Microscopic Bus Performance Analysis Using Real-time Data in Greater Sydney

TINGSEN XIAN

Doctor of Philosophy (Engineering)

Supervisor: Dr. Emily Moylan
Associate Supervisor: Prof. John D. Nelson

A thesis submitted in fulfilment of
the requirements for the degree of
Doctor of Philosophy

School of Civil Engineering
Faculty of Engineering
The University of Sydney
Australia

1 August 2025

Statement of Originality

I hereby declare that, to the best of my knowledge, this thesis is entirely my own work and has not been previously submitted for any degree or any other purpose. I confirm that the intellectual content is the product of my own efforts, and all assistance received in preparing this thesis, as well as all sources used, have been properly acknowledged.

Signature:

Author: Tingsen (Tim) Xian

Author Attribution Statement

This thesis incorporates published and under-review journal papers, which have been included with the agreement of all co-authors. These papers are listed below in the order they appear in the text:

- Xian, T., Chin, T. K., Marks, B., Nelson, J. D., & Moylan, E. (2024). Bus arrival and departure time updates in the Greater Sydney Area. *Scientific Data*, 11(1), 1034. In Chapter 3.
- Xian, T., Nelson, J. D., & Moylan, E. (2025). High resolution bus lane performance evaluation from real time update data. *Transportation Research Interdisciplinary Perspectives*, 32, 101473. In Chapter 4.
- Xian, T., Nelson, J. D., & Moylan, E. (2025). The role of cross-traffic turns in bus performance. In Chapter 5.
- Xian, T., Nelson, J. D., & Moylan, E. (2025). Design evaluation of bus cross-traffic turn priority box. *Journal of Public Transportation (Under Review).* In Chapter 6.

My contributions to each paper include methodology, software, data analysis, investigation, and the preparation of the original draft.

In addition to the statements above, I am the corresponding author for all the published items listed. Permission to include the published material has been granted by all authors.

Signature:

s Author: Tingsen (Tim) Xian

As the supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Signature:

Supervisor: Emily Moylan

Abstract

The traditional approach to measuring bus performance involves conducting manual surveys to record bus arrival and departure times at stops, a process that is both costly and inefficient. However, with the installation of Global Positioning System (GPS) equipment on buses, real-time updates on vehicle positions and bus stop arrival times are now captured at significantly lower costs. The General Transit Feed Specification (GTFS) was developed to manage and utilize this transit data effectively, providing a structured format for public transit agencies to describe their services. Leveraging GTFS-realtime feeds from Public Transport Authorities, bus performance can now be evaluated at a second-level resolution and on a larger spatial scale.

There are several barriers to the widespread adoption of GTFS-Realtime data, such as location-specific data extensions, non-human-readable formats, and data cleaning challenges. This study addresses these obstacles by creating a dataset of actual bus stop arrival and departure times using a data pipeline specifically designed to overcome these challenges. The GTFS-Realtime trip updates are transformed into a 25-month dataset of real-time bus stop arrival and departure times for Sydney, Australia, providing the foundation for microscopic bus performance analysis.

Using GTFS-Realtime Trip Updates, the first study in this thesis assesses the impact of bus priority measures, traffic signals, and cross-traffic turns (left turns in right-hand drive countries) on bus reliability and delays. Traditionally, bus priority measures, such as bus lanes, have been evaluated at an aggregate level. In this study, we employ panel regression models to examine marginal delays across stop-to-stop segments, incorporating variables such as traffic signals, traffic volumes, priority measures, cross-traffic turns, precipitation, and the Coronavirus Disease 2019 (COVID) stringency index.

The model findings show that bus-taxi and bus-HOV lanes reduce stop-to-stop marginal delays and significantly improve service reliability by reducing delay variability. Additionally, traffic signals and cross-traffic turns are shown to substantially impact bus performance. Despite being historically avoided in bus route design, cross-traffic turns continue to contribute significantly to delays and service variability.

Focusing on cross-traffic turns, the second study conducts a microscopic analysis of their effects using both GTFS-Realtime Trip Updates and Vehicle Position data. Statistical analysis reveals that cross-traffic turns not only increase mean delay and variability but also reduce operational speed. The microscopic analysis of vehicle speeds and trajectories around two intersections further illuminates the causes of these delays. Through cross-validation between GTFS-Realtime Trip Updates and Vehicle Position data, the accuracy of the proposed local-environment microscopic analysis is confirmed.

Recognizing the limitations of traditional bus priority measures, which often reduce green time or road space for other vehicles, the final study introduces an innovative solution: the bus cross-traffic turn priority box. This in-lane queue jump lane allows buses to bypass cross-traffic queues, improving both speed and reliability without compromising road space. The bus cross-traffic turn priority box also enhances the performance of general traffic by allowing vehicles to pre-accelerate before the cross-traffic signal turns green, unlike traditional bus priority measures, which often increase delays for other vehicles.

In conclusion, this thesis demonstrates the potential of utilizing GTFS-Realtime data for precise and cost-effective measurement of bus performance. The findings highlight the effectiveness of bus priority measures, such as bus-taxi and Bus-High Occupancy Vehicle (bus-HOV) lanes, in reducing delays and improving reliability. The findings also indicate that bus cross-traffic turns have a major impact by contributing to delays, reducing operational speed, and degrading service reliability. Moreover, the proposed bus cross-traffic turn priority box offers a promising solution to the challenges posed by cross-traffic turns, benefiting both buses and general traffic without increasing delays or reducing road capacity. This research provides valuable insights into the factors affecting bus performance and offers actionable recommendations to enhance bus efficiency worldwide.

Acknowledgements

I would like to begin by expressing my deepest gratitude to my supervisor, Dr. Emily Kate McNeil Moylan, whose guidance, patience, and unwavering support have been instrumental throughout both my Bachelor's and PhD studies. Her mentorship has not only shaped my academic journey but has also enriched my personal growth, for which I am deeply thankful. I also wish to extend my sincere thanks to my co-supervisor, Prof. John D. Nelson, for his continuous support and invaluable contributions during my PhD. Both of my amazing supervisors have become more than just mentors — they are true friends and a second family to me.

I am deeply grateful to Prof. Don MacKenzie from the University of Washington and Prof. Jan-Dirk Schmoecker from Kyoto University for their feedback and participation on my review committee. Their guidance has helped shape this dissertation.

I would like to acknowledge the Faculty of Engineering at The University of Sydney for providing me with the opportunity to pursue my PhD and for the financial support that made this journey possible. I also owe my gratitude to iMOVE and the Cooperative Research Centres program, an Australian Federal Government initiative, for their generous funding of my research.

Special thanks to the Aimsun support team for providing the Aimsun Next software, which was crucial to my study, and to The University of Sydney's High-Performance Computing Cluster, Artemis, for their support in providing the computing power and cloud storage necessary for my research.

I am also thankful to Prof. David Levinson, Prof. Mohsen Ramezani, and Dr. Andres Fielbaum at TransportLab for their guidance during both my Bachelor's and PhD years at The University of Sydney. Their wisdom and support have played a pivotal role in my academic development.

A heartfelt thank you to my family in Shenyang (Mukden) — my father Wei Xian, my mother Jing Xu, my sister Kexin Xian, my grandmother Fengpei Wang, my grandfather Chunbao Xian, my aunt Ying Xian, and my cousin Bo Zhu. Their constant emotional and financial support, especially during the challenging COVID period, has been a pillar of strength and encouragement throughout my studies.

I would also like to express my thanks to my uncle's family — Victor Weige Li, Avery Jinrong Yang, James Ziming Li, and Jerry Zirui Li — as well as my homestay family in Auckland — Cheryl Taylor, Lauren McGregor, Graeme McGregor, Ruby Taylor, and Trent McGregor. Their kindness and unwavering support have been invaluable to me.

I am truly grateful for the company and friendship of my friends in Sydney — Yang Gao, Wenyang Hao, Jingwei Hu, Jason Yuet Fung Ke, Bahman Lahoorpoor, Zhiyuan Lu, Zikuan Lu, Hema Rayaprolu, Junyu Shen, Changle Song, Hao Wang, Haotian Wang, Yadi Wang, Youtian Wang, Zhaohan Wang, Zhexia Wang, Bohao Wen, Zhuopeng Xie, Yue Yang, Dong Zhao, Additionally, I would like to thank my friends from Shenyang (Mukden), Auckland, Brisbane, Melbourne, and other places for their constant encouragement and support.

Finally, I am profoundly thankful to my beloved cats — Nini Xian, Momo Xian, Lala Xian, Keke Xian, Juju Xian, and Dandan Xian — for their companionship and comfort during my studies. Their presence has been a source of peace and joy. I also appreciate the friendship of my cats' friends — Naicha Wang, Xingxing Wang, Miumiu Deng — who brought extra happiness to my life.

Contents

Chapter 4 High Resolution Bus Lane Performance Evaluation from Real-time Update Data 47

Glossary

API: Application Programming Interface. 21, 23, 28, 29, 37, 40, 41, 44, 54, 73
AVL: Automatic Vehicle Location. 10, 22, 24, 47, 73, 88, 112

bus-HOV: Bus-High Occupancy Vehicle. vi, xvi, 16, 53, 57, 64, 66, 70, 111

CBD: Central Business District. 53
COVID: Coronavirus Disease 2019. v, 5, 14, 31, 52, 54, 61, 64–66, 68–70, 74, 111
CSV: Comma-Separated Values. 8, 31, 77

GPS: Global Positioning System. v, 47, 73, 118
GTFS: General Transit Feed Specification. v, vi, x, xvi–xviii, xxi, 3, 4, 8–10, 20–24, 26–38, 40, 42, 44, 45, 47, 54–56, 61, 63, 70–73, 76, 78, 79, 85, 87, 88, 108, 109, 111, 112, 116, 118, 119

HOV: High Occupancy Vehicles. 16, 57, 69, 72

MaaS: Mobility as a Service. 20

NSW: New South Wales. 26, 45, 53, 57, 58, 60, 63, 73

OLS: Ordinary Least Squares. xxi, xxii, 48, 51–53, 64, 66–68, 110
OTR: On-time Running. 23
OxCGRT: Oxford Coronavirus Government Response Tracker. 61

PTA: Public Transport Authoritie. 20, 28–30, 41
PTAs: Public Transport Authorities. 3, 20, 29, 38, 45, 108, 118

SCATS: Sydney Coordinated Adaptive Traffic System. 96, 100

SIRI: Service Interface for Real Time Information. 30, 45, 109, 116

TfNSW: Transport for New South Wales. xvi, 28–31, 42, 44, 53, 55, 57, 58, 73, 116

TSP: Transit Signal Priority. 1, 7, 17, 90, 91

URL: Uniform Resource Locator. 29

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Background

In rapidly growing urban environments, efficient and reliable public transportation systems are essential for reducing congestion, lowering emissions, and enhancing accessibility. Buses, as a major component of public transport networks, play a crucial role in connecting diverse neighborhoods, commercial hubs, and public institutions without the need for specialized infrastructure like tracks. By offering a viable alternative to private vehicles, buses help to reduce traffic volumes and minimize environmental impacts (National Association of City Transportation Officials (NACTO) 2018). However, buses often face significant operational challenges, including delays and unreliable service, which undermine the effectiveness of public transit systems.

Efficient public transportation is vital for mitigating urban congestion and promoting sustainable mobility, making cities more livable. Strategic investments in infrastructure, such as dedicated bus lanes, Transit Signal Priority (TSP) systems, and other bus priority measures, are essential for improving bus performance and service reliability (Transportation Research Board and National Academies of Sciences, Engineering, and Medicine 2013; National Association of City Transportation Officials (NACTO) 2016). These interventions significantly reduce delays, improve service punctuality, and make public transit a more attractive option for commuters. By enhancing the overall speed and reliability of bus services, these measures help to reduce reliance on private vehicles, contributing to sustainable urban mobility and the alleviation of traffic congestion.

Bus performance has been a subject of growing attention since the 1980s as public transit organizations have increasingly faced pressure to meet efficiency goals (Strathman and Hopper 1993). Performance is typically assessed through key metrics such as delays, reliability, and speed. Delays can arise from various factors, including traffic congestion, signal timing, and intersection design (Strathman and Hopper 1993; Gan et al. 2003). Reliability, referring to the consistency of bus arrivals and departures, is another critical indicator of performance (Transportation Research Board and National Academies of Sciences, Engineering, and Medicine 2013). Unpredictable delays, such as those caused by cross-traffic turns (left turns in right-hand traffic environments or right turns in left-hand traffic environments) and heavy traffic volumes, significantly reduce service reliability, making buses less attractive for commuters. Addressing these delays is crucial for increasing ridership, reducing dependence on private vehicles, and promoting environmental sustainability.

The United States Federal Transit Administration's Transit Capacity and Quality of Service Manual categorizes quality of service measurements for fixed-route buses into two main groups: availability measurements (service frequency, span, and accessibility) and comfort and convenience measurements (passenger load, travel reliability, and travel time) (Transportation Research Board and National Academies of Sciences, Engineering, and Medicine 2013). This study focuses primarily on bus delay, service reliability, and travel speed, as these metrics can be significantly improved through interventions such as optimized geometric design (Surprenant-Legault and El-Geneidy 2011).

A key goal of public transit is to prioritize passenger movement over vehicle throughput in order to improve road corridor efficiency (Waterson et al. 2003). By enhancing bus operational speed and reliability, transit systems can attract more passengers, thereby reducing demand for private car travel and contributing to a more sustainable, efficient, and environmentally-friendly urban transportation system.

1.1.1 GTFS data

Cities generate large volumes of data daily through digital services and smart city applications such as sensors, video cameras, traffic management systems, smart meters, vehicles, mobile phones, and internet of things devices. Smart city investments are sometimes justified by the inherent value of the data they generate without comprehensively specifying plans for using that data. Meanwhile, the amount of data generated by smart city applications is growing exponentially. Only a small number of organizations use the data to any significant extent, and most organizations that do employ less than half of the data they have collected (Barrett 2018).

Public Transport Authorities (PTAs) are a key example of underused big data generators. Their daily operations create data on vehicle positions, counts of passengers, user travel patterns, disruptions and real-time conditions, and vehicle arrival times (Google Developers 2022b). This data can be used by transit authorities for service delivery and planning, to improve network performance and safety, and to optimize operational costs and resources (Prommaharaj et al. 2020). There is further opportunity to harness this data for the development of applications that support data-driven decision-making, enable proactive customer engagement, and improve customer experience, thus making cities more efficient and livable.

Traditionally, bus performance has been measured through manual surveys, which are limited in scope and accuracy, and do not cover all bus routes or times of day. However, the development of the newer dataset — GTFS in 2006 by Google revolutionized the way transit data is collected and analyzed. GTFS is now used by all transit agencies that participate in Google Maps (Google Developers 2022b; Wong 2013). GTFS feeds contain detailed transit service data, including stops, routes, trips, and schedules (Google Developers 2022b). With the real-time updates provided by GTFS, micro delays can now be analyzed — delays that were previously undetectable through traditional surveys. Wong showed that GTFS feeds accurately represent the transit network and can be relied upon as a source for measuring transit performance (Wong 2013). This high-resolution, real-time data allows for comprehensive and detailed analysis of transit performance, providing valuable insights that can inform improvements to transit infrastructure.

There are two main barriers to widespread access to GTFS-Realtime information that can be addressed by creating a data processing pipeline. First, the high-volume, realtime feeds are shared using protocol buffers, which serialize structured data for efficient, high-fidelity transfer. Protocol buffer messages require processing before use in most transport applications. Moreover, a typical realtime dataset is constructed from repeated queries to an API providing live trip updates. The GTFS-Realtime Trip Updates information disappears once the bus arrives at its stop, meaning there is no archived source of truth to reconstruct the data after the fact. Queries are made at regular intervals (every minute) and each newer update in a query replaces outdated information from previous queries about each arrival at each stop. These barriers can be addressed with the data processing pipeline that handles data ingestion and processing of the GTFS feeds.

Second, the raw feeds may contain errors such as missing values and data duplicates. The data processing pipeline can detect and clean spurious information. The pipeline addresses this barrier to the widespread use of the realtime information by including data cleaning and transformation modules.

1.1.2 Bus lane performance evaluation

Dedicated bus lanes can accommodate over five times the passenger capacity of mixed traffic lanes, even with a lower frequency of vehicles (National Association of City Transportation Officials (NACTO) 2018). In bustling urban corridors, bus priority schemes have been implemented worldwide to mitigate traffic congestion by improving the appeal of bus services (Shalaby 1999).

Common bus priority schemes include reserved bus-HOV (High Occupancy Vehicle) lanes, bus-taxi lanes, bus-only lanes and bus-only streets (dedicated rights of way, busways), as well as traffic signal priority (Transportation Research Board and National Academies of Sciences, Engineering, and Medicine 2013). These priority schemes claim intuitive benefits in improving bus operational performance such as delay and reliability (National Association of City Transportation Officials (NACTO) 2016). Meanwhile, there are other factors that influence bus operational performance, including passenger demand factors, such as roadworks and incidents, weather effects, weekends, public holidays, school holidays, and COVID lockdowns. Generally, these effects tend to reduce passenger demand, leading to improved bus operational performance (Wang et al. 2011; Tao et al. 2014; Cződörová et al. 2021; Guo et al. 2007). On the other hand, traffic flow, traffic signals, and cross-traffic turns, tend to have a negative impact on bus operational performance (Strathman and Hopper 1993; Gan et al. 2003).

Due to the limits in data availability, only a few studies have examined the impact of bus priority lanes on bus performance using real-time data (Surprenant-Legault and El-Geneidy 2011; Yan et al. 2016), and those that exist typically focus on road corridor level with a limited performance data collection. However, utilizes a significantly larger dataset, providing a more comprehensive analysis of the effects of bus priority lanes. It considers not only corridors with bus priority lanes but also other links without bus priority lanes. Additionally, few researchers have considered the combined effects of multiple variables on bus performance (Tao et al. 2014; Truong et al. 2015; Montero-Lamas et al. 2023), often using low-resolution data. This study addresses this gap by examining the combined effects of bus priority lanes, passenger demand factors, road supply factors, and weather effects at a finer scale. Using data from Sydney, it provides a more detailed understanding of the various factors influencing bus operations.

1.1.3 Bus cross-traffic turn

We use the term “cross-traffic turns" to describe turns which involve crossing opposing traffic while turning — these are right turns in left-hand driving environments including India, the United Kingdom, Australia and Japan and left turns in right-hand driving environments, such as those in continental Europe, North America, and China.

Give-way/yield rules are widely standardized causing cross-traffic turns to pose significant challenges for both cars and buses. Drivers performing cross-traffic turns must wait for either a dedicated signal phase (in the case of protected turns) or a sufficiently large gap in traffic (in the case of permitted turns). Due to these conflicting movements, cross-traffic turns typically receive the least green time at intersections, resulting in larger delays compared with through movements and non-cross-traffic turns (e.g., right turns in right-hand driving environments or left turns in left-hand driving environments).

Many studies have shown that cross-traffic turns contribute to delays at intersections (Newell 1959; DePrator et al. 2017; Manual 2000; Levinson 1998). Due to their larger vehicle size and lower acceleration rates, buses require larger gaps in traffic or longer dedicated signal phases compared to cars, making cross-traffic turns particularly challenging for buses.

While some studies have noted that bus cross-traffic turns negatively impact operational speed (Zhao and Zhou 2018; Shu et al. 2019; Dunne and McArdle 2022), there is a limited number of research providing quantitative assessments using real data to evaluate the effects of cross-traffic turns on bus performance. Although cross-traffic turn delays are typically small and stochastic at individual intersections, their cumulative impact over many intersections along a route or across multiple buses traversing a single intersection can result in significant inefficiencies. These accumulated delays contribute to the overall unreliability and reduced efficiency of bus services.

The outputs aim to inform the design of optimized routes, schedules, and operational strategies to minimize the delays and variability caused by bus cross-traffic turns. Addressing the challenges posed by cross-traffic turns is a vital step toward improving the efficiency and dependability of urban bus operations. This study contributes to the growing body of evidence advocating for targeted, data-driven solutions to enhance public transportation systems, ensuring they remain reliable, efficient, and sustainable in increasingly congested urban environments.

1.1.4 Interventions for bus cross-traffic turns

Traditionally, prioritizing bus cross-traffic turns has relied on solutions such as queue jump lanes, TSP, or bus pre-signals (Transportation Research Board and National Academies of Sciences, Engineering, and Medicine 2013; Smith et al. 2005; Wu and Hounsell 1998). While these methods offer advantages in improving bus performance, they come with several challenges. Queue jump lanes, for example, require additional space at intersections, particularly lateral space. In densely built urban areas, where lateral space is often constrained, converting a traffic lane into a queue jump lane can significantly increase delays for general traffic. This can result in intersection queues spilling back beyond the entrance of the queue jump lane, making it more difficult for buses to access the lane and potentially increasing delays for both buses and general traffic. An alternative, continuous bus priority lanes, could address this issue, but reallocating road space from general traffic to buses might reduce overall corridor efficiency.

TSP can enhance bus performance by prioritizing bus movements at signalized intersections, but this approach reallocates some signal time from general traffic to buses, potentially causing delays for other road users. However, when accounting for the higher occupancy of buses, the overall total delay may be reduced.

Pre-signals can reduce total person hours of delay across the system, making them a viable option for cities looking to enhance bus priority. However, like queue jump lanes, pre-signals require a dedicated bus lane on the approach side of the intersection. In urban areas with limited land and high costs, allocating road space for dedicated bus lanes is often not a feasible solution.

All of these traditional methods require additional space or green time to prioritize buses, which often results in increased delays for general traffic, limiting their practicality in urban environments. This creates a pressing need for innovative solutions that provide bus priority without the drawbacks of existing methods.

To address these limitations, this study introduces and evaluates a novel solution: the Bus Cross-Traffic Turn Priority Box. This priority box not only improves bus speed and reliability by skipping the cross-traffic turn queue but also benefits general traffic by allowing vehicles to pre-accelerate when no bus is present, reducing startup lost time and enhancing overall intersection efficiency. This dual functionality — prioritizing buses while maintaining or even improving general traffic flow— marks a significant advancement over traditional approaches.

1.2 Research objectives

This research aims to develop, store, and validate the GTFS-Realtime Trip Updates data. This involves transforming raw protocol buffer trip updates into user-friendly Comma-Separated Values (CSV) files, enabling microscopic analysis of bus performance. The processed GTFS-Realtime data facilitates high-resolution, second-level assessments of bus delays, operational speeds, and service reliability at the stop-to-stop level. By integrating this data with other available datasets, the research explores the impact of factors such as bus lanes, traffic signals, cross-traffic turns, and traffic volumes on operational delays and reliability. Among these factors, this study focuses on bus cross-traffic turns, which are found to have the most significant impact on bus reliability. Since current priority measures do not effectively reduce delays caused by cross-traffic turns without negatively affecting other traffic or requiring additional road space, this research proposes a novel solution: the Bus Cross-Traffic Turn Priority Box. This innovation aims to substantially reduce bus delays caused by cross-traffic turns while minimizing or even improving the flow of other traffic.

