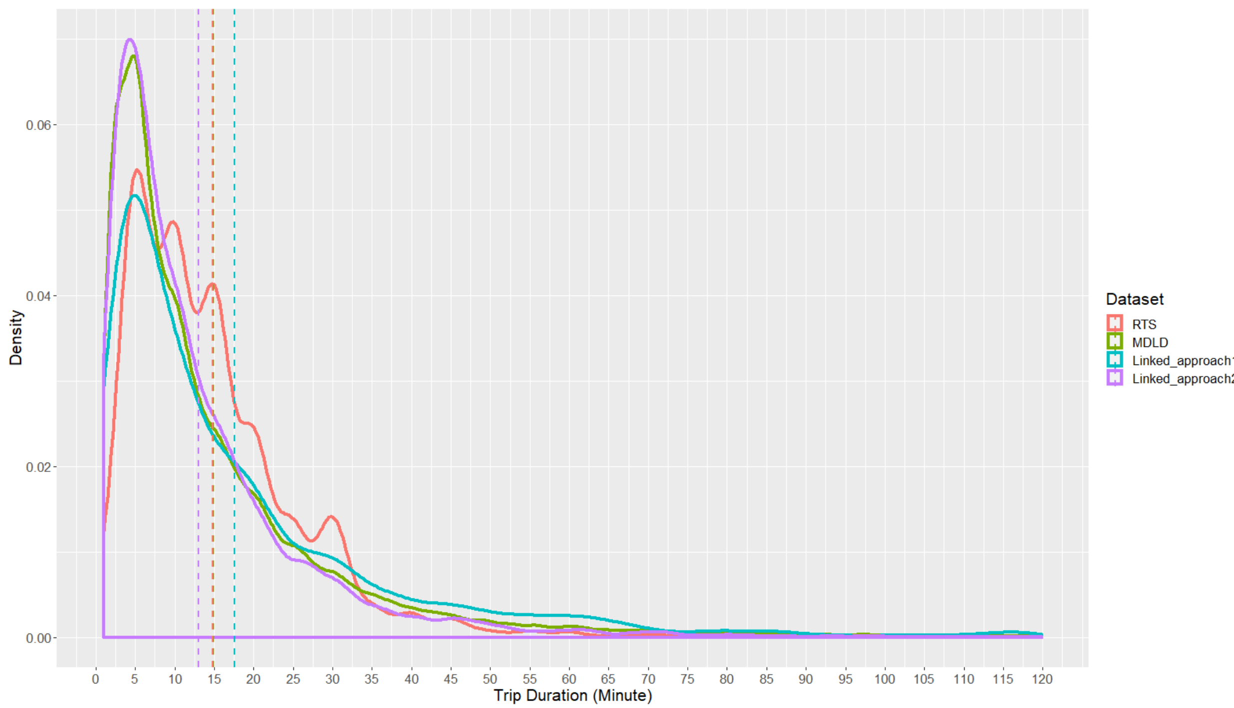


Figure 3.4 Density of trip duration in minute

Figure 3.4 illustrates the distribution of trip durations across the four datasets, with dashed lines representing the respective means. The average trip duration in RTS is 14.8 minutes, closely followed by MDLD with an average of 14.9 minutes, making the red and green dashed lines nearly indistinguishable. However, variations are observed among linked travelers. Those linked by approach 1 display an average trip duration of 17.56 minutes, indicating that although they make fewer trips per day, their individual travel times are longer. Conversely, travelers linked by approach 2 exhibit an average trip duration of 13.00 minutes, indicating a tendency for more frequent short trips per day. This pattern is consistent with the density curves, where the share of short-duration trips (less than 15 minutes) for travelers linked by approach 1 (represented by the blue curve) is significantly lower than that for travelers linked by approach 2 (represented by the purple curve).

Furthermore, Figure 3.4 reveals that trip durations in RTS cluster around multiples of five minutes (e.g., 5, 10, 15...), indicative of "reporting bias" or "rounding bias" commonly encountered in surveys [172]. When respondents are asked to report specific trip details, such as start and end times, they tend to round or approximate their responses to convenient intervals, like 5-minute increments. In contrast, the density curves of trip duration in the other three MDLD-based datasets appear smoother. This is due to the passive collection of travel information in MDLD, which does not rely on self-reported data from survey respondents. Consequently, reporting bias is significantly reduced or eliminated in MDLD, making it a valuable and reliable resource for studying and understanding travel patterns and behaviors.

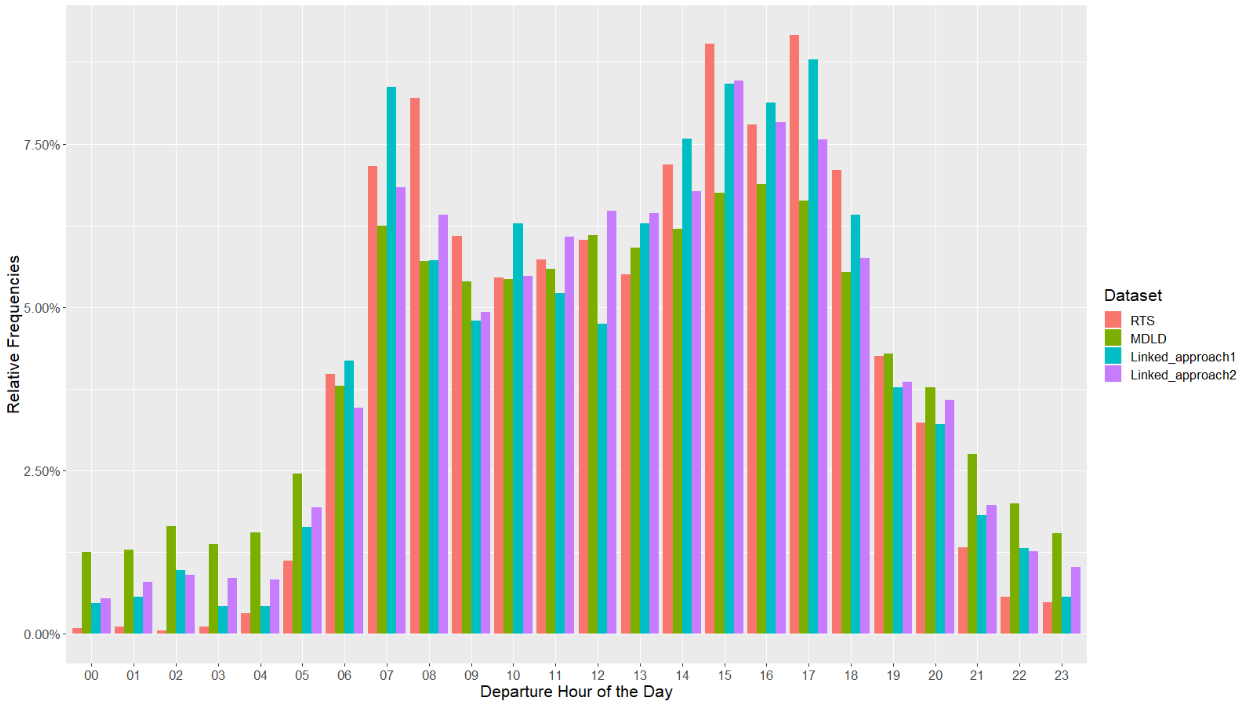
2) Time of day and day of week

The departure hour and arrival hour distributions of recorded trips are presented in Figure 3.5(a) and Figure 3.5 (b), respectively. Across all four datasets, a consistent trend is observed in the temporal distribution, with significantly higher traffic volume during peak periods (7:00-9:00 and 15:00-18:00) compared to other times of the day. However, a noticeable discrepancy exists between RTS and the other three MDLD-based datasets, with RTS reporting fewer trips during off-peak hours, particularly during the night (21:00-5:00). This reveals the presence of off-peak travel underreporting in travel surveys, wherein individuals tend to report fewer off-peak trips compared to what is actually revealed in MDLD. This phenomenon has also been reported in other studies [110].

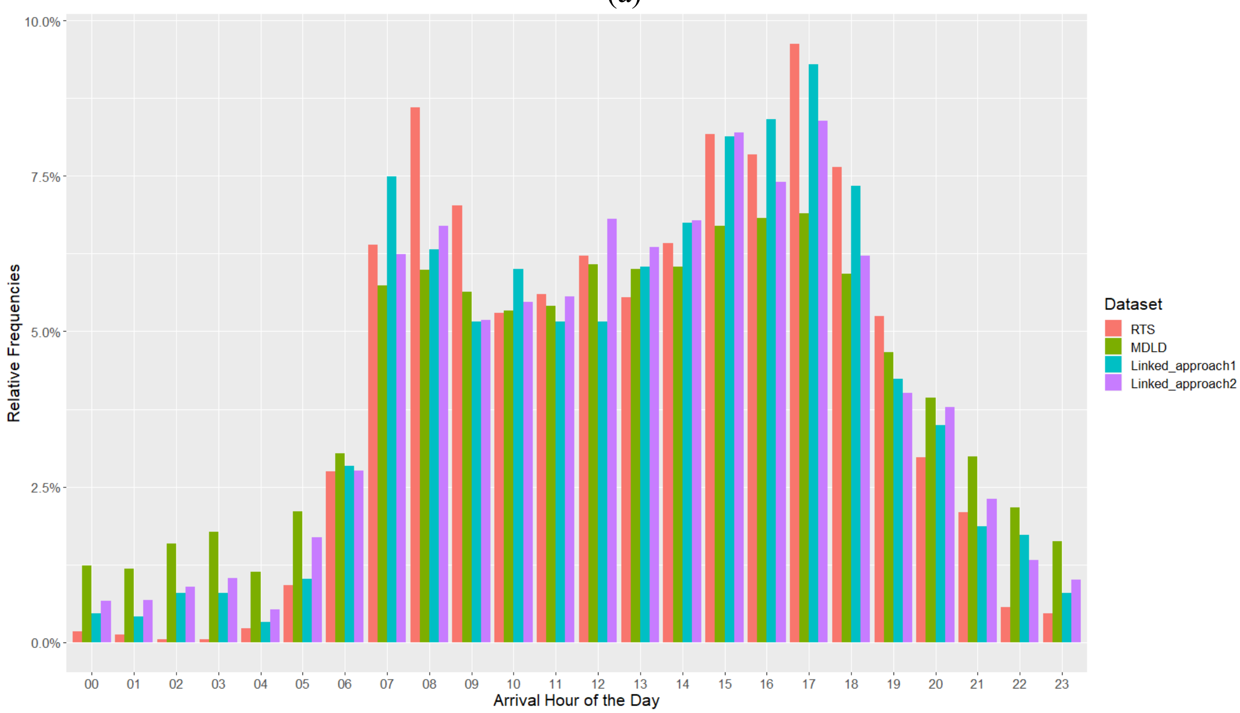
The reasons for off-peak travel underreporting in RTS may be twofold: Firstly, respondents may have difficulty accurately recalling and reporting off-peak trips, especially if they occur infrequently and are not as memorable as peak-hour travel. Secondly, compared to peak-hour trips, respondents might perceive off-peak trips as less relevant or important, especially if they are part of continuous activities or involve short distances, leading to their omission from the reported trips.

This underreporting of off-peak travel in RTS can lead to an imbalanced representation of travel behavior, resulting in an underestimation of off-peak travel patterns. Consequently, RTS may not fully capture the true extent and nature of off-peak travel activities. The higher proportion of trips during midday and late night in MDLD-based datasets indicates that MDLD can be advantageous in addressing this bias. By providing continuous tracking of travel behavior, MDLD offers a more comprehensive and accurate representation of all travel activities, regardless of the time of day.

Regarding the comparison between travelers linked by the two approaches, it is evident that travelers linked by approach 2 exhibit a relatively higher share of off-peak trips and a lower share of peak trips compared to travelers linked by approach 1. This suggests that approach 1 tends to identify individuals following a more regular timetable throughout the day, while approach 2 captures a broader range of travel patterns, including more off-peak travel activities.



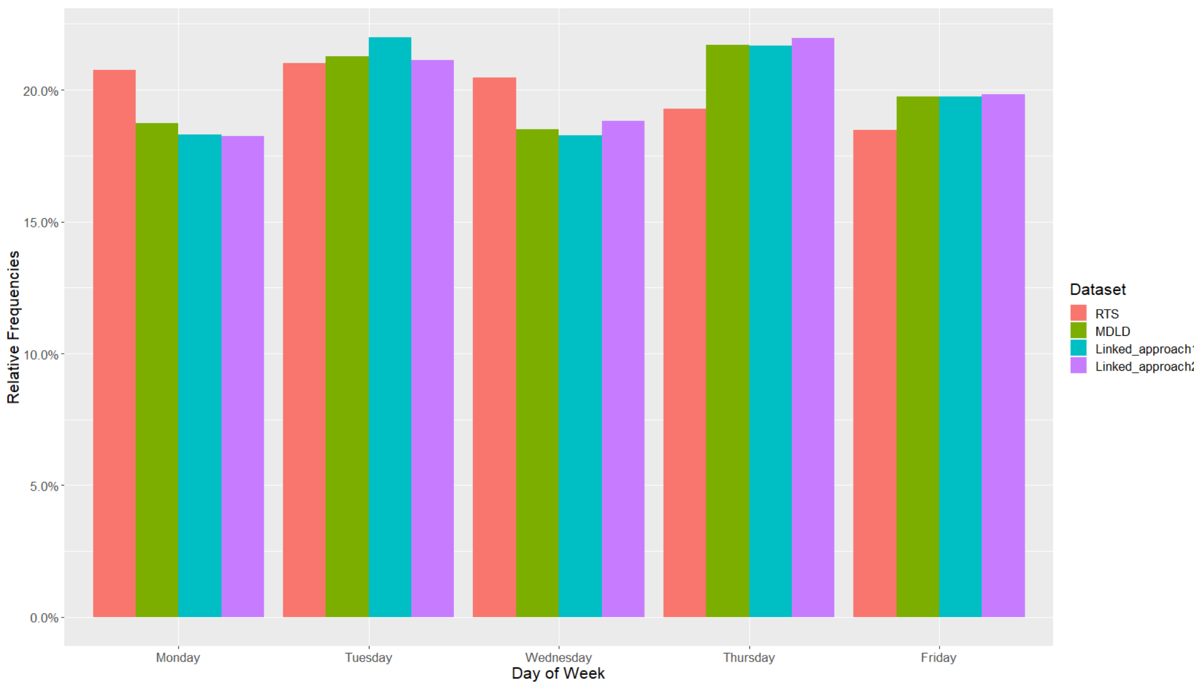
(a)



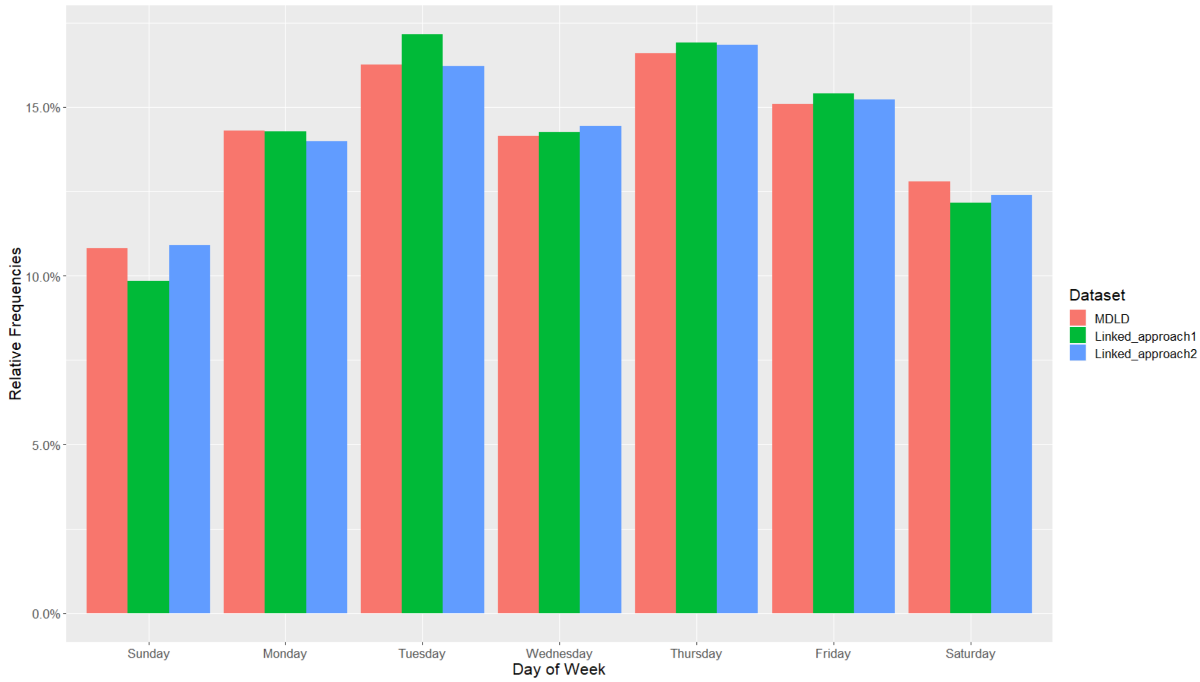
(b)

Figure 3.5 Distribution of trips by the hour of day for (a) the departure time, and (b) the arrival time

Figure 3.6 presents the distribution of trips by day of the week. As RTS only includes weekdays as travel days, Figure 3.6(a) displays the relative frequencies of trips during weekdays for the MDLD-based datasets to facilitate a meaningful comparison. Meanwhile, Figure 3.6(b) illustrates the distribution of trips for all seven days of the week in the three MDLD-based datasets. This distinction highlights one of the major contributions of MDLD, as it overcomes the limitations of self-reported surveys by providing continuous tracking of travel activities.



(a)



(b)

Figure 3.6 Distribution of trips by (a) weekdays, and (b) weekdays and weekends

From Figure 3.6(a), it is evident that RTS records a higher proportion of trips on Monday and Wednesday compared to MDLD, whereas MDLD records a higher proportion of trips on Thursday and Friday compared to RTS. Figure 3.6(b) reveals that travelers linked by approach 1 make slightly more trips during weekdays and fewer trips during weekends compared to travelers linked by approach 2. This finding suggests that approach 1 may tend to link more regular commuters, who primarily travel on weekdays, while approach 2 captures a broader range of travel patterns, including individuals who travel on both weekdays and weekends.

3) Home-based trips and heat map of trip destinations

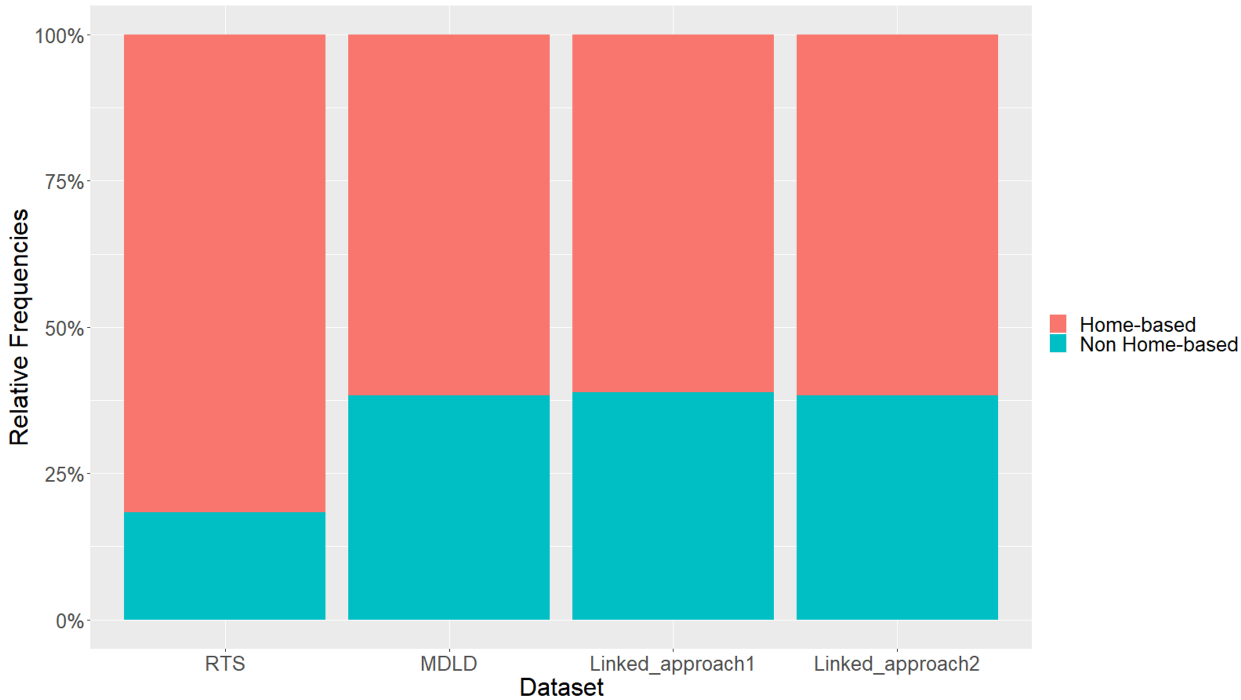
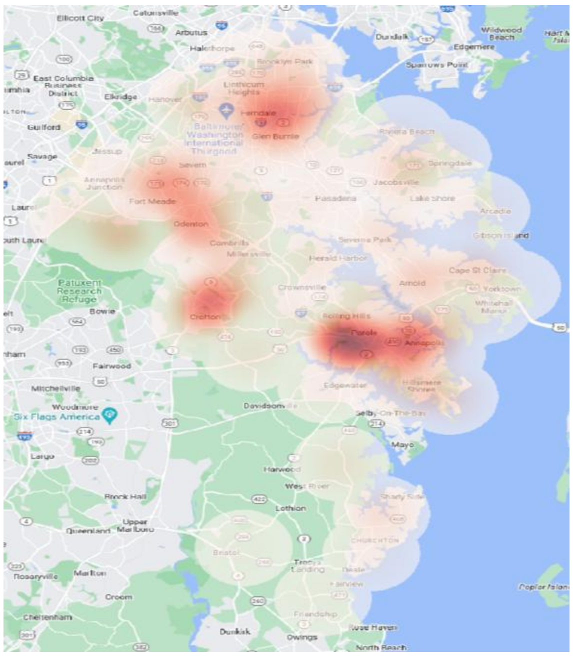
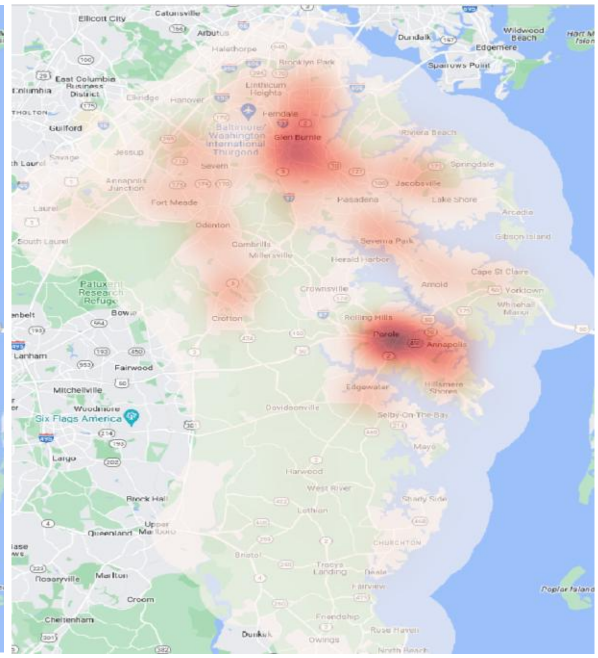


Figure 3.7 Proportion of home-based trips

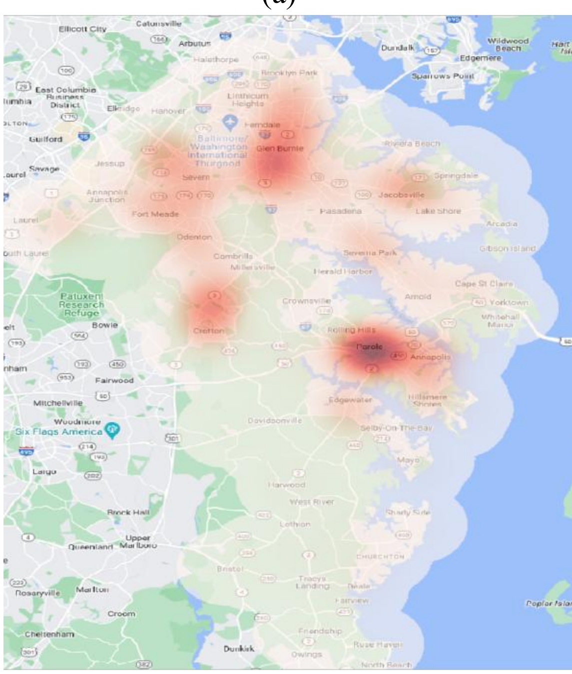
The analysis highlights the variation in the proportion of home-based trips, as depicted in Figure 3.7, which underscores the disparities in travel patterns captured by the RTS and MDLD datasets. In RTS, the proportion of home-based trips is recorded at 81.67%, significantly higher than the three MDLD-based datasets, where home-based trips constitute 61%-62% of the total. This pattern corroborates the discussion in section 4.2.2, indicating that the design and limitations of the RTS data lead to a primary emphasis on trips originating from home and then returning home, which are common and essential travel patterns. As a result, other types of trips, especially those made during off-peak hours for non-home-related purposes, are not adequately represented in the RTS dataset. This omission leads to an incomplete picture of overall transit use and may result in an underestimation of traffic volume during off-peak periods. Therefore, by integrating MDLD into the analysis, researchers can gain valuable insights into the full spectrum of travel patterns, enabling more accurate and inclusive assessments of traffic volume during both peak and off-peak hours. The proportions of home-based trips for travelers linked by approach 1 and approach 2 are both similar to the proportion observed in the entire MDLD dataset. This indicates that the linkage algorithm employed in the study does not introduce bias in terms of home-based trips.



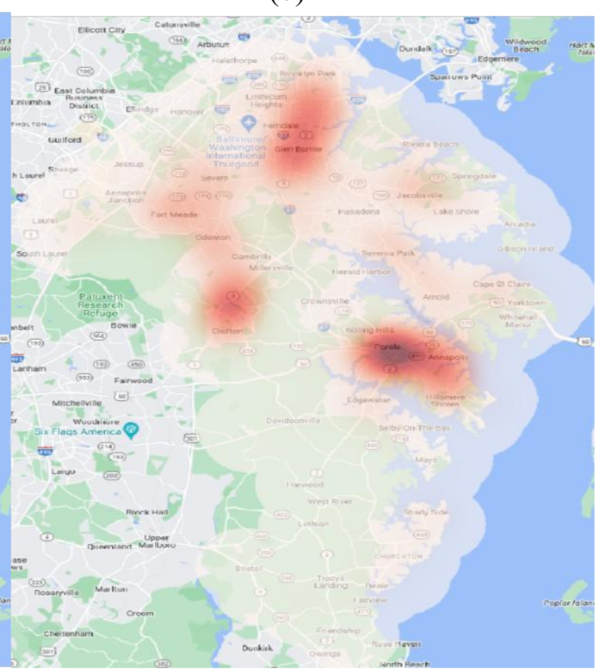
(a)



(b)



(c)



(d)

Figure 3.8 Heat maps of destinations of trips in (a) RTS, (b) MDLD, (c) travelers linked by approach 1, and (d) travelers linked by approach 2

Figure 3.8 presents heat maps of trip destinations from the four datasets, with a focus on trips within Anna Arundel County as mentioned previously. The spatial distribution of trip destinations is found to be generally consistent across all four datasets. Annapolis, the county's largest city, emerges as a prominent hotspot, along with other densely populated towns like Glen Burnie and Crofton. This observation reinforces the reliability and validity of the MDLD data and the linkage approaches employed in the study, as the trip destinations exhibit a reasonable and coherent spatial pattern.

However, there is still a noticeable discrepancy between the heat maps of RTS and MDLD. In RTS, trip destinations appear to be more discrete, with certain rural areas showing no trip records (evident in the bottom left area of the map). Conversely, the MDLD-based datasets exhibit a more widely distributed pattern, covering nearly all areas within Anna Arundel County. This difference may be attributed to the fact that RTS has a smaller sample size compared to MDLD, with trip records collected for a single day for each participant. Consequently, trips made in some rural areas are likely underreported in RTS. MDLD, on the other hand, emerges as a promising data source to address this limitation. By providing continuous movement records over multiple days and encompassing a larger sample size, MDLD offers a more comprehensive representation of travel behavior. This comprehensive coverage ensures that trips made in various areas, including rural regions, are adequately captured, leading to a more accurate depiction of spatial travel patterns within the county.

3.4.3 Socio-Demographics

This section presents the results of socio-demographics for the individuals residing within Anna Arundel County in RTS, individuals linked by approach 1, and individuals by approach 2.

Figure 3.9 illustrates the age distribution across the three datasets, with dashed lines representing the respective means. The average age in RTS is 43.34 years old, which matches the average age of individuals linked by approach 1. Individuals linked by approach 2 exhibit a slightly older average age of 44.13 years old. While there are some slight differences in age distribution among the three datasets, such as more individuals between 45-60 years old and fewer individuals between 25-45 years old in individuals linked by approach 1, the Kolmogorov-Smirnov Test reveals that the p-values for comparing the entire RTS with individuals linked by approach 1 and approach 2 are 0.7851 and 0.9155, respectively. Both p-values are significantly greater than 0.05, indicating insufficient evidence to reject the null hypothesis that these datasets do not come from the same age distribution. Figure 3.10 displays the gender distributions across the three datasets, showing similar proportions with approximately 52% of individuals being female and 48% being male in all datasets. Figure 3.11 shows the race distributions, with the highest percentage of individuals being white, accounting for nearly 80% across all datasets. The distribution of employment status is presented in Figure 3.12, demonstrating comparable patterns among the three datasets. More than 50% of individuals in all datasets are workers. It is notable that the percentage of workers among individuals linked by approach 1 and approach 2 is higher than in the overall RTS. This finding may be attributed to the fact that workers generally make more trips than non-workers, increasing their likelihood of being linked with MDLD, given that the linkage algorithms are based on trip characteristics.