1.3 Thesis structure

This thesis is structured to methodically address each of the research objectives and incorporates four published and under-review journal papers (Figure 1.1). Chapter 2 presents the literature review, synthesizing existing research on bus performance degradation, bus priority measures, public transport performance metrics and the application of GTFS data, while highlighting the critical gaps that this thesis seeks to address. Chapter 3 focuses on

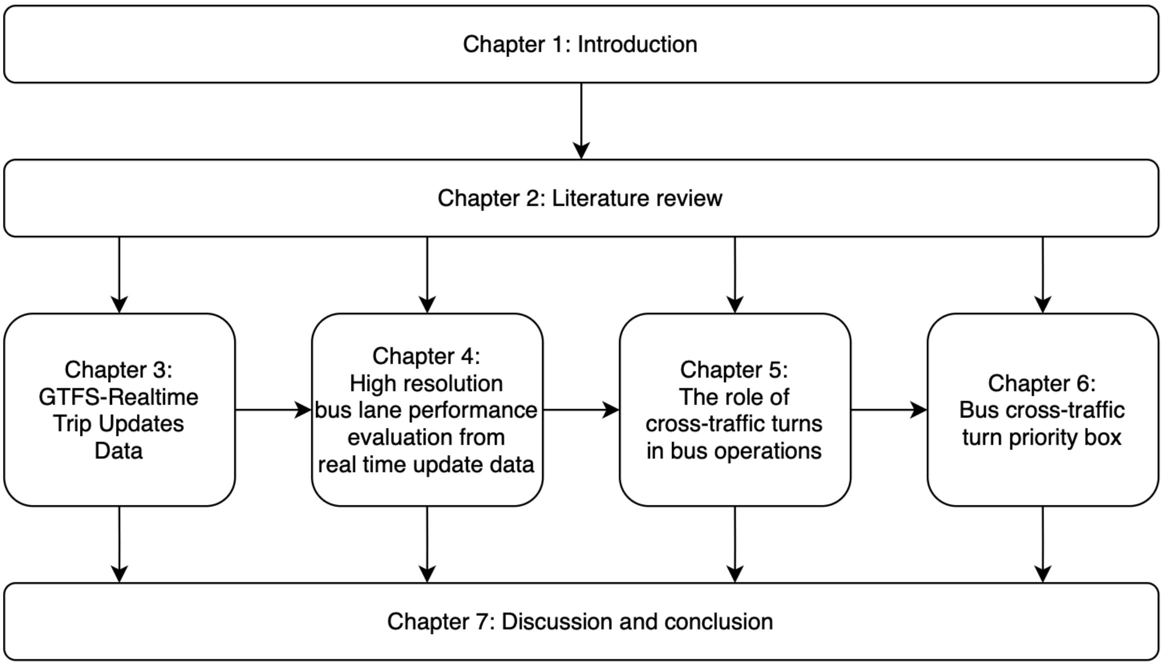


FIGURE 1.1: Thesis outline flowchart. Chapters 3, 4, 5, and 6 are each based on one of the four journal papers. These chapters are interconnected, with each building upon the results of the preceding chapter.

the development, storage, and validation of 25 months of bus arrival and departure times at bus stops in Greater Sydney using GTFS-Realtime Trip Updates data. Chapter 4 examines the combined effects of multiple operational factors at stop-to-stop level on bus performance, utilizing regression models to offer a comprehensive understanding of the variables influencing delays and reliability. Chapter 5 investigates the specific impact of bus cross-traffic turns at microscopic level, quantifying their effects on bus operations and emphasizing their significance as a key challenge to transit performance. Chapter 6 introduces the theoretical framework and simulation evaluation of the Bus Cross-Traffic Turn Priority Box, assessing its effectiveness in real-world applications. The final chapter, Discussion and Conclusion, summarizes the key findings, explores the broader implications of the research, and outlines potential directions for future work.

CHAPTER 2

Literature Review

2.1 Introduction

As cities grow and urbanization accelerates, traffic congestion worsens, leading to slower travel times and strained road capacity. With more vehicles on the road, private cars contribute to delays that also affect buses, which are crucial for alleviating road congestion by carrying large numbers of passengers. However, as congestion increases, buses face their own delays, making their performance even more critical to urban mobility. Efficient bus systems are vital for daily commuting and the overall effectiveness of urban transportation networks. To attract more ridership and improve service efficiency, enhancing bus performance is essential for maintaining a smooth and reliable transportation system in increasingly congested cities.

However, various factors, including traffic volume, signal timing, cross-traffic turns, and passenger demand variations, contribute to the degradation of bus performance, leading to delays and reduced service reliability. Furthermore, the implementation of bus priority measures, including bus lanes, signal priority, and pre-signals, offers opportunities to mitigate delays and improve bus performance. Traditional methods of evaluating bus performance, such as relying on static schedule data or aggregate vehicle counts, are insufficient in capturing the complexities of realtime bus operations. The integration of realtime data sources, such as Automatic Vehicle Location (AVL) and GTFS, has enhanced our ability to assess and address these operational challenges.

This literature review examines the factors contributing to bus performance degradation, evaluates the impact of bus priority measures, and explores how emerging data sources and technologies can optimize public transportation systems. By synthesizing research on these aspects, this review aims to highlight key trends, methodologies, and gaps in the existing literature, ultimately guiding future studies and policy decisions in the domain of bus operations and planning.

2.2 Bus performance degradation

2.2.1 Traffic volume

High traffic volume, leading to congestion, is a major contributor to bus performance degradation. As traffic volume increases, buses face significant delays due to congestion and slower travel speeds, which negatively affect travel time reliability (Sterman and Schofer 1976). Studies have consistently shown that there is an inverse relationship between increasing traffic volume and declining bus performance (Sterman and Schofer 1976; Shalaby 1999; Gan et al. 2003). Collectively, these studies corroborate the inverse relationship between increasing traffic volume and declining bus performance.

2.2.2 Traffic signals

Bus delay often occurs at intersections where public transport signal priority can be implemented. Bus signal priority can take the form of extended green times or dedicated signal phases for buses, which help reduce delays at intersections (Shalaby 1999; Kim 2003; Sterman and Schofer 1976). However, when signal priority is not in place, traffic signals become a major source of delay for buses, particularly during periods of high traffic volume.

In Sydney, signal priority is possible through the adaptive traffic signal control system using a priority-giving algorithm. Although it has been implemented in the past, this system was deactivated in most road corridors during the data collection period for this study. As a result, the traffic signals are anticipated to be a source of unreliability rather than bus priority in the current study.

2.2.3 Bus cross-traffic turns

Cross-traffic turns present significant challenges for both buses and general traffic. Studies have consistently shown that cross-traffic turns reduce intersection capacity, particularly when turning traffic must yield to opposing flows, impacting both turning and through traffic (Newell 1959; DePrator et al. 2017; Manual 2000; Levinson 1998). This reduction in capacity becomes more pronounced with increasing opposing traffic volumes (Messer and Fambro 1977).

Buses, as essential components of urban traffic, are particularly disadvantaged by delays caused by cross-traffic turns. Given their larger size and slower acceleration rates, buses require wider gaps in traffic or longer dedicated signal phases than cars, amplifying delays. Kim (2003) highlights the significant challenges buses face when executing cross-traffic turns on multi-lane arterial roads, noting that these maneuvers not only slow down bus operations but also disrupt overall traffic flow. Other researchers, such as Zhao and Zhou (2018) and Shu et al. (2019), advocate for prioritizing buses at intersections during cross-traffic turns. Their studies demonstrate that such priority measures can reduce delays for both through and turning buses while maintaining private vehicle performance, ultimately reducing total passenger delays at intersections.

Existing countermeasures, such as route adjustments, signal modifications, hook turns, queue jump lanes, and jug handles, are often conceptually motivated and lack rigorous cost-benefit analysis (National Association of City Transportation Officials (NACTO) 2016). Additionally, the individual contributions of each cross-traffic turn to bus delay and reliability have not been systematically quantified, leaving a significant gap in addressing this issue comprehensively.

2.2.4 Passenger demand variation

Passenger demand on buses varies significantly depending on factors such as the time of day, day of the week, public holidays, and weather conditions. Understanding these variations is essential for optimizing bus schedules and ensuring efficient service.

2.2.4.1 Weekends

Similar to demand patterns observed in other transportation modes, bus ridership in most cities — excluding those with a significant proportion of tourist activity — tends to be lower on weekends than on weekdays, primarily due to the reduced number of commuters traveling to work or school (Tao et al. 2014; Balcombe et al. 2004). A study by Wang et al. (2011) which compared weekday and weekend passenger patterns for London's Oyster card users, revealed substantial variations in travel patterns depending on the service catchment areas. Further, Guo et al. (2007) found a heightened sensitivity to weather effects on weekends compared to weekdays. Another study indicated that buses typically have shorter travel times on weekends due to less traffic congestion and fewer stops, resulting in lower dwell times at bus stops (Tao et al. 2014; Suwardo et al. 2009).

2.2.4.2 Public holidays

Similar to the patterns seen among weekend public transport users, bus ridership decreases in many places on public holidays with no notable peak hours (Tao et al. 2014). In Brisbane, Australia, public holidays see the most considerable decline in passenger numbers compared to weekends and school holidays (Tao et al. 2014; Tao et al. 2018). Research by Mendes-Moreira et al. (2015) employed machine learning to evaluate bus schedules, suggesting that public holidays should follow a schedule similar to weekends.

2.2.4.3 School holidays

Weekdays during school holidays have fewer passengers than regular weekdays, but the reduction is minor compared to other passenger effects. This is expected as school students make up only a small proportion of all bus passengers (Tao et al. 2014) and changes in the routines of households with school students are responsible for the rest of the variation in demand. Mendes-Moreira et al. (2015) employed machine learning to evaluate bus schedules and proposed that weekdays during school holidays should follow a different schedule compared to school days.

2.2.4.4 COVID pandemic

The COVID-19 pandemic had major impacts on the global transport network. The implementation of ‘lockdown’ and ‘stay-at-home’ measures has led to both a reduction in travel demand and changes in travel behaviors (Yan et al. 2021; Jenelius and Cebecauer 2020). Public transport ridership diminished owing to increased travel-related risks, while private car use became more prevalent (Yan et al. 2021; Jenelius and Cebecauer 2020). Yan et al. (2021) employed the Oxford Stringency Index to reflect the intensity of stay-at-home directives. They found that as lockdown measures eased, bus delays increased due to a confluence of extended dwell times at stops and heightened traffic congestion (Yan et al. 2021).

2.2.5 Weather effects

Weather affects both public transport service and ridership demand as considered in previous studies (Guo et al. 2007; Tao et al. 2018; Böcker et al. 2013). The impact of weather on transit varies based on the weather severity, weather attribute, mode choice, destination and activity type (Guo et al. 2007; Tao et al. 2018; Böcker et al. 2013; Singhal et al. 2014). Weather conditions including heavy precipitation, low temperatures and strong winds can impact both supply (reduced visibility, reduced freeflow speed) and demand (mode shift driven by weather) in the transport system, which can result in temporary and even long term declines in transit ridership (Tao et al. 2018; Hofmann and O’Mahony 2005; Hine and Scott 2000; Changnon 1996).

Unlike the route design and infrastructure attributes including bus priority lanes and traffic signal, weather is not a decision variable. The studies conducted by Tao et al. (2018) and Kashfi et al. (2013) in Brisbane, Australia, suggest that precipitation has the most significant impact on transit ridership, so this study adopts precipitation as the weather effects indicator. Consequently, the reduction in bus ridership due to precipitation is expected to influence variability of bus operational performance.

2.3 Bus priority measures

2.3.1 Bus priority lanes

Bus priority lanes are a widely adopted strategy to enhance bus service quality because of their low implementation cost and substantial benefits. Proper enforcement ensures these lanes effectively prioritize buses, reducing travel times and increasing reliability with minimal expense and short implementation time (Ma et al. 2014). However, the effectiveness of bus lanes has less often been measured using real-time data. Many previous studies have relied on simulated bus priority scenarios (Shalaby 1999; Zhou et al. 2009; Tsitsokas et al. 2021), with few examining the impact of bus lanes through observational data (Surprenant-Legault and El-Geneidy 2011; Yan et al. 2016). These studies typically show that bus lanes enhance operation speed and reliability. Due to data limitations, these studies often focus on specific road corridors. This study leverages a significantly larger dataset to offer a more comprehensive analysis of bus lane impacts.

FIGURE 2.1: Peak hour only bus-taxi lane on Parramatta Road in Sydney, Australia. The lane is painted red and heavily used by buses. Bus-taxi lanes are commonly located on major bus corridors in Sydney. Source: Google Earth

Previous studies suggest that reliability is a more critical metric for bus performance compared to speed, comfort, and frequency (Zhou et al. 2009; Aemmer et al. 2022). Factors such as route length, intersection control, traffic volumes, and passenger loads can have a significant impact on bus travel reliability (Sterman and Schofer 1976; Abkowitz et al. 1986). Implementing strategies such as reducing route lengths and optimizing intersection controls to prioritize buses can greatly enhance bus travel reliability.

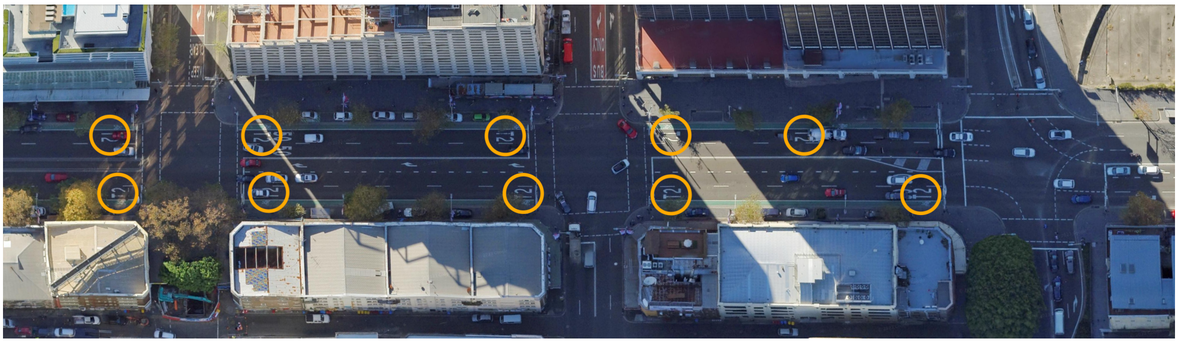


FIGURE 2.2: Day time bus-HOV lane on William Street, Sydney, Australia. ‘T2’ (Transit Lane for vehicles with two or more people) painted on the bus-HOV lane (highlighted in yellow circle). Unlike HOV lanes in North America, HOV lanes in Sydney are more often located on arterial streets instead of motorways. Source: Google Earth

A common approach to implementing bus lanes includes allowing shared use with taxis (see Figure 2.1). Previous simulation study support this policy, indicating that prohibiting taxis from using bus priority lanes causes greater performance deterioration in adjacent traffic than the performance gains achieved for buses in the priority lanes at typical bus density levels (Shalaby 1999).

Another type of bus priority lane incorporates High Occupancy Vehicles (HOV) alongside buses. HOV lanes, designed to reduce travel times for high-occupancy vehicles, including buses, significantly increase the capacity of road corridors by encouraging carpooling (Chang et al. 2008). Enhancing capacity at bottlenecks is particularly effective, leading to improved overall corridor performance (Kwon and Varaiya 2008). While HOV lanes are prevalent in the United States and Canada, they are less common in Europe and Australia. Unlike the highway-median placements typical in other regions, Australian HOV lanes are frequently located in the curbside lane and heavily utilized by buses. A typical urban example is illustrated in Figure 2.2. Although buses are significant users of HOV lanes in Australia, there is a lack of literature evaluating the benefits of HOV lanes for bus operations.

2.3.2 Transit signal priority

Transit Signal Priority (TSP) adjusts signal timings at intersections to grant extra green time or reduce red time for buses, thereby decreasing bus delays and variability in running times (Smith et al. 2005). TSP is considered a cost-effective solution to enhance bus travel speed and reliability, often with minor impacts on general traffic (Smith et al. 2005; Transportation Research Board and National Academies of Sciences, Engineering, and Medicine 2013). Ma et al. (2014) categorize TSP strategies into three types: passive, active, and adaptive control. Passive TSP employs fixed signal timings optimized based on historical bus data, benefiting only on-schedule buses while neglecting real-time bus performance (Ren et al. 2021). Active TSP uses real-time bus arrival information to grant signal priority, benefiting both early and late buses (He et al. 2014). Adaptive TSP optimizes performance for both buses and private vehicles using real-time data collection. By considering traffic flow for both buses and private vehicles, adaptive TSP can minimize total delay while granting priority to buses when they carry many passengers or are significantly delayed (He et al. 2014).

2.3.3 Queue jump lanes

Queue jump lanes (Figure 2.3) are a common bus priority treatment that allows buses to bypass waiting queues at signalized intersections, often employed in conjunction with TSP. They have been proven effective in reducing bus travel times and enhancing reliability (Transportation Research Board and National Academies of Sciences, Engineering, and Medicine 2013; Zhou and Gan 2009). For instance, Zhou and Gan (2005) found that the efficiency of queue jump lanes increases when combined with TSP. However, similar to other bus priority measures, their effectiveness can be diminished when significant traffic queues delay buses from entering the queue jump lane (Zhou and Gan 2005; Truong et al. 2015; Nowlin and Fitzpatrick 1997). To mitigate this issue, Knapp and Jacobson (1995) suggested extending the length of queue jump lanes but this requires additional road space.

FIGURE 2.3: Cross-traffic turn queue jump located near Macquarie University in Sydney, Australia. The cross-traffic turn queue jump lane is positioned on the right side (east side) of the intersection. Large road surface is required to facilitate the cross-traffic turn queue jump lane. Source: Google Earth

2.3.4 Bus pre-signal

Bus pre-signals (Figure 2.4) are traffic signals installed at or near the end of a with-flow bus lane to provide bus priority at downstream junctions (Wu and Hounsell 1998; Guler and Menendez 2014; Guler and Menendez 2015). Implemented in Europe since the 1990s (Wu and Hounsell 1998), bus pre-signals aim to optimize the utilization of the main signal's capacity while prioritizing buses, thereby minimizing total system-wide person hours of delay (Wu and Hounsell 1998; Guler and Menendez 2014; Guler and Menendez 2015). By allowing buses to bypass traffic queues and facilitating easier access to the cross-traffic turn lane (inside lane), bus pre-signals effectively reduce bus delays at signalized intersections (Wu and Hounsell 1998).

While the effectiveness of pre-signals may diminish when intersections become congested — resulting in reduced capacities in the reservoir between the pre-signal and the main signal (Wu and Hounsell 1998) — analytical studies by Wu and Hounsell (1998) and Guler and Menendez (2014) and Guler and Menendez (2015), which consider various traffic scenarios and bus frequencies, indicate that bus pre-signals provide benefits for buses while having minor impacts on general traffic.

FIGURE 2.4: Bus pre-signal in Shepherds Bush, London. Before the far side main signal turns green, general traffics are required to stop at the nearside pre-signal in order to allow buses on the bus lane to advance and skip the traffic queue at the intersection. The pre-signal reduces bus delay and allow easy access to right turn (cross-traffic turn) lane for buses at signalized intersection. Source: Google Earth Street View

2.4 Evaluation

2.4.1 Simulation approaches

Several studies have used simulation models to assess the impact of bus priority measures, particularly bus lanes, on bus and general traffic performance (Shalaby 1999; Gan et al. 2003; Truong et al. 2015; Li et al. 2022; Petit et al. 2021). Shalaby (1999) found that dedicated bus lanes improved bus performance but at the cost of deteriorating the flow of adjacent general traffic. Similarly, Truong et al. (2015) highlighted that continuous bus lanes offer more benefits with less disruption to general traffic than discontinuous lanes, with proportional improvements to bus travel times. Gan et al. (2003) assessed travel speeds for both buses and general traffic, considering factors like traffic volumes, cross-traffic turns and signalization, and found that both proposed and existing bus lanes have a significant impact on the average person travel time. Li et al. (2022) used simulation and machine learning based surrogate models to optimize the bus lanes allocation in large-scale networks. Petit et al. (2021)

presented a framework to optimize bus routes and headways, considering dynamic travel demand and the impact of bus lanes on traffic congestion and overall system performance. Tsitsokas et al. (2021) proposed a dynamic traffic modeling framework to optimally allocate dedicated bus lanes, enhancing urban traffic performance while balancing bus priority and regular traffic flow.

It is worth noting that numerous simulation studies have concentrated on vehicle counts rather than passenger counts in assessing the effectiveness of bus priority measures, thus ignoring the high occupancy within buses. These simulations typically use hourly flow rates to represent traffic demand, which does not take into account the actual arrival times of vehicles. With the realtime data, it is anticipated that analyses employing actual data will provide more accurate comparisons than simulations.

2.4.2 GTFS data

Daily operations of PTAs create data on vehicle positions, counts of passengers, user travel patterns, disruptions and real-time conditions, and vehicle arrival times (MobilityData 2021a). The GTFS data can be used by transit authorities for service delivery and planning, to improve network performance and safety, and to optimize operational costs and resources (Prommaharaj et al. 2020). There is further opportunity to harness this data for the development of applications that support data-driven decision-making, enable proactive customer engagement, and improve customer experience, thus making cities more efficient and livable.

GTFS is a data format that allows multimodal Public Transport Authoritie (PTA) data to be consumed by a wide variety of software applications. Thousands of public transport providers worldwide use the GTFS data format to openly share their data (MobilityData 2021a). This common sharing format allows for the private development of applications (apps) such as those that monitor real-time arrivals, track public transport vehicles, support Mobility as a Service (MaaS), sell commercial advertisements, or share customized messaging (Prommaharaj et al. 2020; Frick et al. 2020).

There are two distinct varieties in the specification. The GTFS-Static (or GTFS-Schedule) format contains planned operational information such as routes, stops, trips, and schedules. Kujala et al. (2018) have previously used this format to publish public transport route and timetable information for 25 cities. In contrast, GTFS-Realtime contains live transit data about vehicle positions, forecasted trip updates, and disruption updates. Due to the large volume of data produced, the realtime data exchange format is based on Protocol Buffers, a language- and platform-neutral mechanism for serializing structured data.

A key type of information contained in the GTFS-Realtime feed is the updated arrival time prediction of every transit vehicle at every planned stop, which is called GTFS-Realtime Trip Updates. Because the realtime data is generated at the time of the query, a dataset constructed from frequent queries will contain redundant, duplicate, and outdated data. To extract the actual arrival information, the protocol buffer messages need to be processed into a readable format, cleaned and the duplicates filtered out. Since the forecasted arrival and departure times are updated continuously, we assume that the most recent update conveys the best estimation of arrival and departure time.

2.4.3 Data approaches

Several studies have utilized real-world data to analyze bus performance and improve transit operations. Early work focused on using static datasets like the GTFS-Static, which primarily reflect impacts of bus schedules. For example, Arias et al. (2021) used GTFS alongside additional datasets to estimate potential travel time reductions afforded by bus preferential treatments. Since GTFS-Static schedules are based on average times between stops, without real-time performance adjustments, their findings were successfully validated through the Google Maps Application Programming Interface (API). Similarly, Prommaharaj et al. (2020) developed an interactive tool using GTFS-Static to visualize transit system operations, both geographically and statistically. While these studies provided valuable insights, they lacked the incorporation of realtime data, which is critical for comprehensively understanding the dynamics of public transportation.

The advent of real-time data has significantly enhanced the precision of bus performance evaluations. AVL data has traditionally been used for microscopic analyses. The AVL system provided real-time vehicle position data, but its availability was limited. AVL data has been used in several studies for more detailed, microscopic analyses of bus operations. For example, Tantiyanugulchai and Bertini (2003) and Yang et al. (2013) relied on AVL data to assess travel speeds, travel times, and reliability at a granular level, such as for individual bus stops or stop-to-stop links. Sun et al. (2021) employed dwell time data from AVL to estimate the demand profile of a bus route using a Bayesian framework. These studies demonstrated the value of AVL data in capturing specific operational challenges, such as delays caused by cross-traffic turns and traffic signals, which could not be captured by broader, aggregate metrics like those from the GTFS-Static dataset.