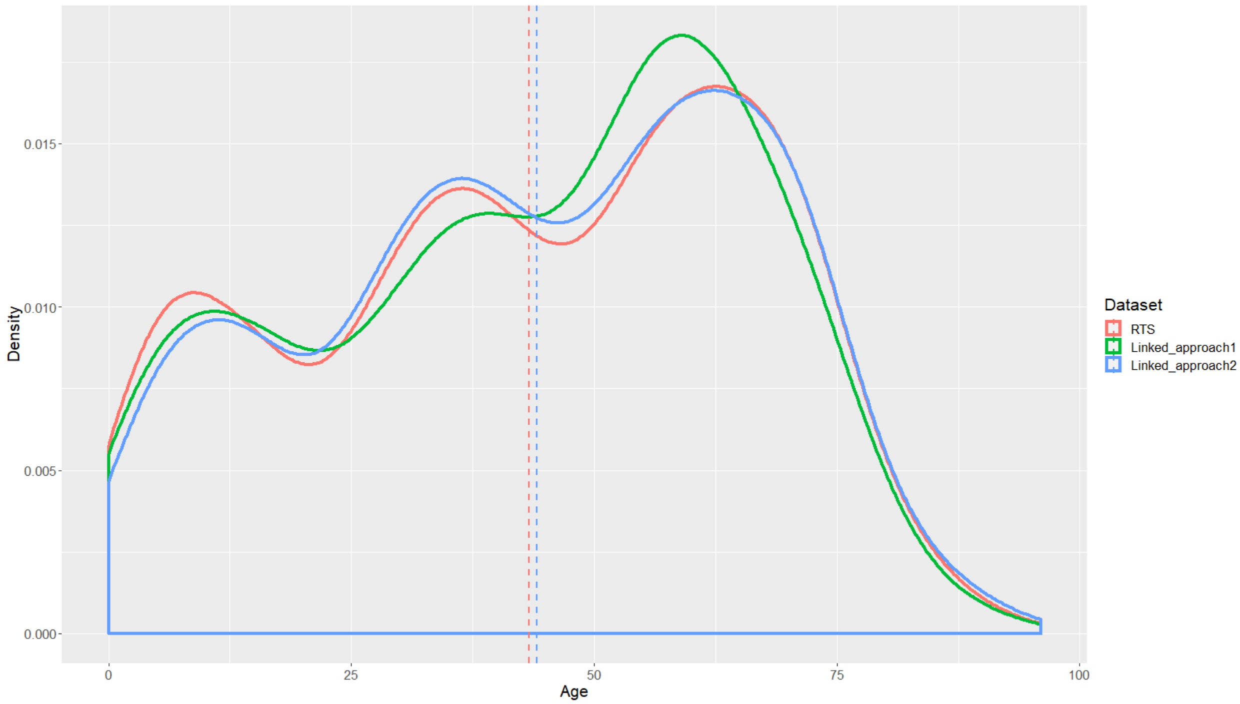


Figure 3.9 Age distribution

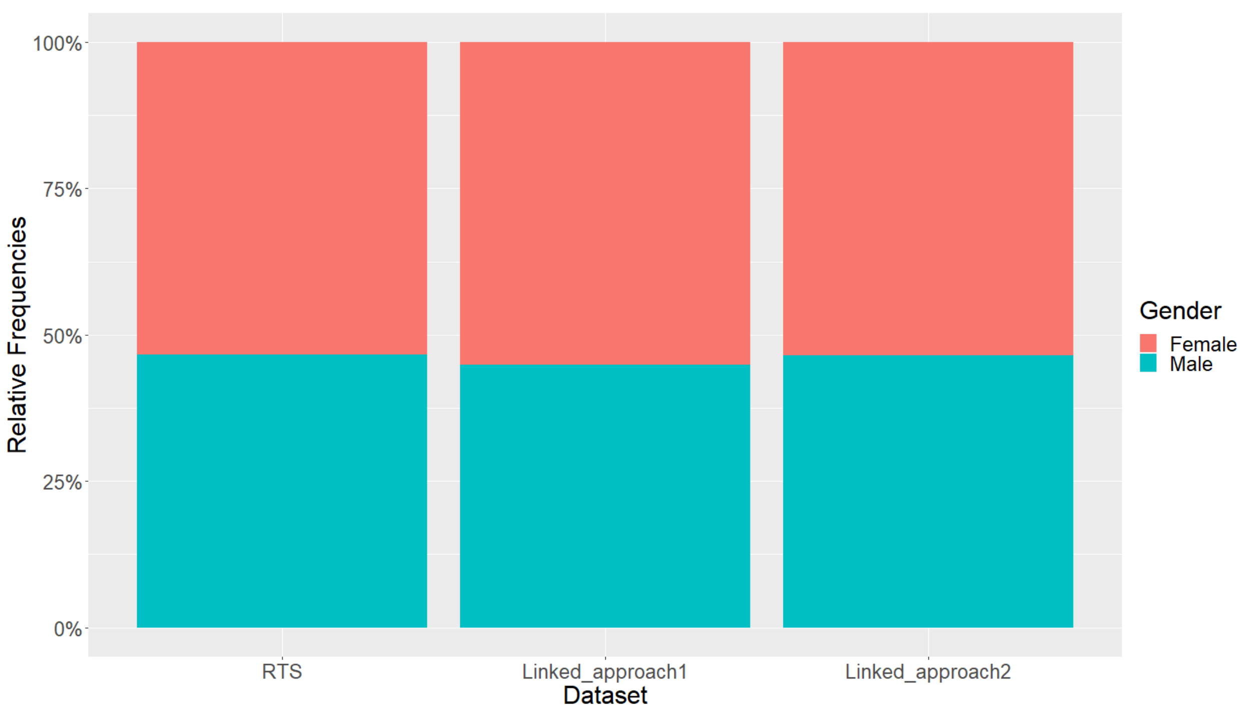


Figure 3.10 Gender distribution

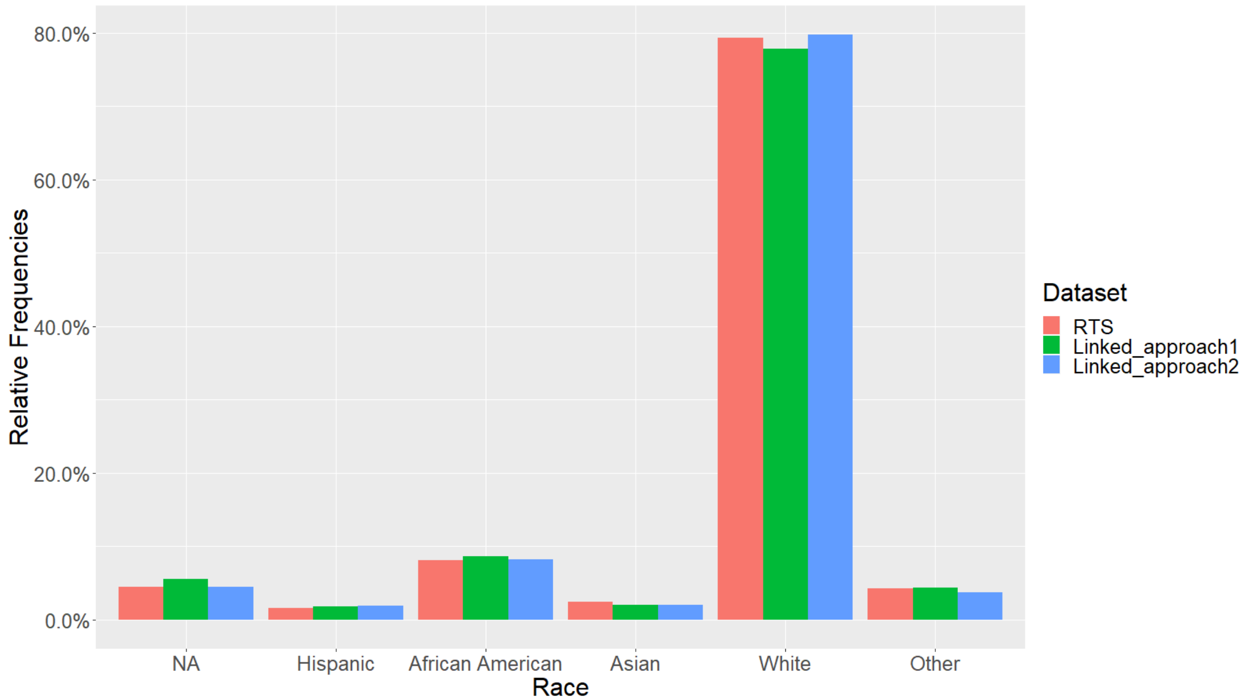


Figure 3.11 Race distribution

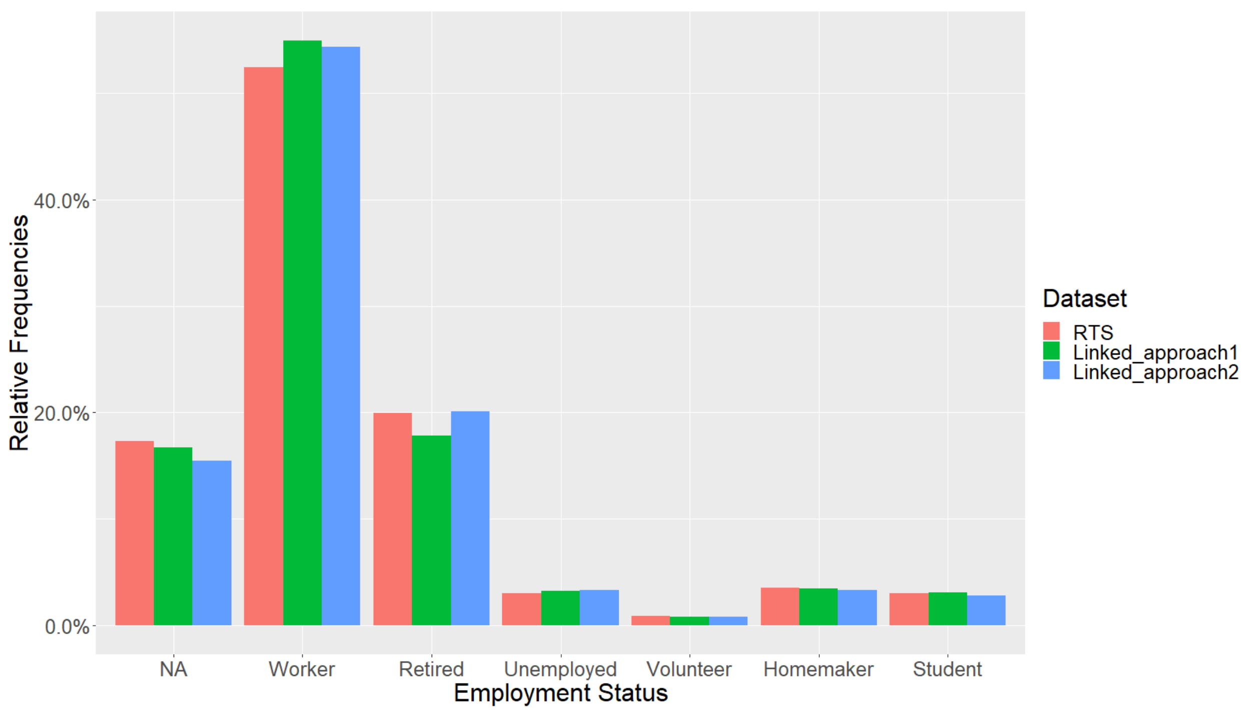


Figure 3.12 Employment status distribution

In conclusion, the distribution of various socio-demographic attributes across the three datasets is generally similar, suggesting that the data panel constructed by the linkage algorithm exhibits comparable demographic characteristics to the original RTS. Thus, it can be considered as representative of the population, similar to the RTS dataset itself. The consistency in socio-demographic profiles enhances the validity and applicability of the data panel for studying travel behavior and making informed decisions in transportation research and planning.

3.4 Discussion

This chapter proposed and implemented a novel approach to enhance travel behavior analysis by integrating Mobile Device Location Data (MDLD) with the Regional Travel Survey (RTS) datasets. Our goal was to address the limitations of traditional travel surveys and leverage the advantages of MDLD to create more comprehensive and accurate representations of individuals' travel behaviors. Two distinct data linkage approaches were utilized to connect individuals across datasets with similar travel characteristics, resulting in the construction of data panels that provide valuable insights into travel patterns.

MDLD offered continuous and high-resolution information on individuals' movements over multiple days, providing a detailed and accurate picture of daily, weekly, and even seasonal travel activities. By combining this geospatially accurate data with the rich demographic information from RTS, our data panels provided comprehensive and reliable representations of travel behaviors. Our approach captured not only peak-hour commutes but also off-peak travel and non-home-related trips, shedding light on previously underreported travel behaviors and offering a more holistic view of individuals' travel patterns. This comprehensive dataset also allowed us to analyze travel behaviors over extended periods, providing a longitudinal perspective on travel habits that was challenging to achieve with traditional survey data.

While our approach showed significant promise, it is essential to acknowledge some limitations. While MDLD provides continuous location data, there may still be gaps in the dataset due to factors such as poor network coverage, battery depletion, or device malfunctions. The effectiveness and accuracy of the home location identification and trip imputation algorithm also needs further validation. The data linkage process also presents challenges, particularly in dealing with multiple possible matches at the trip level. Our algorithm addressed this by leveraging unique matches to establish connections between individuals or devices. While this approach was effective, it is crucial to further explore and refine data linkage methods to ensure accuracy and reliability in large-scale applications.

Despite these challenges, the findings of our study have significant implications for transportation planning and policy-making. By understanding travel behaviors across various time periods and demographic groups, policymakers can tailor transportation services to meet the specific needs of individuals and communities, and cities can develop targeted interventions to enhance mobility and reduce congestion, ultimately contributing to a more livable and sustainable urban environment. As data collection technologies continue to advance, the integration of diverse datasets holds promise for revolutionizing travel behavior analysis and shaping the future of transportation planning in the era of mobile technology and big data.

4. Chapter 4: Analyzing Vehicle Miles Traveled (VMT) Patterns and Gasoline Price Sensitivity Using Mobile Device Location Data

4.1 Problem Statement

In this chapter, we wish to answer two related research questions. First, is it possible to form longitudinal datasets of personal and neighborhood-level VMT based on MDLD data? Implicit in this question is whether the more granular data offers insights and patterns that would be lost when using more aggregate data.

Second, do the more granular data exhibit reliable associations with variables that we would expect to be related with them, such as socio-demographics, economic status, and the price of fuels? To further elaborate on this, do individual and area-level VMT mirror fluctuations in the price of fuels and the local variation in the price of fuels? Conversely, can individual or area-level VMT be used to capture the local demand for gasoline, which should in turn influence the price of gasoline?

While a large literature has examined the impact of socioeconomics and fuel prices on vehicle miles traveled, these studies often rely on conventional travel surveys to produce personal or household-level VMT [108, 111-115], and for these reasons their sample sizes have been small and/or they have lacked a longitudinal component. Another important difference between our work and earlier studies is that we have the weekly price of motor fuels from each gas station in our study area. By contrast, the majority of the earlier studies have deployed state-level [116]or Petroleum Administration for Defense Districts (PADDs)-level prices [117], which assign the same price figure to locales in different states—a situation with a clear measurement error in one key regressor.

We examine MDLD data for 2021 and 2022 from the Washington, DC, metro area, for which no traditional travel survey is available for comparison during the same time frame. We use these data to form a number of longitudinal datasets that track either the daily VMT of individual devices over time, or the average daily VMT in a census block group or tract over time. We discuss their respective advantages and disadvantages, and assess their validity by checking how well they are predicted by local socio-demographics, land use and road infrastructure, and gasoline prices.

We find that with both individual device panels and census area "pseudo panels" the VMT are well predicted by these variables. Importantly, the price of gasoline is negatively related to the VMT, and the implied price elasticities (-0.17 to -0.41), which we interpret to be short-run elasticities, are plausible and consistent with our expectations and previous studies (e.g., −0.013 to −0.577, as in Goetzke & Vance [173]; also see Dimitropoulos et al., [174]). *Total* VMT in an area, on the other hand, are positively associated with gasoline prices, suggesting that they capture demand pressures. Their effect however vanishes when our regression includes population characteristics, land use and infrastructure. Overall, our assessment criteria suggest that MDLD are valuable, and that the device-level longitudinal datasets based on them are a promising source of information for those wishing to explore miles driven and how they are affected by socioeconomic characteristics, land use and infrastructure, and the price of fuel.

The remainder of this chapter is organized as follows. In section 4.2, we explain the MDLD and how they are processed. Section 4.3 provides information about other data utilized in this study. Section 4.4 outlines our methods, and Section 4.5 presents the results. Section 4.6 concludes.

4.2 Mobile Device Location Data

4.2.1 Raw Data Input & Study Area

The raw Mobile Device Location Data (MDLD) utilized in this paper are derived from a consistent national mobility data panel produced by the University of Maryland's NextGen National Household Travel Survey (NHTS) team in collaboration with multiple data vendors. A sample from 2020 reveals that this national mobility panel comprises data from more than 68 million devices, each with a minimum span of 7 days with a minimum of 10 daily sightings within the given month, indicating a sampling rate greater than 20.6% in terms of temporally consistent devices. These devices were observed on average for 24.2 days in a month, with an average of 234.4 sightings per day, which indicates a high level of temporal consistency and comprehensive coverage. Furthermore, the average positioning accuracy, defined as the maximum distance between the recorded location and the true location at the 95% confidence level, is 49.2 feet (15 meters), signifying that the location sightings are very accurate. These high-quality raw data serve as inputs to impute the device owners' home locations and their trip diaries.

Figure 4.1 Geographic Sub-Areas – Core (Inner Ring), Middle Ring, Outer Ring

We use MDLD from the Washington, DC, metro area from May 2021 to the end of 2022. We focus on the District of Columbia plus the "outer ring" of counties around it (except for Loudoun County). We cover five Maryland counties (Calvert, Charles, Frederick, Montgomery, and Prince George's) and five Virginia counties or independent cities (Arlington, the City of Alexandria, the City of Falls Church, Fairfax, and Prince William) (see Figure 4.1).

Among other reasons, we focus on this area because of its varied commuting patterns: The 2019 and 2022 State of the Commute Surveys conducted by the Metropolitan Washington Council of Governments (MWCOG) indicate that 35% of workers commute to Washington, DC, and 58-61% to a workplace in the "middle ring" locations. Moreover, the length of the commute (in miles) has remained stable despite the pandemic (mean 17 miles one-way, median 14 miles one-way). What has changed is however the share of workers that "telecommute" (work from home) at least on an occasional basis. This share was 33% in 2019 and doubled to 65% in 2022 [2, 121].

.

4.2.2 Imputed Home Location and worker status

In the 11 counties in our study area, in 2021 and 2022 the monthly average sampling rate is 8.29%. The sampling rate is calculated as the number of devices divided by the total population residing in each county for each month. Notably, Charles County, MD, records the highest sampling rate at 11.16%, while the city of Alexandria, VA, reports the lowest at 5.62%. The monthly average sampling rate for workers is 7.07% overall, ranging from a low of 4.61% in Arlington County, VA, to a high of 10.80% in Calvert County, MD. Even the lowest sampling rate remains significantly higher than what is typically seen in traditional travel surveys within the field of transportation.

4.2.3 Imputed Trip Roster

The raw sighting data used in this study lacks trip information, a crucial element for understanding changes in human mobility. Trip data is the primary unit of analysis in our research. A tour-based method is employed to infer trips from the sightings [85]. With the home locations identified in the previous step, sightings between two home observations are classified as part of a home-based tour. Within each tour, a decision tree-based recursive algorithm is used to determine the status of each sighting, identifying whether it is continuing an ongoing trip, static, or starting a new trip. By sorting the sightings by timestamps, trips are identified based on the location and time information of their origins and destinations.

Next, the travel mode of each trip is imputed using a machine-learning based algorithm, which allows us to distinguish vehicle trips from all the other travel modes [76]. Air travel is distinguished first due to its unique characteristics of starting and ending at airports, long travel distances and durations, and high speeds, making it easier to separate from other modes. For ground transportation, a random forest model is trained and tested to impute vehicle, bus, rail, walk, and bike trips. This model uses features such as the recording interval of sightings, distance, time, speed-related factors, and multimodal transportation network characteristics. The cross-validation accuracy of the training results reaches 97.1%.

In addition, considering the imputed worker status, the purpose of each trip is determined using a combination of heuristic rules and a decision tree model. Trips that start or end at the workplace by workers on workdays are labeled as work trips. For long-distance trips (more than 50 miles from home), the trip purpose is imputed based on various factors, including trip-related information such as the primary travel mode, destination-related information such as state and tourism visits, and traveler-related information such as worker status. Ultimately, each trip is labeled as either a work trip or a non-work trip.

As a result, we generate a daily trip roster that includes a unique trip identifier, start and end times, coordinates for the origin and destination, imputed travel mode, and the imputed trip purpose (work or non-work) for each trip.

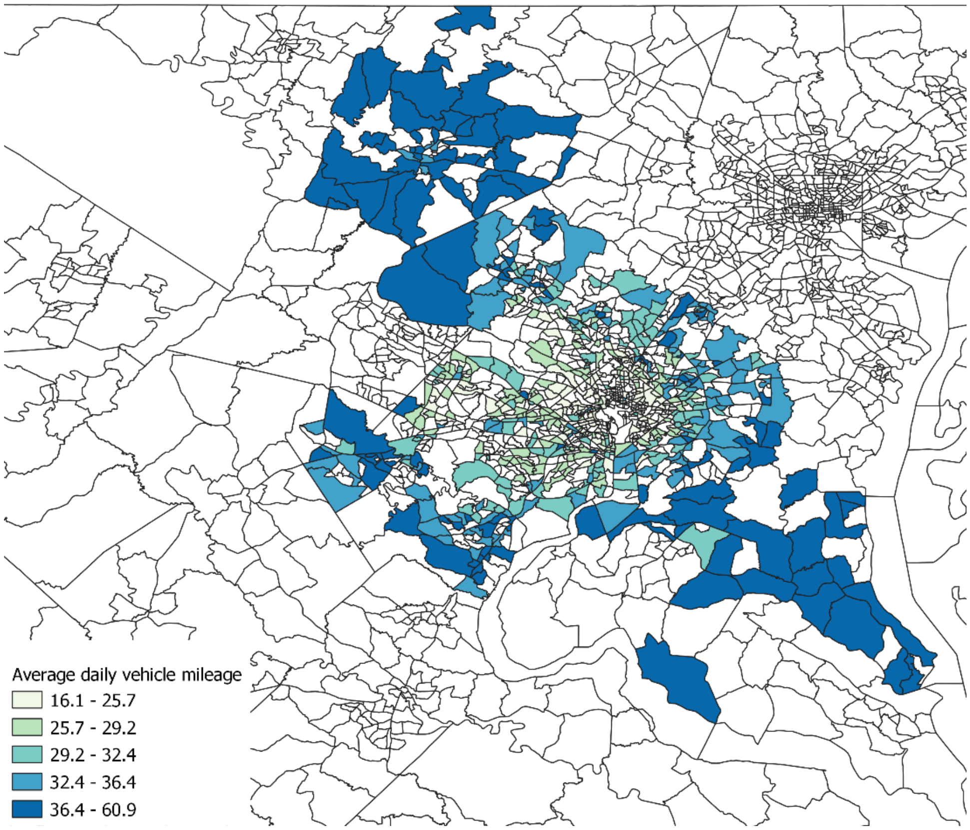


Figure 4.2 Average Daily Vehicle Mileage at the Census Tract Level

In the remainder of this paper, attention is restricted to weekday travel by car or other motor vehicle. During our study period, the average number of vehicle trips remained relatively consistent at 2.5 trips per day and the mileage was 28.95 miles per day. Figure 4.2 presents the spatial distribution of average daily vehicle mileage at the census tract level during the study period. Notably, individuals residing in close proximity to Washington, D.C. tend to log fewer daily vehicle miles, with most covering distances of less than 30 miles per day. Conversely, those living farther from Washington, D.C., such as in Frederick County, MD, Calvert County, MD, and Prince William County, VA, tend to have higher daily vehicle mileage, typically ranging from 36.4 to 60.9 miles per day. This observed pattern aligns with the expectation that individuals living near the city center generally have shorter commute distances and greater access to public transportation options, reducing their reliance on personal vehicle travel, and bodes well for the statistical models described in section 4.5 below.

4.3 Gasoline Prices Data

We purchased gasoline price data from a commercial vendor. Specifically, we have weekly gasoline prices at each of the gas stations within the study area for the entire study period. On average, each week 1,123 gas stations were operating and their prices are documented in our dataset.

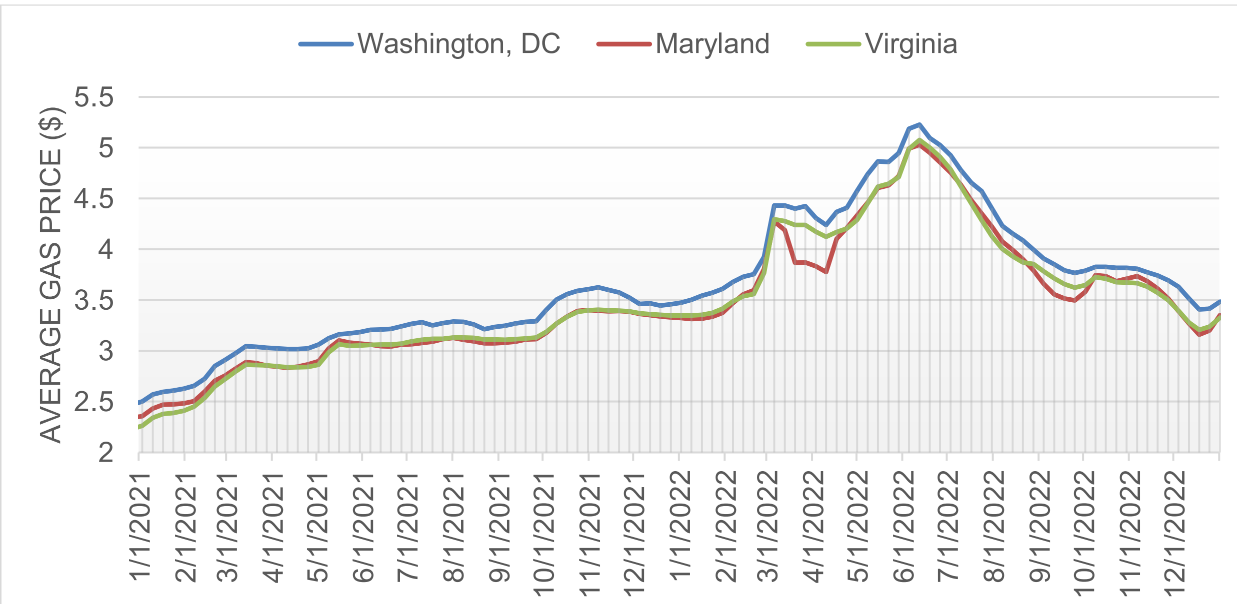


Figure 4.3 The Weekly Variation of Gasoline Price by States

Figure 4.3 illustrates the weekly average gasoline prices in the District of Columbia, and the counties in Maryland and Virginia included in our sample. It is evident that, for the most part, gasoline prices in these states follow a similar trajectory, although they are consistently higher in Washington, DC, than at the other locations. They exhibit a consistent pattern of increasing from the beginning of 2021 to a peak in June 2022, followed by a decline until October 2022, after which they fluctuate. However, a deviation occurred on March 18, 2022, when the General

Assembly of Maryland implemented a 30-day gas tax holiday. Gasoline prices dropped noticeably (if briefly) in Maryland and, to a somewhat lesser extent, in Virginia and Washington D.C., possibly due to competitive forces.

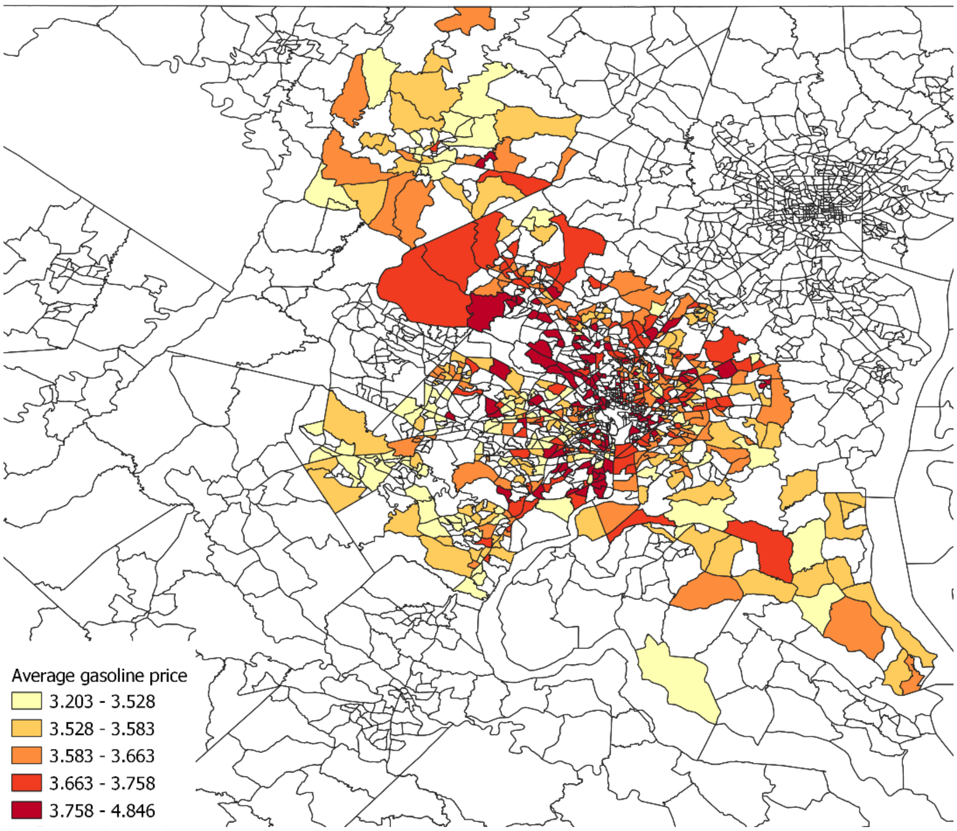


Figure 4.4 Average gasoline price at census tract level

Figure 4.4 depicts the spatial distribution of gasoline prices at the census tract level spanning from January to mid-April 2023. In contrast to the pattern shown in Figure 4.2, it is evident that the closer the area is to Washington, D.C., the higher the average gasoline prices. The disparity between areas with the highest and lowest average gasoline prices exceeds 1.6 dollars per gallon.

4.4 Methods

We examine a variety of descriptive statistics and produce a number of plots for the VMT measured from the MDLD. Since in 2021 the economy was reopening after the early stages of the pandemic, we would expect driving to increase during our study period, although it is not clear whether "more driving" means that more people are on the road, or that each person drives more miles, or both. If the MDLD are a valid measure of mobility, at least the latter tendency should be captured in the data.

We also know, based on the 2022 State of the Commute and personal knowledge, that many workers, including employees of the federal and state governments, continue to work from home several days a week. Does that mean that daily VMTs have stayed low throughout 2021 and 2022? If people are commuting fewer times each week, does that translate into fewer weekly miles, or have they substituted commute miles with driving to other destinations? This question prompts us to compare commute and/or work-related VMT with non-work VMT.

We examine the association between and fuel price by fitting the regression equation

(1) $$y_{it} = \alpha_i + \tau_t + \beta \cdot P_{it} + \varepsilon_{it}$$

where i denotes the cross-sectional unit of observation (a device, or, in alternate regressions, a census tract or block group), t is the week, y is the dependent variable of interest (usually average daily VMT), P is the price of gasoline, and ε_{it} denotes the error term. The right-hand side of equation (1) includes unit-specific fixed effects in hopes of accounting for unobserved heterogeneity and time fixed effects to capture factors that affected the entire study area (e.g., the reopening of the economy, seasonal effects, etc.).

Equation (1) is a two-way fixed effects model, where proper identification of the effect of gasoline price is possible because of the variation within a unit over time and across units in any given period. The appropriate estimation technique is the "within" estimator [175] applied to a model whose right-hand includes gasoline prices and dummies for the week of the sample period.

The price of gasoline started rising in the fall of 2021 and continued to rise until the end of June 2022 (see figure 4.3). At its height, it was over one dollar per gallon more expensive than in the previous year. The demand for gasoline is not very price-elastic, so we expect to detect a modest moderating effect on the tendency to drive during periods of rising prices, accompanied by a slight growth in driving during periods of falling prices. We take advantage of the variation over time as well as across locations within our study area. As shown in figure 4.4, gasoline tends to be more expensive in and around downtown DC, and less expensive in the suburban areas of the "outer ring," but even in the latter the distribution of gasoline prices is uneven and exhibits alternating "pockets" of cheaper and higher prices.

One issue that we must tackle is "which" gasoline price should enter in equation (1). We have weekly gasoline prices for each of the gas stations in the study area, but we do not know where and when the vehicle users in our area buy their gasoline. We experiment with an average price within a short radius of the individual's home (1, 3 and 5 miles), and, for commuting and work-related VMT, with similar averages but around the individual's presumptive place of work. As shown in figures 4.3 and 4.4, the weekly price of gasoline exhibits considerable fluctuations over time (consistently with worldwide trends) and across locations, and tends to be lower in less densely populated area, where, as displayed in figure 4.2, people must and do drive more miles for work and personal reasons. This would result in a mechanically negative coefficient on fuel price in equation (1). Fortunately, the fact that we have repeated observations over time within a unit and that gasoline prices fluctuate over time, addresses this concern and is expected to result in the true causal effect of fuel prices on VMT.