More recently, the GTFS-Realtime dataset, which is based on AVL data, has made real-time bus performance data publicly available. GTFS-Realtime updates, such as Trip Updates and Vehicle Positions, are derived from AVL systems, providing standardized and shareable data. While the GTFS-Realtime dataset is still relatively new and has not been extensively used in prior research, it offers significant advantages over traditional AVL data, particularly in its public accessibility. For instance, Aemmer et al. (2022) conducted an analysis of buses, using GTFS-Realtime data to predictable (systematic) delays and random (stochastic) variation on a segment-by-segment basis. Similarly, Abusalim (2020) investigated the distribution of GTFS-Realtime Trip Updates over a day, while Prommaharaj et al. (2020) and Aemmer et al. (2022) have employed GTFS-Realtime Vehicle Positions data to evaluate bus performance on a microscopic level, examining delays, speeds, flows, densities, and headways.

Despite the advantages of GTFS-Realtime, there remains a gap in its application over extended periods and in conjunction with other datasets, such as road characteristics. However, its growing accessibility marks a significant step forward in the ability to evaluate bus performance in real time, offering an essential tool for addressing localized issues like stop-level delays and cross-traffic turns.

2.4.4 Performance indicators

In exploring bus transit performance, researchers have employed diverse performance indicators, each offering unique insights into transit operations (Yan et al. 2021; Arias et al. 2021; Prommaharaj et al. 2020; Sheth et al. 2007; Chen et al. 2009). Arias et al. (2021) calculated actual travel times between stops, considering dwell time for a nuanced measure of transit performance. Prommaharaj et al. (2020) employed six core indicators, such as mobility and speed, to visually represent transit system operations. Sheth et al. (2007) used variables like headway, intersections, and priority lanes in conjunction with schedule reliability and average travel time for their performance model. Chen et al. (2009) applied a punctuality index, a deviation index, and an evenness index to gauge bus service reliability in Beijing, finding that dedicated bus lanes notably improve reliability.

Historically, reporting has relied on temporal aggregations and accepted margins for bus operational performance. Metrics like On-time Running (OTR) require a threshold of delay (e.g., 3 minutes) and a route or trip will meet that target some fraction of the time (Nakanishi 1997). OTR is convenient because it can be based on delay at time check points, a subset of the stops, and it recognizes that customers have a tolerance for uncertainty in bus arrival (Hickman 2001). However, OTR makes it difficult to pinpoint performance issues on individual stop-to-stop segments — this might include a timetable with a badly timed stop or the advantage of a bus lane on the following stop arrival.

With real-time delays measured to the second, newer data sources like GTFS Realtime enable a range of new performance measures. Based on the availability of both realtime arrival forecasts and vehicle position data, this dataset directly supports performance measures like average headway, variation in headway, average delay, and variation in delay. These measures are determined by the accumulation of delay over the route and may not be suitable for assessing the impact of localized priority measures on the stop-to-stop links. Vehicle speed is also available from the GTFS Realtime — while this is relevant to the local conditions, it refers to the instantaneous speed of the vehicle at the time of the query to the GTFS API and might sample speeds associated with the vehicle during boarding and alighting, at a traffic signal, or when rolling. Aggregated over many vehicles, these values may describe the performance of a stop-to-stop link, but individual measurements are not a good indicator of the impact of priority on bus performance.

Stop-to-stop marginal delay evaluates bus performance on a detailed, microscopic level between successive stops. Both Yan et al. and Kaddoura et al. define this delay as 'the difference in schedule delays for a bus trip between pairs of successive stops' (Yan et al. 2021; Kaddoura et al. 2015). This measure focuses on delays caused by specific factors like bus lanes and traffic signals between two stops. This research utilizes stop-to-stop marginal delay as a microscopic indicator of bus performance, modeling the effects of various independent variables on bus stop-to-stop marginal delays.

2.5 Conclusion

In conclusion, improving bus performance and reliability requires a approach that can addressing both the causes of bus delays and the benefits of bus priority measures. Key contributors to bus performance degradation include high traffic volume, traffic signal delays, and cross-traffic turns. These factors often result in significant delays, reducing bus reliability and efficiency. While previous studies have highlighted these challenges, data constraints have limited our ability to fully quantify their effects and develop effective countermeasures. Real-time data sources such as AVL and GTFS-Realtime have greatly advanced the understanding of these delays, enabling more detailed, micro-level analyses of bus operations and the identification of patterns related to passenger demand.

Bus priority measures, such as dedicated bus lanes, transit signal priority, and queue jump lanes, have shown promise in mitigating these delays. These measures are effective in enhancing bus travel speeds and reliability. However, challenges remain in optimizing these measures, particularly in understanding how various operational factors, such as traffic congestion and passenger demand, interact and impact bus performance. Additionally, while the implementation of bus priority measures has proven beneficial in controlled settings, their real-world effectiveness often varies, and further research is needed to assess their impact across different urban contexts.

Finally, evaluating the effectiveness of these measures remains a critical area for research. Despite the availability of real-time data, there is still a need for more robust evaluation methods that can capture the nuanced effects of bus priority strategies in dynamic environments. Future research should focus on refining data analytical techniques to quantify delays and better integration of bus realtime data. Furthermore, there is a need to explore the combined impact of these measures, considering factors such as weather, operational variability, and environmental changes.

By addressing these gaps, we can enhance the overall performance of bus systems, making them more efficient, reliable, and adaptable to the dynamic needs of urban environments. Ultimately, leveraging these technologies and strategies will contribute to more sustainable and effective public transportation systems, improving urban mobility and quality of life in cities worldwide.

CHAPTER 3

GTFS-Realtime Trip Updates Data

3.1 Introduction

The efficient management of urban bus systems relies heavily on accurate and timely operational data, yet a critical gap exists in the availability of user-friendly bus real-time datasets for large-scale urban areas. This gap hinders the ability of both researchers and transportation authorities to evaluate bus system performance effectively, limiting the potential for informed decision-making and the implementation of targeted interventions. The aim of this chapter is to provide a publicly available dataset that enables the microscopic evaluation of bus operation, current priority measures, and hypothetical interventions that could optimize operational efficiency. By presenting real-time updated GTFS data on bus operations in Sydney, this study addresses the lack of long-term bus realtime arrival and departure data necessary for microscopic analysis of transit performance.

In response to the need for comprehensive, long-term data sources, this chapter introduces a manipulated GTFS-realtime trip updates feed. The feed has been transformed into a general and flexible dataset of real-time bus arrivals, along with the data processing pipeline used for this development. This pipeline, written in Python, enables the generation of 25 months of real departure and arrival information for New South Wales (NSW), Australia. This dataset will serve as a valuable resource for researchers and practitioners working on the optimization of bus priority and bus scheduling.

Additionally, the pipeline's flexibility opens the door for future research that can integrate dynamic and spatial sensitive data, such as weather conditions, roadworks, and traffic volumes.

For example, Chapter 4 demonstrated how traffic volume, signal location, and other spatial and temporal data can be incorporated into the pipeline to evaluate bus lane effectiveness. Additionally, Chapter 5 indicated that the pipeline output can be used to evaluate the effect of cross-traffic turns at a microscopic level between bus stops.

3.2 Methods

Figure 3.1 illustrates an overview of the GTFS data pipeline. Each step discussed in the figure represents a modular function that allows for independent operation of each step in the data pipeline, providing user flexibility. Guides to using the data processing pipeline scripts are provided in the code repository on GitHub (`https://github.com/SCALUT`) (Chin and Xian 2024).

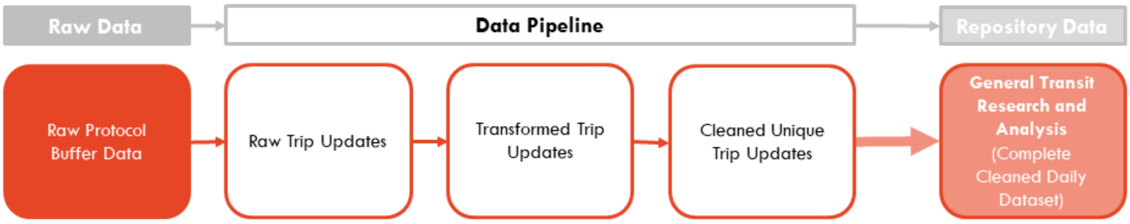


FIGURE 3.1: GTFS data processing pipeline overview. Each module creates intermediate data products that can be used. The final outputs are estimated arrival and departure times for every transit vehicle at every stop.

3.2.1 Data ingestion and processing

The inputs to the data processing pipeline are protocol buffer messages from the GTFS-Realtime Trip Updates feed. Protocol Buffers are a mechanism for serializing structured data (Google Developers 2021b). The data structure is defined in a `.proto` file, the schema governing how the data must be read and written. The `.proto` file is used to generate libraries with classes and functions needed for accurately reading and writing protocol buffer messages. The serialized protocol buffer messages are stored in `.pb` binary files. Protocol buffers are language- and platform-neutral, making them an efficient and flexible approach to sharing high volumes of data.

A trip update includes a forecasted arrival and departure time for every active vehicle at each planned stop, and the most recently published forecast is equated to the actual arrival or departure time. These forecasts are generated by the local PTA, and each authority typically employs its own estimation methods. The data used in this study are the arrival and departure time estimates provided by TfNSW. However, the specific algorithm used to generate these predictions is not publicly available, and no detailed information is provided regarding the underlying prediction methodology. The protocol buffer data are obtained via the relevant PTA's API, which must be queried at short time intervals to ensure timely and accurate forecasts of arrival and departure times. The implications of different query intervals are further discussed in Subsection Query Frequency below.

The data processing workflow begins with a cleaning phase, filtering out rows containing errors flagged within the API. After ensuring that only error-free data remains, the cleaned query outputs are merged, retaining only the most recent update of each vehicle arriving at each stop. If a query at a certain time is incomplete, it will be irrelevant for the stops and vehicles that are missing. This approach guarantees that the final dataset incorporates the most accurate and up-to-date information for each bus arrival. Further information on what fields are contained in the static and realtime specification can be found at `https://gtfs.org`.

After adjusting for local timezones, the data processing pipeline supports the six locations tested in the study, as well as any location worldwide that uses the standard GTFS-Realtime Trip Updates fields. The generalization of the data processing pipeline to a specific region depends on the field structure of the GTFS-Realtime Trip Updates. This dataset was created utilizing the TfNSW standard for the serialization of the protocol buffers– it is very similar to the generic standard outlined at `https://gtfs.org/realtime/feed-examples/trip-updates/`. When using the default pipeline, if a location has additional fields available in the serialization, that information will be ignored. However, if any expected fields are missing compared to the standard structure, an error message will be raised.

As shown in Table 3.1, most testing locales are using version 1 of GTFS-Realtime, and this data processing pipeline uses GTFS-Realtime v1 datasets. The data processing pipeline is primarily developed for the TfNSW realtime bus datasets.

TABLE 3.1: Summary of GTFS Datasets Tested in the data processing pipeline development. Most locations are using version 1. More information about each feed is available through the code repository associated with this paper.

Location	**Transport Authority**	**Version**
Greater Sydney, NSW	Transport for New South Wales	v1 / v2[1]
Greater Brisbane, QLD	TransLink	v2
Greater Adelaide, SA	Adelaide Metro	v1
Canberra, ACT	Transport Canberra	v2
Auckland, NZ	Auckland Transport	v1
Christchurch/Canterbury, NZ	Environment Canterbury (ECan) / Metro	v1

[1] At the time of writing, TfNSW is transitioning the GTFS-Realtime datasets from v1 to v2.

A directory of the feeds used for testing the data processing pipeline, including the version of the GTFS feeds, API Uniform Resource Locator (URL)s, and other associated information, is available in the code repository.

For GTFS, the standard hierarchy of elements and their type definitions are specified in the `gtfs-realtime.proto` file (Google Developers 2021a). Some PTAs generate GTFS-Realtime datasets with extended data contained in optional fields, requiring a PTA-specific `.proto` file. To maintain the generalizability of the data processing pipeline, these extended data are not recorded. The following two datasets tested in this study include extended data and require `.proto` files provided by the relevant PTAs:

- TfNSW protocol file uses extension ID 1007 that includes optional fields for TfNSW VehicleDescriptor and CarriageDescriptor. These apply to the Vehicle Updates feed and are not relevant to the dataset described in this paper.
- Adelaide Metro's protocol file uses extension ID 1999 for wheelchair accessibility information.

3.2.2 Location- and mode-specific structures

The `.proto` for the standard GTFS-Realtime fields is processed to language bindings by MobilityData (MobilityData 2021b) and published as a Python package `gtfs-realtime-bindings`. The PTA-specific `.proto` file can be used in the data processing pipeline if extended data is required.

There are several important considerations when applying the data processing pipeline to alternative locations. First, the optional fields provided in the datasets may be inconsistent among the study locations. Applications requiring information from these optional fields will generally not be portable across different locations. The input to the data processing pipeline must include all of the fields listed in Table 3.2, or an error message will be raised. In areas where the realtime departure time is defined to be the same as the arrival time, the pipeline will compute the same values twice and the dwell time would always be zero. Any additional location- and mode-specific columns will be ignored by the data processing pipeline and will not be read.

Furthermore, the organization of feeds by geography and mode varies by PTA. For example, the Trip Update feed from TfNSW covers the entire state by public transport mode (bus, light rail, ferry, etc.), whereas the Trip Update feed from TransLink in Queensland covers all public transport modes by region. Some locales have combined or excluded some feeds. For example, Canberra combines the Trip Update and Vehicle Position feeds for light rail only but doesn't support Service Alerts. Additionally, for Canberra's buses, realtime information is only available in the Service Interface for Real Time Information (SIRI) standard.

While the data processing pipeline was originally developed to create a user-friendly dataset for one mode and location, there is significant commonality between modes and locations to justify generalizing the code into a data processing pipeline that detects the correct structure. The data processing pipeline outputs have a default configuration to exclude extended data.

3.3 Data records

The dataset covers bus operations across New South Wales, Australia, for a period spanning from June 1, 2020, to June 30, 2022. Data were collected every minute from the TfNSW GTFS realtime Trip Updates API (Transport for NSW 2021b). The spatial extent is shown in Figure 3.2. The dataset covers most of the buses within the Greater Sydney Metropolitan Area. Following the data processing pipeline, the data is ingested, cleaned and processed into bus arrival and departure times. Both Figure 3.2 and Figure 3.3 are created using bus trajectory information contained in the GTFS-static files available from the TfNSW Open Data Hub (Transport for NSW 2021b).

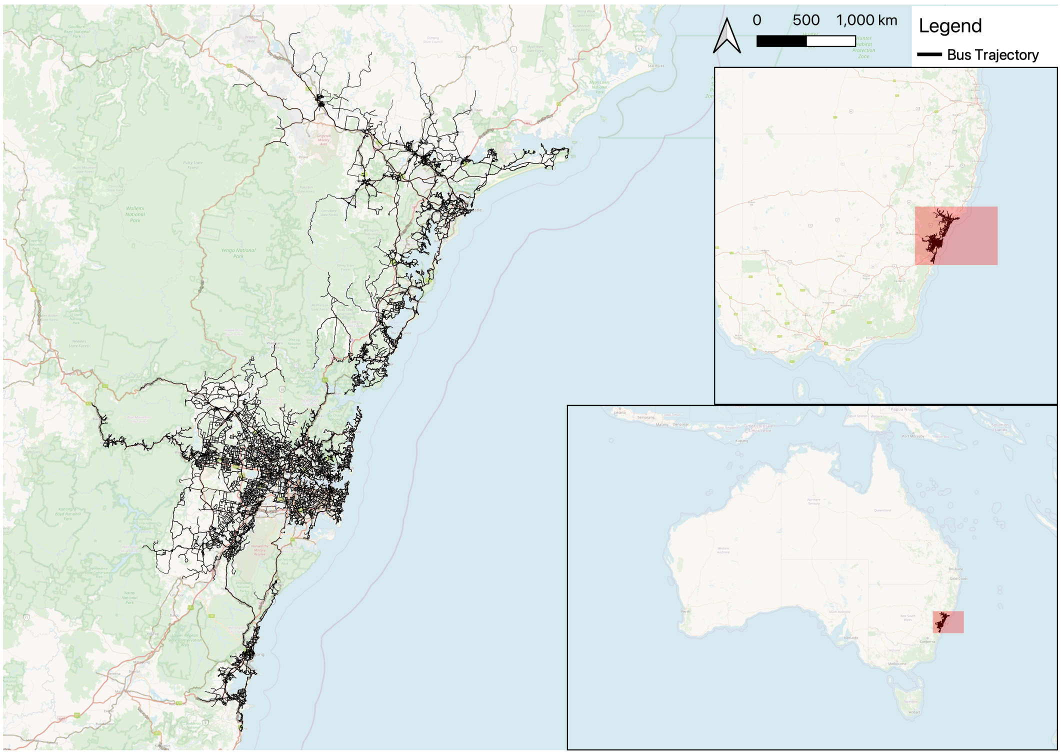


FIGURE 3.2: The TfNSW GTFS-Realtime Trip Updates bus dataset coverage. The feed covers the Greater Sydney Metropolitan Area including Newcastle in the North and Wollongong in the South. Base map: Open Street Map.

A snapshot of the delays reported as the output of the data processing pipeline for a single bus trip is shown in Figure 3.3. Notice that many realised delays are negative (the bus was ahead of the schedule) — this is a common observation when demand was low during and after COVID. The arrival and departure delays for each bus at each bus stop are continuously updated while traveling based on the service status. Only the latest update, which is the most accurate, is retained in the final product so the final reported delay associated with each stop is based on successively later queries going along the route.

3.3.1 Data format

All data are presented in CSV file format to ensure easy accessibility and compatibility with most data analysis tools, avoiding the need for specialized software. The sample of

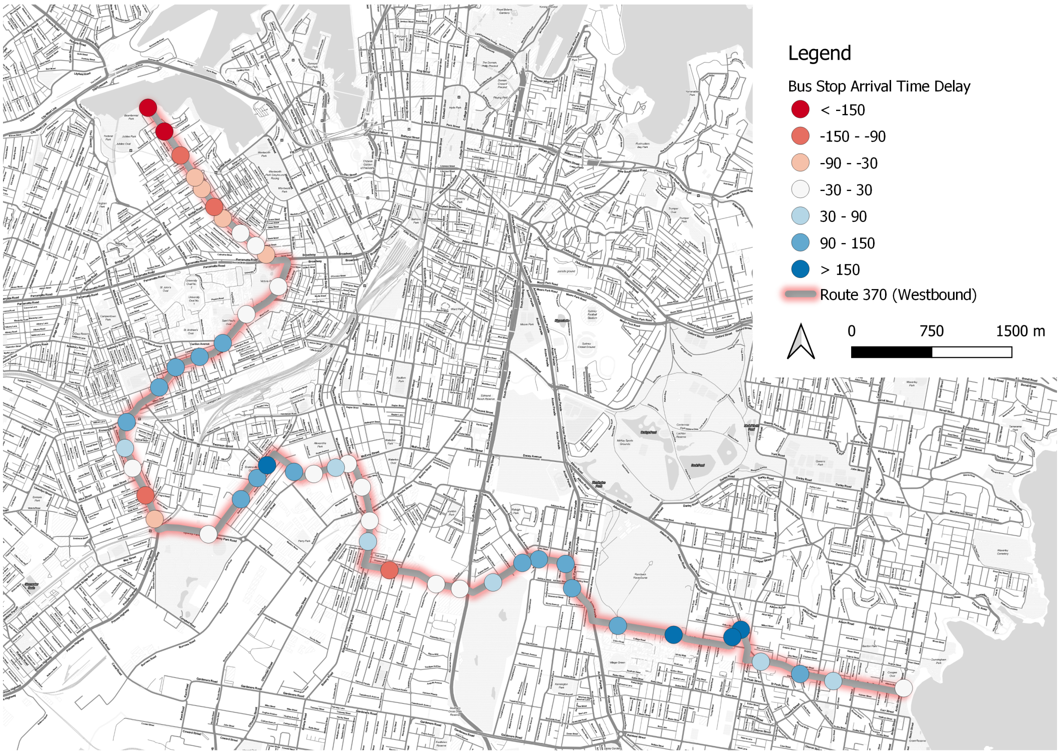


FIGURE 3.3: Latest GTFS-Realtime Trip Updates at each bus stop for a morning peak trip on Route 370 westbound on 01 March 2022. The arrival time estimates vary based on the realtime operational status along the route. Base map: Stamen Toner Lite.

data processing pipeline output, encompassing 25 months of actual bus arrivals/departures and intermediate data products, is stored on the Sydney eScholarship Repository. The data is openly accessible to everyone and can be found at `https://doi.org/10.25910/1pfb-4z05`.

Table 3.2: GTFS-realtime dataset column names and description.

Column Name	Name	Description
id	Vehicle ID	Unique ID associated with each vehicle.
trip_id	Trip ID	ID associated with each bus trip, which can be shared by multiple Vehicle ID across different dates.
trip_schedule_relationship	Trip Schedule Relationship	Relationship to the schedule of each trip. Default is ‘SCHEDULED’. Other values include: ‘ADDED’, ‘NO_DATA’, ‘CANCELED’, ‘0’ and ‘UNSCHEDULED’.
route_id	Route ID	ID associated with each bus route, which is shared by multiple Trip IDs.
vehicle_id	Vehicle ID	Unique ID associated with each vehicle. This matches the ‘id’ when realtime updates are available and is missing when the updates are missing.
stop_sequence	Stop Sequence	The sequence of bus stops, starting from 1, based on the stop order of each trip.
stop_arrival_delay	Estimated Bus Arrival Delay	The bus arrival delay at the bus station measured in seconds. Positive numbers indicate late arrival compared with the schedule, negative numbers indicate early arrival compared with the schedule.
stop_departure_delay	Estimated Bus Departure Delay	The bus departure delay at the bus station measured in seconds. Positive numbers indicate late departure compared with the schedule, negative numbers indicate early departure compared with the schedule.

Column Name	Name	Description
stop_id	Bus Stop ID	ID associated with each bus stop following TfNSW's conventions from the GTFS-static scheduled data.
stop_schedule_relationship	Bus Stop Schedule Relationship	Relationship to the schedule of each bus stop. Default is 'SCHEDULED'. Other values include 'NO_DATA', and 'SKIPPED'.
request_time_dt	Data Request Time	The request time of each data query.
timestamp_dt	Update Time	The time when the information was last updated in the system.
stop_arrival_time_dt	Estimated Bus Arrival Time	The estimated arrival time at the bus stop based on the scheduled time plus the reported delay.
stop_departure_time_dt	Estimated Bus Departure Time	The estimated bus departure time at the bus stop based on the scheduled time plus the reported delay.
trip_start_time_dt	Bus Trip Start Time	The scheduled departure time from the first stop of the trip.

The data product is organized into two formats:
`12_CSV_Transformed_TU` and `13_Cleaned_Daily_TU`, along with a README.txt file, totaling 518 GB.

The `12_CSV_Transformed_TU` format is organized in monthly folders with a naming pattern of `Transformed_TU_YEAR_MONTH` and has a total size of 498 GB. Each month contains daily folders named `gtfs_tu_YEAR_MONTH_date`. These daily folders contain minute-level GTFS-Realtime Trip Updates, with filenames following the pattern `gtfs_tu_YEAR_MONTH_HOUR_MINUTE.csv.gz`. This format is the intermediate data product, primarily designed to track changes to the trip updates every minute.

A further-processed data product is stored in `13_Cleaned_Daily_TU` with a total size of 20GB. The `13_Cleaned_Daily_TU` files include only the most recent GTFS-Realtime Trip Updates recorded before the bus arrived at the stop, ensuring the most accurate updates. This data format is organized into monthly folders named `YEAR_MONTH`. Within each monthly folder, daily Trip Updates files are provided, containing the latest arrival and departure updates for each bus in the system, named `gtfs_tu_YEAR_MONTH_all.csv.gz`. This format is the final data product, which is recommended for uses which pertain to evaluations of the realised bus performance .