In addition, at least with the device-level data, which are drawn at random from a very large universe, it seems reasonable to assume that an individual alone would have no effect on the market price of gasoline in the area where they live or work. At least in the device-level regressions, the price of gasoline where someone works or lives should be exogenous—at least in the short run. Equation (1) does not contain any information about the individual socioeconomic status, local employment conditions, land use, and infrastructure. To examine their effect on VMT, we use variables from the 2021 American Community Survey and TIGER/Line Shapefiles from U.S. Census Bureau. These variables are time-invariant, which forces us to drop the unit-specific fixed effects from equation (1) and re-specify the regression equation as:

$$(2) \qquad y_{it} = \alpha_0 + \tau_t + \beta \cdot P_{it} + \mathbf{x}_i \boldsymbol{\gamma} + \varepsilon_{it},$$

where $\mathbf{x}$ is a vector of socioeconomic, land use, and road infrastructure variables listed in table A in the Appendix.

If the MDLD are to be considered a valid measure of mobility, it is important to check whether the miles driven by individuals are reasonably stable and consistent from one week to the next. One way to see if this is the case is to fit a panel data dynamic model, namely

$$(3) \qquad y_{it} = \alpha_i + \tau_t + \theta \cdot y_{i,t-1} + \varepsilon_{it},$$

where $y_{i,t-1}$ is the VMT in the previous week. We would expect θ to be positive and not too far from 1 (but below one) if the VMT are well behaved and stable over time.

Unless the longitudinal dimension of the panel T is large, the "within" estimator applied to equation (3), however, results in biased estimates of θ. Instrumental variable estimation is necessary to produce asymptotically unbiased of θ (Hsiao, 2014). We deploy the estimation procedure developed by Blundell and Bond [176], which relies on first differencing the data, combining the first differenced data with the original ones, and exploiting orthogonality conditions (of $y_{i,t-2}$ and other past y's and $\mathbf{x}$'s to form instruments for $\Delta y_{i,t-1}$, and of the changes in y and $\mathbf{x}$ to form instruments for $y_{i,t-1}$) in a system method-of-moments estimation setting.

Finally, we re-estimate equations (1) and (2) after replacing the VMT with the price of gasoline at gas station *i* in week *t* as the dependent variable, and amending $\mathbf{x}$ to include variables that capture gasoline demand and supply factors. These include local competition, which we proxy with the number of gas stations within a quarter-mile of any given gas station, and demand for driving, which we measure with total miles driven in the previous week in the area..

<u>4.5 Results</u>

4.5.1 Panels and Pseudo-Panels from the MDLD

We begin with selecting a 10% sample from our universe of devices. Our first longitudinal dataset follows each device and summarizes its VMT on a weekly basis. As shown in the first column of Table 4-1, this approach results in almost 83,000 devices that appear for T weeks, where T ranges from 1 to 68 weeks, averaging 7 weeks. Zero VMT account for about 5% of the observations.

Table 4-1 Description of Panels

	Device Panel	**Pseudo Panel**	
Unit	Device	Census Block Group	Census Tract
Average daily VMT in each week	By device	By all devices residing in the CBG	By all devices residing in the CT
Average gasoline price in each week	Within 5 miles of the home location	Within the CBG	Within the CT
N. units	82,866 (10% sample)	817	623
Avg N. observations (weeks) per unit	7	62	68
% of Obs with 0 VMT	5%	0.2%	0.02%
N. Devices per unit	-	1-2072, 102 on average	1-2642, 237 on average
Weight used in the regression	-	Number of devices in the BG in each week	Number of devices in the CT in each week
Average daily VMT	(w/o zeros) 28.58 miles (w/ zeros) 27.20 miles	32.01 miles	31.47 miles
Standard deviation of daily VMT	(w/o zeros) 31.90 miles (w/ zeros) 31.72 miles	15.41 miles	10.28 miles

Figure 4.5 displays the average daily VMT week by week based on the individual device panel. These exhibit broad swings at the beginning of 2021, increase as consistent with normal seasonal patterns in the summer of 2021, then drop and then bounce up again in 2022. Such swings may be in part due to the composition of the dataset, including the turnover of participating devices, and in fact, as shown in figure 4.5, weighting the data results in much more muted weekly fluctuations, much more stable VMT during much of 2021, and a pronounced increase from the middle of the end of 2022. Since the weighted data are much more stable and consistent with our expectations, in what follows we compute descriptive statistics and plot them using the weights. We do not use the weights in our device-level regressions, as recommended in Solon et al. [177].

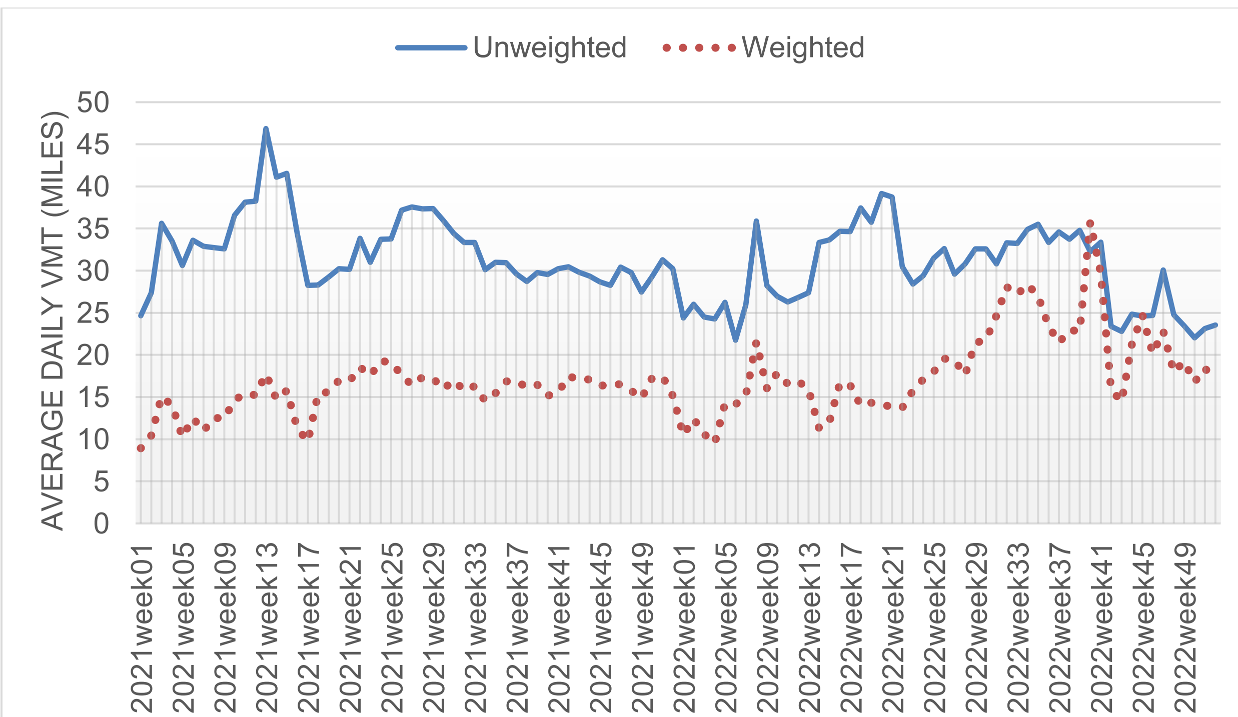


Figure 4.5 Average Daily VMT by week from MDLD

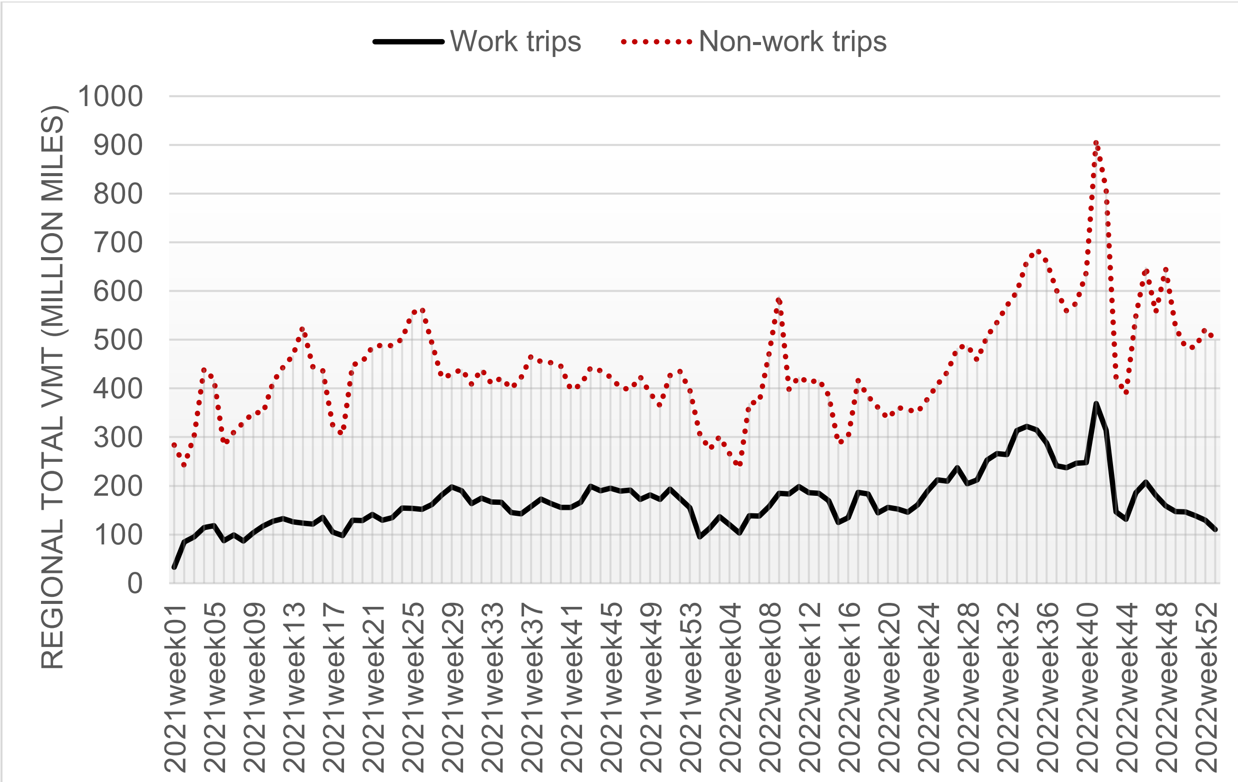


Figure 4.6 Regional Total VMT for work and non-work trips by week from MDLD

Figure 4.6 displays (weighted) total VMT in the study area for both work-related trips and non-work-related trips. Both total weekly commute (or otherwise work-related) VMT and total weekly non-commute VMT actually appear similar overall trend over our study period. Weekly commute miles appear to be less than half of the total non-commute miles, and exhibit much small fluctuations from one week to the next.

One potential problem with longitudinal datasets is attrition, namely the tendency of the participating units to drop out—for good or to reappear after a number of periods [178]. Attrition is not a problem if it is ruled by processes or events unrelated to the dependent variable of interest; it can bias statistical inference if it is ruled by processes or events related to the dependent variable of interest, as might be the case if, for example, high-mileage people tended to drop out of the MDLD faster than low-mileage people.

Unfortunately, it is not possible for us to determine the nature of the attrition in the MDLD, as individuals may themselves change the ID of their device (so that one device would disappear, and another appear) and we are not privy to how the providers pull participants out of the sample and replace them with others. These concerns should however be mitigated when we construct "pseudo panels."

In our "pseudo panels" (shown in the second and third column of table 1) the cross-sectional unit of observation is a geographical area—here, the census block group and the census tract, respectively. In these cases we compute the average of the daily VMT by all of the devices residing in that area. This is a "pseudo" panel because it is possible and likely that different devices have contributed to each week's average daily VMT in that geographical area (Hsiao, 2014, page 408-411) [175]. Drop-outs are presumably replaced by newly recruited devices, which mitigates concerns about non-random attrition. As shown in table 1 this approach produces fewer units (the census block groups or census tracts), but they are followed over a longer period of time (up to T=68 weeks) and the panel is balanced, in that every block group or census tract contributes the same number of observations to the dataset.

It is well known that the mean of a sample of observations has variance σ^2/n, where n is the size of the sample used to compute that mean. Here, n should be replaced by n_t, the number of devices in the block group or census tract in week t. This makes the error terms in equation (1) for the pseudo-panels heteroskedastic, and for this reason we use weights with our pseudo-panel regressions, where the weights are equal to n_t.

4.5.2 The Effect of Gasoline Prices

Tables 4-2 to 4-5 present the results from fitting equation (1) to our panels and pseudo-panels. In table 4-2 the dependent variable is total VMT driven by (the person owning or using) a device, regardless of origin and destination or purpose. In the third and fourth column of this table we average the device-specific VMT over the geographical area where (the owner or user of) the device lives. Tables 4-3, 4-4 and 4-5 are similarly constructed, but the VMT therein are those for specific purposes.

In table 4-2, we used the average gasoline price around the individual's residence or in the specified geographical area (block group or census tract). The coefficients on gasoline price are always negative and strongly statistically significant, implying that people will reduce their driving when gasoline price rises and they will drive more miles when the prices fall, all else the same. The tables also report the price elasticity of VMT, computed at the average price and VMT during the study period.

The estimated price elasticity ranges from -0.17 to -0.41.[1] These values are plausible, and in fact quite strong for (very) short-run elasticities. Note that it was not clear a priori to us whether aggregating the data would result in the appearance of a stronger or weaker elasticity [179]. One reason for the range of elasticities in table 4.3 may be the fact that in the device-level panel, the gasoline price is the average price within 5 miles of the device owner's home, whereas in the census tract and census block group pseudo panels it is the average price in those census areas.

Table 4-2 Estimation results: Equation (1). Dependent variable: Average daily VMT. Gasoline price is the average within 5 miles from home.

	Device-level panel, Weekly	**Device-level panel, incl. zeros, weekly**	**Block group averages, weekly**	**Census tract averages, weekly**
Coefficient on gasoline price	-1.3109 *** (0.1990)	-1.2076 *** (0.1948)	-3.1711 *** (0.1316)	-3.5589 *** (0.1091)
Price Elasticity of daily VMT	-0.1711 (0.0260)	-0.1659 (0.0268)	-0.3593 (0.0149)	-0.4119 (0.0126)
Fixed effects	Device	Device	Block group	Census tract
Time fixed effects	Month, year	Month, year	Month, year	Month, year

[1] Price elasticities at sample averages is $\eta_p = \frac{\text{Coefficient of Gasoline Price} * \overline{\text{Gasoline Price}}}{\overline{\text{VMT}}}$

R square	0.5784	0.5716	0.6444	0.7172
N. Observations	418,147	439,451	49,593	42,258

Note: * = significant at the 10% level; ** = significant at the 5% level; *** = significant at the 10% level.

We also evaluate the effect of gasoline prices on work-related VMT and non-work-related VMT separately, as shown in Tables 4-3 to Table 4-5. In the commuting or work-related VMT regressions, the coefficient on gasoline price has the "wrong" sign, in that it is positive, whether we use the average gasoline price in the area near the presumptive worker's home or work locations (tables 4.3 and 4.4).

We attempted to capture the potentially different sensitivity to gasoline price by adding an interaction between the price of gasoline and a dummy denoting that the owner or user of the device lives in an area where a large share of the residents work for the federal, state or local government. This specification is motivated by the fact that even at the time of this writing non-essential government employees often continue to work for the most part from home, and are required to report to the workplace only once or twice a week. However the coefficient on gasoline price in the work VMT remained positive and that on the interaction was insignificant. We surmise that the model cannot isolate the impact of the gradual end of the COVID-19 pandemic, which has allowed people to resume commuting, from that of the changing fuel prices.

By contrast, with non-work-related VMT, the coefficient on gasoline price is negative in all of the four columns of table 4-5, indicating that while people are returning to commuting, they are still reducing their non-work trips as fuel prices rise. It is noteworthy that the price elasticity, which is always negative in table 4-5, appears to become more pronounced as we aggregate the data to broader and broader geographical areas.

Table 4-3 Estimation results: Equation (1). Dependent variable: Average daily commuting/work-related VMT. Gasoline price is the average within 5 miles from home.

	Device-level panel, Weekly	**Device-level panel, incl. zeros, weekly**	**Block group averages, weekly**	**Census tract averages, weekly**
Coefficient on gasoline price	0.6578 (0.4464)	0.7169 (0.4336)	0.5715*** (0.1180)	0.4267 *** (0.0690)
Fixed effects	Device	Device	Block group	Census tract
Time fixed effects	Month, year	Month, year	Month, year	Month, year
R square	0.7681	0.7658	0.1626	0.2550
N. Observations	112,954	126,892	45,216	42,070

Note: * = significant at the 10% level; ** = significant at the 5% level; *** = significant at the 10% level.

Table 4-4 Estimation results: Equation (1). Dependent variable: Average daily commuting/work-related VMT. Gasoline price is the average within 5 miles from work.

	Device-level panel, Weekly	**Device-level panel, incl. zeros, weekly**	**Block group averages, weekly**	**Census tract averages, weekly**
Coefficient on gasoline price near work locations	1.1601 ** (0.4381)	1.1706 ** (0.4325)	0.8305 *** (0.1097)	0.7040 *** (0.0834)
Fixed effects	Device	Device	Block group	Census tract
Time fixed effects	Month, year	Month, year	Month, year	Month, year
R square	0.7726	0.7739	0.1944	0.2642
N. Observations	110,264	112,290	44,096	41,222

Note: * = significant at the 10% level; ** = significant at the 5% level; *** = significant at the 10% level.

Table 4-5 Estimation results: Equation (1). Dependent variable: Average daily non-commuting/non-work-related VMT. Gasoline price is the average within 5 miles from home.

	Device-level panel, Weekly	**Device-level panel, incl. zeros, weekly**	**Block group averages, weekly**	**Census tract averages, weekly**
Coefficient on gasoline price	-0.7519 *** (0.2026)	-0.6033 ** (0.1968)	-3.8959 *** (0.1662)	-4.1389 *** (0.1244)
Price Elasticity of daily VMT	-0.1000 (0.0269)	-0.0853 (0.0278)	-0.5342 (0.0228)	-0.5904 (0.0177)
Fixed effects	Device	Device	Block group	Census tract
Time fixed effects	Month, year	Month, year	Month, year	Month, year
R square	0.5740	0.5683	0.4532	0.5849
N. Observations	395,897	420,140	44,705	41,540

Note: * = significant at the 10% level; ** = significant at the 5% level; *** = significant at the 10% level.

4.5.3 Socio-Economics, Land Use and Infrastructure, and VMT

Are individual device VMTs or groupwise VMT well predicted by the sociodemographic of the residents of an area, and/or by road infrastructure? The results from estimating equation (2), displayed in Table 4-5, suggest that this is the case: The coefficients on many of the regressors are statistically significant and have the expected signs.

The negative coefficients for population density and road density reveal that individuals residing near city centers tend to have lower vehicle mileage, while those living in areas further from city centers tend to have higher levels of vehicle travel. As expected, people living in areas with a high number of public transit options tend to have lower vehicle mileage. Additionally, individuals with higher incomes living in areas with lower poverty rates are more likely to cover greater daily distances by vehicle. These individuals are generally more likely to own cars, and have substantial commuting needs. Conversely, individuals with lower incomes may favor more cost-effective public transportation options and/or have fewer commuting requirements.

The effect of these variables is not only statistically significant, but also practically important. As shown in table 4-7, our models predict that a one standard deviation increase in population density, job density, road density or the number of public transportation stops (as will occur when moving from one area to another, or because of road infrastructure construction or expansion of public transit options) will change daily miles by amounts ranging from a quarter mile to over one mile. The effect is the most pronounced for public transportation options: Increasing access by one standard deviation reduces daily vehicle miles driven by up to 1.4 miles. The effects are also reasonably stable as we move from the device-level data to the block group and census tract pseudo panels.

The coefficient on gasoline price remains negative and significant, indicating that gasoline price does have an effect on the VMT above and beyond that of the socioeconomics, land use, and infrastructure in the area. Even at a time when the economy was reopening and people were "going places" more, the price of gasoline exhibited a moderating effect on the miles driven.

Table 4-6 Estimation results: Equation (2). Dependent variable: Average daily VMT. Gasoline price is the average within 5 miles from home.

	Device-level panel, weekly	Device-level panel, incl. zeros, weekly	Block group averages, weekly	Census tract averages, weekly
Population density	-3.80E-04 ***	-4.59E-04 ***	-7.45E-05 ***	-4.23E-04 ***
Job density	1.31E-04 ***	8.77E-05 ***	9.26 E-05 ***	1.44 E-04 ***
Road density	-5.91E-02 ***	-6.90E-02 ***	-3.07E+04 ***	-5.65E-02 ***
Area of land	5.99E-08 ***	5.63E-08 ***	1.42E-07 ***	8.85E-08 ***
Number of public transit stops	-8.58E-02 ***	-9.50E-02 ***	-1.49E-01 ***	-1.03E-01 ***
% Population with < $25k/year HH income	-4.82 **	-5.31 ***	-5.02E-01	-1.15E+01 ***
% Population with $25k-50k/year HH income	2. 60E-01	6.46E-01	-1.61 **	-1.43 *
% Population with $50k-75k/year HH income	-4.58E-01	-5.69E-01	4.47 ***	6.64 ***
% Population with $75k-125k/year HH income	2.45 **	3.26 ***	1.12 **	3.73 ***
% Population aged < 17	4.06 ***	5.58 ***	2.52 ***	3.79 ***
% Population aged 18-24	6.42E-01	-2.20 .	1.64 *	6.01E-01
% Population aged 25-34	9.58 ***	9.00 ***	4.12 ***	1.18E+01 ***
% Population aged 35-54	8.11 ***	8.26 ***	-4.00 ***	9.21 ***
% Population aged 55-64	-6.18E-01	3.18E-01	1.86 *	2.54 *
% Population with high school diploma or less	6.91 ***	7.29 ***	5.85 ***	8.02 ***
% Population with some college no degree or associate's degree	1.80 ***	1.79 ***	1.61E+01 ***	1.18E+01 ***
Poverty rate	-1.31E+01***	-1.52E+01 ***	-1.14E+01 ***	-7.89 ***
Gasoline price	-6.49***	-6.67 ***	-4.93 ***	-4.88 ***
Time fixed effects	Month, year	Month, year	Month, year	Month, year
R square	0.03	0.03	0.33	0.46
N. Observations	418,147	439,451	49,683	42,258

Note: * = significant at the 10% level; ** = significant at the 5% level; *** = significant at the 10% level.

Table 4-7 Effect on daily miles of increasing selected independent variables by one standard deviation. Based on the results summarized in table 4.5.

	Device-level panel (no zeros)	Device-level panel (incl. zeros)	Block group pseudo panel	Census tract pseudo panel
Population density	-0.8267 miles	-1.0390 miles	-0.7518 miles	-1.4036 miles
Job density	+0.3392 miles	+0.2454 miles	+0.7248 miles	+0.7325 miles
Road density	-0.4628 miles	-0.5447 miles	-0.3023 miles	-0.4895 miles
Number of public transit stops	-1.2313 miles	-1.3685 miles	-1.0332 miles	-1.4036 miles

4.5.4 Stability of the VMT

In Sections 4.5.2 and 4.5.3 we showed that the device-level panels and the census area panels display (for the most part) the expected associations with the socio-demographics of the area and/or road infrastructure.

As an additional check of the stability of the VMT and their temporal consistency, we report in table 4-8 the results from fitting the dynamic panel model described by equation (3). It is encouraging that the coefficients on the previous week's VMT are always positive; the coefficients are however relatively low, confirming once again the inherent variability in the VMT measured at (relatively speaking) high frequency. As expected, they do increase—suggesting more stability and better predictability—as we aggregate the device-level observation to geographical units. With the census tract pseudo panel, for example, the coefficient on the previous period's daily VMT is 0.25—twice as much as that from the device-level panels.

Table 4-8 Estimation results: Equation (3). Dependent variable: Average Daily VMT.

	Device-level panel, Weekly	**Device-level panel, incl. zeros, weekly**	**Block group averages, weekly**	**Census tract averages, weekly**
Coefficient on VMT in the previous week	0.1431 ***	0.1435 ***	0.1832 ***	0.2548 ***
Time fixed effects	Month, year	Month, year	Month, year	Month, year
N. Observations	373,268	390,786	57,015	47.991

4.5.5 VMT as Predictors of Gasoline Prices

Finally, we "reverse" the question and ask whether total VMT is a valid predictor of gasoline prices at the station level. In the US, the price of gasoline varies dramatically across places and over time. The U.S. Energy Information Administration [180] shows that the price of oil accounts for the majority of the retail price of gasoline. The rest is explained for the most part by refining costs and refineries' margins, marketing and distribution costs, and taxes (17%-22%). The federal gasoline tax is \$0.1840/gallon. As shown in figure 4-3, in our study area the price of gasoline is highest in Washington DC, presumably because of its higher land and business costs than elsewhere, even though its state gasoline tax is lower than the Maryland gasoline tax.[2]

Table 4-9 Estimation results. Dependent variable: Weekly gasoline price from each gasoline station.

	Specification 1	**Specification 2**
Total VMT in census tract in previous week (thou.)	1.28E-04 ***	-3.12E-05
Number of gas stations nearby		-1.72E-02 ***
pop_dens		-1.58E-06 *
Road_dens		3.35E-03 ***
Area of land		4.09E-10 ***
% Population with < \$25k/year HH income		-3.92E-01 ***
% Population with \$25k-50k/year HH income		1.22E-01 ***
% Population with \$50k-75k/year HH income		2.76E-01 ***
% Population with \$75k-125k/year HH income		7.50E-02 ***
% Population aged < 17		1.30E-01 ***
% Population aged 18-24		2.32E-01 ***
% Population aged 25-34		-3.01E-01 ***
% Population aged 35-54		-1.46E-02
% Population aged 55-64		3.02E-01 ***
Poverty rate		3.51E-01 ***
% Population with high school diploma or less		-3.35E-01 ***
% Population with some college no degree or associate's degree		-3.93E-01 ***
Fixed effects	Station, Brand, State	Brand, State
Time fixed effects	Month, Year	Month, Year
R square	0.67	0.65

Note: * = significant at the 10% level; ** = significant at the 5% level; *** = significant at the 10% level.

Table 4-9 presents the results from two alternate specifications of a model where the dependent variable is the average price of gasoline charged by gas station i in our study area in week t. The specification in column (1) includes a rich set of fixed effects, including gasoline station, brand and state fixed effects, as well as month and year fixed effects, but only one regressor, namely the total miles driven by the devices in the census tract around the gas station in the previous week. The coefficient on this variable is positive and significant, and its magnitude

[2] The state tax is \$0.235/gall. in DC, \$0.162/gall. in Virginia, and \$0.37/gall. in Maryland.

implies that increasing the VMT by one standard deviation results in an increase of 0.85 cents/gall. in the price of gasoline.

In column (2) we drop the gas station's fixed effects, and enter the census tract's socio-demographics and road infrastructure, gasoline brand dummies, and a measure of local competition, namely the number of other gasoline stations within 0.25 miles. Taken together, the regressors do a good job explaining the variation in the price of gasoline (R square 0.65)—and render the coefficient on the total VMT driven the previous week statistically insignificant.

Regarding brand dummies, some gasoline brands are significantly more expensive than others. For example, Shell stations sell gasoline at 9.5 cents per gallon more than the "generic" or "no name" blend, while Exxon at 12.2 cents per gallon more. Conversely, Costco and Wawa offer prices that are 25 and 2.4 cents per gallon, respectively, cheaper than the "generic" gasoline. However, removing the brand dummies from the regression only slightly decreases the R-squared value from 0.6478 to 0.6351, indicating that the brand factor contributes only a minor effect in explaining the variation in gasoline prices. Importantly, the presence of three gas stations within a quarter mile is sufficient to reduce the price of gasoline by 5 cents/gallon.

4.6 Conclusions

We have used a 10% sample of MDLD from the Washington, DC, metro area from May 2021 until the end of 2022 to construct a weekly panel that follows individual devices over time, and pseudo panels where each week our metric of interest (daily VMT) is averaged across a specific geographic area—a census block group or a census tract—over our study period.

Since the MDLD data are anonymized, the origin and the destination of each trip must be inferred using certain rules and algorithms. For example, it is assumed that a location is home if it is the location with the highest frequency between 9:00pm and 6:00 the next morning, and work if it is the location with the highest frequency during the daytime on workweek days. We cannot possibly know the individual characteristics of the device owner or user, and they must be imputed from the area's demographics based on data collected from the federal government, such as the American Community Survey and the Census TIGER shapefiles. We also merge the data with infrastructure and land use information. We acquired gasoline prices from each station in the area from a commercial vendor, in sharp contrast with much practice in research, which has used daily gasoline prices—but averaged over the PADD, an area that typically comprises several states [117]—or state-level monthly averages [116]. We argue that deploying data at such a fine level of granularity should keep measurement errors to a minimum.

The quality of the longitudinal datasets constructed in these ways can be assessed using a variety of criteria. The first is the sample size, which depends on the number of units and the number of times each unit contributes an observation to the dataset. Clearly, individual-level panels contain more observations than block-group or census-tract pseudo-panels.

The second is the stability and temporal consistency of the daily miles driven, which we assess by examining the standard deviation of the daily miles driven and the correlation between consecutive daily miles records. In the block-group and census-tract pseudo-panels, the variability of our key variable, daily miles, is less than in the device-level panel, because we take averages across devices.

To obtain the correlation between consecutive daily miles, we fit a dynamic panel data model where each week's miles depend on the miles driven the previous week. Instrumental variable estimation techniques must be used, since the usual estimators are biased with this type of model in the presence of unit fixed effects. The coefficient on the previous week's daily miles traveled is positive, as expected, but relatively low, confirming that with micro-level data it is generally difficult to predict behaviors. The coefficient however gets larger as we aggregate the device-level records to the census tract level.

Our final criteria for assessing the quality of the MDLD and its potential for research purposes is to examine the daily miles' association with variables that are assumed to influence them, such as individual characteristics, the economic circumstances in the area, land use and infrastructure in the area, and the price of gasoline.

We have found that the MDLD daily miles, even at their finest level of granularity, exhibit the expected relationships with community characteristics, infrastructure and land use. Even more important, our regression results show that any changes daily miles are negatively associated with changes in gasoline prices, as predicted by economic theory, even when gasoline prices were rising while the economy was re-opening.

When attention is restricted to presumptive commute or work-related daily miles, the coefficient on gasoline price has the wrong (positive) sign, but it is negative again in the non-commute, non-work miles regressions. We compute the elasticities of the daily VMT with respect to gasoline prices, which we interpret as (very) short-run elasticities, finding that they lie in the range between -0.17 and -0.41, and are thus similar to what reported by other studies [173] and a meta-analysis of recent literature [174].

We conclude that MDLD are valuable on their own (as well as *alongside* conventional surveys), and that the device-level longitudinal datasets based on them are a promising source of information for those wishing to explore miles driven and how they are affected by socioeconomic characteristics, land use and infrastructure, and the price of fuel. In this paper, we have merged device-level records with summary information about the socioeconomic characteristics of the population living in the area around a device owner or user's presumptive home location, but in future research it may be possible to develop datasets with individual-specific information, to use imputation techniques based on the *distribution* of such socioeconomic characteristics, or to deploy methods from Bayesian statistical inference to find what socioeconomic characteristics are most likely to be associated with the observed travel patterns.

In addition, future research may wish to explore the potential of MDLD to construct vehicle- and trip-specific driving cycle patterns—including acceleration, deceleration, time sitting in traffic, etc.—which could be compared with the fuel consumption and emissions information embedded in models like the Motor Vehicle Emission Simulator (MOVES) and others used by policymakers to confirm or update their results.

It may also be possible to use MDLD to observe the response to announcements—by now frequent through cell phone alerts, the vehicle's own navigation system, Google, Waze and other apps—about traffic jams, accidents and the associated road closures, and to model the choice of alternate routes and the intensification of traffic of such alternate routes. This information may assist policymakers and emergency response crews in predicting and redirecting traffic during such events.

5. Chapter 5: Post-Pandemic Road Congestion and Teleworking: Evidence from Mobile Device Location Data

5.1 Problem Statement

The COVID-19 pandemic has produced profound and lasting changes in travel behavior, with teleworking emerging as one of the most widespread and enduring shifts. In the Washington, D.C. metropolitan region, teleworking increased sharply during the pandemic and has remained prevalent in the years that followed. Even as public health restrictions have eased, a substantial share of the workforce continues to telework regularly, reflecting both employee preferences and employer adoption of flexible work arrangements. These structural changes in work patterns have fundamentally altered when, where, and how people travel.

Teleworking is commonly expected to reduce traffic congestion by decreasing the frequency of daily commuting trips, particularly during peak periods. However, observed traffic conditions in many metropolitan regions challenge this expectation. Despite sustained levels of teleworking, travelers increasingly report worsening congestion compared to pre-pandemic conditions. Regional statistics and travel surveys indicate rising traffic volumes on major highways alongside a persistent decline in public transit ridership. Similar patterns have been documented across multiple international contexts, suggesting that post-pandemic congestion dynamics are neither temporary nor region-specific.

This apparent disconnect between reduced commuting and persistent or intensified congestion raises a critical planning question: why has road congestion rebounded or worsened despite a sustained reduction in traditional commute trips? Answering this question requires moving beyond aggregate traffic volumes and examining the underlying behavioral mechanisms that shape congestion, including changes in trip frequency, trip distance, travel timing, and the spatial redistribution of travel demand. In particular, teleworking may enable longer-distance and more discretionary travel, alter residential location choices, and shift travel away from conventional peak periods, all of which can affect congestion in ways not captured by simple trip counts.

Existing data sources face important limitations in addressing this problem. Traditional travel surveys provide rich contextual information on trip purpose and traveler characteristics but are constrained by limited sample sizes, recall bias, and short observation periods. Traffic monitoring systems offer continuous and accurate measurements of speed and volume but are typically restricted to fixed locations and cannot capture origin–destination patterns or broader behavioral changes. As a result, prior analyses often lack the ability to simultaneously observe travel behavior and congestion consistently over time and space.

Mobile Device Location Data (MDLD) offers new opportunities to address these gaps. By providing high-resolution, longitudinal observations of individual travel behavior across large populations, MDLD enables the examination of changes in driving behavior, trip distance, spatial patterns, and congestion using consistent origin–destination pairs over time. However, the potential of MDLD to explain post-pandemic congestion through a behavioral lens remains underexplored, particularly in the context of teleworking and its longer-term implications for travel demand.

This chapter addresses this gap by using MDLD to investigate how driving behavior and road congestion in the Washington, D.C. region evolved before, during, and after the COVID-19 pandemic. By focusing on consistent origin–destination pairs and examining changes in trip frequency, trip distance, travel timing, and spatial structure, the analysis seeks to clarify the mechanisms underlying post-pandemic congestion and to inform transportation planning in an era of sustained teleworking.

5.2 Data and Methodology

5.2.1 Data Source and Study Design

This chapter uses high-resolution Mobile Device Location Data (MDLD) drawn from a national mobility data panel developed by the University of Maryland's Next Generation National Household Travel Survey (NextGen NHTS) team in collaboration with multiple commercial data providers. Detailed descriptions of the MDLD structure, sampling strategy, and data quality are provided in Chapter 3; this section focuses on the aspects of the data relevant to congestion analysis.

The analysis examines driving-related travel behavior in the Washington, D.C. region. Given Washington, D.C.'s role as a regional employment center and its small geographic footprint, a substantial share of vehicle travel involving the District originates from or terminates in neighboring jurisdictions. To capture the full extent of regional travel demand affecting congestion within Washington, D.C., the study includes all driving trips that either originate from or are destined for the District.

Three study periods are selected to represent distinct stages of the COVID-19 pandemic: September 2019 (pre-pandemic), September 2020 (pandemic peak), and September 2022 (late-pandemic). Selecting the same calendar month across years controls for seasonal effects while enabling comparisons of short- and long-term changes in travel behavior and congestion.

5.2.2 Trip Construction and Information Imputation

Raw MDLD sightings were processed using the standardized preprocessing and trip construction pipeline described in Chapter 3. In brief, invalid or low-quality observations were removed, sightings were chronologically ordered by device, and continuous movement trajectories were constructed for each device. Trip identification was performed using a decision tree–based recursive algorithm that distinguishes periods of movement from stationary activity based on distance, time intervals, and implied speed between consecutive sightings. This procedure produces a comprehensive trip roster containing trip start and end times, origins and destinations, and unique trip identifiers. The algorithm and its validation have been documented in prior work and are not repeated here. The resulting trip roster provides the foundation for subsequent congestion analysis.

Because MDLD does not directly record travel mode, travel mode imputation is applied following the framework described in Chapter 3. Trips are first screened to identify air travel based on speed and airport proximity. Remaining trips are classified into ground transportation modes using a supervised machine learning model trained on temporal, spatial, and network-based features. Given the focus of this chapter, only trips classified as driving are retained for analysis.

Accurate estimation of trip distance is essential for distinguishing congestion effects from changes in trip length. Rather than relying on straight-line distances, a map-matching and routing procedure is used to estimate network-based driving distances. Location sightings are snapped to nearby road segments, and routing algorithms reconstruct plausible travel paths between consecutive points. This approach reduces bias associated with sparse sampling and network constraints and ensures consistency in distance estimation across study periods.

5.2.3 Origin–Destination Aggregation and Congestion Measurement

A key challenge in using MDLD to study congestion is distinguishing changes in traffic conditions from changes in the composition of observed trips. Because MDLD represents a sample of the traveling population that may vary over time, simple comparisons of average travel time per trip or per device can be misleading.

To address this issue, trip origins and destinations are spatially joined to census block groups, and driving trips are aggregated into block-group-level origin–destination (OD) tables. Congestion is evaluated by comparing average travel times for the same OD pairs across different study periods. By holding OD pairs constant, this approach controls for spatial heterogeneity in trip patterns and isolates changes in travel time attributable to congestion rather than to shifts in trip length or destination choice. This OD-consistent framework enables a more robust assessment of how congestion evolved before, during, and after the pandemic and supports analysis of differences across weekdays and weekends.

5.3 Results

5.3.1 Number of Driving Trips, Devices, and OD pairs

This study focuses on driving trips that either originate from or are destined for the District of Columbia. After trip identification and travel mode imputation, a total of 622,644 driving trips from 99,037 devices were identified in September 2019, 574,773 trips from 108,333 devices in September 2020, and 858,388 trips from 123,527 devices in September 2022. While the number of observed devices increased steadily over time, reflecting the growing prevalence of smartphones and location-based services, the number of driving trips declined sharply during the pandemic in 2020 and rebounded by 2022.

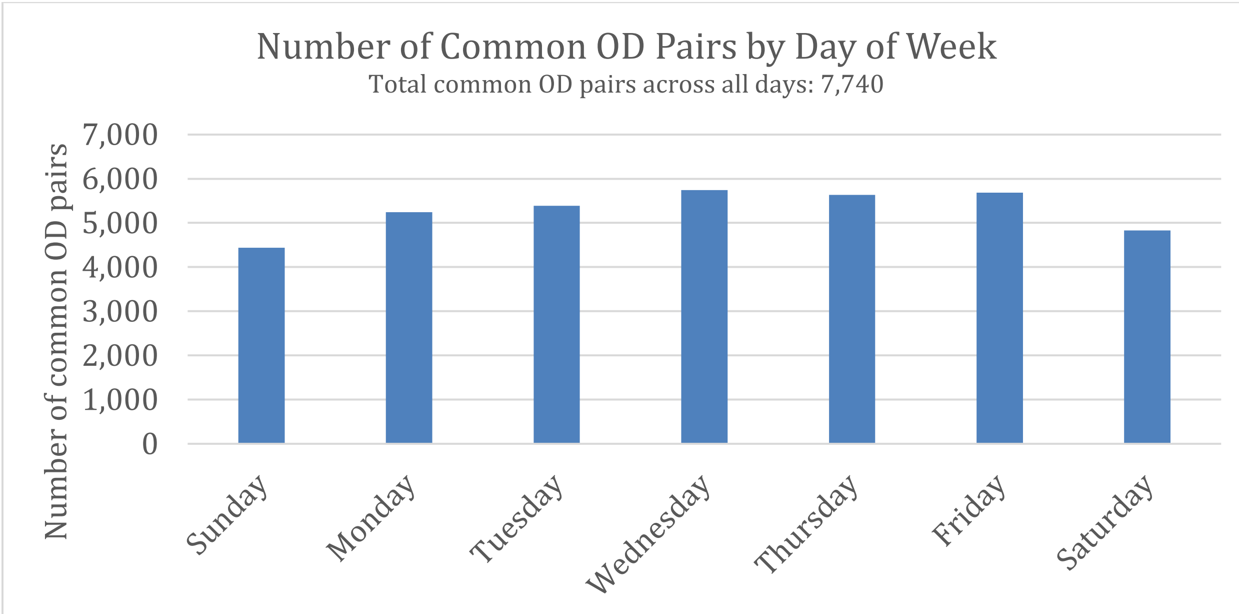


Figure 5.1 Number of Common OD Pairs by Day of Week

As discussed in Section 5.2.3, congestion analysis is conducted by comparing travel times between consistent origin–destination (OD) pairs to mitigate potential sampling bias in mobile device location data. Figure 5.1 illustrates the number of OD pairs that are consistently observed across all three study periods by day of the week. Thousands of common OD pairs are retained for each day, with higher coverage on weekdays than weekends, indicating strong spatial consistency and supporting robust OD-based comparisons.

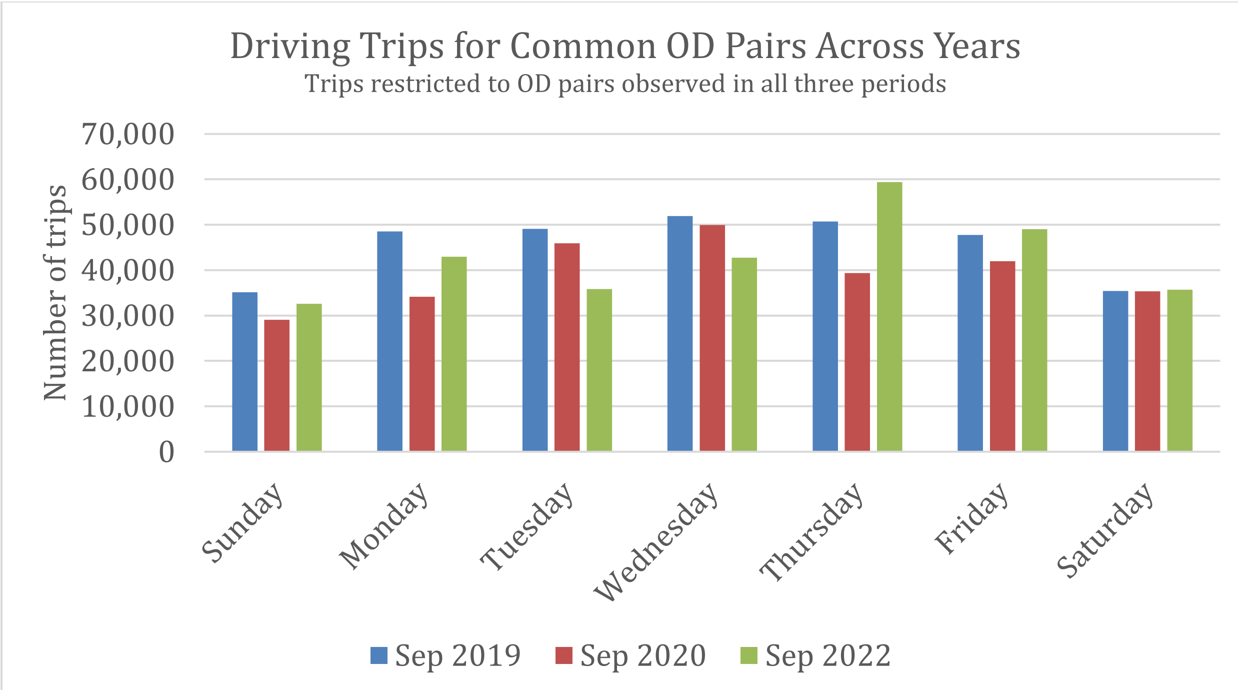


Figure 5.2 Driving Trips for Common OD Pairs Across Years

Figure 5.2 presents the number of driving trips associated with common OD pairs across September 2019, 2020, and 2022. Driving trips declined substantially in 2020 across all days of the week during the height of the COVID-19 pandemic. By 2022, trip volumes rebounded for most days, particularly on weekends and selected weekdays. In contrast, trips on Tuesdays and Wednesdays continued to decline relative to both 2019 and 2020, suggesting sustained reductions in midweek travel that are consistent with persistent teleworking and reduced traditional commuting.

Table 5-1 Number of driving trips & travelers observed between common OD pairs

Day of week	Number of Driving Trips			Number of Travelers		
	Sep 2019	**Sep 2020**	**Sep 2022**	**Sep 2019**	**Sep 2020**	**Sep 2022**
Sunday	35,095	29,063	32,603	17,042	16,393	16,574
Monday	48,559	34,119	42,934	21,920	17,926	20,730
Tuesday	49,073	45,907	35,870	21,712	20,994	18,746
Wednesday	51,931	49,973	42,732	22,968	23,162	21,399
Thursday	50,745	39,373	59,396	22,265	19,845	26,703
Friday	47,786	41,979	49,031	22,569	21,669	24,163

Saturday	35,432	35,321	35,680	18,022	20,236	18,877
Overall	**359,801**	**309,152**	**352,761**	**75,694**	**80,733**	**70,785**

Table 5-1 summarizes the total number of driving trips and travelers retained after restricting the analysis to OD pairs observed in all three periods. Across the three months, more than 300,000 driving trips and over 70,000 unique travelers are retained in each period. These large and stable sample sizes highlight the strength of the MDLD and ensure that subsequent congestion analyses are not driven by changes in sample composition or spatial coverage.

5.3.2 Average Driving Time between OD pairs

After constructing block-group-level OD tables, we first examine whether average driving time between the same OD pairs increased in 2022 relative to 2019. This comparison is a necessary, though not sufficient, condition for assessing whether road congestion worsened after the pandemic.

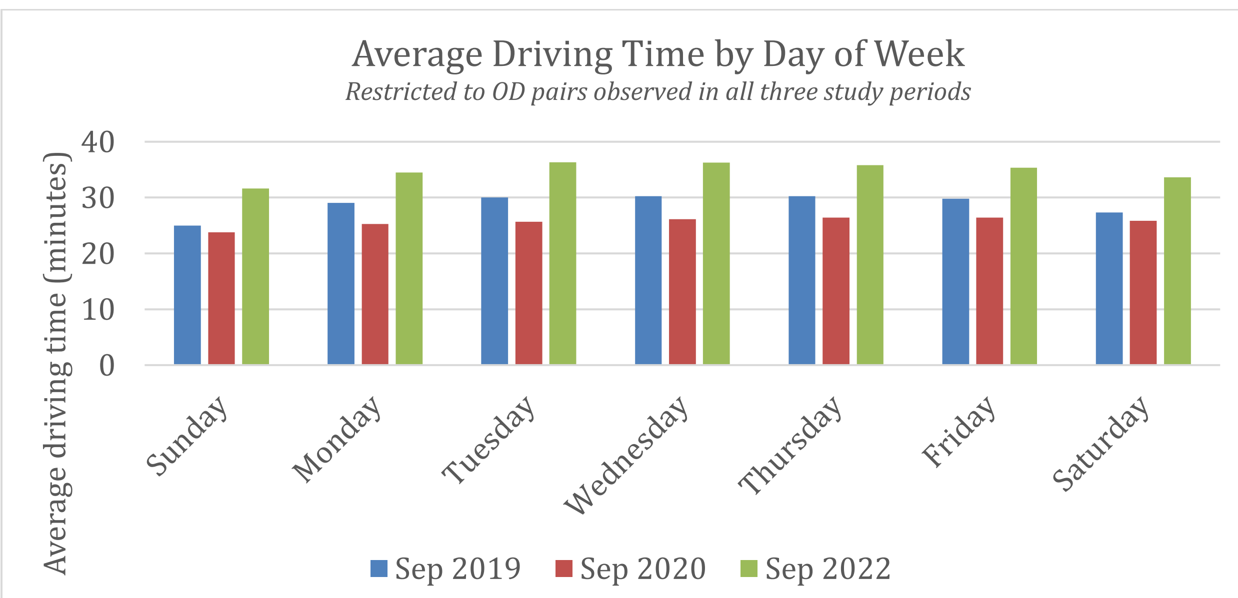


Figure 5.3 Average driving time between common origin–destination (OD) pairs by day of the week

Figure 5.3 summarizes average driving times between common OD pairs overall and by day of the week. In September 2020, average driving time declined substantially relative to September 2019 across all days, reflecting reduced traffic demand during the height of the COVID-19 pandemic. The reduction was more pronounced on weekdays than on weekends, consistent with sharp declines in commuting and work-related travel during this period.

By September 2022, average driving times exceeded pre-pandemic levels for all days of the week. Overall, average driving time increased by 20.55% relative to 2019. Notably, the increase was larger on weekends (24.77%) than on weekdays (19.27%). This pattern suggests a shift in travel behavior toward increased discretionary or non-work-related travel, potentially associated with sustained teleworking arrangements that reduce weekday commuting while enabling greater flexibility for weekend activities, known as the "substitution effect" [181].

While these results indicate that travelers experienced longer travel times in 2022 than before the pandemic, longer travel times alone do not necessarily imply increased congestion. Travel time can be expressed as:

$$Travel\ Time = \frac{Travel\ Distance}{Average\ Speed}$$

An increase in average driving time between OD pairs may result from slower travel speeds (indicative of congestion) or from longer travel distances, reflecting changes in residential location, route choice, or activity patterns. To disentangle these effects, the next section examines changes in average driving distance between OD pairs.

5.3.3 Average Driving Distance between OD pairs

To determine whether longer travel times observed in 2022 were driven by increased travel distances, we next examine changes in average driving distance between the same OD pairs. Figure 5.4 report average driving distances by day of the week for September 2019, 2020, and 2022.

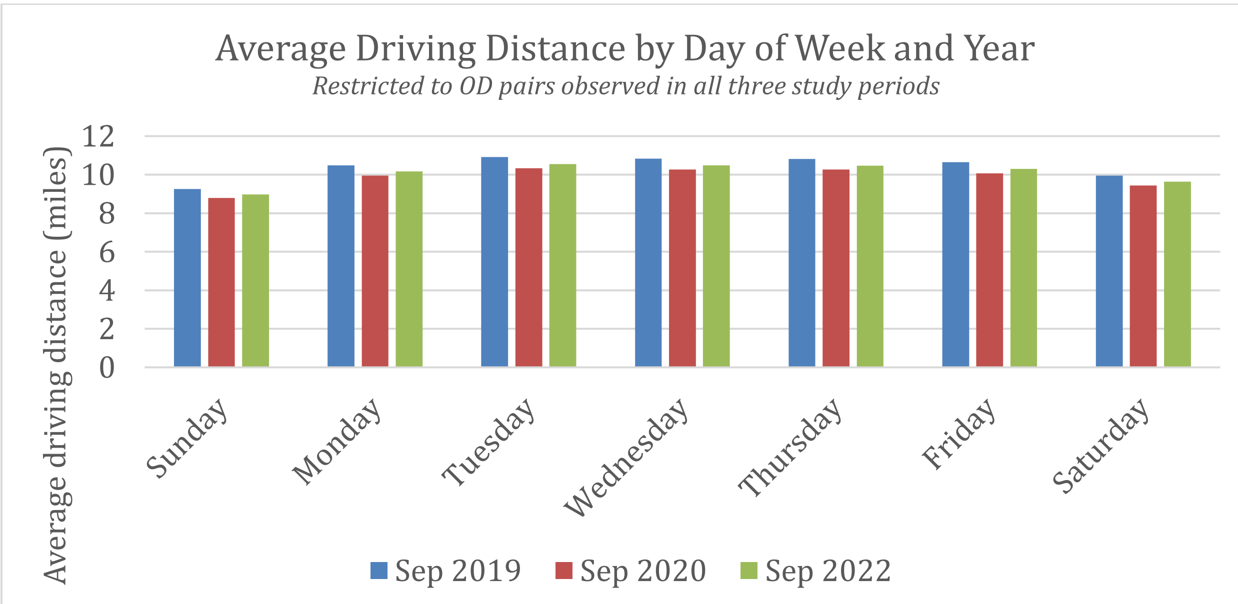


Figure 5.4 Average driving distance between common OD pairs by day of the week and year

In contrast to the substantial increase in average driving time, average driving distance did not increase from 2019 to 2022. Instead, average distance decreased slightly, by approximately 3.26% overall in 2022 relative to 2019. This reduction is consistent across both weekdays and weekends, with only minor variation by day of the week. Average driving distances declined sharply in 2020 and remained below pre-pandemic levels in 2022.

The absence of an increase in average driving distance suggests that longer travel times observed in 2022 are unlikely to be driven by travelers taking longer routes or traveling farther between the same origins and destinations. Possible explanations for the modest decline in distance include improvements in network connectivity, more direct routing, or reduced detouring behavior; however, disentangling these effects is beyond the scope of this study.

Taken together with the results in Section 5.3.2, these findings imply a reduction in average driving speed between common OD pairs from 2019 to 2022. This pattern is consistent with increased road congestion in the Washington region, particularly during weekends, where travel time increases were largest despite stable or declining travel distances. With evidence of increased congestion, the next section explores potential behavioral and structural factors contributing to post-pandemic congestion patterns.

5.3.4 Frequency of Driving Trips

Increased congestion may arise from more frequent travel, longer-distance trips, or slower travel speeds. To assess whether changes in travel frequency contributed to post-pandemic congestion, we examine trends in individual driving trip rates using the trip roster underlying the OD tables. Specifically, we calculate the average number of driving trips per person per day for each study period by day of the week.

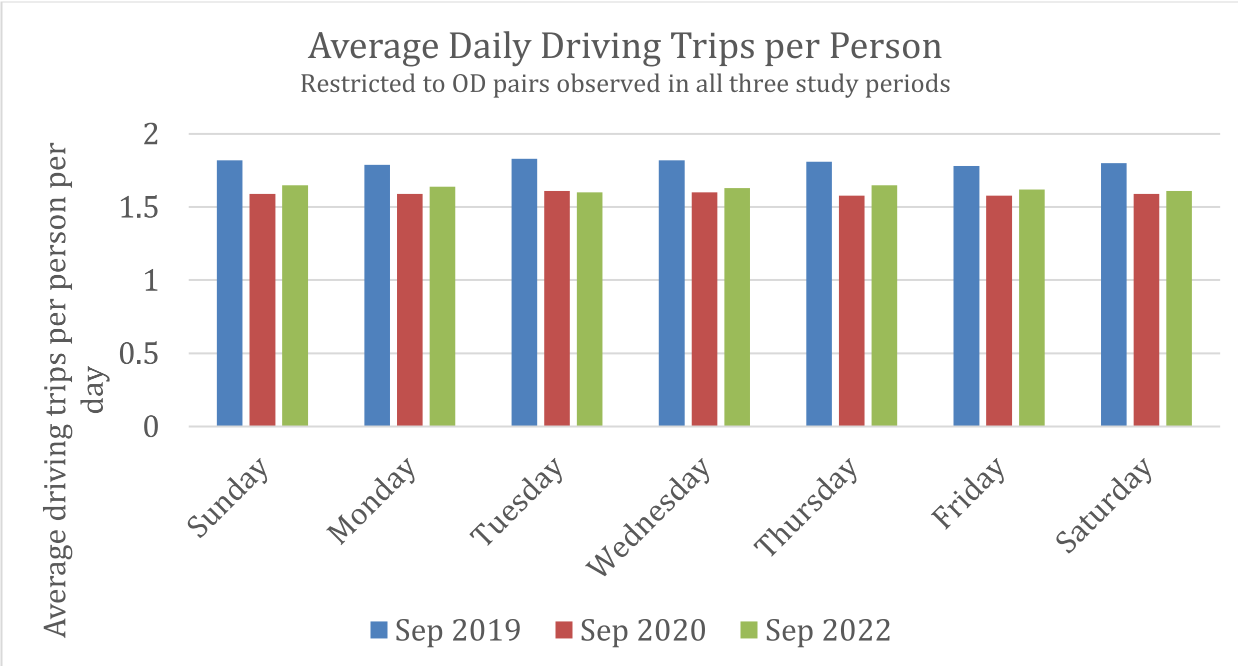


Figure 5.5 Average Daily Driving Trips per Person between Common OD Pairs

Figure 5.5 summarizes average daily driving trip frequencies across September 2019, 2020, and 2022. Consistent with pandemic-related mobility restrictions, the average number of driving trips per person declined substantially from 2019 to 2020 across all days of the week. Although trip rates partially rebounded by 2022, they remained below pre-pandemic levels for every day.

Overall, average daily driving trips decreased from 1.81 trips per person in September 2019 to 1.63 trips per person in September 2022, representing a reduction of approximately 9.7%. Similar patterns are observed across weekdays and weekends, indicating that increased congestion in 2022 is unlikely to be driven by a higher frequency of driving trips. These findings are consistent with results from national surveys such as the National Household Travel Survey , which reported a decline in daily trips per person across all modes from 3.37 in 2017 to 2.28 in 2022 [22, 134]. It is important to note that the driving trip rates reported in Figure 5.5 are lower than those typically reported in household travel surveys. This is because the analysis is restricted to trips occurring between origin–destination pairs observed consistently across all three study periods. As a result, the reported trip rates reflect only a subset of each individual's total daily travel and are not intended to represent total trip-making activity.

5.3.5 Distance of Driving Trips

If congestion is not driven by more frequent travel, a key remaining question is whether individuals are undertaking longer driving trips than before the pandemic. To address this question, we examine changes in the average distance per driving trip using the individual trip roster. Figure 5.6 and Table 5-2 summarize average trip distances by day of the week.

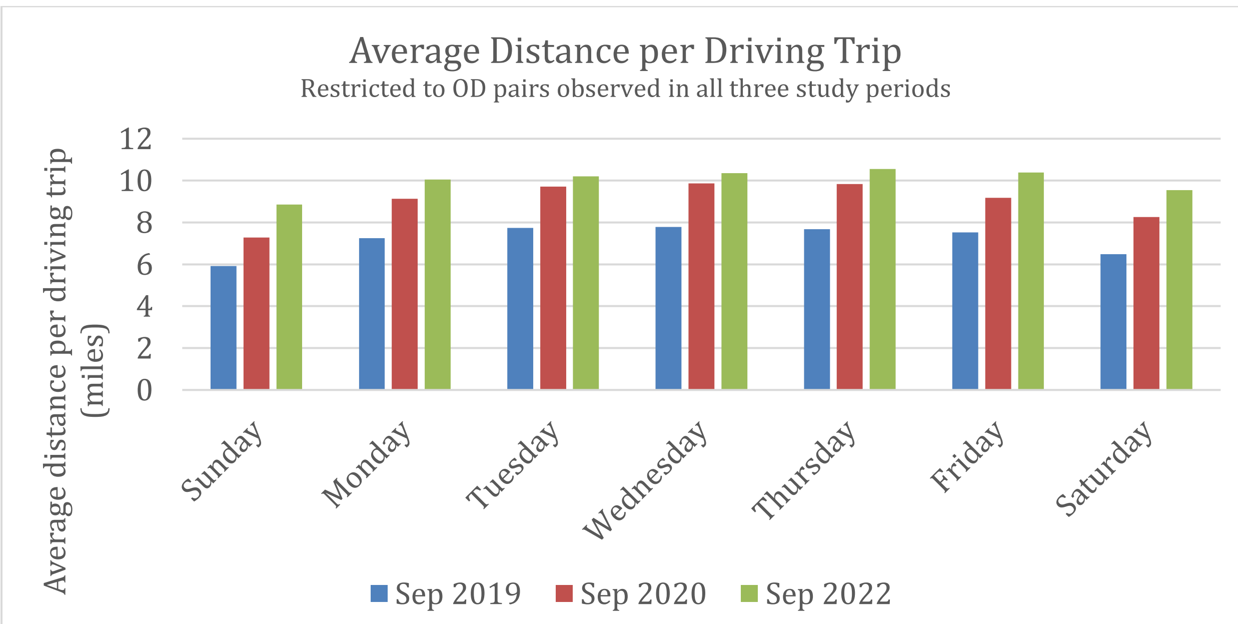


Figure 5.6 Average Distance per Driving Trip

Table 5-2 Average distance per driving trip by day of week

Day of week	Average dist. per driving trip in 2019.9 (Miles)	Average dist. per driving trip in 2020.9 (Miles)	Average dist. per driving trip in 2022.9 (Miles)	Difference from 2019.9 to 2020.9	Difference from 2019.9 to 2022.9
Sunday	5.92	7.28	8.86	22.95%	49.59%
Monday	7.25	9.13	10.05	25.91%	38.60%
Tuesday	7.74	9.72	10.21	25.55%	31.82%
Wednesday	7.78	9.86	10.35	26.74%	32.97%

Thursday	7.68	9.84	10.55	28.10%	37.42%
Friday	7.53	9.18	10.39	21.97%	37.97%
Saturday	6.48	8.26	9.55	27.40%	47.32%
Overall	7.41	9.33	10.35	25.89%	39.66%

The results show a clear and sustained increase in the average distance per driving trip over time. In September 2019, the average driving trip length was 7.41 miles. This value increased to 9.33 miles in September 2020 and further to 10.35 miles in September 2022, representing an overall increase of 39.66% relative to pre-pandemic levels. Increases are observed consistently across all days of the week.

Importantly, these results contrast with the OD-pair-level analysis in Section 5.3.3, which shows no increase in average driving distance between the same origin–destination pairs. Together, these findings indicate that longer travel times are not driven by increased detouring or longer routes between fixed locations, but rather by a shift toward trips connecting more distant origins and destinations. In other words, while travelers are not traveling farther *between the same OD pairs*, a growing share of driving trips involves longer-distance OD pairs.

Several mechanisms may contribute to this shift. First, the expansion of teleworking has enabled greater residential flexibility, allowing individuals to live farther from workplaces while commuting less frequently. As a result, although commuting trips occur less often, the distance traveled on in-office days is longer. Second, the substitution effect suggests that teleworkers may be more willing to undertake longer discretionary trips, reflecting a higher tolerance for travel distance compared to traditional daily commuters.

Day-of-week patterns provide further insight. In September 2020, average trip distances increased similarly on weekdays (25.65%) and weekends (25.18%), reflecting widespread mobility constraints during the pandemic. By September 2022, however, the increase became more pronounced on weekends (48.46%) than on weekdays (35.76%), suggesting a shift toward longer leisure-oriented travel as pandemic restrictions were lifted and teleworking became more established.

Taken together, these results indicate that post-pandemic travel behavior is characterized by fewer but longer driving trips, which, combined with reduced travel speeds, contributes to heightened congestion despite lower trip frequencies.

5.3.6 Patterns of Origins and Destinations

Beyond changes in trip frequency and distance, shifts in the spatial distribution of travel may also contribute to post-pandemic congestion. One indicator of such shifts is the proportion of interstate driving trips involving Washington, D.C. Figure 5.7 summarizes changes in the share of trips that cross state boundaries across the three study periods.