All of the data products have the same columns, which are listed in Table 3.2. The column 'id' and the column 'vehicle_id' have the same value when the realtime update is provided. When realtime updates are missing, the 'vehicle_id' is blank. Each route, trip, and vehicle have their own ID. A route ID maps to multiple trip IDs. Trip IDs repeat over days if trips have the same features, such as stopping patterns, stopping times, start times, and other elements. Each vehicle can service different trips or routes. The arrival delay and departure delay are estimated based on an undisclosed algorithm from TfNSW, which considers both historical arrival and departure times as well as realtime operation status. It is worth noting that the arrival delay and departure delay include an estimated dwell time within the dataset. This can differ from other regions where the arrival and departure times are identical in both scheduled and real-time data.

Besides delay, there are two ways the data indicates deviation from the intended schedule. The 'trip_schedule_relationship' indicates whether the scheduled trip occurred, whereas the 'stop_schedule_relationship' indicates whether an individual stop was skipped in that trip. There can be columns with 'SCHEDULED' in 'trip_schedule_relationship', but 'SKIPPED' in 'stop_schedule_relationship', indicating that the trip occured according to the schedule, but this stop is skipped. The 'timestamp_dt' is the reporting time of the most recent information in the trip update, whereas the 'request_time_dt' is the time when the relevant query was made to the API. The 'request_time_dt' should always be later than or equal to the 'timestamp_dt'.

3.3.2 Data cleaning

After reading the protocol buffer messages with the appropriate bindings, the data needs to be cleaned of errors. There are three types of errors found and cleaned: missing columns, missing values, and data duplicates. Missing columns and data duplicates are found to be extremely rare. When there are missing columns in a protocol buffer message, the file is skipped. When data duplicates exist, only the last row of the duplicated data is kept.

The most common error is missing data. This is first checked by the column ‘stop schedule relationship’. If the cell in this column indicates ‘NO_DATA’, then this data entry is removed. The other column ‘trip schedule relationship’ is checked afterward. The cells in this column have six types: ‘NO_DATA’, ‘CANCELED’, ‘UNSCHEDULED’, ‘0’, ‘SCHEDULED’, and ‘ADDED’. The types ‘NO_DATA’, ‘CANCELED’, ‘UNSCHEDULED’, and ‘0’ are considered missing values and are removed.

The most frequent error type is ‘NO_DATA’ in the column ‘stop schedule relationship’, which accounts for 40-50% of each protocol buffer message. This high percentage is because the GTFS-Realtime data reports delays over a longer time span than the actual trip time, including periods before the trip start time when there is no relevant realtime information available. Since the main data product (‘13_Cleaned_Daily_TU’) only contains the latest update, these pre-start records are irrelevant for data size and quality. The other three missing data types (‘CANCELED’, ‘UNSCHEDULED’, ‘0’) in the column ‘trip schedule relationship’ generally account for less than 0.05% of the total data missing from the whole file. The remaining records are processed using the data processing pipeline.

3.3.3 One-day data sample

The dataset presented in this paper has 25 months of data, and is approximately 0.5 TB. For ease of access, a one-day data sample is provided to allow potential users to explore the data structure without downloading all of the data. This one-day sample contains data from 1 June 2021 for both formats `12_CSV_Transformed_TU` and `13_Cleaned_Daily_TU`. The `12_CSV_Transformed_TU` sample is stored under the folder `one_day_sample_Transformed_TU_2021_06_01`, and the `13_Cleaned_Daily_TU` sample is stored under the folder `one_day_sample_Transformed_13_cleaned_daily_TU`. The one-day data sample is uploaded to the same repository and follows the same structure as the original dataset.

3.4 Technical validation

The input to the data processing pipeline is a set of query results from the GTFS Trip Updates API. As the Trip Updates contain short-term forecasts of arrival and departure times, the quality of the data processing pipeline output depends on both the quality of these forecasts as well as the accuracy of the data processing pipeline.

To ensure the quality of the forecasts as inputs to the data processing pipeline, the data processing pipeline output is manually checked against in-person observations of bus arrivals. Bus arrivals were observed at two bus stops during the evening peak hour on 3 April 2024. The stops are located on a busy arterial corridor serving both local and express buses, and 70 buses were observed in the data collection period. A summary of the differences between the observations and the data processing pipeline output is presented in Table 3.3.

Overall the observed arrival and departure times are close to the forecasted arrival and departure time from the GTFS-Realtime Trip Updates. With the one minute update frequency which is used to generate the dataset, it is found that on average the observed arrivals were around 8 seconds later and the departures were around -5 seconds earlier than the forecasts reported in the data processing pipeline outputs. The difference in update frequency will be discussed in the section below.

The standard deviation of the difference is less than one minute for all update frequencies less than 7 minutes. The lower standard deviation in the arrival time differences indicates that the data processing pipeline output is more reliable at estimating arrival time than departure time, which is expected because of the difficulty of forecasting dwell time associated with boarding and alighting. During the observation period, there were ticket inspectors at one

TABLE 3.3: Summary of the differences between observed and pipeline-reported stop arrival and departure times for 70 buses. Positive values in the difference imply that the observation occurred after the forecast. Both arrival and departure time forecasts are within a tolerance expected by most bus users.

	Observed - Pipeline Output			
	Arrival Time Estimates (s)		Departure Time Estimates (s))	
Update Frequency (mins)	Mean	Std	Mean	Std
1	7.96	15.13	-4.59	33.79
2	11.81	21.50	-0.74	37.29
3	12.07	27.35	-0.35	44.96
4	15.24	35.34	3.04	50.68
5	18.78	40.54	6.71	56.46
6	11.19	44.12	-0.63	57.73
7	14.22	49.66	1.99	64.75
8	16.16	43.15	4.54	56.21
9	22.41	53.52	10.72	63.85
10	29.79	81.97	19.91	90.94

stop and three buses were held for ticket checking— this delay would not have been reflected in the GTFS Trip Updates forecast.

Three out of the total 70 buses could not be matched during the observation period. These three buses were marked as ‘NO_DATA’ in the ‘stop schedule relationship’ column and were removed during the data cleaning process. This illustrates that incomplete data from the PTAs is a key weakness in the data processing pipeline. The validation against observed bus arrivals and departures demonstrates that the data processing pipeline outputs are reliable for constructing bus performance measures at the stop level (such as average waiting time) within a tolerance that would be acceptable for most bus users.

3.4.1 Query frequency

The frequency of queries is a crucial variable for scholars and practitioners utilizing the data processing pipeline. Despite its importance, the related discussion on query frequency is missing in the scholarly literature on GTFS-Realtime datasets. Setting a high query frequency can enhance data accuracy, ensuring that real-time updates are timely and precise.

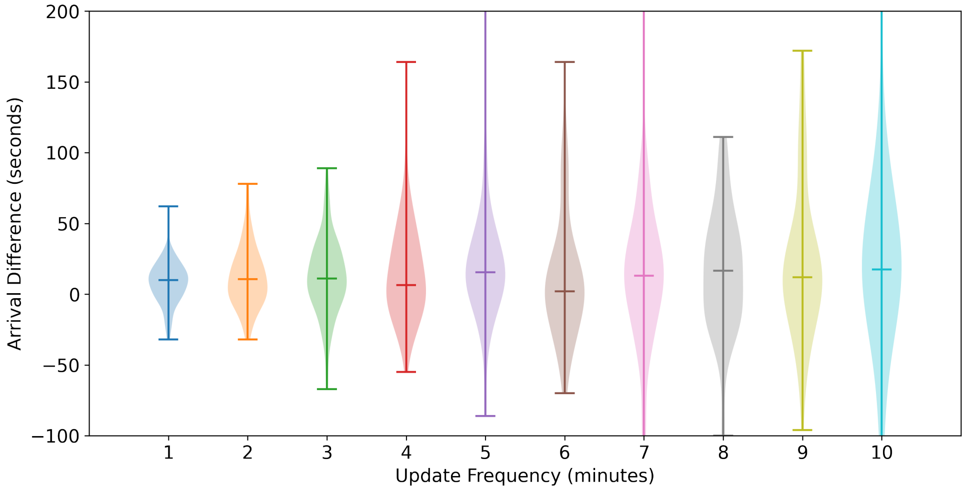


FIGURE 3.4: Violin plot of the difference between observed and pipeline-reported stop arrival time estimates based on different query frequencies. The variation in stop arrival time estimates increases as the query interval increases. The stop departure time estimates show a similar pattern.

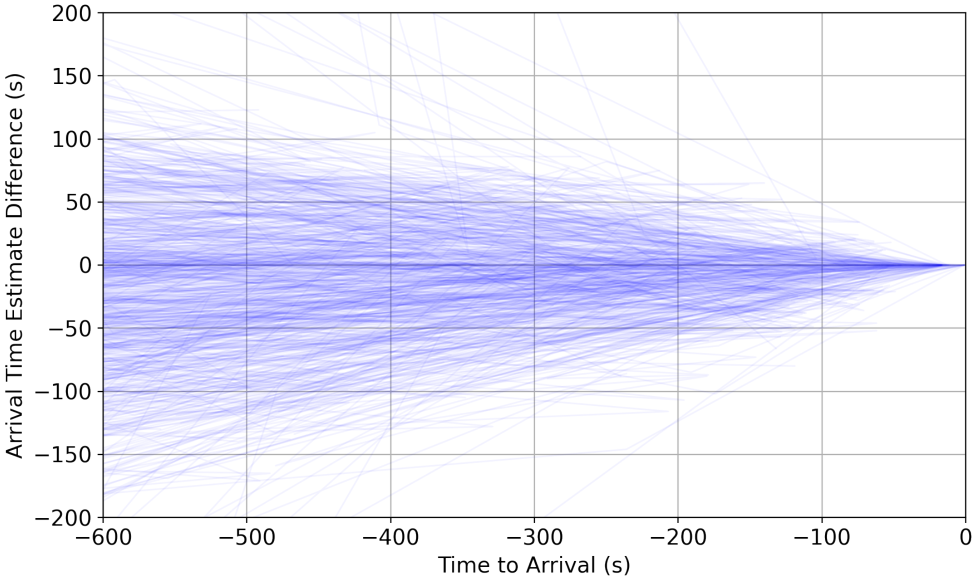


FIGURE 3.5: Stop arrival time estimate variation approaching the bus stop. As the buses approach the studied bus stop, the variation in stop arrival time estimates significantly reduces compared to the last stop arrival time estimates. The stop departure time estimates indicate a similar pattern.

However, this approach also generates a significant amount of redundant data, which can strain computational resources and data storage capacities.

This section aims to fill this gap by comparing different levels of query frequencies and their impact on data accuracy. A 1-minute query frequency refers to sending an API request every minute, whereas a 10-minute query frequency refers to sending a request every 10 minutes. It is important to note that query frequency does not correspond to the forecast horizon of the arrival or departure prediction; rather, it determines how often the system checks for the most recent updates. For example, a forecast may still predict arrival times 10 or more minutes into the future, regardless of whether the data is pulled every minute or every 10 minutes.

Lower query frequencies increase the likelihood of missing intermediate updates, particularly in rapidly changing operational conditions, and may result in fewer samples or outdated information being recorded. In contrast, higher query frequencies offer finer temporal granularity, capturing more update snapshots and improving the reliability of arrival and departure estimates. This comparison is intended to help users determine an appropriate query frequency that balances computational efficiency, data storage needs, and the desired level of temporal accuracy in the resulting dataset.

Table 3.3 and Figure 3.4 show how the difference between the observed bus arrivals and the pipeline output varies with query frequency. As the query interval increases, both the mean and standard deviation values increase. Even at the largest frequency, the mean difference is still less than 30 seconds, but the uncertainty increases to a minute and a half. This indicates that with fewer updates, the arrival times become noisy.

The changes in stop arrival and departure times are also compared internally within the GTFS-Realtime Trip Updates dataset using the intermediate data product from `12_CSV_Transformed_TU`. Figure 3.5 illustrates how arrival time estimates for stops evolve across multiple query timestamps, comparing each with the final prediction obtained just before the vehicle's actual arrival. This allows an assessment of how prediction accuracy improves as the vehicle approaches the stop. A key observation from the figure is that the arrival time estimates tend to stabilize as the bus nears the stop. This stabilization reflects a decrease in prediction uncertainty when the bus is physically closer to its destination, aligning with operational expectations. For example, under a 1-minute query frequency, the average time between when a prediction is recorded and the actual arrival time is approximately -30 seconds. At this short horizon, the predictions made within 30 seconds of arrival are highly accurate relative to the inherent variability in bus operations, demonstrating the accuracy of the dataset.

The selection of an appropriate API query frequency involves balancing data quality with computational and storage costs. In this study, a 1-minute query interval — the highest feasible frequency — is employed to ensure maximum temporal resolution and to support applications requiring precise arrival and departure estimates. While long-term data archiving is generally not a constraint — given the scalability of modern cloud storage solutions — storage still incurs financial costs. More critically, data processing imposes significant computational demands. In contrast, lower-frequency queries reduce the computational burden but risk diminishing the potential for retrospective analyses that rely on high-resolution data.

This exploration of query frequency considers 1 minute as the smallest interval and demonstrates acceptable data accuracy with that frequency. In fact, higher frequency queries are possible, but they are limited by the update frequency of the PTA's feeds. Vehicles push their updated locations to the PTA at semi-regular intervals of several seconds. This information is rapidly processed through a model to create the updated vehicle arrival times at future stops. This update becomes available for querying at the update time. The average gap between this update time and the query that retrieves it is 9 seconds for the presented dataset. This calculation implies that the query interval could be reduced to roughly twice the gap (18 seconds) in an attempt to improve accuracy, but anything more frequent would result in redundant queries to the date update information. It remains for future work to determine whether queries more often than once per minute would offer worthwhile improvements in the accuracy of the forecast.

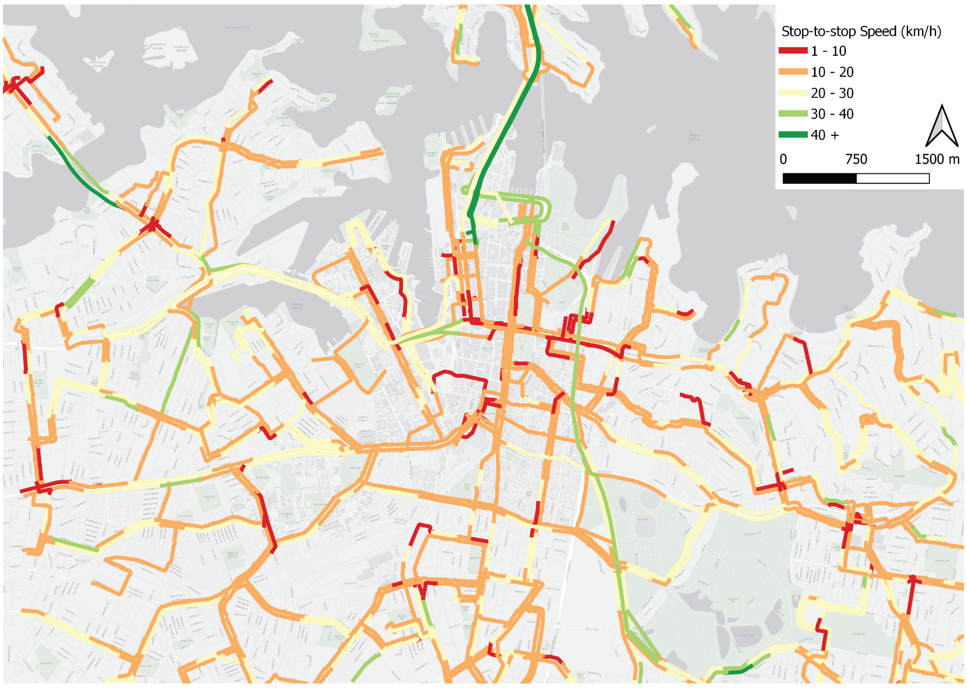
3.4.2 Data description

In the data processing pipeline output, each observation contains the arrival and departure times of a single vehicle at a single stop. This is a flexible format that acts as a foundation for many types of bus performance analysis. For example, the difference between arrival and departure time from the previous stop represents the travel time for that stop-to-stop segment. The distance between two stops can be obtained from the bus route trajectory shapefiles through the GTFS-Static data, which is available from the TfNSW Open Data Hub (Transport for NSW 2021b). Travel time and distance are combined to calculate the bus performance measures shown in Figure 3.6.

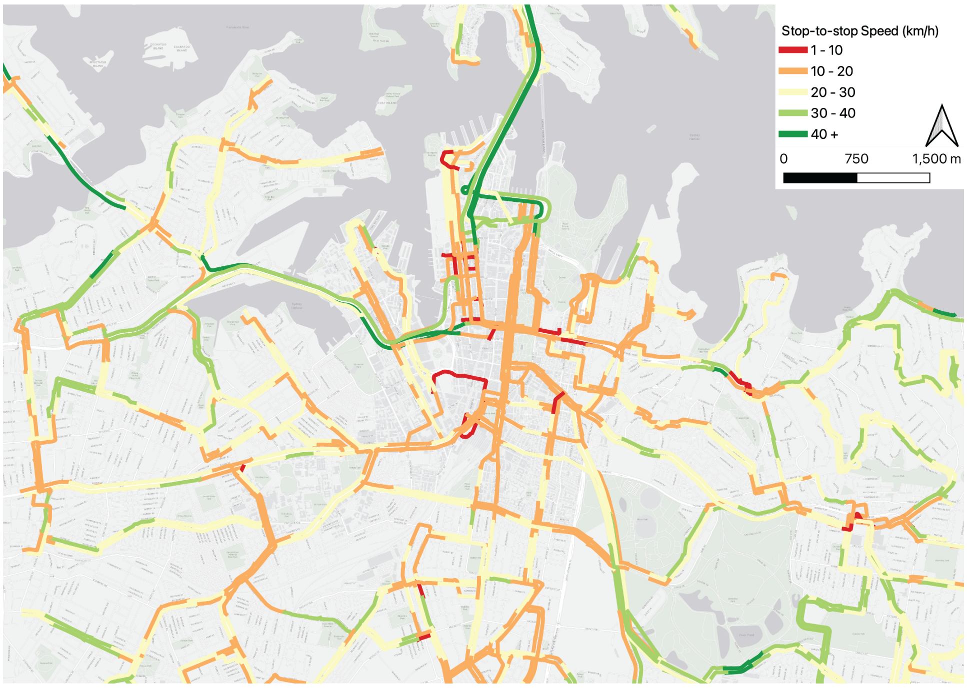
Figure 3.6 compare AM peak (6-10 AM) and off-peak (7 PM - 6 AM) speeds in March 2022 to visualise the dynamics and spatial resolution of the data. As expected, the AM peak period shows the slower bus operation speeds compared to Inter-Peak (10 AM - 3 PM), PM Peak (3 - 7 PM) and Off Peak (7 PM - 6 AM).

As buses approach the city center, the data shows a notable decrease in operational speeds with faster travel times in the contraflow direction. Notably, stop-to-stop segments equipped with bus lanes, like the southbound direction on the Sydney Harbour Bridge at the top-centre of the map, demonstrate faster operational speeds. Conversely, segments that involve turns typically display reduced speeds.

The off-peak period shows the highest bus operation speeds among all four time periods analyzed. Most stop-to-stop segments during this time show higher operation speeds than the AM peak as expected due to lighter traffic conditions. However, the city center still displays reduced operational speeds compared with other areas, reflecting the impact of higher traffic density, dense traffic signals, and lower speed limits. Consistent with observations from the previous plot, longer stop-to-stop segments that utilize motorways or bus lanes also demonstrate higher speeds. Overall, the spatial distribution of bus operation speeds, as estimated from the data processing pipeline output, aligns well with expectations. This suggests that GTFS-Realtime data processed in the pipeline supports detailed spatio-temporal analysis of bus performance.



(a) AM peak bus operation speed near Sydney CBD. Motorways and bus lane facilities show high operational speeds, whereas intersections and segments with turns have slower speeds.



(b) Off peak bus operation speed near Sydney CBD. Only the busiest parts of the CBD have slow speeds, and most of the longer segments are operating at the speed limit.

FIGURE 3.6: Comparison of bus operation speeds near Sydney CBD during AM peak and Off peak hours. Link-to-link comparison shows how performance changes over the day. Except in the CBD, the bus network is sparse compared to the road network. Base map: Esri Gray Light.

3.5 Usage notes

The data processing pipeline outputs can be used to evaluate bus delays at each stop measured to the second. For many public transport system performance interventions such as bus lanes or transit signal priority, the benefit arises from small time savings over a large number of vehicles and passengers. To measure these accurately and robustly, we need high resolution, vast data on speeds and delays. The data processing pipeline outputs offer a flexible, user-friendly format for high-resolution transit system performance measurement, which is beneficial for public transit operators and authorities.

This dataset was created by processing queries to the TfNSW GTFS-realtime Trip Update v1 API. The API is under a Creative Commons Attribution license which allows users to re-use and re-distribute the data.

3.5.1 Opportunities for improvement

The development of this data processing pipeline relied on the researchers' understanding of the engineering and operations of the public transport system. Data cleaning and processing was based on relevant insights into why the data could be recorded in misleading ways. Future development could take advantage of machine learning-powered anomaly detection techniques to improve data quality without relying on domain expertise in public transit operations.

The GTFS-Realtime Trip Updates do not provide information on passenger boarding, meaning the dataset cannot examine the effects of boarding time or crowding. However, the GTFS-Realtime Vehicle Positions feed includes a high-level indicator of passenger occupancy, which can be processed and correlated with the GTFS-Realtime Trip Updates to create a more comprehensive dataset.

Some issues identified in the GTFS-Realtime v1 datasets as discussed in the GTFS data processing pipeline process above are expected to be resolved in the GTFS-Realtime v2. Most of these issues are caused by the optional fields in GTFS-Realtime v1. For instance, only seven (7) data fields were required out of the 63 data fields. The optional fields allowed some critical information to be missing but still reported in the data feed, which affects data quality (Barbeau 2017). As PTAs undertake the shift towards GTFS-Realtime V2, it will be necessary to refine the current GTFS data processing pipeline to take into account the improvements.

The current GTFS data processing pipeline could be expanded to include other data standards, such as SIRI European Committee for Standardisation (2021) to serve more locations and PTAs.

The data processing pipeline has been primarily developed using NSW's bus data, and tested against six cities in Australasia. The data processing pipeline could be tested on a wide range of feeds to ensure it works elsewhere in the world. This process could result in the creation of additional datasets to share within the data warehouse.

3.6 Conclusion

This study demonstrates the development and technical validation of GTFS-Realtime Trip Updates data in Greater Sydney, which includes bus arrival and departure time estimates. The outputs provide a reliable and flexible foundation for high-resolution bus performance analysis, offering valuable insights into operational efficiency at the stop level. Validation against in-person observations shows that the pipeline's forecasts closely align with actual observed bus arrivals and departures, with small tolerances (8 seconds for arrival and -5 seconds for departure), making it suitable for assessing bus performance at the stop-to-stop and second-level granularity.

In conclusion, the data processing pipeline and the associated dataset offer a powerful tool for public transport authorities and researchers to evaluate bus delays, monitor service reliability, and inform the design of interventions such as bus lanes and transit signal priority. The implications of these results are discussed in Chapter 7. Building on the output of this chapter, Chapter 4 examines the impact of route characteristics — such as bus lanes, traffic volume, bus cross-traffic turns, and traffic signals — on stop-to-stop marginal delay at the second level.

Chapter 5 further investigates the effect of bus cross-traffic turns on delay, reliability, and speed. This study serves as a step toward more data-driven approaches to improving urban transport systems globally.

CHAPTER 4

High Resolution Bus Lane Performance Evaluation from Real-time Update Data

---

4.1 Introduction

This chapter employs the GTFS-Realtime data from the Chapter 3 to measure the effects of these variables at a microscopic level, specifically analyzing each bus-stop-to-bus-stop link to gain deeper insights. Quantifying the microscale benefits of bus priority treatments has historically relied on AVL data that was not publicly available. The emergence of GTFS-Realtime data now allows for the validation of expected benefits from bus lanes, such as micro-delay reduction at the seconds level. This data, which combines on-board GPS equipment with geofenced stops (Transport for NSW (TfNSW) 2018), enables the evaluation of delays relative to the schedule for every bus arrival. Such a detailed level of data is essential for a thorough evaluation of bus operational performance and for quantifying the advantages or disadvantages of priority measures.

Delay relative to schedule is an important indicator of bus operational performance, directly affecting both the experience of riders and the management of transit operations. The variability in delay is one measure of reliability, allowing passengers to gauge the need for buffer times in their travel plans. In this chapter, the mean and standard deviation values of delay are used as performance indicators to evaluate bus operational performance on stop-to-stop segment basis within time rolling windows of 30 bus arrivals. Regression models are employed to analyze the impact of various bus stop-to-stop characteristic factors on these performance indicators, enabling a comprehensive analysis of their effects.

4.2 Methodology

In order to quantify the benefit of bus lanes on bus performance, this chapter defines a relevant measure of delay that varies spatially and temporally in the dataset. This measure is combined with other properties of the system in panel regression models to predict delay and variability with and without bus lanes.

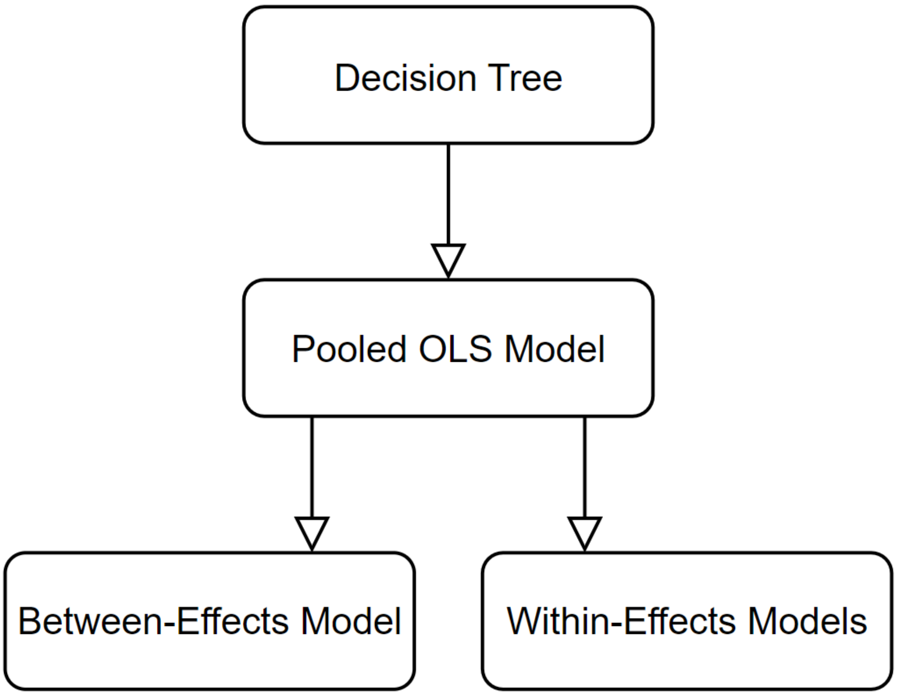


FIGURE 4.1: Modelling flowchart illustrating the logic of how the techniques build upon each other.

Four different model techniques are utilized in this chapter (Figure 4.1). Initially, a decision tree is employed to understand the data and to guide the selection of variables for the linear regression analysis. Pooled OLS models are then used to interpret the effects of various design elements on bus performance. Given the expected correlation in error terms arising from repeated measures for the same stop-to-stop segments or time periods, panel regression techniques, including between-effects and within-effects models, are applied.

4.2.1 Stop-to-stop marginal delay

The performance measure used in this chapter is marginal delay — the incremental delay accrued between one stop and the next. Delay is the difference between the scheduled and actual arrival time and can be positive (behind schedule) or negative (ahead of schedule). The marginal delay at stop i is defined in Equation 4.1.

$$M_i = A_i - D_{i-1} \tag{4.1}$$

where:

M is the stop-to-stop marginal delay is seconds, A_i is the arrival delay at the next stop and D_{i-1} is the departure delay at the previous stop.

TABLE 4.1: Stop-to-stop marginal delay calculation example. Marginal delay refers to the delay increase between two successive stops. Marginal delay reflects the incremental delay changes at the stop-to-stop link due to the link characteristics.

<table>
<tr><th>Bus stop ID</th><th>Arrival delay (s)</th><th>Departure delay (s)</th><th>Marginal delay (s)</th></tr>
<tr><td rowspan="2">200000</td><td rowspan="2">5</td><td rowspan="2">8</td><td>-</td></tr>
<tr><td rowspan="2">-8</td></tr>
<tr><td rowspan="2">200001</td><td rowspan="2">0</td><td rowspan="2">0</td></tr>
<tr><td rowspan="2">12</td></tr>
<tr><td rowspan="2">200002</td><td rowspan="2">12</td><td rowspan="2">-5</td></tr>
<tr><td>-</td></tr>
</table>

This is the appropriate measure because it directly reflects the operational characteristics of individual stop-to-stop segments — specifically, the portion of the bus route after departure from the previous stop and prior to arrival at the following stop. By isolating this segment, the marginal delay can be more accurately attributed to features along the link, such as the number of traffic signals, the presence or absence of dedicated bus lanes, or the effects of cross-traffic turns. For example, if the departure delay at the previous stop is 8 seconds and the arrival delay at the next stop is 0 seconds, the marginal delay for that segment is –8 seconds, indicating that the bus made up time along that segment. Conversely, if the departure delay is 0 seconds and the arrival delay is 12 seconds, the marginal delay is 12 seconds, reflecting a deterioration in schedule adherence over that segment (see Table 4.1). As such, marginal delay serves as a key performance indicator that quantifies how each segment contributes to the overall reliability and on-time performance of the bus service. This link-level perspective is particularly valuable for understanding localized sources of delay and identifying potential infrastructure or operational improvements.

In order to evaluate how the attributes of a stop-to-stop link contribute to delay, we can consider both the mean and the standard deviation of delay for the set of buses traversing that link in an interval. Since timetables should be adjusted to reflect the benefits of priority measures like bus lanes, the mean of the marginal delay should not be well-explained by the presence of these measures. However, other priority measures such as transit signal priority might result in a positive or negative net delay due to the dynamic nature of traffic lights. A better measure of the impact of bus lanes is the standard deviation of marginal delay which measures the unreliability of that segment. The goal is to find attributes, such as the presence of bus lanes, that accurately predict reliable performance.

4.2.2 Regression tree

Regression tree is a machine learning approach to predicting continuous variables such as the mean or standard deviation of marginal delay. The unsupervised process identifies splits in the independent variables that result in the highest distinction between associated dependent variables. Regression trees are generally more accurate predicters than linear regression analysis but don't always have interpretable results. Therefore, regression trees are used before the linear regression analysis to estimate the importance of each variable in predicting the dependent variable. This approach allows for the quantitative comparison of the importance of various factors on bus performance and testing their significance. In the analysis below, regression trees were used to select the variables for the panel regression model specification and to support the interpretation by grouping variables with similar importance.

TABLE 4.2: Variables used in panel regression models. Large number of variables are used to reflect time and spatially varying attributes.

Symbol	Name	Units
Δ_{it}	Marginal delay change at link i time t	s
α_i	Link-specific effects	
β	Fitted coefficients	
B_{it}	Bus-taxi lane length on link i at time t	km
H_{it}	Bus-HOV lane length on link i at time t	km
S_i	Number of traffic signals on link i	
D_t	Traffic flow at time t	10^3 veh/(hour× lane)
C_i	Cross-traffic turn on link i	
T_{it}	Scheduled travel time on link i at time t	s
V_{it}	Scheduled travel speed on link i at time t	km/h
L_i	Length on link i	km
R_t	Precipitation at time t	mm
W_t	Weekend at time t	
P_t	Public holiday at time t	
S_t	School holiday at time t	
I_t	COVID stringency index at time t	
ϵ_{it}	Pooled OLS model error term	
ϵ_i	Between OLS model error term	
ϵ_t	Within OLS model error term	

4.2.3 Regression analysis

Using regression, this chapter quantifies the factors affecting bus performance and their statistical significance. Link-level performance relative to the schedule and link-level reliability are modeled using the mean and standard deviation of marginal delay, respectively, as dependent variables. These metrics are calculated within rolling windows of 30 bus arrivals. The reliability is especially important for evaluating the importance of bus priority measures since the timetable should be appropriately adjusted to account for the net effect of any priority measure, so the reliability of that link is a more important measure of the priority measure's benefit.

The Pooled OLS model specification for the mean value of stop-to-stop marginal delay is described in Equation 4.2, and the Pooled OLS model specification for the standard deviation value of stop-to-stop marginal delay is described in Equation 4.3. Both models consider link attributes, including bus lanes in operation, traffic lights, traffic flow, scheduled travel time, scheduled travel speed, stop-to-stop link length, cross-traffic turns, precipitation, school holidays, public holidays, and COVID stringency index (Table 4.2). The traffic flow, scheduled travel time, and scheduled travel speed are designed to capture the schedule effects. Due to the intercorrelation between these three variables, the top two most significant variables are kept to eliminate the schedule effect in the two Pooled OLS models. These Pooled OLS models assume independence of the error terms, ϵ_{it}.

$$\begin{aligned}\Delta_{it,mean} = {} & \beta_B B_{it} + \beta_H H_{it} + \beta_S S_i + \beta_D D_t + \beta_C C_i + \beta_T T_{it} + \beta_L L_i \\ & + \beta_R R_t + \beta_W W_t + \beta_P P_t + \beta_S S_t + \beta_I I_t + \epsilon_{it}\end{aligned} \tag{4.2}$$

$$\begin{aligned}\Delta_{it,std} = {} & \beta_B B_{it} + \beta_H H_{it} + \beta_S S_i + \beta_D D_t + \beta_C C_i + \beta_T T_{it} + \beta_V V_{it} \\ & + \beta_R R_t + \beta_W W_t + \beta_P P_t + \beta_S S_t + \beta_I I_t + \epsilon_{it}\end{aligned} \tag{4.3}$$

Panel regression techniques address correlation in the error terms arising from repeated measures for the same stop-to-stop segments or time periods (Baltagi 2008). By incorporating similar variables to those used in an OLS regression, panel techniques address correlated errors through modeling between effects and within effects. Although panel regression can yield better results when correlated error terms are significant, many variables have to be excluded due to the limits in data, potentially leading to specification bias.

A between-effects model regresses the mean of the standard deviation value of each stop-to-stop segment, removing the time-series aspect of the data, as shown in Equation 4.4. Consequently, all time-varying attributes are excluded from the model. In the context of bus lanes, this model quantifies how the spatial extent and operational duration of the bus lane impact the average performance of the segments.

$$\Delta_i = \alpha_i + \beta_B B_{it} + \beta_H H_{it} + \beta_C C_i + \beta_T T_{it} + \beta_V V_{it} + \epsilon_i \tag{4.4}$$

A within effects model is an OLS regression within each segment, converting the modelling to a set of time series without cross-sectional modeling. The results should be analyzed through the distribution of the resulting coefficients. In this chapter, a within model has been created for each stop-to-stop link with the standard deviation value as the dependent variable, as shown in Equation 4.5. Only the time-varying attributes are retained, including the time-varying bus lanes. It is worth noting that stop-to-stop segments with no time-varying bus lanes, including those with 24-hour bus lanes and no bus lanes, are dropped. In the context of bus lanes, this model quantifies how delay changes during versus outside the bus lanes' hours of operation while controlling for other time-varying attributes like traffic volume.

$$\begin{aligned}\Delta_t = \; & \beta_B B_{it} + \beta_H H_{it} + \beta_D D_t + \beta_T T_{it} + \beta_R R_t + \beta_W W_t \\ & + \beta_P P_t + \beta_S S_t + \beta_I I_t + \epsilon_t\end{aligned} \tag{4.5}$$

4.2.4 Study area

This chapter analyzes the buses in Sydney, New South Wales between 1 November 2020 and June 2022. The extent of the study area, shown in Figure 4.2 isolates the majority of the bus lane implementations and focuses on those regions where diurnal tidal flow from and towards Sydney Central Business District (CBD) dominates the bus services. Areas with potential variable influences on this flow, such as secondary employment hubs like Macquarie Park, Parramatta, and Chatswood, are avoided. Specifically excluded is the orange zone within Sydney CBD, due to its significantly different traffic conditions. Within the designated study area, numerous road corridors feature a variety of bus priority measures, including segments with bus-taxi lanes and bus-HOV lanes. This research investigates all stop-to-stop segments in the area where both scheduled and actual speeds range between 1 and 100 km/h.

4.3 Data for bus lane performance evaluation

This research utilizes various datasets from NSW to assess the impact of bus priority lanes on stop-to-stop marginal delay (Table 4.3). The primary dataset is derived from the TfNSW

GTFS-Realtime API, which provides real-time trip updates for calculating stop-to-stop marginal delay. The GTFS-Static dataset combines spatial information (such as bus trajectories and stop locations) with temporal details (such as schedules and stop sequences), enabling the analysis of GTFS-Realtime data and defining the characteristics of stop-to-stop routes.

The Clearway and Signalised Intersections datasets, which contain spatial route details, are used to determine the route characteristics of between bus stops. Hourly data is sourced from the Traffic Volume Viewer and Precipitation datasets, while daily information is drawn from the School Holiday, Public Holiday, and COVID Stringency Index datasets. Both hourly and daily data are employed to analyze temporal effects on bus performance within the study area.

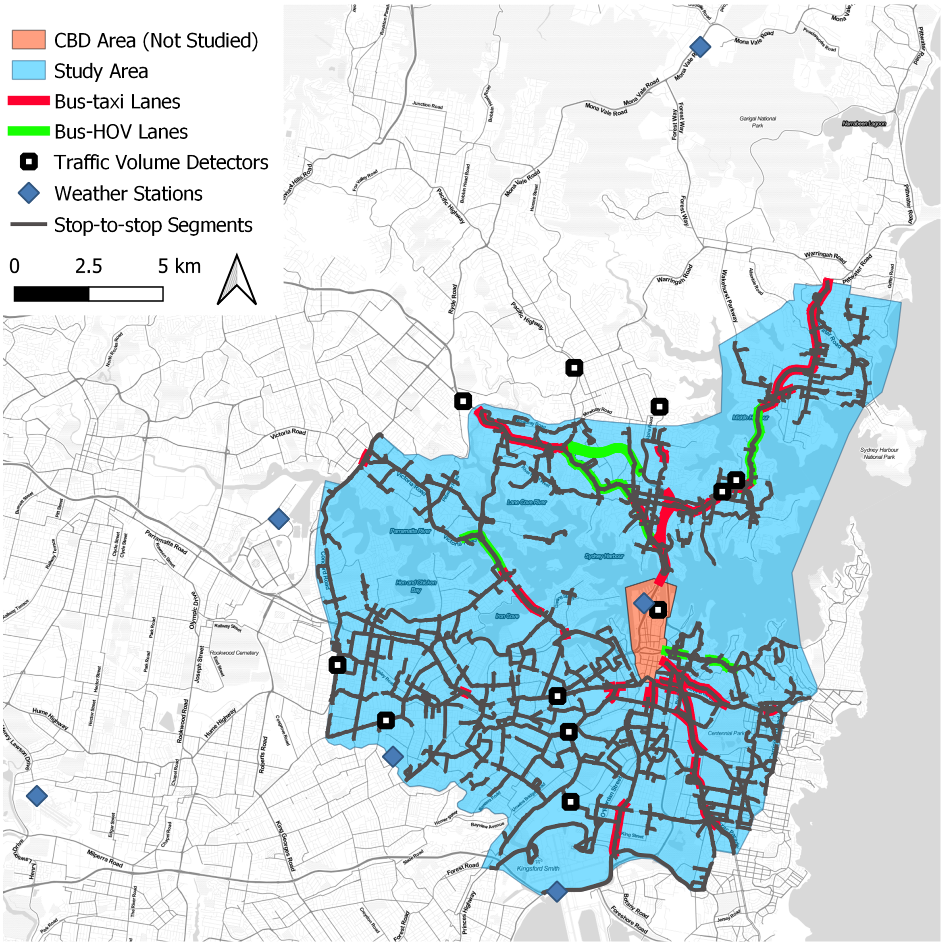


FIGURE 4.2: Studied area: Suburbs around Sydney CBD. Tidal flows are significant during peak hours within the studied area.

TABLE 4.3: Data source and information contained. Various data sources are used in this chapter to support the analysis. These data sets are conjoined to perform the regression analysis.

Data Set	Source	Information Contained
GTFS-Static	Transport for NSW	Route Trajectory Schedule Bus Stop Location and Sequence
GTFS-Realtime	Collected using Transport for NSW's Open API	Arrival Delays
Clearway Data	Transport for NSW	Bus Lane Location Operating Hours
Signalised Intersections Data	Transport for NSW	Location of traffic signals
Traffic Volume Viewer	Transport for NSW	Network Hourly Flow
Precipitation	Australian Government Bureau of Meteorology	Precipitation precipitation
School Holiday	NSW Government	NSW Public School Holidays
Public holiday	Python-Holidays Library	NSW Public Holidays
COVID Stringency Index	Oxford Coronavirus Government Response Tracker (OxCGRT)	COVID Effects

4.3.1 GTFS

4.3.1.1 GTFS-Static

As mention in Chapter 3, the GTFS, commonly referred to as GTFS-Static, is a standardized format designed for public transport schedules and related geographical data. This dataset encompasses time-varying information, such as bus stop arrival schedules and stop sequences, along with spatial data, including bus route trajectories and stop locations (Google Developers 2022b). In this chapter, GTFS-Static data, which aligns with the real-time data, was obtained from the TfNSW's Open Data Hub. This dataset is crucial for processing GTFS-Realtime data, ensuring an accurate spatial match between the calculated marginal delay and the corresponding stop-to-stop segments.

4.3.1.2 GTFS-Realtime

The output from Chapter 3, the GTFS-Realtime dataset, is designed to be used alongside GTFS-Static and contains real-time operational information for public transport. This chapter utilized GTFS-Realtime Trip Updates, which provide arrival time estimates at the time of the query (Google Developers 2022a). By comparing this real-time data with the scheduled bus performance, the stop-to-stop marginal delay can be assessed. The data was subsequently cleaned and processed into a daily record of actual bus arrival and departure times at each bus stop for this chapter.

In this chapter, both realtime arrival and scheduled travel time data are extracted from the GTFS-Realtime dataset. The realtime arrival data is used to calculate the stop-to-stop marginal delay, while the scheduled travel time, which provides the expected travel time between bus stops based on the official schedule, is included as one of the variables in the panel regression analysis.

The GTFS-realtime query results were processed into 136,849,872 updated arrival times, which were then aggregated by stop-to-stop segments using a time rolling window of 30 consecutive readings. Consequently, each result represents the average value of the 30 nearest readings in time for each stop-to-stop link. Given the large sample size, every fifth reading was retained to reduce data volume. Each reading provides both mean marginal delay and standard deviation measurements within the rolling window. These mean and standard deviation values serve as dependent variables in regression analysis. The data analysis process in this chapter involves the merging of multiple datasets. Many of the datasets used in this chapter contain some missing data. To ensure that each row includes data from all sources, any row with a missing data cell has been dropped, resulting in a reduction of the total data size.

The mean marginal delay quantifies the bus's on-time performance for each specific stop-to-stop link and within designated time rolling windows. The marginal delays presented in this section are calculated in seconds, representing the incremental delay accumulated between two consecutive stops.

The standard deviation of marginal delay serves as an indicator of bus travel reliability for each unique stop-to-stop link and time rolling window. Positive coefficient values suggest that a factor increases the standard deviation of marginal delay, indicating higher unreliability in bus operations. Conversely, negative coefficient values imply that a factor reduces the standard deviation of marginal delay, signifying improved reliability in bus operations.

The studied data contains repeated observations of both the cross-section (stop-to-stop links) and time periods (the same time rolling window of the same day), so we expected correlation in the error terms of the regression. As a result, panel regression is used to model the studied data to address the panel effects.

4.3.2 Time varying attributes

4.3.2.1 Clearway data

In NSW, clearways are sections of road where stopping and parking are prohibited to improve traffic flow, typically during peak times (Transport for NSW 2021a). The clearway dataset available from TfNSW's Open Data Hub includes information about bus lanes, as curbside parking restrictions are enforced when bus lanes are active. This dataset contains details about two types of bus lanes: Bus-taxi lanes and bus-HOV lanes. Bus-taxi lanes can be used by buses, taxis, and emergency vehicles. bus-HOV lanes, designated for HOVs such as T2 and T3 lanes (lanes for vehicles with at least two or three passengers, respectively), also permit buses. Although HOV lanes primarily serve high-occupancy vehicles, they provide significant benefits to buses by offering dedicated transit space.

The bus lane operation hours for each studied time period were recorded for analysis. These hours, along with clearway operating hours, are stored in unstructured text fields that often contain duplicate entries and conflicting information. The clearway dataset was cross-checked with Google Maps Street View in the studied area to ensure its reliability and accuracy. The hours of operation were re-coded to binary variables for four study time periods for weekdays: Morning Peak (6 - 10 AM), Between Peak (10 AM - 3 PM), Evening Peak (3 - 7 PM), and 24/7. These bus lane operation time periods are based on TfNSW's definition. Off-peak bus lanes are designated bus lanes that remain operational at all times. Since there were no bus lanes that only operate on the weekends in the study area, any marginal delay measurements from the weekend are matched to the Off Peak bus lane operation status for that location.

4.3.2.2 Signalized intersections data

The signalized intersections data, sourced from the TfNSW Open Data Hub, includes details such as the geographical locations and installation dates of signalized intersections in NSW. This dataset is spatially joined to the route shapes segmented into stop-to-stop sections, providing a count of the traffic lights encountered along each stop-to-stop link.

4.3.2.3 Traffic volume viewer

Demand for the bus system is highest during the morning and evening peak periods as measured by frequency of service or number of boardings. This high demand can have a detrimental impact on on-time performance including unreliability related to boarding and alighting times. But the peak period is also associated with high traffic volumes, which reduce rolling speed between stops especially on segments without bus lanes. Figure 4.3 indicates the distribution of traffic flows across a set of fixed-counter locations in peak and off-peak hours. Since many bus lanes only operate in the peak hours, it's clear that a model specification must include variables for bus lane operation as well as traffic volume in order to avoid an ambiguous interpretation. Moreover, because the distributions overlap, it's clear that a peak-hour indicator variable is not sufficient to capture detailed impact of traffic congestion on bus performance.

The Traffic Volume Viewer provides archived traffic volumes at sparse traffic counter locations within NSW. In this chapter, the average hourly traffic flows per lane across the studied area in Sydney are computed using data from 11 monitoring sites (Figure 4.2). The traffic counters are grouped by whether they record traffic volumes in the citybound, outbound or both directions.