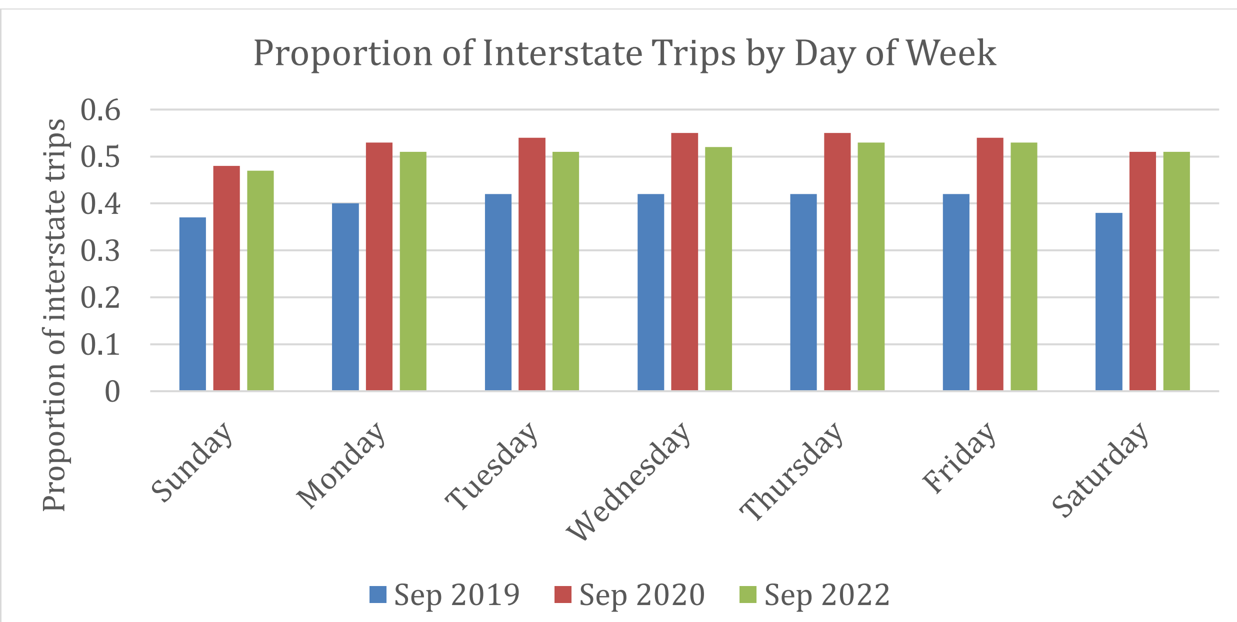


Figure 5.7 Proportion of Interstate Trips by Day of Week

Prior to the pandemic, fewer than half of the driving trips originating from or destined for Washington, D.C. were interstate, with an overall proportion of 0.42 in September 2019. During the pandemic, this share increased substantially, reaching 0.54 in September 2020. Notably, the proportion of interstate trips remained elevated in September 2022 at 0.53, indicating that the spatial reconfiguration of travel patterns persisted beyond the initial pandemic period.

This increase in interstate travel suggests a shift toward longer-distance origin–destination pairs involving surrounding states, such as Maryland and Virginia. One plausible explanation is increased residential relocation or extended residential flexibility enabled by teleworking, allowing individuals to live farther from the District while continuing to travel into the city when necessary. In addition, greater schedule flexibility associated with teleworking may have encouraged more discretionary interstate travel, particularly for leisure and social activities.

Day-of-week patterns reinforce this interpretation. While the proportion of interstate trips increased across all days, the rise was especially pronounced on weekends, consistent with earlier findings of larger weekend increases in travel time and trip distance. These results highlight the role of spatial redistribution of travel demand, rather than increased trip frequency, in shaping post-pandemic congestion patterns.

In summary, the results indicate that increased road congestion in the Washington region, particularly on weekends, is not driven by more frequent driving or longer routes between the same origins and destinations. Instead, congestion appears to be associated with a combination of reduced travel speeds, longer per-trip distances, and a persistent shift toward more interstate and longer-distance travel patterns, likely influenced by widespread teleworking.

5.4 Conclusion and Discussion

This study examined changes in traffic patterns in the Washington, D.C. region across different stages of the COVID-19 pandemic using high-resolution Mobile Device Location Data

(MDLD). By tracking driving behavior before, during, and after the pandemic, the analysis provides new evidence on how teleworking has reshaped travel patterns and road congestion.

The results indicate that road congestion in the Washington region has intensified relative to pre-pandemic conditions, with the increase particularly pronounced on weekends. Importantly, this rise in congestion did not coincide with a return to pre-pandemic levels of driving frequency. Instead, the findings reveal a fundamental shift in travel behavior: individuals are making fewer driving trips overall, but those trips are longer and increasingly concentrated between more distant origins and destinations. This pattern suggests that post-pandemic congestion is driven less by the number of trips and more by changes in trip length, spatial structure, and traffic flow efficiency.

A key contribution of this study is the distinction between distance changes at the origin–destination (OD) pair level and at the individual trip level. While average driving distance between the same OD pairs did not increase, the average distance per driving trip rose substantially. This indicates that travelers are not taking longer routes between fixed locations, but are instead increasingly engaging in trips that connect more distant locations. Such shifts are consistent with greater residential flexibility enabled by teleworking, allowing individuals to live farther from workplaces while commuting less frequently, as well as with increased willingness to undertake longer discretionary trips.

The spatial redistribution of travel demand is further reflected in the rising share of interstate trips involving Washington, D.C. The persistence of elevated interstate travel into 2022 suggests that pandemic-induced changes, such as residential relocation or extended residential flexibility, have had lasting impacts on regional travel patterns. These spatial shifts help explain why congestion has worsened despite lower trip rates, particularly during weekends when discretionary travel is more prevalent.

Together, these findings highlight a telework–congestion paradox: although teleworking reduces commuting frequency, it can simultaneously contribute to longer trips, altered travel timing, and spatially redistributed demand, ultimately exacerbating congestion under certain conditions. This underscores the limitations of relying solely on trip counts or commuting metrics to assess transportation system performance in the post-pandemic era.

Several limitations of this study should be acknowledged. While MDLD provides extensive spatial and temporal coverage, it lacks detailed information on traveler characteristics. Future research could integrate MDLD with traditional travel surveys or administrative datasets to better link observed mobility patterns with socio-demographic attributes and teleworking behavior. In addition, further work is needed to examine long-term trends beyond the immediate post-pandemic period and to assess how evolving work policies may continue to reshape travel demand.

From a policy perspective, the results suggest that congestion mitigation strategies must adapt to changing travel behaviors. Traditional approaches focused on peak-hour commuting may be insufficient in an environment characterized by longer-distance travel, increased weekend congestion, and greater interstate connectivity. Enhancing regional transit options, coordinating transportation planning across state boundaries, and accommodating more dispersed and flexible travel patterns will be critical for managing congestion in the post-pandemic Washington region.

6. Chapter 6: Evaluating Greenhouse Gas Emissions from Mobile Device Location Data

6.1 Problem Statement

Transportation is the largest source of greenhouse gas (GHG) emissions in the United States, making accurate and timely estimation of transportation-related emissions essential for climate policy, regional climate action planning, and monitoring progress toward decarbonization goals. Beyond measuring total emissions, effective mitigation strategies increasingly require insight into who generates emissions and why—that is, how emissions are distributed across travel behaviors, trip purposes, and populations.

Current emissions inventory frameworks provide limited support for such behavioral attribution. The U.S. Environmental Protection Agency's MOtor Vehicle Emission Simulator (MOVES), the primary tool used for on-road emissions estimation in the United States, produces detailed, spatially resolved estimates of emissions based on vehicle activity occurring within defined geographic regions. While this location-based, operating-oriented framework is well suited for regulatory and planning applications, it does not explicitly track trip origins and destinations, traveler residence, or trip purpose. As a result, emissions are quantified by where they occur on the roadway network but remain weakly connected to the underlying travel behaviors and populations that generate them.

This limitation constrains the design and evaluation of targeted mitigation strategies. Policies such as land-use interventions, demand management, transit investment, and equity-focused climate initiatives depend on understanding how emissions are linked to travel behavior, rather than solely where emissions are produced. Without a behavioral perspective, it is difficult to assess which travel patterns contribute most to regional emissions or how changes in behavior—

such as shifts in commuting, non-work travel, or teleworking—affect transportation-related carbon footprints.

Recent advances in Mobile Device Location Data (MDLD) offer new opportunities to address this gap. MDLD provides high-resolution, longitudinal observations of individual travel behavior, including trip origins and destinations, travel distances, timing, and spatial patterns, across large populations and extended time periods. These data have been widely used to study travel behavior, mobility responses to external shocks, and spatial patterns of activity. However, their potential to support behaviorally grounded estimation and monitoring of transportation-related GHG emissions remain underdeveloped.

A key challenge lies in integrating behaviorally rich mobility data with established emission modeling frameworks in a way that complements, rather than replaces, existing inventories. Such integration must address uncertainty in unobserved vehicle attributes and reconcile differences between resident-based and roadway-based accounting perspectives. Without such a framework, the analytical potential of MDLD for informing transportation climate policy remains largely untapped.

This chapter addresses this challenge by developing a trip-based, behaviorally grounded framework for estimating on-road CO_2 emissions using MDLD. By linking observed travel behavior to vehicle emission factors and explicitly accounting for uncertainty, the framework seeks to enhance understanding of how transportation emissions are generated across space, time, and travel behavior, thereby supporting more targeted and equitable climate policy analysis.

.

6.2 Data

This chapter integrates three primary datasets to support behaviorally grounded estimation of transportation-related CO_2 emissions: (1) Mobile Device Location Data (MDLD), (2) vehicle ownership and fleet composition data from the Regional Travel Survey (RTS), and (3) vehicle-specific fuel economy and emission factors from the U.S. Environmental Protection Agency (EPA). Each dataset provides complementary information—human mobility, vehicle characteristics, and emission rates—that together enable spatially and temporally resolved, trip-based emission estimation.

6.2.1 Mobile Device Location Data (MDLD)

The MDLD used in this analysis was provided by the University of Maryland and compiled from multiple commercial location data vendors. The structure, sampling characteristics, and processing of MDLD are described in detail in Chapters 2 and 3; this section focuses on elements relevant to emission estimation.

Using an established MDLD processing pipeline, raw location sightings were converted into individual trips, with trip origins, destinations, timing, and distances inferred from reconstructed trajectories. Home locations for each device were inferred based on nighttime stay patterns and spatially joined to Census tracts and counties. Travel mode imputation was applied following the procedures described earlier, and only trips classified as vehicle trips were retained for emission analysis. Trip-level vehicle miles traveled (VMT) were computed and aggregated across temporal dimensions such as hour of day and day of week.

The analysis covers travel by residents across 11 counties in the Washington, D.C. metropolitan region, including the District of Columbia, selected counties in Maryland, and counties and independent cities in Virginia. The study period corresponds to April 2019, representing typical pre-pandemic travel conditions and capturing both weekday and weekend travel. On average, approximately 82,700 devices per day were observed across the region, yielding more than 7.2 million vehicle trips during the study period. The large sample size and continuous temporal coverage support robust analysis of emissions across space, time, and travel behavior.

6.2.2 Regional Travel Survey (RTS) Vehicle Data

Vehicle ownership and fleet composition information were obtained from the National Capital Region Transportation Planning Board's 2018 Regional Travel Survey (RTS) [182]. The RTS provides vehicle-level data including make, model, model year, household expansion weights, and geocoded home locations at the Census tract and county levels. Within the study area, the RTS contains an unweighted total of 18,394 vehicles. After applying household expansion weights, the RTS vehicles represent approximately 3.18 million vehicles in the study region.

These data are used to characterize the spatial distribution of vehicle technologies and emission intensities across Census tracts. For each tract, weighted distributions of vehicle emission rates were derived by combining RTS vehicle records with EPA-tested fuel economy and emission data. In tracts with limited RTS observations, county-level weighted distributions were used as fallback estimates to ensure complete spatial coverage while preserving regional fleet characteristics.

6.2.3 EPA Fuel Economy and Emission Data

Vehicle-level CO_2 emission rates were obtained from the EPA's official fuel economy database (fueleconomy.gov), which provides laboratory-tested tailpipe emission data for passenger vehicles sold in the United States. The database reports standardized CO_2 emission rates under controlled testing conditions, allowing consistent comparison across vehicle makes, models, and model years. In this study, the variable co2tailpipegpm (grams of CO_2 per mile) was used as the primary emission factor.

EPA emission records were merged with RTS vehicle data based on vehicle make, model, and model year, yielding vehicle-specific CO_2 emission rates for the observed household fleet. This integration produces tract-level weighted distributions of emission rates that reflect spatial variation in vehicle composition and technology adoption, including differences in vehicle size, fuel economy, and drivetrain characteristics. These tract-level emission rate distributions form the basis for assigning CO_2 emission factors to MDLD-inferred trips using both deterministic fleet-average and probabilistic Monte Carlo approaches described in Section 3.

Across all RTS vehicles successfully matched to EPA emission records (N = 18,394), tailpipe CO_2 emission rates exhibit substantial heterogeneity. Emission rates range from near zero for electric vehicles to approximately 890 g CO_2 per mile for high-emission vehicles, with a median of 373 g CO_2 per mile and an interquartile range of approximately 317–433 g CO_2 per mile. This wide distribution underscores the importance of accounting for fleet heterogeneity when estimating transportation emissions and motivates the use of probabilistic vehicle assignment in the MDLD-based framework.

6.3 Methodology

This study proposes a framework for estimating greenhouse gas (GHG) emissions from travel activity derived from MDLD. The approach integrates spatially resolved vehicle composition and emission rates from RTS and EPA datasets with trip-level distance and temporal information from MDLD. Figure 6.1 illustrates the overall workflow.

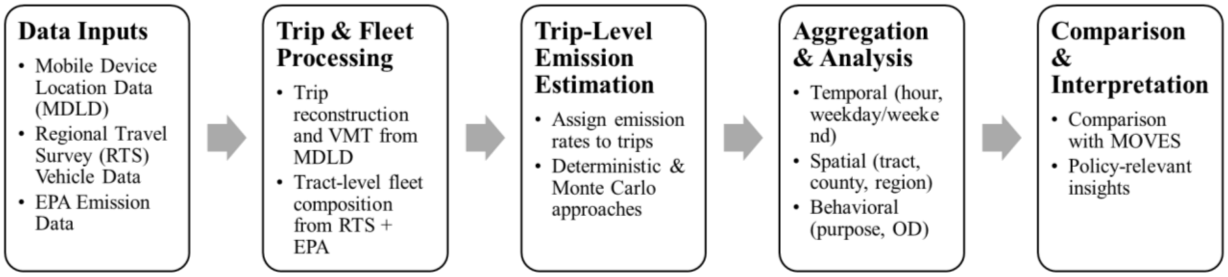


Figure 6.1 Overall workflow of the MDLD-based CO_2 emission estimation framework

6.3.1 Trip Construction and Activity Measures from MDLD

Trip-level travel activity was derived from Mobile Device Location Data (MDLD) using the standardized processing framework described in earlier chapters. Raw location records were converted into individual trips, defined as movements between stationary activity locations. Each trip includes a start and end time, origin and destination, and travel duration. To support resident-based emission analysis, a home location was inferred for each device based on nighttime stay patterns and assigned to Census tracts and counties. Trips were filtered to retain only vehicular travel based on trajectory and speed characteristics. Vehicle miles traveled (VMT) were then calculated at the trip level from reconstructed movement paths. Because MDLD represents a sample of the traveling population, expansion weights were applied to trips to align aggregated travel activity with regional population benchmarks [85]. The resulting weighted trip dataset provides the behavioral basis for subsequent emission estimation and aggregation across space and time.

6.3.2 Estimating Trip-Level CO_2 Emissions

To examine different approaches to estimate trip-level CO_2 emissions, six methods were evaluated:

- Approach 0: EPA Average Passenger Vehicle Benchmark

As a baseline reference, trip-level CO_2 emissions were first estimated using a constant emission factor of 393 g CO_2 per mile, representing the average gasoline-powered passenger vehicle. This value is derived from the U.S. Environmental Protection Agency's Greenhouse Gas Equivalencies Calculator [183] and reflects the weighted-average fuel economy and tailpipe CO_2 emissions of passenger cars and light trucks in the United States. This benchmark provides a simple and transparent point of comparison that ignores variation in vehicle characteristics, operating conditions, and travel behavior. As such, it is not intended to represent realistic fleet heterogeneity, but rather to serve as a common reference for evaluating the relative performance of more detailed emission estimation methods.

- Approach 1: Trip-average speed–based emission model

This approach estimates trip-level CO_2 emissions as a function of the average travel speed of each trip. Following Zhang et al. [184], CO_2 emission intensity (grams per kilometer) is modeled as a polynomial function of average trip speed, derived from laboratory and on-road emission tests of Japanese vehicles conducted by the National Institute for Land and Infrastructure Management. Trip-average speed is computed as total trip distance divided by travel time, and total trip emissions are obtained by multiplying the estimated emission intensity by trip distance. This method captures the non-linear relationship between speed and fuel consumption but assumes homogeneous vehicle characteristics and relies on parameters calibrated for Japanese vehicle fleets and operating conditions.

- Approach 2: Trajectory-level speed-based emission model

This approach estimates trip-level CO_2 emissions using speed information at the trajectory-point level rather than relying on trip-average speed. Following Zhao et al. [185], emissions are computed as a function of travel distance, travel time, and instantaneous speed changes between successive GPS sampling points. The model accounts for speed-dependent fuel consumption and incorporates the effect of acceleration by distinguishing between speed-increasing and non-increasing segments along the trajectory. Trip-level emissions are obtained by aggregating emissions across all trajectory points within a trip. Compared with trip-average speed models, this approach better captures short-term speed fluctuations and driving dynamics, but it assumes homogeneous vehicle characteristics and relies on emission coefficients calibrated using taxi data and non-U.S. contexts.

- Approach 3: Trajectory-based emission model incorporating stationary time

This approach estimates trip-level CO_2 emissions by explicitly accounting for both moving and stationary vehicle operation, following Kan et al. [186]. Fuel consumption is first estimated using a COPERT-based model, where consumption during moving activity is expressed as a non-linear function of average speed for each space–time path segment, and fuel consumption during stationary activity (engine-on idling) is modeled as a linear function of stop duration. Total fuel consumption for a trip is computed as the sum of fuel used during moving and stationary periods along the inferred trajectory. CO_2 emissions are then obtained by applying a constant conversion factor of 8,887 grams of CO_2 per gallon of gasoline consumed, consistent with EPA and IPCC guidelines [183]. Compared with speed-only models, this approach better captures operational driving conditions such as congestion and idling, but it relies on emission parameters calibrated using European COPERT standards and assumes homogeneous gasoline vehicle characteristics.

- Approach 4: Tract-level fleet-average emission assignment

This approach estimates trip-level CO_2 emissions using tract-specific average vehicle emission rates derived from the RTS vehicle fleet. Vehicle-level CO_2 emission factors (grams per mile) from EPA-tested fuel economy data are merged with RTS vehicle records, and a weighted mean emission rate is computed for each Census tract using household expansion weights. All devices residing in the same tract are assigned the same average emission rate, representing a deterministic baseline that reflects spatial variation in fleet composition while ignoring within-tract heterogeneity.

For each trip i made by device d on day t, CO_2 emissions are calculated as:

$$\mathrm{CO}_{2(i,d,t)} = \bar{r}_d \times \mathrm{VMT}_{(i,d,t)}$$

where $\bar{r}_d$ is the tract-level average emission rate (grams of CO_2 per mile) assigned based on the device's home tract, and $\mathrm{VMT}_{(i,d,t)}$ is the trip distance measured from MDLD. This approach preserves consistency with regional fleet characteristics but suppresses variability in vehicle emissions among households within the same tract.

- Approach 5: Monte Carlo Assignment.

To represent uncertainty and heterogeneity in vehicle ownership that cannot be directly observed in MDLD, a Monte Carlo (MC) framework was implemented. Vehicle emission-rate pools were constructed by merging RTS vehicle records with EPA-certified tailpipe CO_2 emission rates (grams per mile). Pools were defined at the Census tract level where sufficient RTS vehicle observations were available; for tracts with limited vehicle records, county-level pools were used as a fallback to ensure complete spatial coverage. Within each pool, emission rates were sampled with replacement using household expansion weights, preserving the observed distribution of vehicle types and technologies.

For each Monte Carlo iteration k, every device d residing in tract j was independently assigned a vehicle emission rate $r_d^{(k)}$ drawn from the corresponding tract- or county-level pool. Trip-level CO_2 emissions were then computed by multiplying the sampled emission rate by the trip distance derived from MDLD. Device-level emissions were aggregated to tract, county, and regional totals **within each simulation**, yielding one emission realization per spatial unit per iteration. A total of 1,000 Monte Carlo simulations were performed, producing empirical distributions of total emissions that explicitly reflect uncertainty in unobserved vehicle characteristics. Summary statistics (e.g., mean, median, and percentile ranges) were then derived across simulations for subsequent analysis and comparison.

Together, Approaches 4 and 5 provide complementary fleet-aware baselines that differ only in their treatment of within-tract vehicle heterogeneity, allowing the added value of probabilistic vehicle assignment to be isolated and evaluated.

6.3.3 Aggregation by Temporal and Spatial Dimensions

Trip-level CO_2 emissions were aggregated across multiple temporal and spatial dimensions to support comparison across methods and alignment with policy-relevant reporting scales. For temporal aggregation, each trip was assigned to an hourly bin based on its local start time. Trips were further classified as weekday or weekend travel to capture systematic differences in travel patterns and vehicle operation between workdays and non-workdays.

For spatial aggregation, emissions were attributed to the residential locations of travelers rather than to roadway segments. Specifically, trips were aggregated by the home Census tract and home county of each device, as inferred from MDLD. This origin-based accounting framework interprets emissions as the carbon footprint generated by residents and neighborhoods, in contrast to conventional location-based inventories that assign emissions to where vehicles operate on the road network. This distinction enables analysis of who generates emissions and how travel demand contributes to regional carbon outcomes.

For each temporal unit (hour × day type), or for each spatial unit, total CO_2 emissions were computed as the weighted sum of trip-level emissions:

$$CO_{2,agg} = \sum_{i \in h \ or \ s} CO_{2(i,d,t)} \times w_{(i,d,t)}$$

Where h denotes a specific hour–day-type combination, s denotes a spatial unit, and $w_{i,d,t}$ is the MDLD expansion weight associated with trip i.

6.3.4 Comparison with MOVES

To benchmark the MDLD-based framework, emissions were compared against outputs from the EPA's Motor Vehicle Emission Simulator (MOVES5), the standard tool used for on-road emissions inventory in the United States. MOVES was configured for the same 11-county Washington, D.C. metropolitan region using the default national database with county-level geographic selection.

Hourly CO_2 emissions and VMT were extracted from MOVES for on-road roadway types and aggregated to the same temporal and spatial resolution used in the MDLD analysis. Because MOVES attributes emissions to vehicles operating within a geographic area regardless of vehicle residency, whereas the MDLD framework attributes emissions to trips generated by residents, comparisons were conducted at regional and county scales where these two accounting perspectives are most comparable.

The benchmarking focused on the following metrics:

- Average daily total CO_2 emissions,
- Average CO_2 emissions per mile (CO_2 / VMT),
- Hourly emission profiles by weekday and weekend,
- Weekday versus weekend contrasts at regional and county levels.

Relative differences and ratios between MDLD-based estimates and MOVES outputs were computed to assess consistency, identify systematic differences, and evaluate the extent to which activity-based, origin-attributed emissions align with conventional roadway-based inventories.

6.3.5 Variance Decomposition and Uncertainty Analysis

To quantify uncertainty in trip-based CO_2 estimation and identify dominant sources of variability, a variance decomposition analysis was conducted at the device level. Total variability in estimated emissions was decomposed into two components:

- Across-method variability (within devices):

For each device, variation in estimated emissions across alternative emission estimation approaches reflects methodological uncertainty arising from different modeling assumptions.

- Across-device variability (within methods):

For each estimation approach, variation in emissions across devices reflects heterogeneity in travel behavior, trip intensity, and (implicitly) vehicle characteristics.

This decomposition enables comparison of the relative importance of modeling choices versus behavioral heterogeneity in shaping emission outcomes, providing insight into whether uncertainty is driven primarily by methodological assumptions or by real-world variation in travel activity.

6.3.6 Trip Purpose Inference and Origin–Destination Emission Attribution

Trip purposes were inferred from MDLD using a rule-based classification framework that integrates spatial context, activity duration, time of day, and known land-use patterns surrounding trip endpoints [85]. Each trip was assigned to one of several mutually exclusive purpose categories, including home-based work (HBW), home-based other (HBO), work-based other (WBO), and other-based other (OBO).

Trip-level CO_2 emissions were subsequently linked to inferred trip purposes, enabling emissions to be aggregated and compared across purpose categories and between weekdays and weekends. This structure supports examination of how different types of travel demand, i.e. commuting versus discretionary travel, contribute to total emissions under varying temporal contexts.

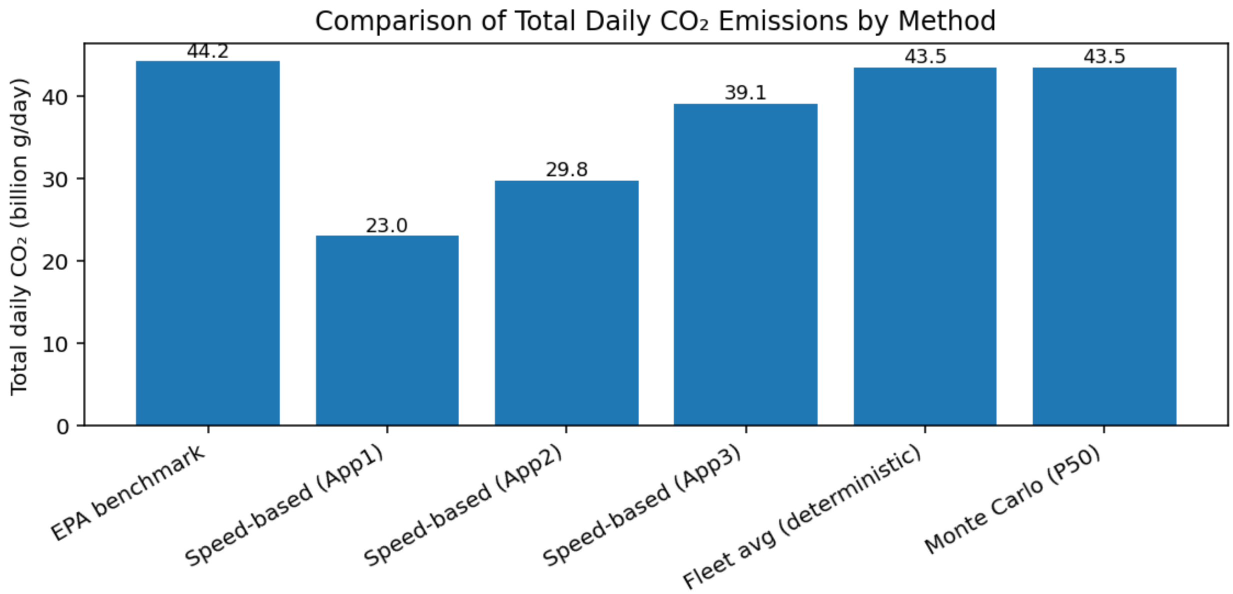
Origin–destination (OD) matrices were constructed by aggregating emission-weighted trip flows between geographic units. For each trip, the origin corresponds to the inferred trip start location and the destination to the inferred trip end location. OD emissions were aggregated at the county level to support regional-scale analysis and visualization. These OD emission matrices form the basis for identifying dominant inter-county emission flows and for visualizing spatial interaction patterns using heatmaps.

6.3 Results

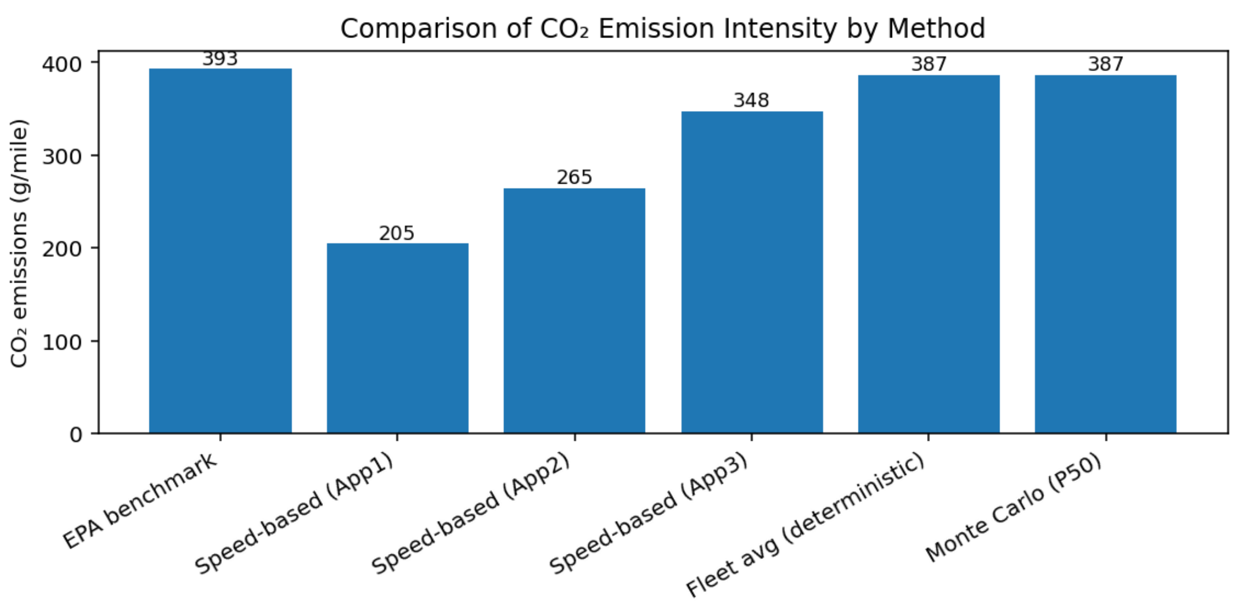
6.3.1 Comparison of Emission Estimation Methods

- Differences in total emissions and emissions per mile

Figure 6.2 compares total daily CO_2 emissions and emission intensity (grams per mile) estimated using alternative trip-level emission modeling approaches. Speed-based approaches (Approaches 1–3) yield substantially lower emission intensities, ranging from approximately 205 to 348 g CO_2 per mile, and correspondingly lower total daily CO_2 emissions (23.0–39.1 billion g/day), representing reductions of 12–48% relative to the EPA benchmark. These differences reflect the fact that the speed–emission relationships used in these approaches were calibrated using non-U.S. vehicle fleets and operating conditions, which tend to underestimate emissions for U.S. passenger vehicles.



(a)

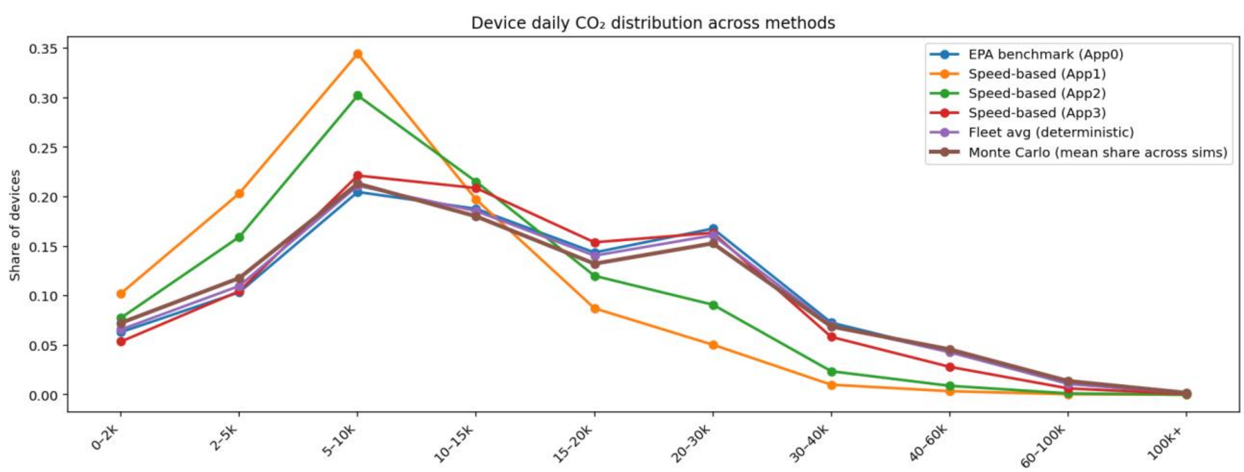


(b)

Figure 6.2 Average daily CO_2 emissions (a) and emission intensity (grams per mile) (b)

In contrast, methods incorporating U.S.-specific vehicle fleet composition produce emission intensities of approximately 387 g CO_2 per mile, closely aligning with the EPA benchmark value of 393 g CO_2 per mile. The deterministic fleet-average approach produces a total daily CO_2 estimate of approximately 43.5 billion grams per day, closely aligned with the EPA benchmark estimate of 44.2 billion grams per day. The Monte Carlo approach, summarized using the median (P50) across simulations, yields a nearly identical estimate (43.5 billion grams per day), indicating that stochastic vehicle assignment does not materially affect aggregate regional emissions. These results demonstrate that emission estimates are highly sensitive to modeling assumptions, and that fleet-aware approaches are critical for producing realistic CO_2 estimates in U.S. contexts.

- Distribution of Device-Level Daily CO_2 Emissions Across Methods



Daily CO_2 emissions (g/ day)

Figure 6.3 Distribution of device-level daily CO_2 emissions across emission estimation methods

Figure 6.3 compares the distribution of daily CO_2 emissions per device across alternative emission estimation methods. Device-level emissions are summarized using daily averages and grouped into emission bins to highlight differences in the shape of the distributions rather than differences in aggregate totals.

Across all methods, the distribution of device-level daily CO_2 emissions is right-skewed, with the majority of devices emitting between approximately 5,000 and 20,000 grams of CO_2 per day. However, substantial differences emerge in how methods allocate devices across lower and higher emission ranges. Speed-based approaches (Approaches 1–3) systematically shift mass toward lower emission bins, particularly in the 2–5k and 5–10k g/day ranges, while assigning fewer devices to higher-emission bins above 20k g/day. This pattern is consistent with the lower emission intensities produced by these models and reflects their calibration to non-U.S. vehicle fleets and operating conditions.

In contrast, the EPA benchmark, deterministic fleet-average approach, and Monte Carlo method yield closely aligned distributions across most bins, particularly in the mid-range (5–30k g/day). These methods produce substantially heavier upper tails than the speed-based approaches, indicating a higher prevalence of devices with large daily emissions. The Monte Carlo approach closely tracks the deterministic fleet-average distribution while exhibiting slightly higher shares in both lower and higher emission bins. This similarity reflects the fact that variation in device-level daily CO_2 emissions is dominated by differences in travel activity (vehicle miles traveled), rather than by variation in vehicle emission rates. As a result, probabilistic vehicle assignment does not materially alter the distribution of daily emissions at the individual-device level. However, the Monte Carlo framework provides a more realistic representation of fleet composition uncertainty and becomes relevant for analyses that focus on emission-rate distributions examined in subsequent sections.

- Distribution of Device-Level Emission Rates

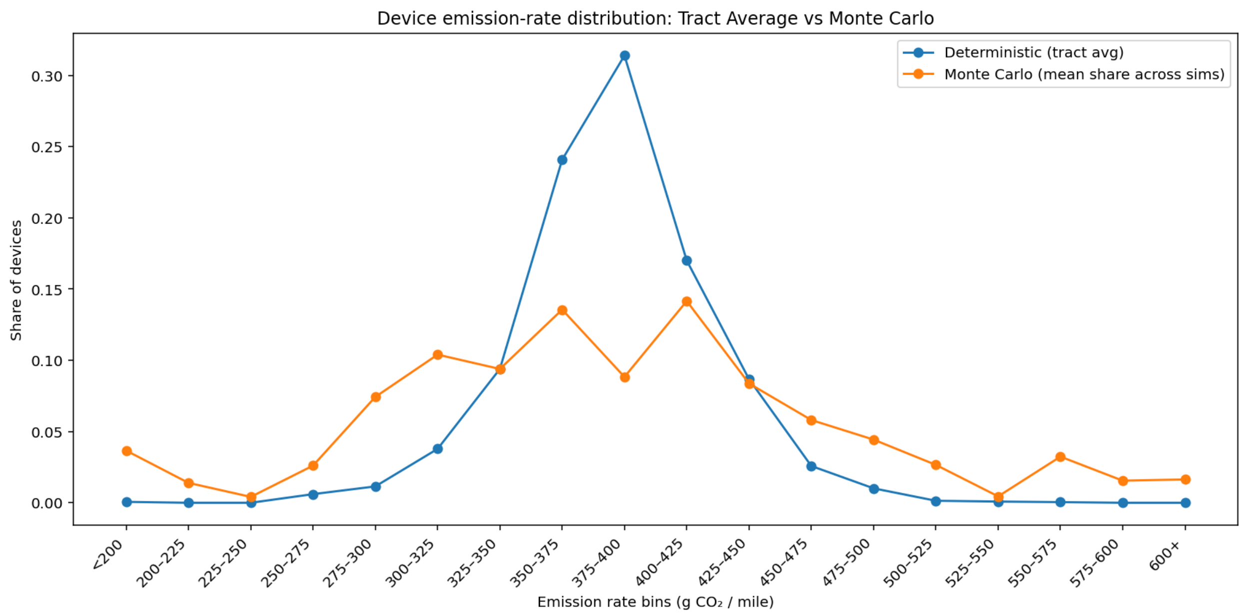


Figure 6.4 Distribution of per-mile CO_2 emission rates assigned to devices under the deterministic tract-average approach and the Monte Carlo framework

Figure 6.4 compares the distribution of per-mile CO_2 emission rates assigned to devices under the deterministic tract-average approach and the Monte Carlo framework. While the distributions of device-level daily CO_2 emissions are largely similar across fleet-based methods, the emission-rate distributions reveal substantial differences in the representation of vehicle heterogeneity.

Under the deterministic approach, all devices within a given census tract are assigned the same average emission rate, resulting in a highly concentrated distribution centered around mid-range values (approximately 350–425 g CO_2 per mile). As a consequence, the deterministic method assigns very small shares of devices to the lowest and highest emission-rate bins. In contrast, the Monte Carlo approach produces a markedly broader distribution of emission rates, with higher shares of devices assigned to both low-emission and high-emission bins. This reflects the probabilistic assignment of vehicle emission rates drawn from the tract-level fleet pool, allowing devices within the same tract to represent a diverse mix of vehicle technologies and efficiencies. Importantly, the central tendency of the Monte Carlo emission-rate distribution remains close to the tract average, explaining why aggregate emissions and device-level daily CO_2 totals are similar across fleet-based methods. However, the heavier tails produced by the Monte Carlo approach indicate that it preserves within-tract fleet heterogeneity that is suppressed under deterministic averaging.

Taken together with the device-level daily emission results, these findings demonstrate that variability in total daily CO_2 emissions is primarily driven by differences in travel activity, while the Monte Carlo framework adds value by more realistically representing uncertainty in vehicle characteristics.

6.3.2 Variance Decomposition of CO_2 Estimates

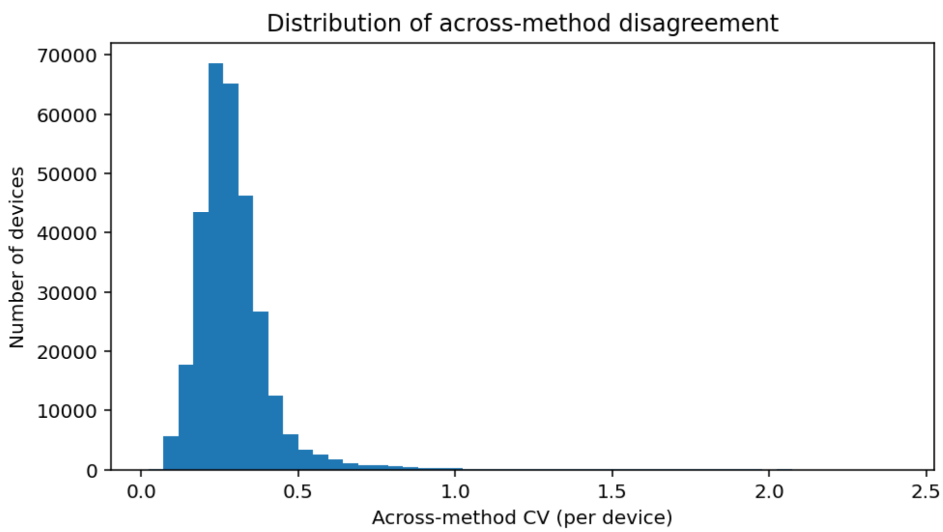


Figure 6.5 Distribution of across-method disagreement in daily CO_2 estimates

Figure 6.5 illustrates the distribution of across-method disagreement in estimated daily CO_2 emissions at the device level, quantified by the coefficient of variation (CV) across alternative emission estimation methods for the same traveler. The distribution is right-skewed but concentrated at relatively low values, with a median across-method CV of 0.27 and a 90th percentile of 0.41, indicating that sensitivity to modeling assumptions is moderate for most devices and extreme disagreement is uncommon.

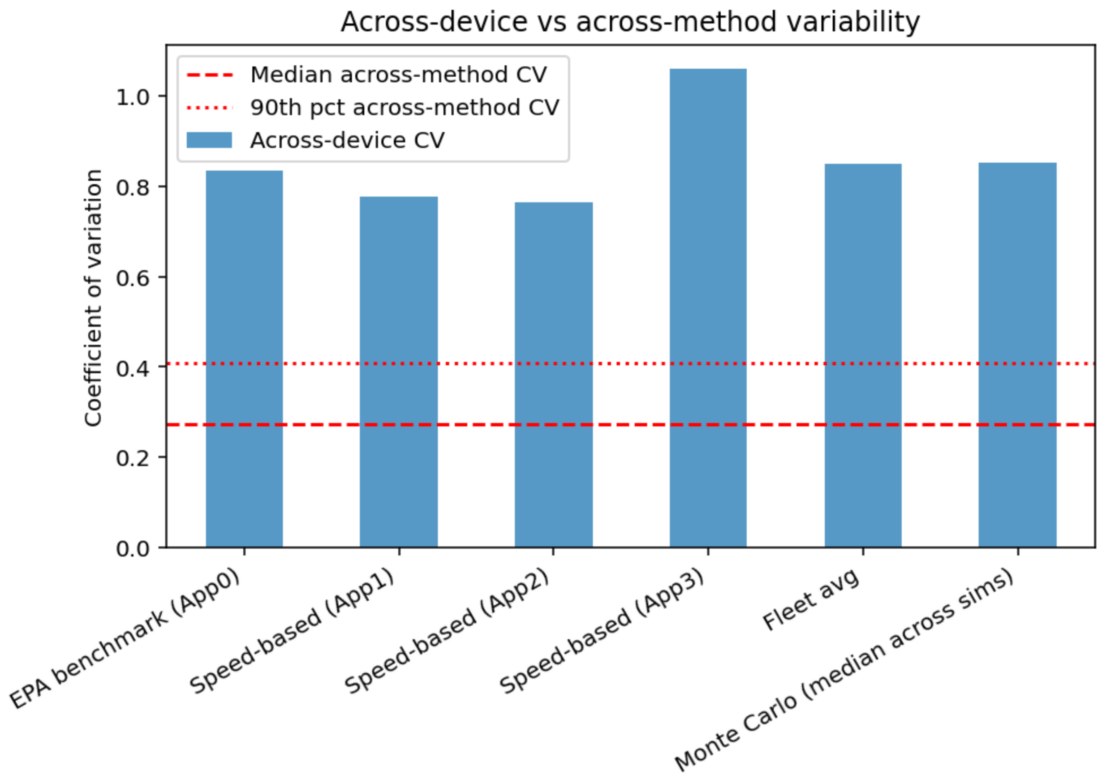
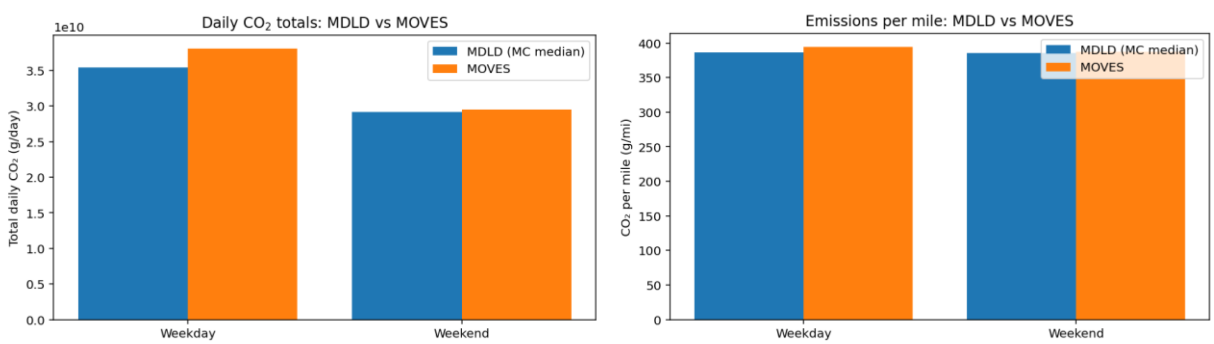


Figure 6.6 Comparison of across-device and across-method variability in daily CO_2 emissions

In contrast, Figure 6.6 compares across-device variability under each fixed method with across-method disagreement. Across-device CVs, computed as the coefficient of variation of daily CO_2 emissions across devices for each method, are consistently large across all methods, ranging from 0.77 to 1.06. These values are approximately three to four times larger than the median across-method CV and remain well above even the 90th percentile of across-method disagreement. This pattern indicates that heterogeneity in travel behavior across individuals, such as differences in trip frequency and distance, is the dominant source of variability in resident-based CO_2 emissions.

Taken together, these results demonstrate that while emission estimates are sensitive to methodological assumptions, behavioral heterogeneity across travelers overwhelmingly drives variability in CO_2 emissions. The consistency of across-device CVs across all methods further suggests that this conclusion is robust to emission modeling choice. Consequently, uncertainty-aware approaches such as Monte Carlo vehicle assignment are best viewed as complementary tools that capture emission-factor uncertainty, rather than as sources of dominant variability relative to real-world differences in travel behavior.

6.3.3 Comparison of MDLD-based CO_2 Emissions with MOVES



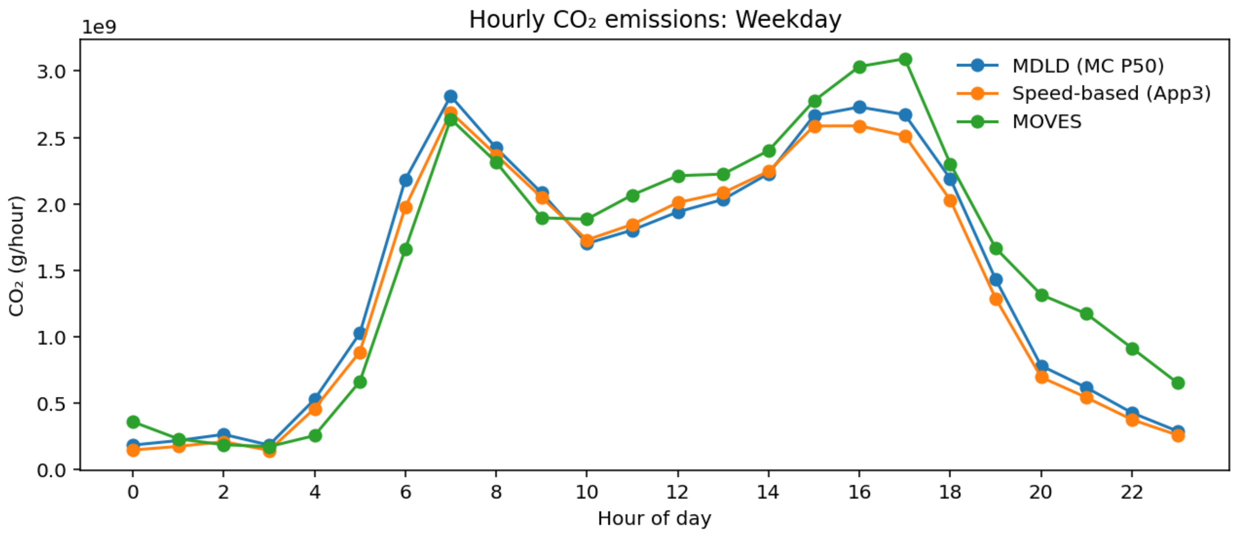
(a) (b)

Figure 6.7 Daily CO_2 totals estimated by MDLD and MOVES (a) and Average CO_2 emissions per mile from MDLD and MOVES (b)

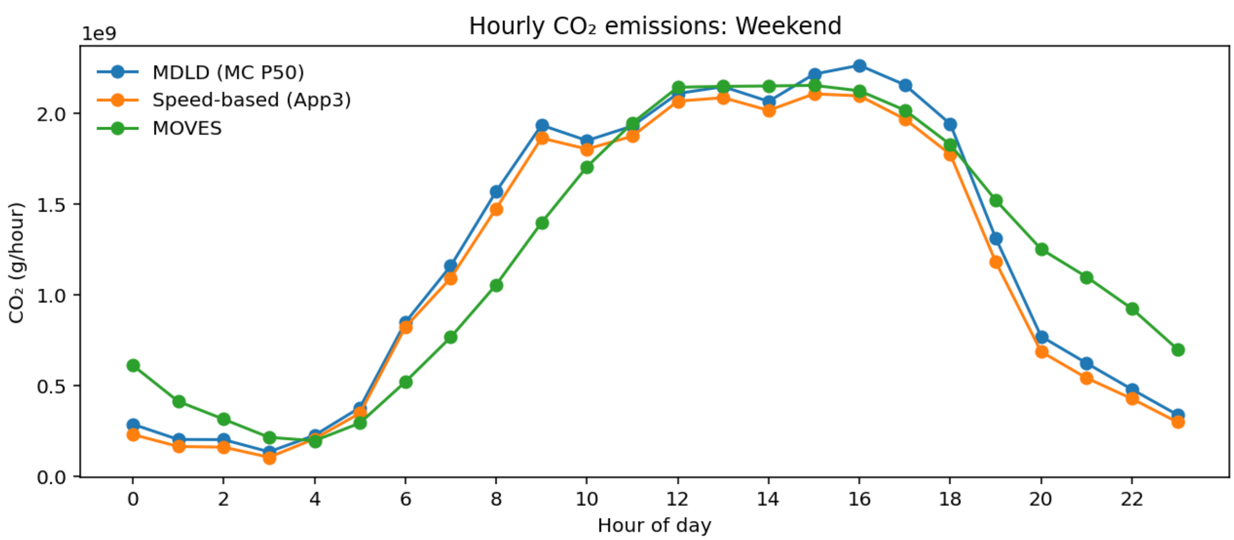
Figure 6.7 (a) compares total daily CO_2 emissions estimated using the MDLD-based Monte Carlo median approach and EPA's MOVES model for weekdays and weekends. MDLD-based estimates (3.54×10^{10} g/day) are slightly lower than MOVES (3.81×10^{10} g/day) on weekdays, with a ratio of 0.93, and closely aligned on weekends, with a ratio of 0.99. In both cases, the magnitude of total daily emissions is comparable, indicating consistency between the resident-based MDLD framework and the activity-based MOVES inventory at the regional scale.

Figure 6.7 (b) compares average CO_2 emissions per mile. Emission intensities from MDLD and MOVES are highly consistent for both weekdays (387 vs 394 g/mi) and weekends (386 vs 387 g/mi), with differences on the order of a few percent. This agreement suggests that differences in total emissions are driven primarily by differences in estimated travel activity rather than systematic discrepancies in emission factors.

Together, these results demonstrate that the MDLD-based framework reproduces regional-scale emission levels and intensities comparable to MOVES. Differences between MDLD-based and MOVES emission estimates reflect fundamental differences in accounting perspective rather than discrepancies in emission factors. MOVES estimates emissions from vehicles **operating** within a geographic boundary, including trips generated by non-residents and through traffic, whereas the MDLD framework attributes emissions to trips originating from different locations within the study area. As a result, MDLD-based estimates may be lower than MOVES estimates in regions with substantial inbound commuting or pass-through traffic, particularly on weekdays. The close agreement in emissions per mile across the two approaches indicates that observed differences in total emissions are driven primarily by differences in attributed travel activity rather than emission rates.



(a)



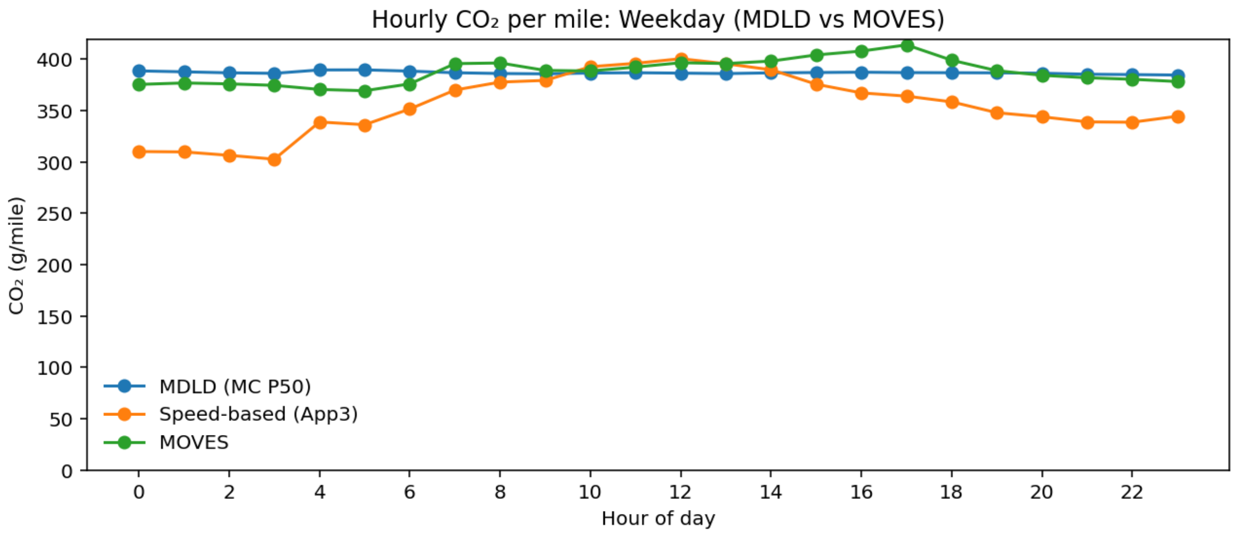
(b)

Figure 6.8 Hourly CO_2 emissions on weekdays (a) and weekends (b): MDLD vs. MOVES

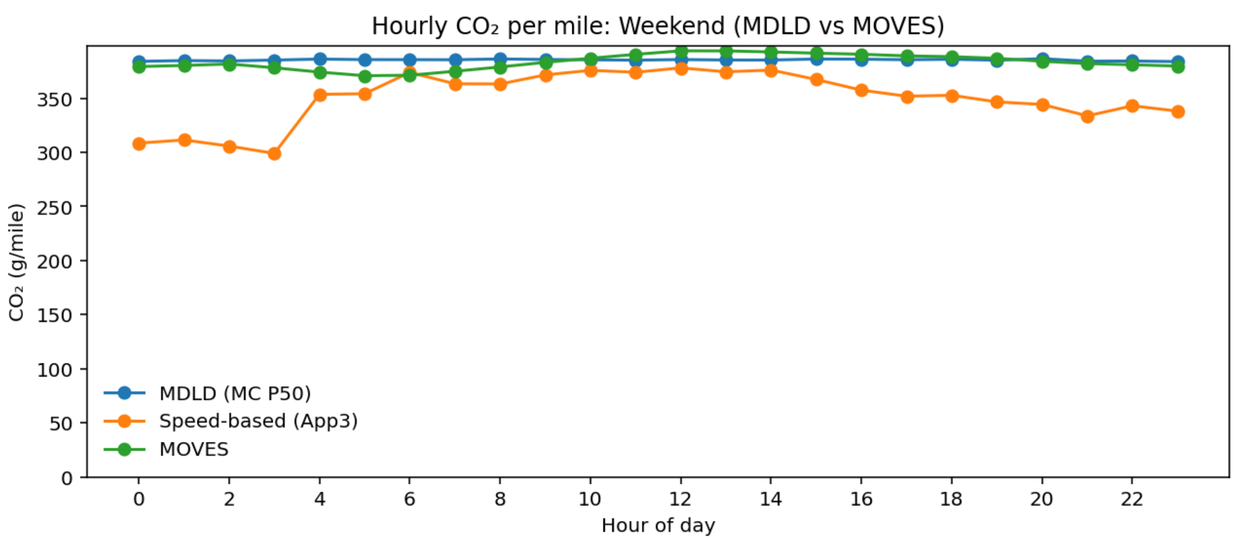
Figure 6.8 compares hourly CO_2 emissions estimated using the MDLD-based Monte Carlo approach (P50), a representative speed-based method (Approach 3), and EPA MOVES for weekdays and weekends. Across all methods, the diurnal profiles exhibit consistent temporal patterns, including pronounced morning and afternoon peaks on weekdays and a flatter, extended midday peak on weekends. This consistency indicates that MDLD-derived travel activity captures realistic temporal dynamics of regional travel demand. The timing of peak emissions aligns closely between the two models, reflecting similar representations of daily travel rhythms.

In terms of magnitude, MDLD-based estimates using the Monte Carlo median (P50) consistently exceed those from the speed-based approach (App3), particularly during peak hours, while remaining modestly lower than MOVES estimates. This pattern reflects differences in underlying modeling assumptions: speed-based methods rely on simplified emission–speed relationships calibrated on non-U.S. fleets, whereas the Monte Carlo framework incorporates U.S.-specific fleet composition derived from survey data.

Notable differences arise during off-peak hours. MOVES generally reports higher emissions during late evening and nighttime hours, whereas MDLD-based estimates are lower during these periods. During weekday peak periods, MDLD estimates are slightly higher in the morning commute hours, while MOVES tends to produce higher emissions during the late afternoon peak. These differences likely reflect conceptual distinctions between the two frameworks: MOVES estimates emissions from vehicles operating within a geographic boundary, whereas the MDLD approach attributes emissions to trips generated by residents, emphasizing travel origin and behavioral patterns.



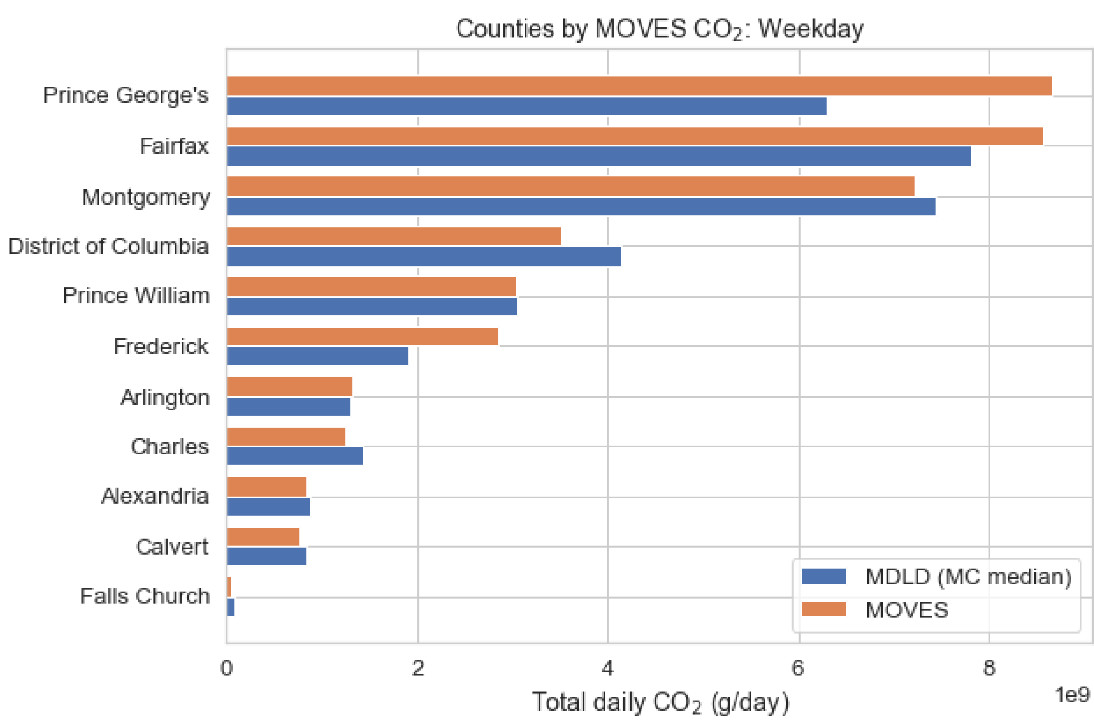
(a)



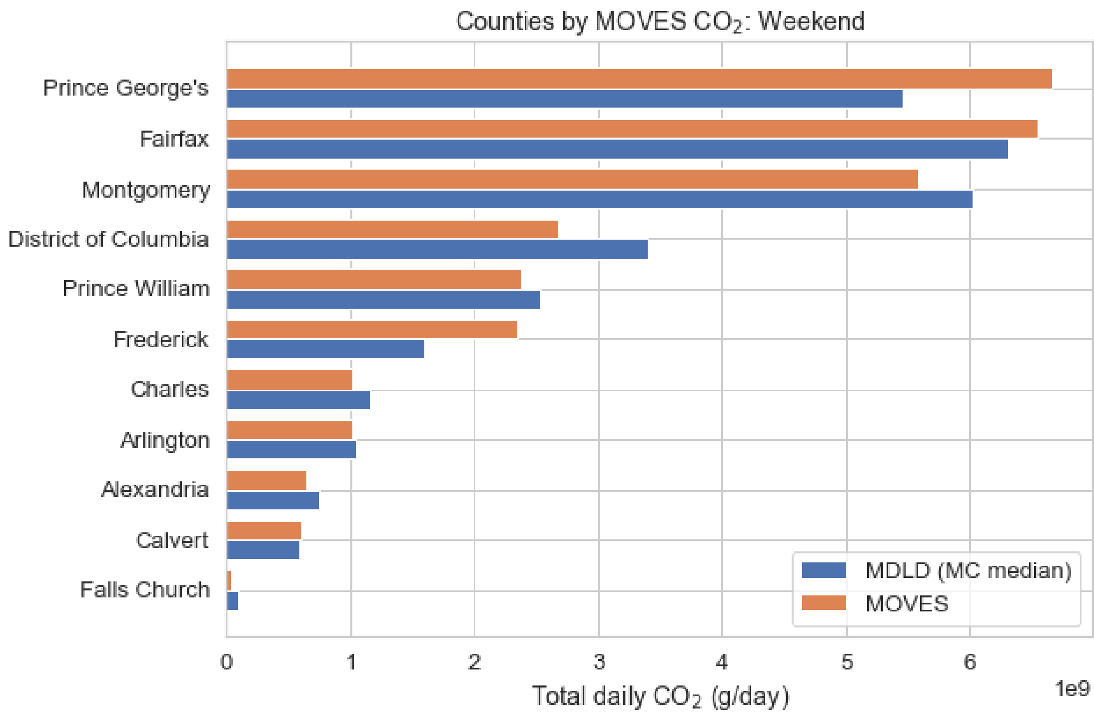
(b)

Figure 6.9 Hourly CO_2 emission intensities on weekdays (a) and weekends (b): MDLD vs. MOVES

Figure 6.9 presents hourly CO_2 emission intensities (grams per mile) for the same comparison. While the Monte Carlo (fleet-aware) and MOVES estimates produce relatively stable CO_2 intensities across hours, the speed-based approach exhibits pronounced diurnal variation, with lower emission rates during late night/early morning and higher values during daytime and peak periods. This pattern likely reflects time-of-day changes in traffic operating conditions (e.g., free-flow versus congested stop-and-go regimes), suggesting that speed-based methods can provide useful sensitivity to congestion dynamics. However, the speed-based approach remains systematically lower in magnitude than both the fleet-aware and MOVES estimates across most hours, consistent with the fact that its emission–speed relationship was calibrated under non-U.S. fleets and operating contexts. Taken together, these results indicate that fleet-aware and speed-based approaches emphasize different aspects of real-world emissions: the former improves realism in absolute levels and vehicle-mix heterogeneity, whereas the latter captures temporal variation associated with traffic conditions.