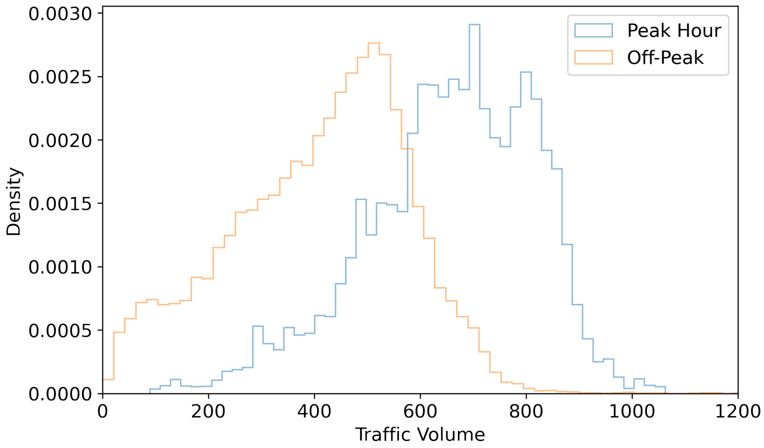


FIGURE 4.3: Traffic flow distribution during peak hours and off-peak Hours. Peak hours and off peak hours have different traffic flow distribution. This ensures that traffic flow used is a suitable indication for traffic demand within the studied area.

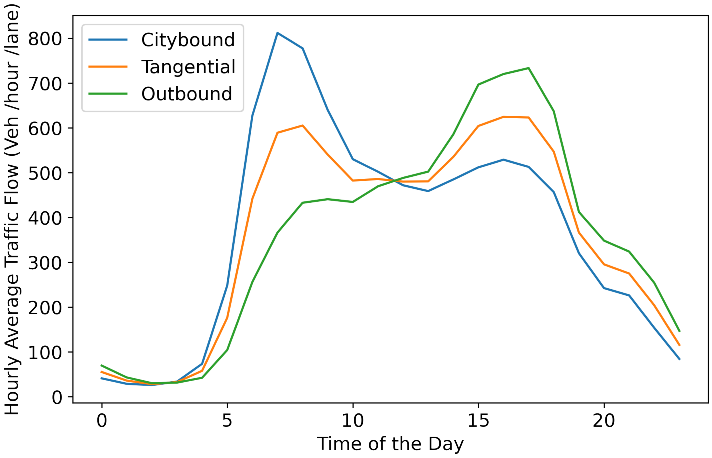


FIGURE 4.4: Traffic volume data on 01 June 2022. Traffic flow is more critical during peak hours, particularly in the expected direction of travel for that peak. This sample demonstrates that the traffic volume captures expected demand patterns in the study area.

Each stop-to-stop segment is flagged as belonging to either citybound, outbound or tangential, and is matched with the average of the corresponding traffic counters. For tangential stop-to-stop segments, the average traffic flow between the two directions is considered. This is represented by the orange line depicted in Figure 4.4. To facilitate further analysis, the average hourly flows are joined with the stop-to-stop marginal delay data, based on time and direction. Figure 4.3 showcases a notable disparity between peak hour and non-peak hour traffic volumes. This observation suggests that the Traffic Volume Viewer data used in this project is essential to capture the time-varying ambient conditions on the road network.

4.3.2.4 Precipitation

The precipitation data comes from the Australian Bureau of Meteorology. This dataset captures precipitation measurements at five weather stations located within or in close proximity to the area of interest. For each hour, the average precipitation value from the selected weather stations is averaged and used as an indicator of precipitation. The region-wide hourly precipitation values are matched with the stop-to-stop marginal delay data for that hour.

4.3.2.5 School holiday

School holidays are defined as periods when schools are closed and students do not commute to school. School holidays generally result in a reduction in traffic volume and changes in behavior associated with varied routines for families (Tao et al. 2018). The school holiday data utilized in this chapter is generated based on information sourced from the NSW Department of Education website. In NSW, the school holidays include two weeks in approximately April, July and October plus an extended holiday from late December to the end of January. This includes both weekdays and weekends within the designated school holiday period. The school holiday data is joined with the remaining datasets by date.

4.3.2.6 Public holiday

The public holiday data used in this chapter is obtained from the Python-holiday library. This library provides functionality for generating sets of holidays designated by the government for different countries and subdivisions. In this chapter, the NSW subdivision is used (*python-holidays — holidays documentation* 2023). The public holiday data is joined with the remaining dataset by date.

4.3.2.7 COVID stringency index

The study period includes times when COVID lockdowns caused significant drops in travel demand, leading to adjustments in bus services. To account for these variations, the COVID Stringency Index is used as a proxy for the intensity of lockdowns. This index, sourced from the Oxford Coronavirus Government Response Tracker (OxCGRT), is a composite measure derived from nine government response metrics. These metrics include school closures, workplace closures, cancellation of public events, restrictions on public gatherings, closures of public transport, stay-at-home requirements, public information campaigns, restrictions on internal movements, and international travel controls. Each metric is rated from 0 to 100, with higher scores indicating stricter responses. The Stringency Index is the mean score of these metrics on a given day, reflecting the severity of government measures against COVID (Mathieu et al. 2020). This data is integrated with the studied dataset by date to analyze the impacts of lockdown intensity on travel patterns.

4.3.3 Spatially varying attributes

Bus stop locations and route information are essential for generating trajectories for each stop-to-stop link. Shape files, which outline the complete travel path of each bus trip from origin to destination, are sourced from the GTFS-static dataset. To analyze bus performance between stops, these shape files are segmented at bus stop locations, creating multiple sections representing paths between adjacent stops. Each section forms a stop-to-stop trajectory. These trajectories are then matched with marginal delay measurements from the GTFS-Realtime dataset, enabling the inclusion of road link characteristics as detailed in the following subsections and illustrated in Figure 4.5.

4.3.3.1 Bus lane locations

Bus lane information for each stop-to-stop segment is obtained by spatially aligning priority lane data with the corresponding bus trajectories. To prevent errors in the spatial joins, such as merging westbound trips with eastbound bus lanes, the bearing is considered for both the bus

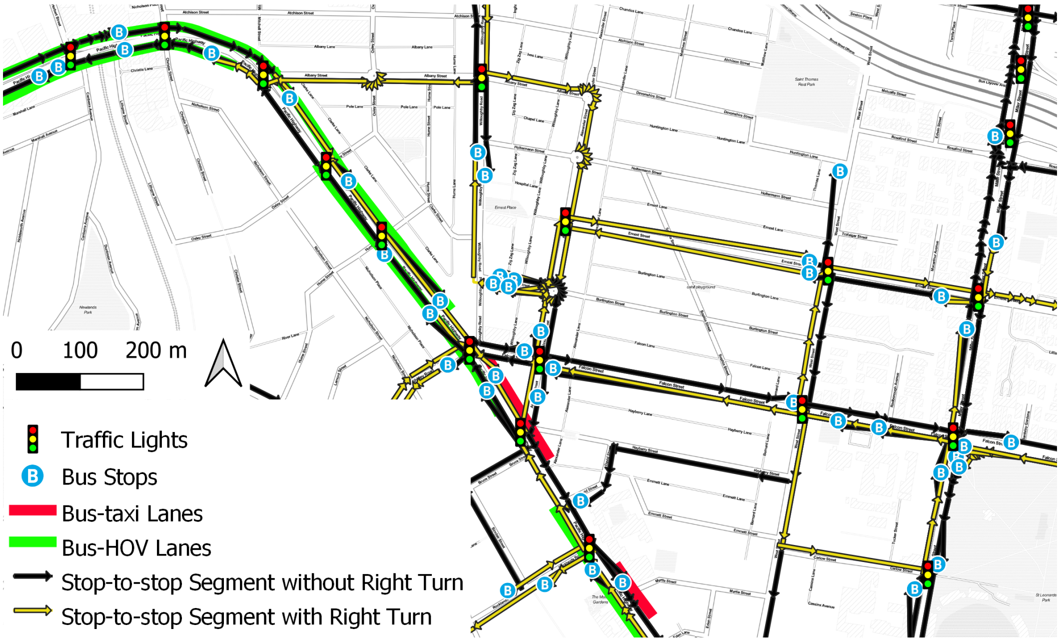


Figure 4.5: Road characteristics information at Crows Nest, located northwest of Sydney CBD. Spatial elements shown in this figure are spatially joined together, representing the stop-to-stop link characteristics. Due to the variety in bus directions and stopping patterns, overlaps of the stop-to-stop segments commonly occur on this map.

lane shape files and the stop-to-stop trajectories. Additionally, the operational hours of each bus lane are taken into account during this alignment to ensure they match the bus service times accurately. This process ensures that the effective length and type of bus lane for each stop-to-stop segment are recorded, reflecting the relevant operational hours.

4.3.3.2 Traffic lights

Traffic lights can impact on-time performance by introducing variability in travel times depending on the traffic signal phase and causing potential queuing during periods of high demand. To determine the number of traffic lights on each stop-to-stop segment, traffic light locations from the signalized intersections dataset are spatially joined with the stop-to-stop trajectories, and the total number of traffic lights encountered along each segment is counted. Transit signal priority, a common bus priority measure, necessitates a detailed assessment of the impact of traffic signals on bus performance. Although not implemented during this study period, the findings from this analysis can quantify the effect of unprioritized traffic signals on bus performance.

4.3.3.3 Segment length

Attributes of the stop-to-stop segment including length, scheduled travel speed and cross-traffic turn information are not directly available in any existing dataset. These variables are derived from the trajectories of each stop-to-stop link generated from GTFS-static dataset.

The length of each stop-to-stop link is determined based on the stop-to-stop trajectory. This length is calculated by measuring the length of the corresponding shapefile segment for the specific stop-to-stop link. The measurement is done in meters using the NSW Lambert (EPSG:8058) projection. Generally, it is expected that as segment length increases, there will be more opportunity to accumulate delay, but, depending on the cause of the delay, there might be more opportunity to catch up to the schedule on longer links.

4.3.3.4 Schedule travel speed

Urban buses operate in a variety of contexts ranging from motorways to local roads, which correspond to a range of expectations about delay and reliability. Scheduled speed is a proxy that captures the type of roads being travelled in each stop-to-stop segment that encompasses road design, speed limit, and expected congestion level. The scheduled travel speed V_{it} is calculated using the stop-to-stop length L_i and the scheduled travel time T_{it}. The calculation can be represented by the equation:

$$V_{it} = \frac{L_i}{T_{it}} \tag{4.6}$$

4.3.3.5 Cross-traffic turns

In countries that drive on the left side of the road, cross-traffic turns are right turns. These introduce delay and unreliability because they can only occur during a gap in the parallel traffic or a dedicated signal phase. Bus cross-traffic turns have major impacts on the bus performance (Zhao and Zhou 2018). The existence of cross-traffic turns on each stop-to-stop link is determined based on the stop-to-stop trajectory. To identify whether a cross-traffic turn occurs within the stop-to-stop link, the change in bearing is analyzed along the stop-to-stop trajectory. A stop-to-stop link is considered to have a cross-traffic turn if there is a change in the bearing degree between 45 degrees and 135 degrees, which corresponds to an approximate cross-traffic turn of 90 degrees. This threshold helps to differentiate cross-traffic turns from other directional changes that may occur along the trajectory. To ensure data accuracy, results were spot checked within the studied area and found to be accurate.

4.4 Results

4.4.1 Regression tree

The regression tree presented in Table 4.4 predicts standard deviation and mean value of marginal delay through binary splits according to 13 variables. The presented tree has 22 branches, which cannot be visualized succinctly, but the table of importance for each factor in the regression tree indicates how important each variable is in determining the variability in performance (the column of importances of both values sums to 1.0). Three out of the top four most important variables are directly related to the bus schedule. Traffic flow is the most significant variable apart from the variables directly related to the bus schedule, followed by COVID stringency index and cross-traffic turns. Due to the limited length of bus priority lane within the studied area, both bus-taxi lane length and bus-HOV lane length are relatively low in importance in the regression tree model. These importance values in this regression tree are used for variable selection rather than prediction and interpretation.

4.4.2 Pooled regression model

The pooled regression model, presented in Equations 4.2 and 4.3, does not consider any panel effects. Table 4.5 shows the Pooled OLS model for the mean value of stop-to-stop marginal delay (Equation 4.2). All coefficients from the fitted model have intuitive signs and

TABLE 4.4: Variable importance in the regression tree for standard deviation and mean of stop-to-stop marginal delay. Three of the top four variables are directly related to the bus schedule. These importance values are used for variable selection.

Variable	Std	Mean
Scheduled Travel Time	0.276	0.209
Traffic Flow	0.166	0.111
Link Length	0.159	0.190
Scheduled Travel Speed	0.138	0.196
COVID Stringency Index	0.083	0.063
Cross-traffic Turn	0.041	0.071
Number of Traffic signals	0.040	0.053
Bus-taxi lane Length	0.037	0.059
Precipitation	0.028	0.014
Weekend	0.014	0.014
School Holiday	0.008	0.006
Bus-HOV lane Length	0.007	0.012
Public Holiday	0.003	0.002
R^2	0.4505	0.8215

are significant. Negative relationships are observed between marginal delay and both types of bus lanes. Each traffic signal increases the marginal delay by 2.59 s. Every hundred vehicles per lane-hour of the studied corridors increases the marginal bus delay by 0.50 s. Cross-traffic turns increase the stop-to-stop marginal delay by 26.23 s. The model indicates that making a cross-traffic turn is a major concern during bus operation.

The mean marginal delay captures the performance of buses relative to their schedule. As the data used in this chapter are drawn from an interval encompassing both COVID lockdowns and post-COVID reduced demand for transit, it is expected that the schedule may be more generously padded than the demand level warranted. This would be observed as a tendency towards negative coefficients for the scheduled travel time variable. Both length and scheduled travel time are used in the models in order to disambiguate the impact of link magnitude and expected congestion, even though there is correlation found between scheduled travel time and link length.

Every mm of precipitation per hour reduces stop-to-stop marginal delay by 0.16 s, which could be explained by the reduced travel demand on buses during raining conditions. The

TABLE 4.5: Pooled OLS model for the mean value of stop-to-stop marginal delay. All parameters are statistically significant and the signs are intuitive. The presence of bus lane measures results in a reduction in mean stop-to-stop marginal delay, indicating both decreased delays and over-scheduling padding.

Variable	Coefficient	Std. Err.	P-value
Bus-taxi lane Length	-6.1557	0.0516	0.0000
Bus-HOV lane Length	-7.1653	0.0690	0.0000
Number of Traffic signals	2.5920	0.0076	0.0000
Traffic Flow	5.0509	0.0337	0.0000
Cross-traffic Turn	26.232	0.0208	0.0000
Scheduled Travel Time	-0.4373	0.0002	0.0000
Link Length	30.384	0.0339	0.0000
Precipitation	-0.1582	0.0087	0.0000
Weekend	-0.6852	0.0185	0.0000
Public Holiday	-1.7229	0.0501	0.0000
School Holiday	-2.2104	0.0187	0.0000
COVID Stringency Index	-0.1354	0.0003	0.0000
Number of observations	15962428		
R^2	0.5263		

weekend marginal delay is 0.69 s less than weekdays, and public holidays and school holidays reduce average marginal delay by 1.72 s and 2.21 s respectively. These values indicate that, even controlling for the overall traffic volume, buses tend to be more on-time or even ahead of schedule outside the typical weekday peaks. The stop-to-stop marginal delay difference between the most COVID effected day observed and the least COVID affected day is 8.71 s based on the model result. A R^2 value of 0.53 indicates that the model explains slightly over half of the variation in mean marginal delay, and most attributes of the links are statistically significant but still have a small impact.

The standard deviation of marginal delay indicates how reliably each stop-to-stop link performs. This measure is essential to model if we assume that the timetable has been appropriately updated to reflect the relevant conditions along the route. Table 4.6 shows the Pooled OLS model for the standard deviation value of stop-to-stop marginal delay (Equation 4.3). All coefficients have intuitive signs and are significant. Negative relationships are observed between standard deviation of marginal delay and both bus-taxi and bus-HOV lanes. Bus-taxi lanes, which are the more restrictive priority measure, have a stronger reliability improvement

TABLE 4.6: Pooled OLS model of the standard deviation of stop-to-stop marginal delay. All parameters are statistically significant and the signs are intuitive. The presence of bus lane measures reduce the standard deviation, which is synonymous with improved reliability.

Variable	Coefficient	Std. Err.	P-value
Bus-taxi lane Length	-12.564	0.0407	0.0000
Bus-HOV lane Length	-5.5584	0.0560	0.0000
Number of Traffic signals	0.9709	0.0065	0.0000
Traffic Flow	14.195	0.0281	0.0000
Scheduled Travel Time	0.1772	0.0001	0.0000
Scheduled Speed	0.4437	0.0009	0.0000
Cross-traffic Turn	12.902	0.0171	0.0000
Precipitation	0.4112	0.0072	0.0000
Weekend	-1.8418	0.0154	0.0000
Public Holiday	-0.6829	0.0414	0.0000
School Holiday	0.4149	0.0155	0.0000
COVID Stringency Index	0.0854	0.0003	0.0000
Number of observations	15962428		
R^2	0.7345		

than bus-HOV lanes, which are more permissive and more vulnerable to violation. Each traffic signal increases the standard deviation of marginal delay by almost one second which is statistically significant but not substantial. Every hundred vehicles added on each lane during each hour increases the standard deviation of marginal delay by 1.42 s. Cross-traffic turns increase the standard deviation of marginal delay by 12.90 s. Similar to the mean marginal delay models, this model indicates that cross-traffic turns are a major concern during bus operation. Both scheduled travel time and scheduled speed increase unreliability of marginal delay. Both parameters are used in the models in order to capture the effects of the set timetable.

Every mm of precipitation per hour increases the standard deviation of marginal delay — a relatively heavy rain of 5 mm per hour could result in 2.06 s of unreliability of delay on each segment. Both weekend and public holiday tend to be more reliable, which is possibly due to more distributed destinations than the weekday peak tidal flow. Combining the results with the mean marginal delay model, school holiday increase unreliability while reducing expected marginal delay. The difference in the standard deviation of marginal delay between the most COVID affected day and the least COVID affected day is 5.5 s per segment based on the model results. If most bus routes have 20-50 stops, most of these factors, despite being statistically significant, would not accumulate to a substantial amount of unreliability over the route. The exceptions are bus lanes and cross-traffic turns, which have the potential to add up to several minutes over the length of the route. A R^2 value of 0.74 indicates that the model explains over half of the variation in standard deviation of marginal delay.

4.4.3 Between-Effects panel regression model

In panel data analysis, the impact of a given variable can differ across clusters, *i.e.* stop-to-stop links in this chapter. For instance, while a bus priority measure might improve performance within a specific location, it could be primarily implemented in areas with exceptionally poor performance, leading between-effects models to suggest worse overall performance in the presence of the measure. However, this chapter finds that bus priority lanes provide more reliability benefits in the between-effects model, indicating that these lanes are more effective when comparing stop-to-stop segments with bus priority measures to those without, where time variations are ignored.

The between-effects panel regression model is an OLS model that averages the values for each stop-to-stop segment. This model aggregates the data by stop-to-stop links, which means each stop-to-stop link forms one observation, and any time variations are averaged out. The between-effects coefficients describe the relationships between segment attributes and marginal delay between clusters but loses the variety within the clusters. This also means that all stop-to-stop segments are equally weighted in this model even though some might be based on many more observations.

The fitted between-effects model (Table 4.7) indicates both types of bus lanes improve reliability (reduce standard deviation of marginal delay). The analysis reveals a greater reduction in the variation of marginal delay for bus-taxi lanes compared to bus-HOV lanes, which is expected since bus-taxi lanes are the more restrictive priority measure. Because the between effects model has no consideration of the time variation, none of the time specific

TABLE 4.7: The between-effects panel regression model for the standard deviation of stop-to-stop marginal delay. This models further aggregates the data by maintaining the mean value of each stop-to-stop segment. The results further substantiate that bus lanes enhance service reliability.

Variable	Coefficient	Std. Err.	P-value
Bus-taxi lane length	-19.732	3.1892	0.0000
Bus-HOV lane length	-16.583	3.9274	0.0000
Scheduled travel time	0.2474	0.0049	0.0000
Scheduled travel speed	1.0186	0.0370	0.0000
Cross-traffic turns	18.817	1.0906	0.0000
Number of clusters	2737		
R^2	0.8401		

variables including, traffic flow, precipitation, weekend, public holiday, school holiday and COVID Stringency Index, are included.

In the between-effects model, unreliability increases with the number of cross-traffic turns, the scheduled travel time and the scheduled travel speed. While statistically significant, the between effects coefficients once again indicate that most factors will not contribute enough unreliability to be meaningful to a bus operator or a passenger. The exception would be cross-traffic turns — each cross-traffic turn can cause significant service level reduction in most of the road corridors.

Both scheduled travel time and scheduled speed appear to contribute to an increase in the standard deviation of marginal delay. Consistent with the other models, these variables are included to account for scheduling effects. This model explains 84% of the variation in unreliability between segments.

4.4.4 Within-Effects panel regression models

The within-effects regression models divide the full dataset by stop-to-stop segments, with each segment being modeled independently. Only stop-to-stop links featuring time-varying bus lanes are analyzed to determine the effects of bus lanes, as variables that do not change within a segment cannot contribute to within-entity effects. The within-effects regression models control for the effect of bus priority lane (bus-taxi lanes and bus-HOV lanes) length, traffic volume, scheduled travel time, precipitation, COVID stringency index, weekend, public holiday and school holiday.

Omitting all time-invariant attributes of the segment allows this model to isolate the effect on reliability of turning on and off the bus lanes. The key result is illustrated as the histogram of coefficients on the two bus lane variables in Figure 4.6. The set of segments shows a distribution of coefficients ranging from roughly -200 to 200 seconds. It can been seen that most of the parameters are distributed around zero, which indicates that the bus operational reliability overall is similar when bus priority lanes are on or off. Considering the fact that bus priority lanes are generally in operation during peak hour when traffic conditions are worse along the major corridors (see Figure 4.2), this stability indicates that bus priority lanes are able to maintain service reliability despite peak-hour congestion. This suggests that the set of bus lanes is operating where and when it should be.

4.5 Conclusion

This study demonstrates the effectiveness of using GTFS-Realtime data from Chapter 3 to assess bus performance on a microscopic level by examining stop-to-stop marginal delay and utilizing detailed spatial information about bus priority infrastructure. Analyzing 20 months of data from the Sydney metropolitan area, the results show that bus-taxi lanes and bus-HOV lanes significantly enhance bus performance and travel reliability. Cross-traffic turns in bus operations are identified as a major concern across all models. Therefore, cross-traffic turns should generally be avoided within bus routes. Given the substantial impact of cross-traffic turns on bus operations, Chapter 5 will further investigate the effects of cross-traffic turns on bus performance. The study also indicates that the presence of traffic signals increases the standard deviation of stop-to-stop marginal delay, suggesting a lack of public transport signal priority in the studied area. This highlights an opportunity to enhance bus performance through the implementation of transit signal priority in the area. Further discussion of these results can be found in Chapter 7. In conclusion, this study emphasizes the significant benefits

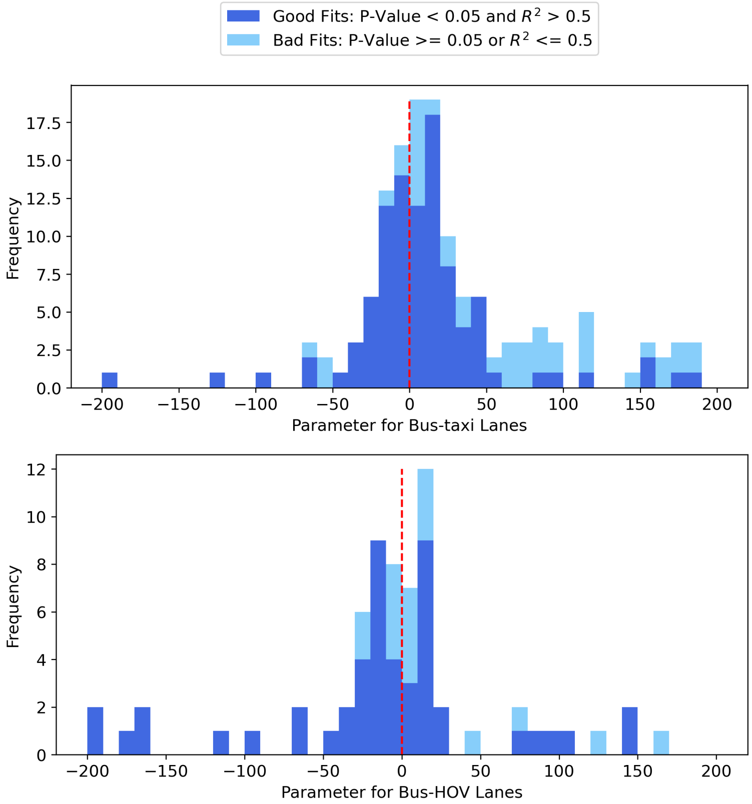


FIGURE 4.6: Bus priority lane length parameters distribution from the within-effects panel regression of standard deviation of marginal delay. The upper figure shows the parameter distribution for bus-taxi lanes and the lower figure shows the parameter distribution for Bus-HOV lanes. Higher values indicate that bus lanes cause unreliability whereas lower values indicate that bus lanes improve unreliability.

of bus lanes in improving bus performance and reliability and underscores the potential of GTFS-Realtime data in providing a comprehensive evaluation of bus priority infrastructure.