(a)



(b)

Figure 6.10 County-level total daily CO_2 emissions on weekdays (a) and weekends (b): MDLD vs. MOVES

Figures 6.10 compare county-level total daily CO_2 emissions estimated by the MDLD-based Monte Carlo framework and MOVES for weekdays and weekends, respectively. The two methods exhibit strong agreement in the relative ranking of counties, with Prince George's, Fairfax, Montgomery, and the District of Columbia consistently emerging as the largest contributors. This consistency indicates that MDLD-derived travel activity reproduces the dominant spatial structure of regional emissions.

Despite this agreement in spatial patterns, systematic differences in emission levels are observed across counties. In employment- and residence-intensive areas such as the District of Columbia and Montgomery County, MDLD-based estimates exceed MOVES, reflecting emissions generated by resident trips that subsequently leave the county. In contrast, MOVES reports higher emissions in corridor-oriented counties such as Prince George's, where pass-through and non-resident traffic contribute substantially to on-road emissions. These differences highlight the conceptual distinction between origin-based and operating-based emission accounting.

Differences between the two approaches are notably smaller on weekends, when travel is more likely to be locally generated and commuter flows are reduced. Together, these results demonstrate that MDLD and MOVES provide complementary perspectives on spatial emission allocation, with MDLD offering additional insight into the residential and behavioral drivers of transportation emissions. The county-level comparison also highlights that while MOVES and MDLD produce similar regional totals, the spatial allocation of emissions depends critically on whether emissions are attributed to locations of vehicle operation or to origins of travel demand.

6.3.4 Emissions by Trip Purpose

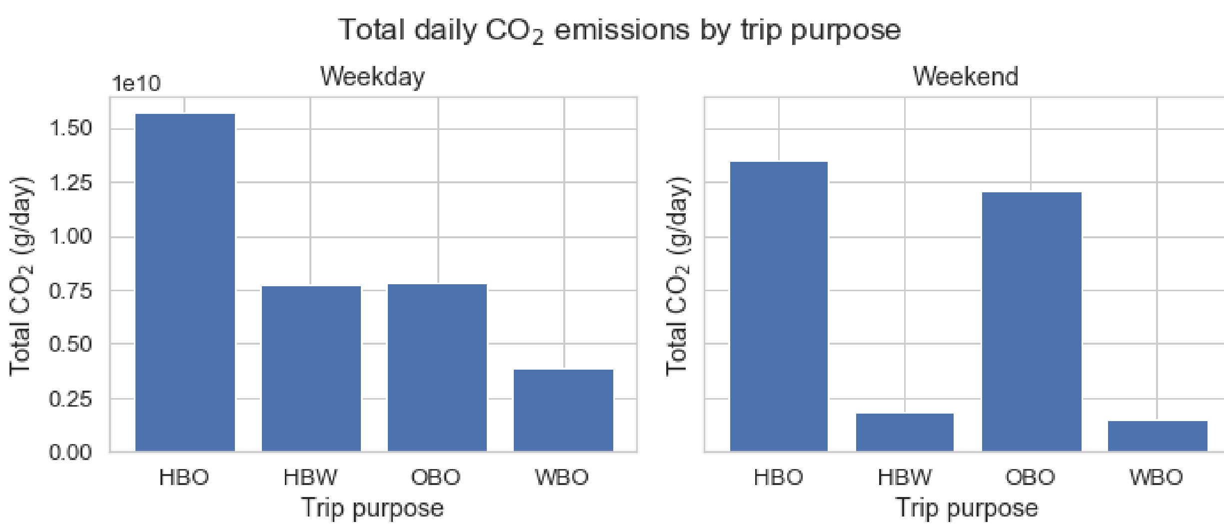


Figure 6.11 Total CO_2 by trip purpose on weekdays and weekends

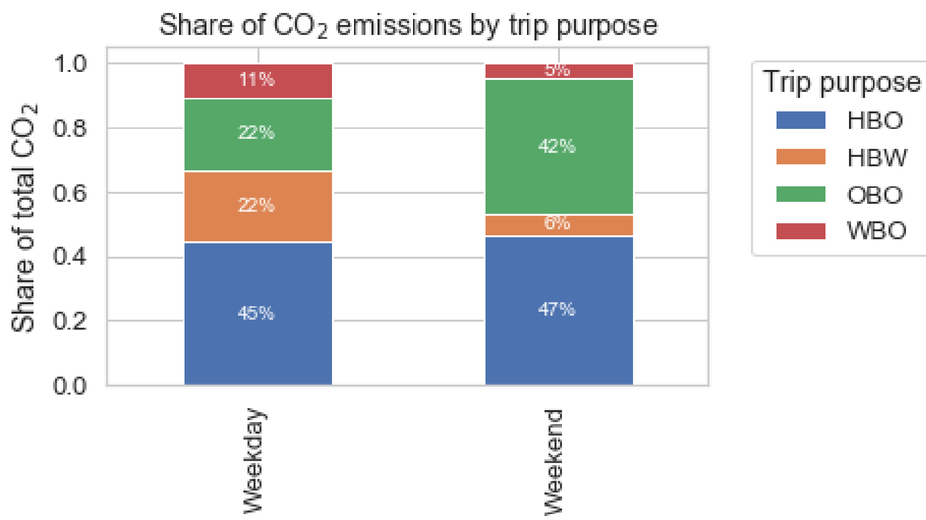


Figure 6.12 CO_2 shares by trip purpose on weekdays and weekends

Figure 6.11 presents total daily CO_2 emissions by trip purpose for weekdays and weekends, and Figure 6.12 shows the shares. On weekdays, home-based other (HBO) trips account for the largest share of emissions, exceeding those associated with home-based work (HBW) travel. Other-based other (OBO) trips also contribute substantially, indicating that non-work activities are a dominant driver of transportation emissions even during weekdays.

On weekends, the contribution of work-related travel declines sharply, while HBO and OBO trips dominate total emissions. Despite lower overall travel volumes, weekend emissions remain substantial, reflecting continued high levels of discretionary travel.

These results demonstrate the value of the MDLD-based framework for linking transportation emissions to underlying activity patterns, enabling analyses that are not possible using traditional roadway-based emission models such as MOVES. Policies targeting non-work travel, such as land-use planning, local accessibility improvements, and travel demand management, may yield emissions reductions comparable to or greater than those focused exclusively on commuting.

6.3.5 Origin–Destination Emission Flows

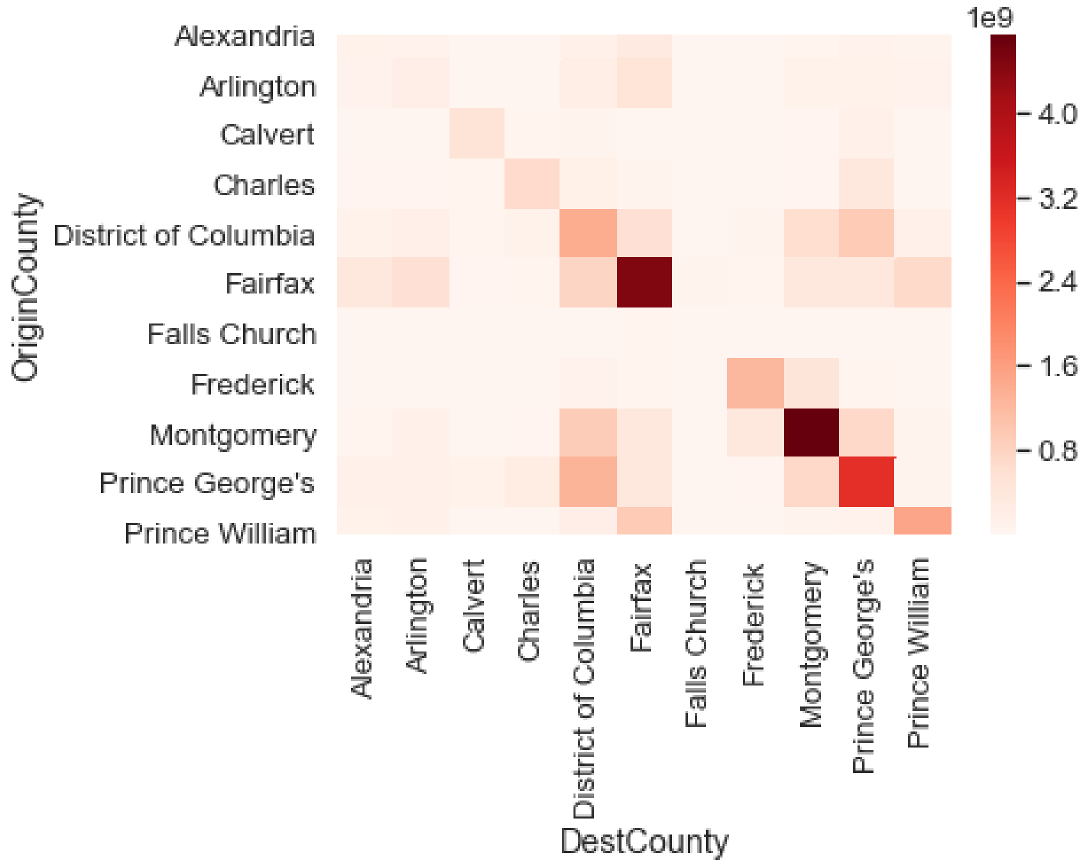


Figure 6.13 County-level origin–destination (OD) CO_2 emissions estimated using the MDLD-based framework for weekday travel

Figure 6.13 presents county-level origin–destination (OD) CO_2 emissions estimated using the MDLD-based framework for weekday travel. Emissions are dominated by within-county trips, as indicated by the strong diagonal in the OD heatmap, reflecting the substantial contribution of local travel to total regional emissions. Cross-county emissions are concentrated in a limited number of OD pairs, primarily connecting major suburban counties with the District of Columbia, highlighting the role of commuting and regional travel corridors.

The relative contributions of within- and cross-county travel to total CO_2 emissions are also calculated: On weekdays, within-county trips account for 51.3% of total emissions, while cross-county trips contribute 48.7%. A nearly identical split is observed on weekends, with 51.5% of emissions occurring within counties and 48.5% associated with cross-county travel. The stability of this split across day types suggests that local travel remains a dominant source of transportation emissions even outside traditional commuting periods. More than half of transportation CO_2 emissions originate from trips that begin and end within the same county, indicating that policies targeting local travel behavior may yield emission reductions comparable to those focused on long-distance commuting.

These results underscore the importance of local accessibility and short-distance travel in shaping regional emissions, while also revealing the persistence of inter-county travel dependencies. By attributing emissions to trip origins rather than roadway locations, the MDLD-based framework provides insight into spatial emission responsibility and regional interconnections that cannot be captured by traditional operating-based emission models.

6.4 Discussion

This study proposes and evaluates a trip-based framework for estimating on-road CO_2 emissions using mobile device location data (MDLD), integrated with spatially resolved vehicle composition and emission factors derived from regional travel survey (RTS) and EPA datasets. By linking trip-level distance and timing information with tract-level fleet characteristics and explicitly accounting for uncertainty in vehicle assignment, the framework enables fine-grained spatial, temporal, and behavioral analysis of transportation emissions that is not possible using traditional roadway-based models alone.

6.4.1 Key Findings

Several key findings emerge from the analysis. First, comparisons across alternative trip-level emission estimation methods demonstrate that modeling assumptions influence estimated emissions, but these effects are modest relative to variability driven by differences in travel behavior across individuals. Across-device coefficients of variation are consistently three to four times larger than across-method disagreement, indicating that heterogeneity in travel activity—rather than emission modeling choice—is the dominant driver of variability in resident-based CO_2 estimates. This finding highlights the importance of behaviorally rich data sources such as MDLD for understanding transportation emissions.

Second, the Monte Carlo vehicle assignment approach provides a principled way to represent uncertainty arising from unobserved vehicle characteristics while preserving consistency with regional emission totals. Although Monte Carlo median and tract-level weighted-average estimates are similar at aggregate scales, the Monte Carlo framework adds value by more realistically representing uncertainty in vehicle characteristics.

Third, MDLD-based emission estimates are broadly consistent with EPA's MOVES model at the regional scale. Daily total emissions and emissions per mile are closely aligned for both weekdays and weekends, and both approaches reproduce similar diurnal emission profiles. Systematic differences between MDLD and MOVES are attributable to differences in accounting perspective rather than discrepancies in emission factors. MOVES attributes emissions to vehicles operating within geographic boundaries, including non-resident and pass-through traffic, whereas the MDLD framework attributes emissions to trips generated by residents. This distinction explains why MDLD estimates are generally lower than MOVES estimates in commuter-intensive counties on weekdays and why the two approaches converge more closely on weekends.

Fourth, the MDLD framework enables attribution of emissions to travel purposes and origin–destination flows, revealing insights that are not accessible in roadway-based inventories. Non-work travel, particularly home-based other and other-based other trips, accounts for a substantial share of total emissions on both weekdays and weekends, underscoring the importance of discretionary travel in shaping transportation-related greenhouse gas emissions. County-level OD analysis further shows that while cross-county travel remains significant, slightly more than half of emissions originate from within-county trips across both day types. This result highlights the critical role of local travel and accessibility in regional emission outcomes.

6.4.2 Implications for Policy and Planning

The findings have several important implications for transportation and climate policy. First, the dominance of behavioral heterogeneity and non-work travel suggests that emission reduction strategies focused exclusively on commuting may overlook a large share of potential reductions. Policies aimed at improving local accessibility, supporting mixed-use development, and reducing the need for discretionary auto travel may yield emission benefits comparable to those targeting work trips alone.

Second, the distinction between resident-based and operating-based emission accounting has implications for equity and responsibility. By attributing emissions to the origins of travel demand, the MDLD framework enables identification of which communities generate emissions and how travel patterns differ across neighborhoods and jurisdictions. This perspective is particularly relevant for designing equitable climate policies, evaluating household-level burdens, and aligning transportation investments with emission reduction goals.

Third, the OD analysis highlights the importance of regional coordination. Although within-county travel accounts for a majority of emissions, cross-county flows are concentrated in a limited number of major corridors, reflecting interdependencies among jurisdictions. Effective emission reduction strategies therefore require coordination across local and regional agencies rather than isolated, jurisdiction-specific interventions.

6.4.3 Limitations and Future Research

This study has several limitations that point to directions for future research. First, MDLD does not directly observe vehicle type, fuel, or occupancy, requiring probabilistic assignment based on aggregate fleet distributions. While the Monte Carlo framework explicitly represents this uncertainty, future work could integrate additional data sources—such as vehicle registration records or connected vehicle data—to refine vehicle attribution. Second, the analysis focuses on tailpipe CO_2 emissions and does not account for upstream emissions or non-exhaust pollutants. Extending the framework to a broader set of environmental impacts would further enhance its policy relevance. Finally, although the framework is demonstrated for a large metropolitan region, applying it to other contexts would help assess its generalizability and sensitivity to regional travel patterns.

Chapter 7: Conclusion

7.1 Overview of the Dissertation

This dissertation investigates how Mobile Device Location Data (MDLD) can be leveraged to advance transportation analysis in an era of rapid technological change, evolving travel behavior, and growing sustainability challenges. Traditional transportation data sources, such as household travel surveys, traffic counts, and roadway-based emission inventories, have long formed the backbone of transportation planning and policy analysis. While these tools remain indispensable, they face fundamental limitations related to sample size, temporal coverage, behavioral resolution, and the ability to capture emerging mobility patterns.

By exploiting the scale, continuity, and behavioral richness of MDLD, this dissertation demonstrates how new data infrastructures can complement conventional approaches and enable a more nuanced understanding of travel behavior, congestion dynamics, and environmental impacts. Across four interconnected studies, the dissertation develops methods to integrate MDLD with traditional datasets, validates MDLD against established behavioral relationships, applies MDLD to diagnose post-pandemic congestion, and extends its use to behaviorally grounded estimation of transportation-related greenhouse gas emissions.

Together, these chapters illustrate how transportation outcomes, such as congestion and emissions, are increasingly shaped not only by infrastructure supply or trip frequency, but by deeper behavioral and spatial transformations in how, when, and why people travel.

7.2 Summary of Key Findings

7.2.1 Advancing Transportation Data Infrastructure through Data Fusion

A central contribution of this dissertation is the development of methods to integrate MDLD with traditional travel survey data. By linking MDLD with the Regional Travel Survey (RTS), Chapter 3 demonstrates how the complementary strengths of these datasets can be combined to overcome long-standing data constraints. While travel surveys provide rich socio-demographic context, they are typically limited in duration and sample size. MDLD, in contrast, offers continuous, multi-day observations of travel behavior but lacks direct demographic information.

The data fusion approaches developed in this dissertation show that it is possible to construct longitudinal data panels that retain the behavioral richness of MDLD while incorporating demographic and contextual information from surveys. These fused datasets capture travel behaviors that are often underrepresented in surveys, including off-peak travel, non-work trips, and multi-day variability in individual travel patterns. As a result, the dissertation establishes a stronger empirical foundation for analyzing travel behavior over extended periods and across population groups.

7.2.2 Behavioral Validity of MDLD for Travel Demand Analysis

Chapter 4 addresses a critical question underlying the use of MDLD in transportation research: whether passively collected mobility data can recover economically meaningful behavioral relationships. Using longitudinal MDLD panels from the Washington, D.C. metropolitan region, the study examines the relationship between vehicle miles traveled (VMT) and gasoline prices during a period of substantial economic and social disruption.

The findings demonstrate that MDLD-based measures of daily VMT respond to changes in gasoline prices in a manner consistent with economic theory and prior empirical studies. Estimated short-run elasticities fall within the range reported in the literature, even as gasoline prices rose during the post-pandemic reopening period. Importantly, the analysis highlights how behavioral signal strength varies with the level of spatial aggregation, with aggregated pseudo-panels exhibiting greater stability and predictability than individual-level records.

These results provide strong evidence that MDLD captures meaningful behavioral responses to economic incentives, reinforcing its credibility as a data source for travel demand analysis. This validation is essential for interpreting the subsequent congestion and emissions analyses, which rely on MDLD to infer behavioral mechanisms rather than merely describe observed traffic patterns.

7.2.3 Post-Pandemic Congestion as a Behavioral and Spatial Phenomenon

Building on the behavioral foundations established earlier, Chapter 5 examines how teleworking and post-pandemic behavioral shifts have reshaped road congestion in the Washington, D.C. region. Contrary to the expectation that reduced commuting would alleviate congestion, the analysis reveals that congestion has intensified relative to pre-pandemic conditions, particularly during weekends.

A key insight of this chapter is that post-pandemic congestion is not primarily driven by a recovery in trip frequency. Instead, congestion is shaped by longer average trip distances, increased spatial dispersion of travel demand, and shifts in travel timing. By distinguishing changes at the origin–destination (OD) pair level from changes at the individual trip level, the study shows that travelers are not taking longer routes between the same locations; rather, they are increasingly engaging in trips that connect more distant locations.

These patterns are consistent with greater residential flexibility enabled by teleworking, reduced reliance on daily commuting, and increased discretionary travel. The persistence of elevated interstate travel further indicates that pandemic-induced changes in residential location and travel behavior have had lasting regional impacts.

Collectively, these findings highlight a telework–congestion paradox: policies and practices that reduce commuting frequency may simultaneously contribute to congestion through longer and more spatially dispersed trips. This challenges traditional congestion metrics and underscores the need to rethink how transportation system performance is evaluated in the post-pandemic era.

7.2.4 Behaviorally Grounded Estimation of Transportation Emissions

Chapter 6 extends the behavioral perspective of the dissertation to environmental outcomes by developing a trip-based framework for estimating transportation-related CO_2 emissions using MDLD. By integrating trip-level mobility data with spatially resolved vehicle fleet characteristics and emission factors, the framework enables emissions to be attributed to travel behavior, trip purpose, and residential geography.

A central finding is that variability in emissions is driven primarily by heterogeneity in travel behavior rather than by differences in emission modeling assumptions. Across-device variability in emissions substantially exceeds across-method disagreement, highlighting the dominant role of behavior in shaping transportation-related carbon footprints. This underscores the value of behaviorally rich data sources for climate analysis.

The chapter also demonstrates that MDLD-based emission estimates are broadly consistent with EPA's MOVES model at aggregate scales, with remaining differences attributable to contrasting accounting perspectives. While MOVES attributes emissions to where vehicles operate, the MDLD framework attributes emissions to the origins of travel demand. This distinction enables new insights into who generates emissions and why, revealing the substantial contribution of non-work and local travel to regional emissions.

7.3 Methodological Contributions

Collectively, this dissertation makes several methodological contributions to transportation research. First, the dissertation demonstrates how MDLD can be systematically integrated with traditional travel surveys to create longitudinal, behaviorally rich datasets that overcome the limitations of either source alone. Second, by recovering theory-consistent responses to gasoline prices, the research establishes MDLD as a credible tool for analyzing travel demand and behavioral responses to economic factors. Third, the congestion analysis introduces an OD-consistent framework that separates changes in congestion from changes in trip composition, providing a more robust basis for evaluating system performance over time. Finally, the GHG framework shifts emissions analysis from a purely roadway-based perspective to a behaviorally grounded, resident-based perspective, enabling attribution of emissions to travel behavior and population groups.

These contributions highlight the value of viewing MDLD not as a replacement for existing data sources and models, but as a complementary tool that expands the scope of questions that transportation researchers and planners can address.

7.4 Implications for Transportation Planning and Policy

The findings of this dissertation have important implications for transportation planning and policy. First, they suggest that strategies focused narrowly on peak-hour commuting are increasingly insufficient in a post-pandemic context characterized by flexible work arrangements, longer-distance travel, and elevated discretionary demand. Effective congestion management must account for changes in trip length, spatial distribution, and weekend travel patterns.

Second, the results underscore the importance of regional coordination. The growing prominence of interstate and cross-jurisdictional travel highlights the need for planning approaches that transcend administrative boundaries and address regional travel systems as integrated networks.

Third, from an environmental perspective, the dominance of behavioral heterogeneity and non-work travel in shaping emissions suggests that climate strategies focused exclusively on commuting may miss significant opportunities for emission reduction. Policies that improve local accessibility, reduce the need for discretionary auto travel, and align land use with travel demand can play a critical role in achieving transportation-related climate goals.

Finally, the resident-based perspective enabled by MDLD provides a foundation for more equitable climate and transportation policy by linking emissions and congestion outcomes to the origins of travel demand rather than solely to roadway locations.

7.5 Limitations and Future Research

Despite its contributions, this dissertation has several limitations. MDLD does not directly observe individual socio-demographic characteristics, vehicle attributes, or teleworking status, requiring imputation or probabilistic assignment based on aggregate data. While these approaches are carefully designed and validated, they introduce uncertainty that future research could reduce through integration with additional data sources.

The empirical analyses focus on specific regions and time periods, and findings may not fully generalize to regions with different travel patterns, land-use structures, or policy environments. In addition, the emissions analysis considers tailpipe CO_2 emissions and does not account for upstream emissions or non-exhaust pollutants.

This dissertation opens several promising avenues for future research. Further the integration of MDLD with surveys, administrative records, and emerging connected vehicle data could enhance demographic and vehicle attribution. Extending the analyses to longer post-pandemic horizons would improve understanding of the durability of observed behavioral changes.

Future work could also leverage MDLD to study route choice, real-time responses to information, and traffic propagation dynamics. Expanding the emission framework to include upstream emissions, criteria pollutants, and exposure-based metrics would further strengthen its policy relevance. Finally, applying the methods developed here to additional metropolitan regions would help assess their generalizability and inform broader transportation and climate policy debates.

Reference

[1] F. H. Administration. *2022 NextGen National Household Travel Survey Core Data*,

[2] MWCOG, "2022 State of the Commute Survey Report," Metropolitan Washington Council of Governments, Commuter Connections, 2023. [Online]. Available: https://www.mwcog.org/file.aspx?D=HcOqbzivuFayTfyAlhvUJhe72nkkosOrz2TZl%2bOlFXE%3d&A=3b5jlNJv7k8i9DmLKmqJ5c9bgLZ451b3R0E2zs1pReQ%3d

[3] P. Stopher, Y. Zhang, J. Armoogum, and J.-L. Madre, "National household travel surveys: The case for Australia," in *34th Australasian Transport Research Forum (ATRF), Adelaide, South Australia*, 2011: Citeseer.

[4] S. Hong, F. Zhao, V. Livshits, S. Gershenfeld, J. Santos, and M. Ben-Akiva, "Insights on data quality from a large-scale application of smartphone-based travel survey technology in the Phoenix metropolitan area, Arizona, USA," *Transportation Research Part A: Policy and Practice,* vol. 154, pp. 413-429, 2021.

[5] P. R. Stopher, *Methods for household travel surveys*. Transportation Research Board, 1996.

[6] B. Pan, "Online travel surveys and response patterns," *Journal of Travel Research,* vol. 49, no. 1, pp. 121-135, 2010.

[7] A. Frei, T. Kuhnimhof, and K. W. Axhausen, "Long distance travel in Europe today: Experiences with a new survey," *Arbeitsberichte Verkehrs-und Raumplanung,* vol. 569, 2009.

[8] S. W. Lau, "Truck Travel Surveys: A review of the Literature and State-of-the-Art," 1995.

[9] H.-L. Hwang and J. Rollow, "Data processing procedures and methodology for estimating trip distances for the 1995 American Travel Survey (ATS)," Oak Ridge National Laboratory. Center for Transportation Analysis, 2000.

[10] M. Clarke, M. Dix, and P. Jones, "Error and uncertainty in travel surveys," *Transportation,* vol. 10, no. 2, pp. 105-126, 1981.

[11] A. Richardson, E. Ampt, and A. Meyburg, "Nonresponse issues in household travel surveys," in *Conference proceedings*, 1996, vol. 10, pp. 79-114.

[12] P. R. Stopher, K. Kockelman, S. P. Greaves, and E. Clifford, "Reducing burden and sample sizes in multiday household travel surveys," *Transportation Research Record,* vol. 2064, no. 1, pp. 12-18, 2008.

[13] M. Yang *et al.*, "Big-Data Driven Framework to Estimate Vehicle Volume Based on Mobile Device Location Data," *Transportation Research Record,* p. 03611981231174240, 2023.

[14] S. Hu, C. Xiong, P. Chen, and P. Schonfeld, "Examining nonlinearity in population inflow estimation using big data: An empirical comparison of explainable machine learning models," *Transportation Research Part A: Policy and Practice,* vol. 174, p. 103743, 2023.

[15] C. Ratti, D. Frenchman, R. M. Pulselli, and S. Williams, "Mobile landscapes: using location data from cell phones for urban analysis," *Environment and planning B: Planning and design,* vol. 33, no. 5, pp. 727-748, 2006.

[16] D. Bachir, G. Khodabandelou, V. Gauthier, M. El Yacoubi, and J. Puchinger, "Inferring dynamic origin-destination flows by transport mode using mobile phone data," *Transportation Research Part C: Emerging Technologies,* vol. 101, pp. 254-275, 2019.
[17] D. Ashbrook and T. Starner, "Using GPS to learn significant locations and predict movement across multiple users," *Personal and Ubiquitous computing,* vol. 7, pp. 275-286, 2003.
[18] M. Fontaine, S. Olariu, and M. Weigle, "Traffic monitoring," *Vehicular Networks from Theory to Practice,* 2009.
[19] N. Caceres, L. Romero, and F. G. Benitez, "Exploring strengths and weaknesses of mobility inference from mobile phone data vs. travel surveys," *Transportmetrica A: Transport Science,* vol. 16, no. 3, pp. 574-601, 2020.
[20] G. Zhao, M. B. Al-Khasawneh, T. Tuoto, and C. Cirillo, "Data Fusion for Travel Analysis: Linking Travel Survey and Mobile Device Location Data," 2024.
[21] G. Leduc, "Road traffic data: Collection methods and applications," *Working Papers on Energy, Transport and Climate Change,* vol. 1, no. 55, pp. 1-55, 2008.
[22] N. McGuckin and A. Fucci, "Summary of travel trends: 2017 national household travel survey," United States. Department of Transportation. Federal Highway Administration, 2018.
[23] P. Chen, S. Hu, Q. Shen, H. Lin, and C. Xie, "Estimating traffic volume for local streets with imbalanced data," *Transportation research record,* vol. 2673, no. 3, pp. 598-610, 2019.
[24] L. Zhang and X. He, "Feasibility and advantages of estimating local road vehicle miles traveled on basis of global positioning system travel data," *Transportation research record,* vol. 2399, no. 1, pp. 94-102, 2013.
[25] D. P. Wagner, D. M. Neumeister, and E. Murakami, "Global positioning systems for personal travel surveys," in *National Traffic Data Acquisition Conference (1996: Albuquerque NM) National Traffic Data Acquisition Conference (NATDAC'96): proceedings Vol. 1*, 1996.
[26] M. Ojah and D. Pearson, "Austin/San Antonio GPS-Enhanced Household Travel Survey," *Texas Transportation Institute,* 2008.
[27] J. Wolf and M. Lee, "Synthesis of and statistics for recent GPS-enhanced travel surveys," in *Proceedings of the International Conference on Survey Methods in Transport: Harmonization and Data Comparability, International Steering Committee for Travel Survey Conferences. Annecy, France*, 2008.
[28] P. Frank, "Chicago regional household travel inventory," 2010.
[29] P. NuStats, "Regional travel survey: final report," *Atlanta Regional Commission, Atlanta,* 2011.
[30] H. Safi, B. Assemi, M. Mesbah, L. Ferreira, and M. Hickman, "Design and implementation of a smartphone-based travel survey," *Transportation Research Record,* vol. 2526, no. 1, pp. 99-107, 2015.
[31] A. Haghani, M. Hamedi, and K. F. Sadabadi, "I-95 Corridor coalition vehicle probe project: Validation of INRIX data," *I-95 Corridor Coalition,* vol. 9, 2009.
[32] D. Schrank, B. Eisele, and T. Lomax, "2014 Urban mobility report: powered by Inrix Traffic Data," 2015.