CHAPTER 5

The Role of Cross-traffic Turns in Bus Operations

5.1 Introduction

In Chapter 4, we found that bus lanes and HOV lanes (known as transit lanes in Australia) were effective priority measures for reducing bus delays and improving bus reliability. When controlling for the presence of bus lanes, we observed that higher traffic volumes, the number of traffic lights, and cross-traffic turns all significantly contributed to bus unreliability during operations. Chapter 4 shows that a single cross-traffic turn could lead to more performance degradation than ten traffic lights or an additional 900 vehicles per lane per hour, in terms of both delay and reliability. This highlights the importance of implementing additional bus priority measures, particularly at cross-traffic turns, to reduce bus delays and improve reliability at intersections.

This chapter leverages the availability of realtime data from Chapter 3 to make several contributions toward addressing the issue of bus performance degradation caused by cross-traffic turns. First, stop-to-stop links within Greater Sydney are categorized and compared based on the presence of cross-traffic turns at stop-controlled intersections, signalized intersections, and those without cross-traffic turns, highlighting their impacts on increasing mean delay and delay variability as well as reducing operational speed. Subsequently, vehicle speeds and trajectories within the local environment of cross-traffic turns at two intersections are analyzed to provide a deeper understanding of the effects of cross-traffic turns. Cross-validation between GTFS-Realtime Trip Updates and Vehicle Positions data confirms the accuracy of this local environment microscopic analysis.

5.2 Methodology

5.2.1 Data for bus cross-traffic turn analysis

The GTFS-Realtime dataset provides realtime operational details for public transit systems through three data feeds: Trip Updates, Service Alerts, and Vehicle Positions. This study utilizes the GTFS-Realtime Trip Updates from Chapter 3 and Vehicle Positions data feeds. The Trip Updates feed provides real-time estimates of arrival and departure times at bus stops, while the Vehicle Positions feed provides operational details, including GPS location and speed (Google Developers 2022a). Trip Updates are used to compare stop-to-stop links with and without cross-traffic turns by three indices: delay, reliability, and speed. Meanwhile, Vehicle Positions are used to compare the average bus operational speed at intersections between buses traveling straight through and those making cross-traffic turns.

Both the Trip Updates and Vehicle Positions datasets were archived from TfNSW's GTFS-Realtime API at one-minute intervals during the study period. The stop-to-stop comparison relies on the Trip Updates dataset, which is used to calculate the three indices. The Trip Updates data was developed as part of our previous study on buses within NSW in Chapter 3. Additionally, the GTFS-Static/Schedule dataset is utilized to provide trajectory information, which is used to identify the presence of cross-traffic turns and to calculate the distance between two bus stops for speed analysis.

The GTFS-Realtime Vehicle Positions data is utilized to conduct the intersection level speed comparison. The GTFS-Realtime Vehicle Positions updates based on the AVL data from buses, which includes the vehicle's location, operational speed, and bearing. The GTFS-Realtime Trip Updates is used to filter the Vehicle Positions by direction and the GTFS-Static dataset is used to provide trajectory for speed map plotting. Both GTFS-Realtime datasets are employed to validate the use of vehicle position data for estimating bus travel time at intersections between two bus stops, with comparisons made against the Trip Updates feed.

5.2.2 Stop-to-stop links comparison

For the comparison of stop-to-stop links, this chapter follows a process similar to that outlined in Chapter 4. This chapter focuses on buses in Greater Sydney, New South Wales (Figure 5.1) over the period from 1 November 2020 to 30 June 2022, avoiding times with major COVID impacts. Three indices are considered in the stop-to-stop links comparison: delay, reliability, and speed. Both delay and reliability are based on the changes in stop delay time between two bus stops, reflecting the impact of cross-traffic turns. The stop delay change between two stops is defined as the stop-to-stop marginal delay, as outlined in Equation 4.1 which represents the delay increase from the previous departure to the next arrival. This stop-to-stop marginal delay aims to reflect the incremental delay accrued due to cross-traffic turns by comparing the stop-to-stop links with and without cross-traffic turns. Delay is calculated as the difference between the scheduled and actual arrival times and can be either positive (behind schedule) or negative (ahead of schedule).

The delay measurement captures the increase in delay between two stops, isolating the effect of cross-traffic turns and reflecting their impact on delay. The reliability measurement quantifies the variability in stop-to-stop marginal delay between two stops. This variability is measured within a rolling time window of the 30 nearest readings. This method is specifically used to assess the effects of cross-traffic turns on reliability, as indicated by fluctuations in delay. Unlike the approach in Chapter 4, no average value is taken for delay, and no sampling is performed for delay and reliability measures to reduce the data size.

Speed is measured as the distance between two stops (L) divided by the travel time (t), and is used to reflect the impact of cross-traffic turns on bus operational speed between stops.

$$V = L/t \tag{5.1}$$

The stop-to-stop links are categorized into three groups: No Cross-Traffic Turn, Cross-Traffic Turn at Sign-Controlled Intersections, and Cross-Traffic Turn at Traffic Signals. Links classified as No Cross-Traffic Turn include only straight movements and non-cross-traffic

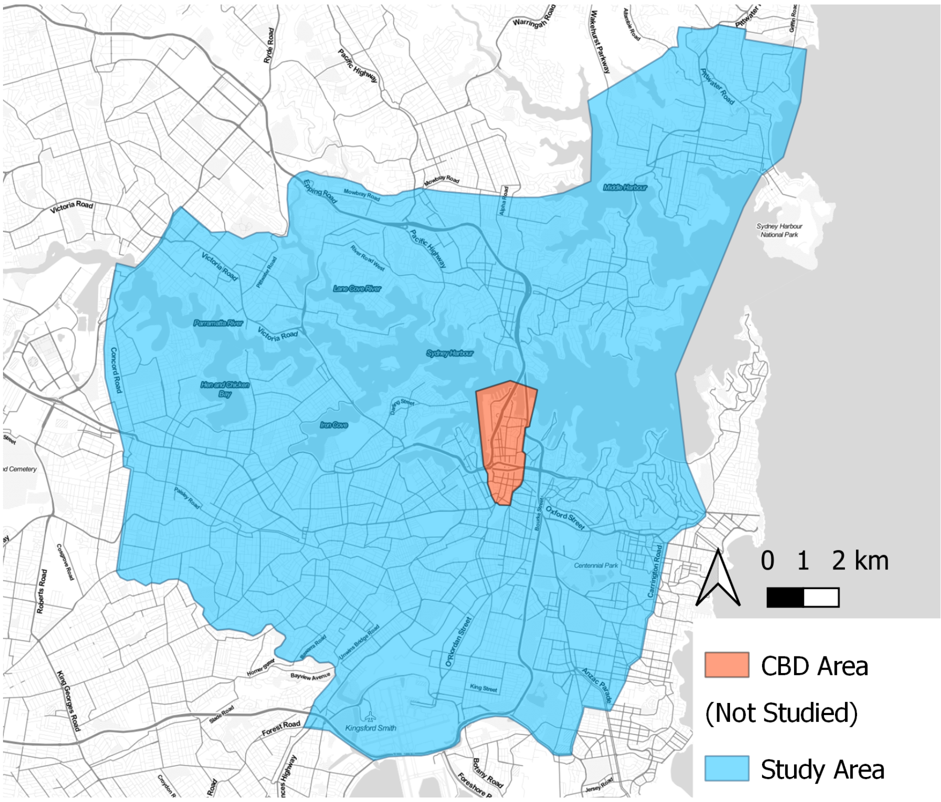


FIGURE 5.1: Study area for stop-to-stop link comparison. Stop-to-stop links within the study area are used for cross-traffic turn analysis. Suburbs around Sydney CBD were selected as the study area, while Sydney CBD was avoided due to major differences in traffic conditions.

turns (e.g., left turns in left-hand driving, as considered in this study) between two stops. A stop-to-stop link is classified as Cross-Traffic Turn at Sign-Controlled Intersections if it involves only cross-traffic turns at intersections regulated by yield signs, stop signs, or roundabouts, and does not include any cross-traffic turns at signalized intersections. Links that involve a cross-traffic turn at a signalized intersection are categorized as Cross-Traffic Turn at Traffic Signals. If a stop-to-stop link includes both cross-traffic turns at sign-controlled intersections and at signalized intersections, it is categorized under Cross-Traffic Turn at Traffic Signals, as signalized intersections typically exert a greater impact on bus performance.

To compare the differences between cross-traffic turns at yield or stop signs and at roundabouts, the category cross-traffic turn at sign-controlled intersections is further divided into two subcategories: Cross-traffic turn at yield/stop sign and cross-traffic turn at roundabout. The cross-traffic turn at yield/stop sign category includes those links where the cross-traffic turn only occurs at a yield or stop sign between two stops. If a cross-traffic turn is detected at a roundabout between two stops, the link is categorized as cross-traffic turn at roundabout. If a stop-to-stop link contains both cross-traffic turns at yield/stop signs and at roundabouts, it is categorized as cross-traffic turn at roundabout because roundabouts have a greater impact on performance due to the increased difficulty of bus maneuvering.

5.2.3 Intersection comparison

The intersection-level comparison provides a more localized and detailed understanding of the effects of cross-traffic turns compared to the stop-to-stop level analysis. Two intersections within Greater Sydney, Australia, were selected for in-depth study: the intersection of Pitt Street and Eddy Avenue in Haymarket and the intersection of Parramatta Road and Norton Street in Leichhardt (Figure 5.2). These intersections are heavily utilized by buses, with an average frequency of at least six buses per hour during the studied period (Table 5.1). Both intersections exhibit significant delays attributable to cross-traffic turns and are presented as illustrative case studies to explore these effects in detail.

TABLE 5.1: Vehicle Positions data at the two studied intersections. Both intersections are heavily utilized by buses during the morning peak period for both through and cross-traffic turn movements. The number of Vehicle Positions readings depends on the volume of buses and the time spend traversing the studied area at each intersection.

Intersection	Direction	Number of Routes	Buses / Hour	Vehicle Positions Readings / Hour	Total Vehicle Positions Readings
Pitt Street and Eddy Avenue	Through	18	79.19	1368	120423
	Cross-Traffic Turn	1	6.13	137	12080
Parramatta Road and Norton Street	Through	4	7.58	26	2345
	Cross-Traffic Turn	4	16.84	164	14411

This study focuses on the morning peak period, which is considered more critical in the Greater Sydney Area (shown in Figure 4.4). The GTFS-Realtime Vehicle Positions data from 6 am to 10 am is used, which is the morning peak hours definition by Transport for New South Wales. The analyzed dataset spans the period from 1st June 2022 to 30th June 2022,

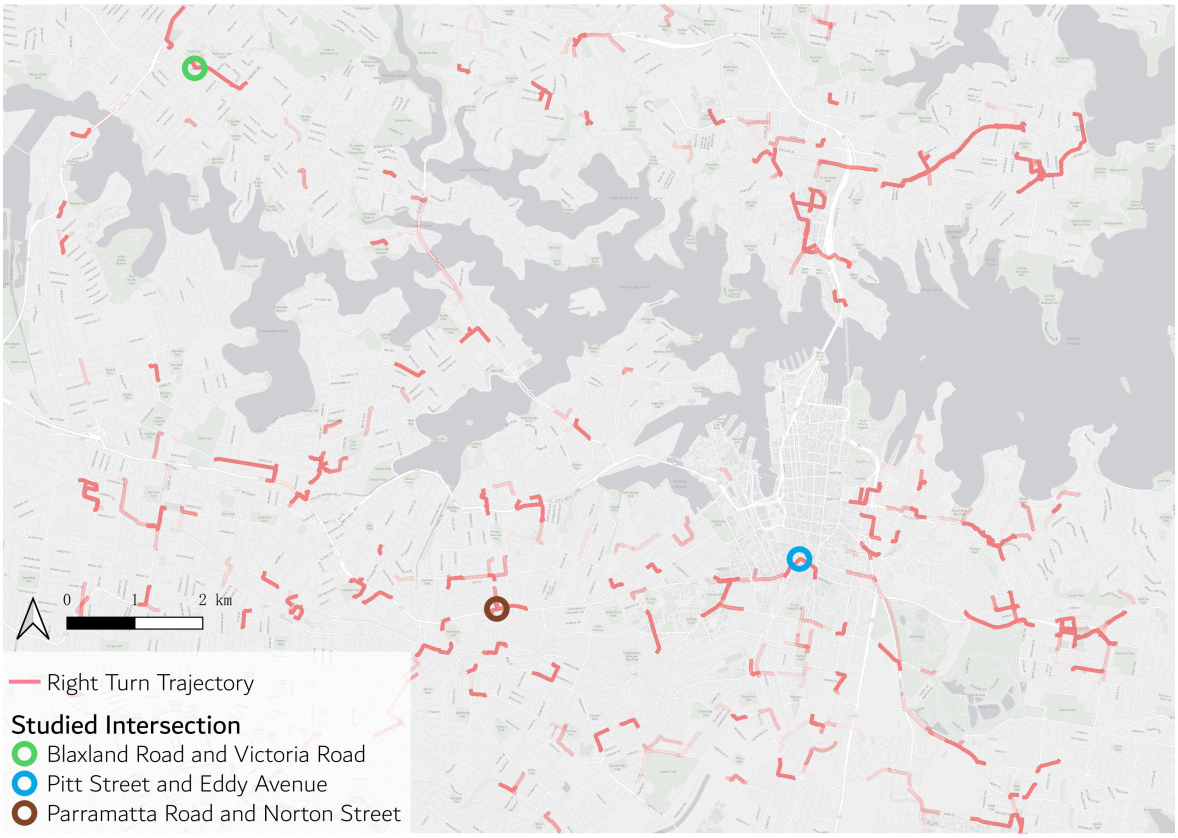


FIGURE 5.2: Trajectory plot of stop-to-stop links includes cross-traffic turns at traffic signals, highlighting the 10% slowest operational speeds across all observations. Darker trajectories indicate segments with a higher concentration of low-speed bus movements. The two intersections studied in this chapter, along with the intersection of Blaxland Road and Victoria Road examined in chapter 6, all exhibit significant delays due to cross-traffic turns and are presented as illustrative case studies.

resulting in at least 2,300 measurements for each direction under analysis (Table 5.1). At each studied intersection, the critical cross-traffic turn direction and the through movements within the same approaching direction are examined for comparison.

Before the analysis, the raw Vehicle Positions data is converted into CSV files. The Vehicle Positions data is first separated by direction at the studied intersection. Speed maps are then generated for the studied directions. Lane-level trajectories for both directions are estimated by analyzing typical bus lane usage patterns. These trajectories are divided into 5-meter segments for each direction. The vehicle position data is spatially joined to the nearest 5-meter segment to compute the local average speed for each interval. To smooth out rapid speed

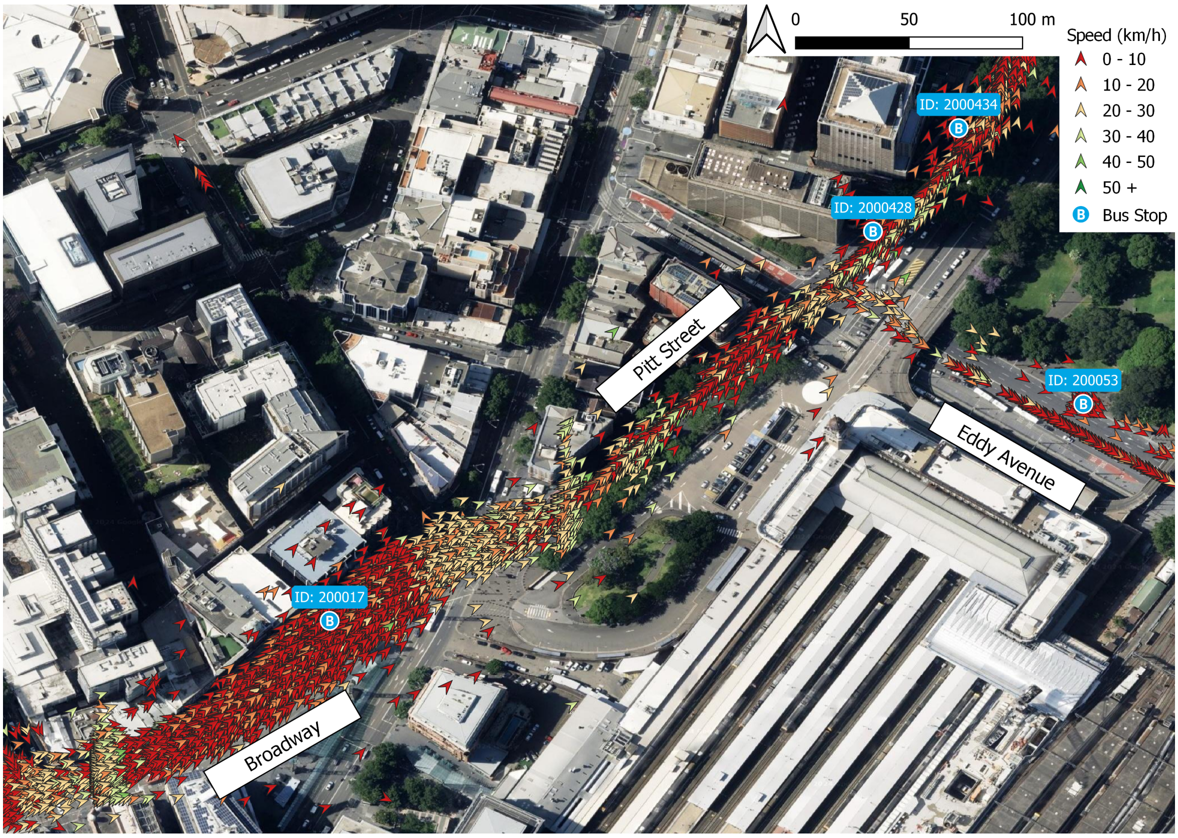


FIGURE 5.3: GTFS-Realtime Vehicle Positions data at the intersection of Pitt Street and Eddy Avenue, Haymarket, Greater Sydney, Australia. This intersection serves as a major bus corridor due to its proximity to Sydney Central Station. The speed map highlights substantial slowdowns, driven by frequent traffic congestion and bus bunching, particularly during peak hours.

fluctuations, the local average speed for each interval is further averaged with the speeds of its two neighboring intervals. The resulting speed maps for the two intersections are presented in the Results section below.

At the intersection of Pitt Street and Eddy Avenue in Haymarket (Figure 5.3), the critical cross-traffic turn direction is from the southwest (Broadway) to the east (Eddy Avenue). This route serves as a major bus corridor, heavily utilized by city-bound buses during peak hours. With the high volume of road users, including light rail, this intersection frequently operates at full capacity. As a result, bus bunching is common during peak periods, and bus operation speed is relatively lower compare the other studied intersection (Table 5.4).

FIGURE 5.4: GTFS-Realtime Vehicle Positions data at the intersection of Parramatta Road and Norton Street, Leichhardt, Greater Sydney, Australia. Due to the proximity of a bus depot to the north of Norton Street, many buses make right turns at this intersection. During the morning peak, heavy eastbound traffic towards the city leads to significant speed reductions for right-turning vehicles, including buses.

At the intersection of Parramatta Road and Norton Street in Leichhardt (Figure 5.4), the critical cross-traffic turn direction is from the east (Parramatta Road) to the north (Norton Street). This cross-traffic turn is frequently used by outbound buses during peak hours, due to the bus depot located further north along Norton Street. The cross-traffic turn direction conflicts with the heavy eastbound city-bound traffic on Parramatta Road during the morning peak, resulting in a significant speed difference between buses making the cross-traffic turn and those traveling straight through the intersection.

5.3 Results

5.3.1 Stop-to-stop links comparison

Tables 5.2 and 5.3 describe the distribution of each category of stop-to-stop links using mean, standard deviations, two sample t-tests and K-S tests. It can be observed that cross-traffic turns at sign-controlled intersections negatively impact all three measurements: delay, standard deviation, and speed, when compared to stop-to-stop links with no cross-traffic turns. These negative impacts are more pronounced for the category of cross-traffic turns at traffic signals, which are expected to result in greater delays. Although only minor differences are observed between cross-traffic turns at yield/stop signs and those at roundabouts, the results suggest that roundabouts result in more difficulties for buses performing cross-traffic turns due to the tight turning curve.

TABLE 5.2: Mean and standard deviation values for each category of stop-to-stop links. Major differences are observed between links with and without cross-traffic turns. Cross-traffic turns at sign-controlled intersections show less impact on performance compared to those at traffic signals, while roundabouts exhibit slightly worse performance than sign-controlled cross-traffic turns.

Measurement	Values	No Cross-traffic Turn	Cross-traffic Turn at Sign Controlled	Cross-traffic Turn at Traffic Signal	Cross-traffic Turn at Yield/Stop Sign	Cross-traffic Turn at Roundabout
Stop-to-stop Marginal Delay (s)	Mean	-26.64	-23.93	-12.35	-24.51	-22.06
	Standard Deviation	50.02	71.70	78.40	74.28	62.83
Stop-to-stop Standard Deviation (s)	Mean	31.21	43.78	47.71	42.43	48.16
	Standard Deviation	24.55	42.27	40.26	38.30	52.96
Stop-to-stop Speed (km/h)	Mean	39.34	33.36	24.15	33.69	32.34
	Standard Deviation	22.43	19.32	15.81	19.85	17.52

TABLE 5.3: Results of two-Sample t-tests and Kolmogorov-Smirnov (K-S) tests. Larger t-statistics indicate greater differences between group means, while larger K-S statistics suggest greater differences between the distributions. Given the large sample size, all tests returned p-values below 0.001, indicating statistically significant differences across all comparisons.