[33] C. Chen, J. Ma, Y. Susilo, Y. Liu, and M. Wang, "The promises of big data and small data for travel behavior (aka human mobility) analysis," *Transportation research part C: emerging technologies,* vol. 68, pp. 285-299, 2016.
[34] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *nature,* vol. 453, no. 7196, pp. 779-782, 2008.
[35] C. Kang, Y. Liu, X. Ma, and L. Wu, "Towards estimating urban population distributions from mobile call data," *Journal of Urban Technology,* vol. 19, no. 4, pp. 3-21, 2012.
[36] F. Calabrese, M. Diao, G. Di Lorenzo, J. Ferreira Jr, and C. Ratti, "Understanding individual mobility patterns from urban sensing data: A mobile phone trace example," *Transportation research part C: emerging technologies,* vol. 26, pp. 301-313, 2013.
[37] J. Leber, "How wireless carriers are monetizing your movements," *URL: https://www. technologyreview. com/s/513016/how-wireless-carriers-are-monetizing-your-movements,* 2013.
[38] C. J. Riederer, S. Zimmeck, C. Phanord, A. Chaintreau, and S. M. Bellovin, "" I don't have a photograph, but you can have my footprints." Revealing the Demographics of Location Data," in *Proceedings of the 2015 ACM on Conference on Online Social Networks*, 2015, pp. 185-195.
[39] C. P. Barros and J. Prieto-Rodriguez, "A revenue-neutral tax reform to increase demand for public transport services," *Transportation Research Part A: Policy and Practice,* vol. 42, no. 4, pp. 659-672, 2008.
[40] M. B. Rojas IV, E. Sadeghvaziri, and X. Jin, "Comprehensive review of travel behavior and mobility pattern studies that used mobile phone data," *Transportation Research Record,* vol. 2563, no. 1, pp. 71-79, 2016.
[41] R. Su, E. C. McBride, and K. G. Goulias, "Pattern recognition of daily activity patterns using human mobility motifs and sequence analysis," *Transportation Research Part C: Emerging Technologies,* vol. 120, p. 102796, 2020.
[42] R. Amoh-Gyimah and E. N. Aidoo, "Mode of transport to work by government employees in the Kumasi metropolis, Ghana," *Journal of Transport Geography,* vol. 31, pp. 35-43, 2013.
[43] R. Hawkins and P. R. Stopher, "Collecting data with GPS: Those who reject, and those who receive," 2004.
[44] P. R. Stopher and S. P. Greaves, "Household travel surveys: Where are we going?," *Transportation Research Part A: Policy and Practice,* vol. 41, no. 5, pp. 367-381, 2007.
[45] M. Janzen, "Population synthesis for long-distance travel demand simulations," in *6th symposium of the European association for research in transportation (hEART 2017)*, 2017: ETH Zurich, Institute for Transport Planning and Systems.
[46] A. Bwambale, C. F. Choudhury, S. Hess, and M. S. Iqbal, "Getting the best of both worlds: a framework for combining disaggregate travel survey data and aggregate mobile phone data for trip generation modelling," *Transportation,* vol. 48, pp. 2287-2314, 2021.
[47] M. Flamm and V. Kaufmann, "The concept of network of usual places as a tool for analysing human activity spaces," in *11th World conference on transport research Berkeley*, 2007, pp. 1-28.
[48] L. Gong, T. Morikawa, T. Yamamoto, and H. Sato, "Deriving personal trip data from GPS data: A literature review on the existing methodologies," *Procedia-Social and Behavioral Sciences,* vol. 138, pp. 557-565, 2014.

[49] P. R. Stopher, Q. Jiang, and C. FitzGerald, "Processing GPS data from travel surveys," *2nd international colloqium on the behavioural foundations of integrated land-use and transportation models: frameworks, models and applications, Toronto,* 2005.

[50] W. Chen, M. Ji, and J. Wang, "T-DBSCAN: A Spatiotemporal Density Clustering for GPS Trajectory Segmentation," *International Journal of Online Engineering,* vol. 10, no. 6, 2014.

[51] C. Zhou, D. Frankowski, P. Ludford, S. Shekhar, and L. Terveen, "Discovering personally meaningful places: An interactive clustering approach," *ACM Transactions on Information Systems (TOIS),* vol. 25, no. 3, pp. 12-es, 2007.

[52] S. Phithakkitnukoon, T. Horanont, G. Di Lorenzo, R. Shibasaki, and C. Ratti, "Activity-aware map: Identifying human daily activity pattern using mobile phone data," in *Human Behavior Understanding: First International Workshop, HBU 2010, Istanbul, Turkey, August 22, 2010. Proceedings 1*, 2010: Springer, pp. 14-25.

[53] K. Xie, K. Deng, and X. Zhou, "From trajectories to activities: a spatio-temporal join approach," in *Proceedings of the 2009 International Workshop on Location Based Social Networks*, 2009, pp. 25-32.

[54] L. Huang, Q. Li, and Y. Yue, "Activity identification from GPS trajectories using spatial temporal POIs' attractiveness," in *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on location based social networks*, 2010, pp. 27-30.

[55] L. Pappalardo, L. Ferres, M. Sacasa, C. Cattuto, and L. Bravo, "Evaluation of home detection algorithms on mobile phone data using individual-level ground truth," *EPJ data science,* vol. 10, no. 1, p. 29, 2021.

[56] R. Ahas, S. Silm, O. Järv, E. Saluveer, and M. Tiru, "Using mobile positioning data to model locations meaningful to users of mobile phones," *Journal of urban technology,* vol. 17, no. 1, pp. 3-27, 2010.

[57] F. Calabrese, G. Di Lorenzo, L. Liu, and C. Ratti, "Estimating Origin-Destination flows using opportunistically collected mobile phone location data from one million users in Boston Metropolitan Area," 2011.

[58] L. Alexander, S. Jiang, M. Murga, and M. C. González, "Origin–destination trips by purpose and time of day inferred from mobile phone data," *Transportation research part c: emerging technologies,* vol. 58, pp. 240-250, 2015.

[59] Y. Pan *et al.*, "Residency and worker status identification based on mobile device location data," *Transportation Research Part C: Emerging Technologies,* vol. 146, p. 103956, 2023.

[60] H. Huang, Y. Cheng, and R. Weibel, "Transport mode detection based on mobile phone network data: A systematic review," *Transportation Research Part C: Emerging Technologies,* vol. 101, pp. 297-312, 2019.

[61] P. Stopher, C. FitzGerald, and J. Zhang, "Search for a global positioning system device to measure person travel," *Transportation Research Part C: Emerging Technologies,* vol. 16, no. 3, pp. 350-369, 2008.

[62] K. W. Axhausen, S. Schönfelder, J. Wolf, M. Oliveira, and U. Samaga, "80 weeks of GPS-traces: Approaches to enriching the trip information," *Arbeitsberichte Verkehrs-Und Raumplanung,* vol. 178, 2003.

[63] N. Schuessler and K. W. Axhausen, "Processing raw data from global positioning systems without additional information," *Transportation Research Record,* vol. 2105, no. 1, pp. 28-36, 2009.

[64] L. Gong, H. Sato, T. Yamamoto, T. Miwa, and T. Morikawa, "Identification of activity stop locations in GPS trajectories by density-based clustering method combined with support vector machines," *Journal of Modern Transportation,* vol. 23, pp. 202-213, 2015.

[65] C. Zhou, H. Jia, Z. Juan, X. Fu, and G. Xiao, "A data-driven method for trip ends identification using large-scale smartphone-based GPS tracking data," *IEEE Transactions on Intelligent Transportation Systems,* vol. 18, no. 8, pp. 2096-2110, 2016.

[66] L. Gong, T. Yamamoto, and T. Morikawa, "Identification of activity stop locations in GPS trajectories by DBSCAN-TE method combined with support vector machines," *Transportation research procedia,* vol. 32, pp. 146-154, 2018.

[67] Y. Ye, Y. Zheng, Y. Chen, J. Feng, and X. Xie, "Mining individual life pattern based on location history," in *2009 tenth international conference on mobile data management: Systems, services and middleware*, 2009: IEEE, pp. 1-10.

[68] Z. Yao, F. Yang, Y. Guo, P. J. Jin, and Y. Li, "Trip end identification based on spatial-temporal clustering algorithm using smartphone positioning data," *Expert Systems with Applications,* vol. 197, p. 116734, 2022.

[69] F. Wang, J. Wang, J. Cao, C. Chen, and X. J. Ban, "Extracting trips from multi-sourced data for mobility pattern analysis: An app-based data example," *Transportation Research Part C: Emerging Technologies,* vol. 105, pp. 183-202, 2019.

[70] H. Gong, C. Chen, E. Bialostozky, and C. T. Lawson, "A GPS/GIS method for travel mode detection in New York City," *Computers, environment and urban systems,* vol. 36, no. 2, pp. 131-139, 2012.

[71] L. Stenneth, O. Wolfson, P. S. Yu, and B. Xu, "Transportation mode detection using mobile phones and GIS information," in *Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems*, 2011, pp. 54-63.

[72] G. Xiao, Z. Juan, and C. Zhang, "Travel mode detection based on GPS track data and Bayesian networks," *Computers, environment and urban systems,* vol. 54, pp. 14-22, 2015.

[73] M. A. Shafique and E. Hato, "Travel mode detection with varying smartphone data collection frequencies," *Sensors,* vol. 16, no. 5, p. 716, 2016.

[74] B. Wang, L. Gao, and Z. Juan, "Travel mode detection using GPS data and socioeconomic attributes based on a random forest classifier," *IEEE Transactions on Intelligent Transportation Systems,* vol. 19, no. 5, pp. 1547-1558, 2017.

[75] R. Brunauer, M. Hufnagl, K. Rehrl, and A. Wagner, "Motion pattern analysis enabling accurate travel mode detection from GPS data only," in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, 2013: IEEE, pp. 404-411.

[76] M. Yang, Y. Pan, A. Darzi, S. Ghader, C. Xiong, and L. Zhang, "A data-driven travel mode share estimation framework based on mobile device location data," *Transportation,* vol. 49, no. 5, pp. 1339-1383, 2022.

[77] M. H. Nguyen, J. Armoogum, J.-L. Madre, and C. Garcia, "Reviewing trip purpose imputation in GPS-based travel surveys," *Journal of Traffic and Transportation Engineering (English Edition),* vol. 7, no. 4, pp. 395-412, 2020.

[78] L. Shen and P. R. Stopher, "A process for trip purpose imputation from Global Positioning System data," *Transportation Research Part C: Emerging Technologies,* vol. 36, pp. 261-267, 2013.

[79] J. Wolf, R. Guensler, and W. Bachman, "Elimination of the travel diary: Experiment to derive trip purpose from global positioning system travel data," *Transportation Research Record,* vol. 1768, no. 1, pp. 125-134, 2001.
[80] J. Wolf, S. Schönfelder, U. Samaga, M. Oliveira, and K. W. Axhausen, "Eighty weeks of global positioning system traces: approaches to enriching trip information," *Transportation Research Record,* vol. 1870, no. 1, pp. 46-54, 2004.
[81] Y. Cui, C. Meng, Q. He, and J. Gao, "Forecasting current and next trip purpose with social media data and Google places," *Transportation Research Part C: Emerging Technologies,* vol. 97, pp. 159-174, 2018.
[82] G. Xiao, Z. Juan, and C. Zhang, "Detecting trip purposes from smartphone-based travel surveys with artificial neural networks and particle swarm optimization," *Transportation Research Part C: Emerging Technologies,* vol. 71, pp. 447-463, 2016.
[83] L. Montini, N. Rieser-Schüssler, A. Horni, and K. W. Axhausen, "Trip purpose identification from GPS tracks," *Transportation Research Record,* vol. 2405, no. 1, pp. 16-23, 2014.
[84] Z. Deng and M. Ji, "Deriving rules for trip purpose identification from GPS travel survey data and land use data: A machine learning approach," in *Traffic and Transportation Studies 2010*, 2010, pp. 768-777.
[85] L. Zhang *et al.*, "Next generation National Household Travel Survey National Origin Destination Data Passenger Origin-Destination Data Methodology Documentation," *Federal Highway Administration.[Google Scholar],* 2021.
[86] D. Pickrell , D. Pace , J. Wishart , and C. Roycroft "Development of VMT Forecasting Models for Use by the Federal Highway Administration," ed: Federal Highway Administration, 2023.
[87] A. E. Lee and S. L. Handy, "Leaving level-of-service behind: The implications of a shift to VMT impact metrics," *Research in Transportation Business & Management,* vol. 29, pp. 14-25, 2018.
[88] Caltrans, "Vehicle Miles Traveled-Focused Transportation Impact Study Guide," ed: Caltrans, 2020.
[89] R. Ewing and R. Cervero, "“Does compact development make people drive less?” the answer is yes," *Journal of the American Planning Association,* vol. 83, no. 1, pp. 19-25, 2017.
[90] J. Yeo, S. Park, and K. Jang, "Effects of urban sprawl and vehicle miles traveled on traffic fatalities," *Traffic injury prevention,* vol. 16, no. 4, pp. 397-403, 2015.
[91] J. E. Hughes, D. Kaffine, and L. Kaffine, "Decline in traffic congestion increased crash severity in the wake of COVID-19," *Transportation research record,* vol. 2677, no. 4, pp. 892-903, 2023.
[92] K. A. Small and K. V. Dender, "Fuel efficiency and motor vehicle travel: the declining rebound effect," *The energy journal,* vol. 28, no. 1, pp. 25-52, 2007.
[93] C. McCahill, "The Amount We Drive Could Make or Break Clean Energy Plans," State Smart Transportation Initiative 2021. [Online]. Available: https://bit.ly/3ifHouH
[94] T. Klein, "What U.S. Cities Are Doing to Decarbonize Transport: Focus on Fuels & VMT Reduction," Future Fuel Strategies, 2019. [Online]. Available: https://bit.ly/3jOiRM0

[95] J. Nakamya, E. Moons, S. Koelet, and G. Wets, "Impact of data integration on some important travel behavior indicators," *Transportation research record,* vol. 1993, no. 1, pp. 89-94, 2007.
[96] D. Deka and D. Fei, "A comparison of the personal and neighborhood characteristics associated with ridesourcing, transit use, and driving with NHTS data," *Journal of Transport Geography,* vol. 76, pp. 24-33, 2019.
[97] W. Zhang and M. Zhang, "Short-and long-term effects of land use on reducing personal vehicle miles of travel: Longitudinal multilevel analysis in Austin, Texas," *Transportation Research Record,* vol. 2500, no. 1, pp. 102-109, 2015.
[98] D. Salon, M. G. Boarnet, S. Handy, S. Spears, and G. Tal, "How do local actions affect VMT? A critical review of the empirical evidence," *Transportation research part D: transport and environment,* vol. 17, no. 7, pp. 495-508, 2012.
[99] D. Heres-Del-Valle and D. Niemeier, "CO2 emissions: Are land-use changes enough for California to reduce VMT? Specification of a two-part model with instrumental variables," *Transportation Research Part B: Methodological,* vol. 45, no. 1, pp. 150-161, 2011.
[100] R. Li and K. M. Kockelman, "Predicting a Vehicle’s Distance Traveled from Short-duration Data," *Findings,* 2019.
[101] J. Fan, C. Fu, K. Stewart, and L. Zhang, "Using big GPS trajectory data analytics for vehicle miles traveled estimation," *Transportation research part C: emerging technologies,* vol. 103, pp. 298-307, 2019.
[102] A. Nasri *et al.*, "Advanced vehicle miles traveled estimation methods for non-federal aid system roadways using GPS vehicle trajectory data and statistical power analysis," *Transportation research record,* vol. 2673, no. 11, pp. 296-308, 2019.
[103] S. Gurram, V. Sivaraman, J. T. Apple, and A. R. Pinjari, "Agent-based modeling to simulate road travel using Big Data from smartphone GPS: An application to the continental United States," in *2019 IEEE International Conference on Big Data (Big Data)*, 2019: IEEE, pp. 3553-3562.
[104] Y. Pan *et al.*, "National-Level Multimodal Origin–Destination Estimation Based on Passively Collected Location Data and Machine Learning Methods," *Transportation Research Record,* vol. 2678, no. 5, pp. 525-541, 2024.
[105] Q. Su, "How does fuel economy of vehicles affect urban motor vehicle travel in the USA?," *Energy Efficiency,* vol. 8, pp. 339-351, 2015.
[106] P. Barla, B. Lamonde, L. F. Miranda-Moreno, and N. Boucher, "Traveled distance, stock and fuel efficiency of private vehicles in Canada: price elasticities and rebound effect," *Transportation,* vol. 36, pp. 389-402, 2009.
[107] C. N. Hatz II, "Panel data analysis of fuel price elasticities to vehicle-miles traveled for first year participants of the national evaluation of a mileage-based road user charge study," The University of Iowa, 2011.
[108] T. Wang and C. Chen, "Impact of fuel price on vehicle miles traveled (VMT): do the poor respond in the same way as the rich?," *Transportation,* vol. 41, pp. 91-105, 2014.
[109] M. Taiebat, S. Stolper, and M. Xu, "Forecasting the impact of connected and automated vehicles on energy use: a microeconomic study of induced travel and energy rebound," *Applied Energy,* vol. 247, pp. 297-308, 2019.

[110] R. Chapleau, P. Gaudette, and T. Spurr, "Strict and deep comparison of revealed transit trip structure between computer-assisted telephone interview household travel survey and smart cards," *Transportation research record,* vol. 2672, no. 42, pp. 13-22, 2018.

[111] H. S. Dillon, J.-D. Saphores, and M. G. Boarnet, "The impact of urban form and gasoline prices on vehicle usage: Evidence from the 2009 National Household Travel Survey," *Research in Transportation Economics,* vol. 52, pp. 23-33, 2015.

[112] F. Goetzke and C. Vance, *Is gasoline price elasticity in the United States increasing? Evidence from the 2009 and 2017 national household travel surveys* (no. 765). Ruhr Economic Papers, 2018.

[113] T. P. Wenzel, "Elasticity of Vehicle Miles of Travel to Changes in the Price of Gasoline and the Cost of Driving in Texas," 2018.

[114] K. Gillingham, "Identifying the elasticity of driving: evidence from a gasoline price shock in California," *Regional Science and Urban Economics,* vol. 47, pp. 13-24, 2014.

[115] S. L. Puller and L. A. Greening, "Household adjustment to gasoline price change: an analysis using 9 years of US survey data," *Energy economics,* vol. 21, no. 1, pp. 37-52, 1999.

[116] K. Gillingham, A. Jenn, and I. M. Azevedo, "Heterogeneity in the response to gasoline prices: Evidence from Pennsylvania and implications for the rebound effect," *Energy Economics,* vol. 52, pp. S41-S52, 2015.

[117] L. T. Burra, C. Shen, A. Alberini, and C. Cirillo, "Do households substitute mileage across their cars?," *Transportation Research Part D: Transport and Environment,* vol. 123, p. 103914, 2023.

[118] A. Bick, A. Blandin, and K. Mertens, "Work from home before and after the COVID-19 outbreak," *American Economic Journal: Macroeconomics,* vol. 15, no. 4, pp. 1-39, 2023.

[119] G. Buomprisco, S. Ricci, R. Perri, and S. De Sio, "Health and telework: New challenges after COVID-19 pandemic," *European Journal of Environment and Public Health,* vol. 5, no. 2, p. em0073, 2021.

[120] J. S. F. Chow, D. Palamidas, S. Marshall, W. Loomes, S. Snook, and R. Leon, "Teleworking from home experiences during the COVID-19 pandemic among public health workers (TelEx COVID-19 study)," *BMC Public Health,* vol. 22, no. 1, p. 674, 2022.

[121] MWCOG, "2019 State of the Commute Survey Report," Metropolitan Washington Council of Governments, Commuter Connections, 2020. [Online]. Available: https://www.mwcog.org/file.aspx?D=HcOqbzivuFayTfyAlhvUJhe72nkkosOrz2TZl%2bOlFXE%3d&A=%2bkIjc%2fnIQiqtav9hkV%2b7cN%2fnZ1nVfMkbtPLYAPGMWIU%3d

[122] M. S. Nemțeanu, D. C. Dabija, and L. Stanca, "The influence of teleworking on performance and employees' counterproductive behaviour," *Amfiteatru Economic,* vol. 23, no. 58, pp. 601-619, 2021.

[123] J.-V. Alipour, H. Fadinger, and J. Schymik, "My home is my castle–The benefits of working from home during a pandemic crisis," *Journal of Public Economics,* vol. 196, p. 104373, 2021.

[124] C. Borkowski and R. Kaynas, "Telework trends," in "Beyond the Numbers: Employment & Unemployment," U.S. Bureau of Labor Statistics, 2025, vol. vol. 14 / no. 2. [Online]. Available: https://www.bls.gov/opub/btn/volume-14/telework-trends.htm

[125] F. Wöhner, "Work flexibly, travel less? The impact of telework and flextime on mobility behavior in Switzerland," *Journal of transport geography,* vol. 102, p. 103390, 2022.
[126] N. Bloom, "Hybrid is the future of work," *Stanford Institute for Economic Policy Research (SIEPR): Stanford, CA, USA,* 2021.
[127] P.-H. Wong, K. Kourtit, and P. Nijkamp, "The teleworking paradox: The geography of residential mobility of workers in pandemic times," *The Annals of Regional Science,* vol. 74, no. 1, p. 39, 2025.
[128] D. A. Hensher, M. J. Beck, J. D. Nelson, and C. Balbontin, "Reducing Congestion and Crowding with Working from Home," in *Transport and Pandemic Experiences*, vol. 17: Emerald Publishing Limited, 2022, pp. 235-255.
[129] M. J. Delventhal, E. Kwon, and A. Parkhomenko, "JUE Insight: How do cities change when we work from home?," *Journal of Urban Economics,* vol. 127, p. 103331, 2022.
[130] G. Currie, T. Jain, and L. Aston, "Evidence of a post-COVID change in travel behaviour–Self-reported expectations of commuting in Melbourne," *Transportation Research Part A: Policy and Practice,* vol. 153, pp. 218-234, 2021.
[131] C. Rojas *et al.*, "Short run "rebound effect" of COVID on the transport carbon footprint," *Cities,* vol. 131, p. 104039, 2022.
[132] I. Duncan, L. Lazo, J. George, and M. Laris, "New habits are making more commutes miserable," in *The Washington Post*, ed. Transportation, 2023.
[133] O. o. B. a. Policy and F. T. Administration, "Effects of the COVID-19 Pandemic on Transit Ridership and Accessibility," 2024. [Online]. Available: https://www.transit.dot.gov/research-innovation/effects-covid-19-pandemic-transit-ridership-and-accessibility-report-0268
[134] S. Bricka, T. Reuscher, P. Schroeder, M. Fisher, J. Beard, and X. L. Sun, "Summary of Travel Trends: 2022 National Household Travel Survey," United States. Federal Highway Administration. Office of Policy and …, 2024.
[135] B. P. Loo and Z. Huang, "Spatio-temporal variations of traffic congestion under work from home (WFH) arrangements: Lessons learned from COVID-19," *Cities,* vol. 124, p. 103610, 2022.
[136] P. Xu, W. Li, X. Hu, H. Wu, and J. Li, "Spatiotemporal analysis of urban road congestion during and post COVID-19 pandemic in Shanghai, China," *Transportation research interdisciplinary perspectives,* vol. 13, p. 100555, 2022.
[137] D. L. Schrank, L. Albert, K. Jha, and W. L. Eisele, "2025 Urban Mobility Report," in "Urban Mobility Report (TTI)," Texas A & M Transportation Institute College Station, Texas 2025. [Online]. Available: https://static.tti.tamu.edu/tti.tamu.edu/documents/mobility-report-2025.pdf
[138] D. L. Schrank, L. Albert, K. Jha, and W. L. Eisele, "2023 Urban Mobility Report," in "Urban Mobility Report (TTI)," Texas A & M Transportation Institute, College Station, Texas, 2024. [Online]. Available: https://static.tti.tamu.edu/tti.tamu.edu/documents/umr/archive/mobility-report-2023.pdf
[139] J. Anable, L. Brown, I. Docherty, and G. Marsden, "Less is more: Changing travel in a post-pandemic society," *CREDS Report Available online. last accessed 18th April,* 2022.
[140] EPA, "Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2022," U.S. Environmental Protection Agency, 2024. [Online]. Available: https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks-1990-2022

[141] B. Girod, D. P. van Vuuren, and B. de Vries, "Influence of travel behavior on global CO2 emissions," *Transportation Research Part A: Policy and Practice,* vol. 50, pp. 183-197, 2013.
[142] E. J. Wilson, R. Wilson, and K. J. Krizek, "The implications of school choice on travel behavior and environmental emissions," *Transportation Research Part D: Transport and Environment,* vol. 12, no. 7, pp. 506-518, 2007.
[143] G. Marsden, C. Mullen, I. Bache, I. Bartle, and M. Flinders, "Carbon reduction and travel behaviour: Discourses, disputes and contradictions in governance," *Transport Policy,* vol. 35, pp. 71-78, 2014.
[144] U. EPA, "Overview of EPA's MOtor Vehicle Emission Simulator (MOVES5)," ed: US Environmental Protection Agency: Washington, DC, USA, 2025.
[145] EPA, *MOVES5 Technical Guidance: Using MOVES to Prepare Emission Inventories for State Implementation Plans and Transportation Conformity*. United States Environmental Protection Agency, 2024.
[146] EPA, "MOVES Greenhouse Gas Guidance: Using MOVES for Estimating State and Local Inventories of Onroad Greenhouse Gas Emissions and Energy Consumption," vol. EPA-420-B-24-023, ed, 2024.
[147] Y. Guo, Z. Chen, A. Stuart, X. Li, and Y. Zhang, "A systematic overview of transportation equity in terms of accessibility, traffic emissions, and safety outcomes: From conventional to emerging technologies," *Transportation research interdisciplinary perspectives,* vol. 4, p. 100091, 2020.
[148] E. Feitelson, "Introducing environmental equity dimensions into the sustainable transport discourse: issues and pitfalls," *Transportation Research Part D: Transport and Environment,* vol. 7, no. 2, pp. 99-118, 2002.
[149] A. Karner, J. London, D. Rowangould, and K. Manaugh, "From transportation equity to transportation justice: within, through, and beyond the state," *Journal of planning literature,* vol. 35, no. 4, pp. 440-459, 2020.
[150] C. Beckx, L. I. Panis, J. Vankerkom, D. Janssens, G. Wets, and T. Arentze, "An integrated activity-based modelling framework to assess vehicle emissions: approach and application," *Environment and Planning B: Planning and Design,* vol. 36, no. 6, pp. 1086-1102, 2009.
[151] J. Y. Hao, M. Hatzopoulou, and E. J. Miller, "Integrating an activity-based travel demand model with dynamic traffic assignment and emission models: Implementation in the Greater Toronto, Canada, area," *Transportation Research Record,* vol. 2176, no. 1, pp. 1-13, 2010.
[152] W. Poortinga, A. Spence, C. Demski, and N. F. Pidgeon, "Individual-motivational factors in the acceptability of demand-side and supply-side measures to reduce carbon emissions," *Energy Policy,* vol. 48, pp. 812-819, 2012.
[153] M. Ben-Akiva *et al.*, "The behavioral dimension of transport decarbonization," *npj Sustainable Mobility and Transport,* vol. 3, no. 1, p. 3, 2026.
[154] A. Mathez, K. Manaugh, V. Chakour, A. El-Geneidy, and M. Hatzopoulou, "How can we alter our carbon footprint? Estimating GHG emissions based on travel survey information," *Transportation,* vol. 40, no. 1, pp. 131-149, 2013.
[155] Y. Yang, C. Wang, and W. Liu, "Urban daily travel carbon emissions accounting and mitigation potential analysis using surveyed individual data," *Journal of Cleaner Production,* vol. 192, pp. 821-834, 2018.

[156] Q. Qiao, F. Zhao, Z. Liu, X. He, and H. Hao, "Life cycle greenhouse gas emissions of Electric Vehicles in China: Combining the vehicle cycle and fuel cycle," *Energy,* vol. 177, pp. 222-233, 2019.
[157] Y. Li, N. Ha, and T. Li, "Research on carbon emissions of electric vehicles throughout the life cycle assessment taking into vehicle weight and grid mix composition," *Energies,* vol. 12, no. 19, p. 3612, 2019.
[158] J. Zielinski, R. Andreucci, N. Rajagopalan, and C. B. Aktas, "Prospects for meeting the corporate average fuel economy standards in the US," *Resources, Conservation and Recycling,* vol. 136, pp. 466-472, 2018.
[159] N. R. Council *et al.*, *Effectiveness and impact of corporate average fuel economy (CAFE) standards*. National Academies Press, 2002.
[160] S. Zhu, *Integration of commercial vehicle GPS and roadside intercept survey data*. University of Toronto (Canada), 2017.
[161] M. Fleming, B. Kirby, and K. I. Penny, "Record linkage in Scotland and its applications to health research," *Journal of clinical nursing,* vol. 21, no. 19pt20, pp. 2711-2721, 2012.
[162] A. Sayers, Y. Ben-Shlomo, A. W. Blom, and F. Steele, "Probabilistic record linkage," *International journal of epidemiology,* vol. 45, no. 3, pp. 954-964, 2016.
[163] T. Enamorado, B. Fifield, and K. Imai, "Using a probabilistic model to assist merging of large-scale administrative records," *American Political Science Review,* vol. 113, no. 2, pp. 353-371, 2019.
[164] M. A. Jaro, "Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida," *Journal of the American Statistical Association,* vol. 84, no. 406, pp. 414-420, 1989.
[165] J. de Bruin, "Record Linkage Toolkit Documentation," 2022.
[166] I. P. Fellegi and A. B. Sunter, "A theory for record linkage," *Journal of the American Statistical Association,* vol. 64, no. 328, pp. 1183-1210, 1969.
[167] G. J. Bauman, *Computation of weights for probabilistic record linkage using the EM algorithm*. Brigham Young University, 2006.
[168] M. Muller, "Dynamic time warping in information retrieval for music and motion," *Dynamic time warping Information retrieval for music and motion,* pp. 69-84, 2007.
[169] C.-C. Hung, W.-C. Peng, and W.-C. Lee, "Clustering and aggregating clues of trajectories for mining trajectory patterns and routes," *The VLDB Journal,* vol. 24, pp. 169-192, 2015.
[170] L. Sun, W. Zhou, B. Jiang, and J. Guan, "A real-time similarity measure model for multi-source trajectories," in *2017 International Conference on Computing Intelligence and Information System (CIIS)*, 2017: IEEE, pp. 257-262.
[171] X. Gong, Z. Huang, Y. Wang, L. Wu, and Y. Liu, "High-performance spatiotemporal trajectory matching across heterogeneous data sources," *Future Generation Computer Systems,* vol. 105, pp. 148-161, 2020.
[172] P. Rietveld, "Rounding of arrival and departure times in travel surveys: an interpretation in terms of scheduled activities," *Journal of transportation and statistics,* vol. 5, no. 1, pp. 71-82, 2002.
[173] F. Goetzke and C. Vance, "An increasing gasoline price elasticity in the United States?," *Energy Economics,* vol. 95, p. 104982, 2021.
[174] A. Dimitropoulos, W. Oueslati, and C. Sintek, "The rebound effect in road transport: A meta-analysis of empirical studies," *Energy Economics,* vol. 75, pp. 163-179, 2018.

[175] C. Hsiao, *Analysis of panel data* (no. 64). Cambridge university press, 2022.
[176] R. Blundell and S. Bond, "Initial conditions and moment restrictions in dynamic panel data models," *Journal of econometrics,* vol. 87, no. 1, pp. 115-143, 1998.
[177] G. Solon, S. J. Haider, and J. M. Wooldridge, "What are we weighting for?," *Journal of Human resources,* vol. 50, no. 2, pp. 301-316, 2015.
[178] J. M. Wooldridge, *Econometric analysis of cross section and panel data*. MIT press, 2010.
[179] M. Miller and A. Alberini, "Sensitivity of price elasticity of demand to aggregation, unobserved heterogeneity, price trends, and price endogeneity: Evidence from US Data," *Energy Policy,* vol. 97, pp. 235-249, 2016.
[180] U. U.S. Energy Information Administration. "What do we pay for per gallon of retail regular grade gasoline?" https://www.eia.gov/energyexplained/gasoline/factors-affecting-gasoline-prices.php (accessed.
[181] P. Zhu, "Are telecommuting and personal travel complements or substitutes?," *The Annals of Regional Science,* vol. 48, no. 2, pp. 619-639, 2012.
[182] N. C. R. T. P. B. TPB, "2017/2018 REGIONAL TRAVEL SURVEY - TECHNICAL DOCUMENTATION," 2021.
[183] EPA. "Greenhouse Gas Equivalencies Calculator - Calculations and References." https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator-calculations-and-references (accessed.
[184] H. Zhang, J. Chen, W. Li, X. Song, and R. Shibasaki, "Mobile phone GPS data in urban ride-sharing: An assessment method for emission reduction potential," *Applied Energy,* vol. 269, p. 115038, 2020.
[185] P. Zhao, M.-P. Kwan, and K. Qin, "Uncovering the spatiotemporal patterns of CO2 emissions by taxis based on Individuals' daily travel," *Journal of Transport Geography,* vol. 62, pp. 122-135, 2017.
[186] Z. Kan, L. Tang, M.-P. Kwan, and X. Zhang, "Estimating vehicle fuel consumption and emissions using GPS big data," *International journal of environmental research and public health,* vol. 15, no. 4, p. 566, 2018.