Measurement	Category		No Cross-traffic Turn		Cross-traffic Turn at Traffic signal	Cross-traffic Turn at Yield/Stop Sign
	Testing Against		Cross-traffic Turn at Sign Controlled	Cross-traffic Turn at Traffic Signal	Cross-traffic Turn at Sign Controlled	Cross-traffic Turn at Roundabout
Stop-to-stop Marginal Delay	Two Sample T-Tests	T-statistic	96.35	853.90	249.82	27.25
		P-value	<0.001	<0.001	<0.001	<0.001
	K-S Tests	K-S Statistic	0.06	0.15	0.09	0.05
		P-value	<0.001	<0.001	<0.001	<0.001
Stop-to-stop Standard Deviation	Two Sample T-Tests	T-statistic	209.61	452.32	37.17	25.21
		P-value	<0.001	<0.001	<0.001	<0.001
	K-S Tests	K-S Statistic	0.09	0.28	0.09	0.06
		P-value	<0.001	<0.001	<0.001	<0.001
Stop-to-stop Speed	Two Sample T-Tests	T-statistic	488.65	2349.99	922.64	55.84
		P-value	<0.001	<0.001	<0.001	<0.001
	K-STests	K-S Statistic	0.13	0.32	0.25	0.04
		P-value	<0.001	<0.001	<0.001	<0.001

The analysis presented in Tables 5.2 and 5.3 and Figure 5.5 shows that stop-to-stop segments with cross-traffic turns at traffic signals experience the highest delays compared to the other

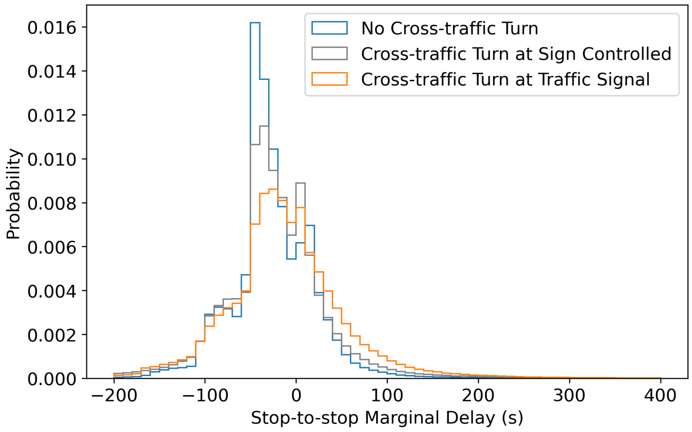


FIGURE 5.5: Comparison of stop-to-stop marginal delay distribution. Stop-to-stop links with cross-traffic turns exhibit greater marginal delays. Cross-traffic turns at signalized intersections causing more significant delays than those at sign-controlled intersections.

two categories, as confirmed by the larger K-S statistic. Cross-traffic turns at sign-controlled intersections result in less delay than those at signal-controlled intersections. Furthermore, the difference in delays between cross-traffic turns at sign-controlled intersections and those at traffic signals is more pronounced than the difference between no cross-traffic turns and cross-traffic turns at sign-controlled intersections. Stop-to-stop links without cross-traffic turns exhibit a tighter distribution of delays, while those with cross-traffic turns, especially at traffic signals, show a higher mean delay closer to zero. It is important to note that the data for this study were collected during the COVID-19 pandemic, a period characterized by relatively low transit ridership and reduced traffic congestion. This likely contributed to over schedule padding in bus timetables, resulting in the negative mean delays observed in the stop-to-stop links. These results provide substantial evidence to support the claim that cross-traffic turns, whether at signalized or signalized intersections, contribute to an increase in average stop-to-stop marginal delay during bus operation.

In addition to delays, cross-traffic turns are a significant source of unreliability for buses. The standard deviation of stop-to-stop marginal delay follows a similar trend as stop-to-stop marginal delay measurements. As shown in Figure 5.6 and Tables 5.2 and 5.3, cross-traffic turns at both traffic signals and sign-controlled intersections exhibit notably higher standard deviation distributions compared to stop-to-stop segments without cross-traffic

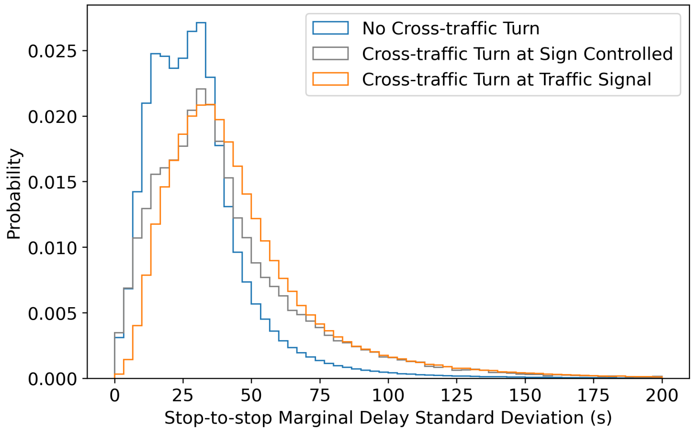


FIGURE 5.6: Comparison of stop-to-stop marginal delay standard deviation distribution. Stop-to-stop links without cross-traffic turns demonstrate the highest reliability, with a lower standard deviation of marginal delay. Consistent with the marginal delay results, cross-traffic turns at traffic signals have a greater negative impact on performance compared to those at sign-controlled intersections.

turns. This statistically significant difference supports the conclusion that cross-traffic turns increase variability of delay and, therefore, decrease reliability in bus operations. With a lower mean standard deviation, sign-controlled intersections result in less unreliability compared to signalized intersections. Furthermore, difference between cross-traffic-turn and no-cross-traffic-turn reliability metrics are higher than those for marginal delay, indicating that cross-traffic turns have a greater impact on unreliability than on delays.

Tables 5.2 and 5.3 and Figure 5.7 compare bus operational speed for stop-to-stop segments under different cross-traffic turn conditions. According to the T-statistic and K-S statistic values, the differences in speed distributions between the categories are statistically significant among all turn categories. Segments without cross-traffic turns, represented by the blue line, exhibit the highest average operational speed, followed by segments with cross-traffic turns at sign-controlled intersections. Some fluctuations are observed at higher speeds for cross-traffic turns at sign-controlled intersections, likely due to inaccurate travel time estimates when two stops are close to each other — this often occurs on local roads where sign-controlled intersections are more common. The statistical results confirm that performing cross-traffic turns significantly reduces the operational speed of buses between stops, with sign-controlled

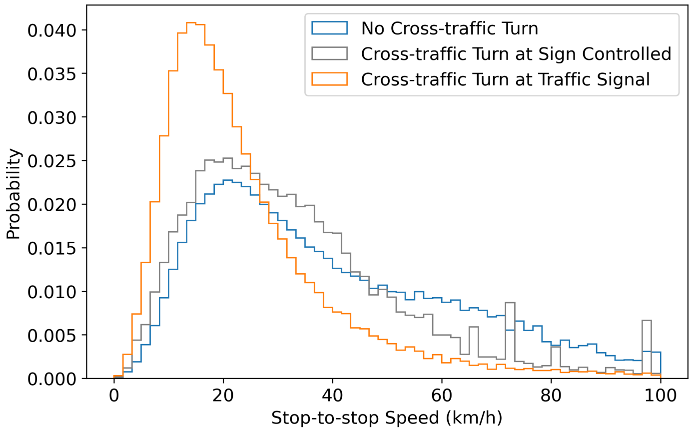


FIGURE 5.7: Stop-to-stop bus operation speed distribution comparison. Cross-traffic turn at traffic signal indicates a significantly lower speed distribution. Cross-traffic turn at sign controlled intersections has a lower speed distribution than no cross-traffic turn but a higher speed distribution than cross-traffic turn at traffic signal.

intersections causing a smaller reduction in speed compared to signalized intersections, though they still contribute to a noticeable decrease in bus travel speed.

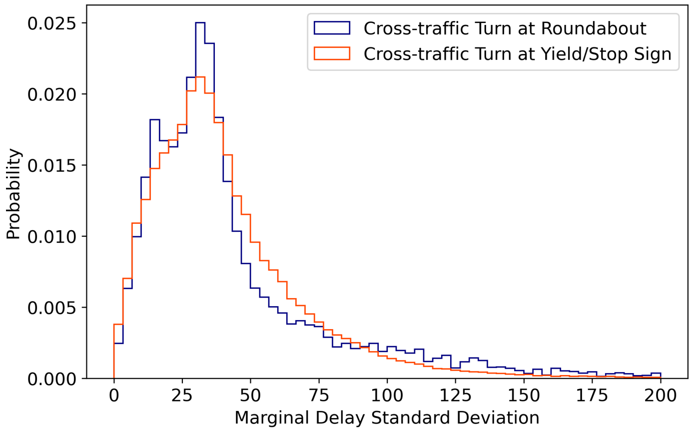


FIGURE 5.8: Comparison of stop-to-stop marginal delay standard deviation distribution between yield/stop signs and roundabouts. The two curves exhibit similar distributions but differ in mean and standard deviation values. The other performance measurements show a similar trend when comparing yield/stop signs with roundabouts.

Tables 5.2 and 5.3 and Figure 5.8 compare the distribution of stop-to-stop marginal delay, standard deviation, and operational speed for stop-to-stop segments with cross-traffic turns at roundabouts versus those at yield/stop signs. The stop-to-stop marginal delay shows similar patterns between the two intersection types, with roundabouts exhibiting slightly higher delays. Similarly, the distribution of marginal delay standard deviations also reflects similar trends. Roundabouts introduce slightly greater unreliability due to tighter turning conditions. Additionally, the bus operational speed distribution shows that yield/stop signs generally allow for higher speeds compared to roundabouts, likely due to the better turning conditions at unsignalized intersections. Overall, roundabouts tend to cause higher delays and slightly lower speeds, making them more challenging for bus maneuverability than yield/stop sign-controlled intersections.

5.3.2 Intersection comparison

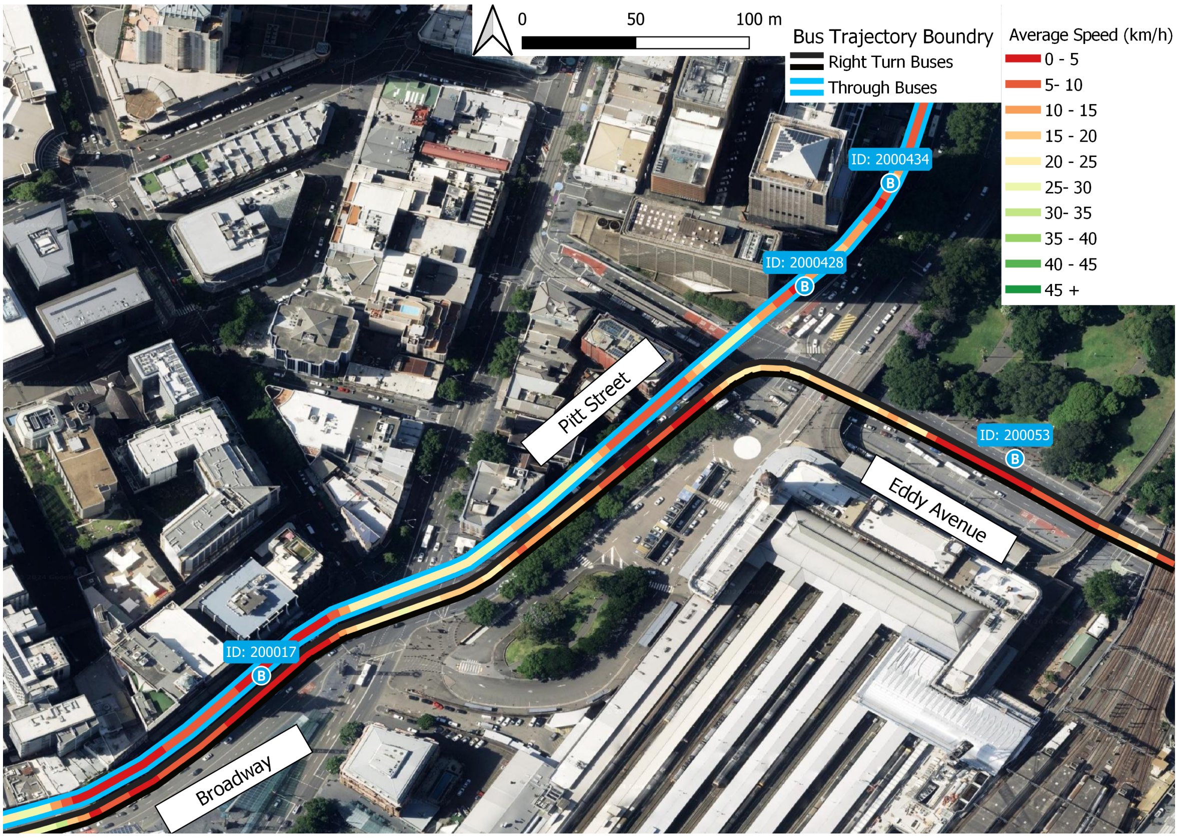


FIGURE 5.9: Microscopic speed map at the intersection of Pitt Street and Eddy Avenue, Haymarket, Greater Sydney, Australia. Both through and cross-traffic buses experience relatively slow average speeds due to high congestion. A noticeable speed difference is observed between through buses and buses making cross-traffic turns.

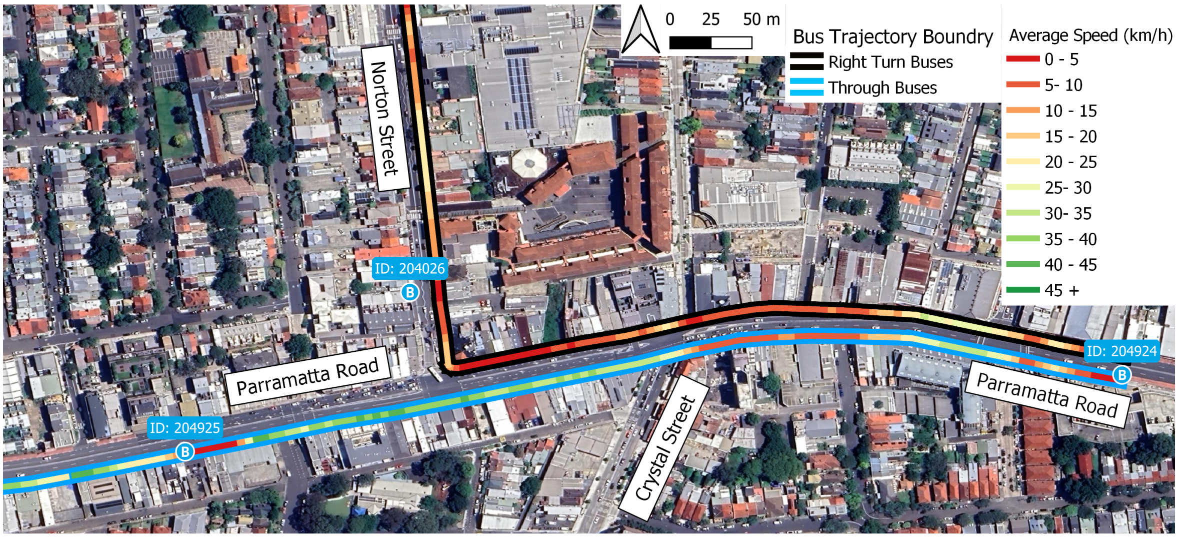


FIGURE 5.10: Microscopic speed map at the Intersection of Parramatta Road and Norton Street, Leichhardt, Greater Sydney, Australia. The through buses only experience reduced average speeds before reaching Crystal Street, as the intersections at Crystal Street and Norton Street are coordinated. A significant speed difference is observed between the through buses and the cross-traffic turn buses at this intersection.

The distributions of delay, reliability, and speed mask the diverse, idiosyncratic impacts at each intersection. The microscopic intersection comparison analysis offers a detailed examination of bus performance in relation to specific design attributes. Figure 5.9 presents the estimated speed map at the intersection of Pitt Street and Eddy Avenue, derived from GTFS-Realtime Vehicle Positions data. Given that this intersection is adjacent to Central Station, and is heavily used by both vehicular and light rail traffic during peak hours, it is expected that buses operate at a lower average speed in this area. Significant delays are evident at this intersection. The data shows that buses making cross-traffic turns experience a more rapid and earlier speed reduction compared to through buses, at the entrance of Pitt Street before Eddy Avenue. This location is typically the start of the cross-traffic turn queues, as observed. The speed map demonstrates that cross-traffic turn buses face significantly more delay than through buses.

Figure 5.10 shows the estimated speed map at the intersection of Parramatta Road and Norton Street, derived from GTFS-Realtime Vehicle Positions data. A significant difference is observed between the two trajectories. Buses traveling westbound through the intersection experience minimal delays, particularly before reaching Crystal Street, and encounter almost no delays at the Parramatta Road and Norton Street intersection, since the two intersection signal are coordinated. In contrast, buses making westbound cross-traffic turns face considerable delays before Crystal Street and maintain reduced operational speeds up to the Parramatta Road and Norton Street intersection. Lower speeds are observed on Norton Street, as it is a local street with a lower speed limit of 40 km/h and narrower lateral space. Typically, the northbound direction of Norton Street is not congested, so Norton Street is not the primary cause of delays for cross-traffic turns at the intersection of Parramatta Road and Norton Street. Similar to the previous intersection, bus speeds drop noticeably upon entering the traffic queue. The slower average speed for cross-traffic turn buses is anticipated, given the lower priority assigned to cross-traffic turns at this intersection.

The results outlined in Table 5.4 reveal that executing cross-traffic turns significantly extends travel times and reduces travel speed, underscoring the impact of bus cross-traffic turns as a major concern at intersections. At intersections with substantial bus cross-traffic turn volumes, it is advisable for transit authorities to consider introducing bus priority measures, such as queue jump lanes, to enable buses to bypass queues and enhance operational efficiency. However, implementing such priority lanes in urban settings presents considerable challenges due to space constraints. Transforming an existing traffic lane into a bus priority lane could reduce the overall capacity of the intersection, potentially leading to congestion that obstructs buses and affects both general traffic flow and bus service efficiency. Moreover, the potential for increased delays for other vehicles necessitates careful consideration when introducing these measures. In the context of the intersection of Parramatta Road and Norton Street, where road space is even more constrained than at the intersection of Pitt Street and Eddy Avenue, traditional bus priority measures may not be sufficient to improve bus performance without significantly worsening delays for general traffic.

5.4 Validation

Given the novelty of GTFS-Realtime microscopic analysis, cross-validation is employed to assess the accuracy of the intersection comparison microscopic analysis. This validation process involves comparing the results from two GTFS-Realtime datasets: Trip Updates and Vehicle Positions. While Trip Updates provide travel time estimates for individual stop-to-stop segments, which are typically shorter than the full trajectory lengths estimated by Vehicle Positions data, the travel time and speed comparisons are thus confined to each stop-to-stop segment within the directions under study.

Due to some inaccurancy in the bus stop locations provided by the GTFS-Static data, a more precise stop location is determined using Google Maps street view. The exact position of the stop post is used to calculate the stop-to-stop segment trajectory length, which is then applied in Vehicle Positions calculations. These calculations are based on the average speeds estimated in the speed maps (Figure 5.9 and 5.10).

TABLE 5.4: Cross-validation results of travel time estimates between Trip Updates and Vehicle Positions. The travel time and speed estimates from the two datasets are generally consistent. However, Vehicle Positions tend to slightly overestimate travel time and underestimate speed, as slower-moving buses generate more location data within the studied area.

Intersection	Direction	Stop ID		Travel Time Estimates (s)		Speed Estimates (km/h)	
		Start	End	Trip Updates	Vehicle Positions	Trip Updates	Vehicle Positions
Pitt Street and Eddy Avenue	Through	200017	2000428	110.30	112.34	10.40	10.21
	Cross-Traffic Turn	200017	200053	215.06	224.02	6.85	6.57
Parramatta Road and Norton Street	Through	204924	204925	165.21	154.96	12.15	12.95
	Cross-Traffic Turn	204924	204026	211.79	251.26	8.13	6.85

Comparing the two datasets affirms the effectiveness of the microscopic analysis because the average stop-to-stop travel times and speeds are very similar when estimated from the two datasets. As indicated in Table 5.4, the Vehicle Positions dataset tends to provide slightly higher travel time and lower travel speed estimates compared to the Trip Updates dataset. This discrepancy is anticipated due to the different methodologies used in their calculations: Vehicle Positions are calculated using space-mean speeds, whereas Trip Updates are based on time-mean speeds. Consequently, faster vehicles have a greater influence on the time-mean calculations. The precise and intuitive variation of the microscopic speed maps combined with the similar travel times from the cross-validation, support the reliability of the analysis. Furthermore, this consistency suggests that GTFS-Realtime Trip Updates and Vehicle Positions datasets can be effectively combined in future research.

5.5 Conclusion

This chapter utilized GTFS-Realtime Trip Updates and Vehicle Positions data to analyze the impact of cross-traffic turns on bus operations. The findings revealed that cross-traffic turns increase schedule delays, reduce operational speeds, and impair service reliability. The results emphasize the importance of minimizing or eliminating cross-traffic turns in route design, as stop-to-stop links without cross-traffic turns consistently outperform those with such maneuvers. Additionally, GTFS-Realtime Vehicle Position data is employed for a microscopic analysis, aggregating bus AVL data and spatially averaging it by the nearest 5-meter trajectory interval lines. This novel application of GTFS-Realtime data for evaluating route and geometry design highlights its potential for future studies focused on micro-level improvements to bus performance at intersections. Since traditional bus priority measures cannot effectively reduce the bus cross-traffic turn delay at the studied intersections without significantly delaying other traffic, a novel approach — the bus cross-traffic turn priority box — is introduced in Chapter 6. Further discussion on the impact of bus cross-traffic turns and their operational implications can be found in Chapter 7.

CHAPTER 6

Bus Cross-traffic Turn Priority Box

6.1 Introduction

Cross-traffic turns have traditionally been avoided in bus route design due to their contribution to service unreliability. The results in Chapter 4 indicate that a single cross-traffic turn can result in more negative effects than ten traffic lights or a traffic volume of 900 vehicles per lane per hour, in terms of both delay and reliability. In Chapter 5, we find that cross-traffic turns significantly impact bus operating speed, delay, and reliability when analyzing these metrics at the second level between stop-to-stop segments. Despite these significant impacts, there is a lack of effective, space-efficient solutions to mitigate the negative effects of bus cross-traffic turns.

Traditionally, prioritizing bus cross-traffic turns requires solutions such as queue jump lanes or transit signal priority systems. Both of the solutions requires additional later space to prioritize buses and results in delay increase for general traffic. Therefore, there is a pressing need for innovative solutions that can provide bus priority without the drawbacks of existing methods.

To address this gap, we propose an innovative, intersection-focused priority measure — the bus cross-traffic turn priority box. This in-lane cross-traffic turn queue jump lane allows buses to skip the cross-traffic turn queue using through lanes, similar to how a bicycle box offsets the traffic queue upstream from the intersection. The bus turning priority box enables buses to bypass the right-turn traffic queue and pre-accelerates traffic when there is no bus present, increasing green time efficiency. Importantly, it benefits bus operations while also reducing delays for general traffic, without requiring extensive lateral space.

This chapter focuses on evaluating the effectiveness of the bus cross-traffic turn priority box in reducing the negative effects associated with bus cross-traffic turns using Aimsun microscopic traffic simulation. We aim to assess the limitations of existing bus cross-traffic turn priority measures to highlight the necessity for alternative solutions, develop the principles and theoretical framework of the bus cross-traffic turn priority box, evaluate its performance by applying the proposed solution to a real-world intersection using real-time data.

6.2 Methodology

6.2.1 Problem statement

In densely built urban environments, geometric constraints often make it impractical to allocate additional space to alleviate traffic congestion. Prioritizing buses under these limitations typically requires converting existing road space into dedicated bus areas — such as queue jump lanes — or allocating specific signal times to buses through TSP. However, these conventional bus priority measures inevitably penalize general traffic to favor buses, leading to unintended consequences.

When bus priority measures are improperly designed, they can increase delays for both buses and general traffic. For example, converting a general-purpose traffic lane into a bus queue jump lane at a critical intersection may reduce the throughput for other vehicles. This reduction can cause queues to spill back upstream from the intersection. If the bus queue jump lane is only a short distance from the intersection, buses may become blocked by the traffic queue, resulting in increased delays for both buses and general traffic.

Transit agencies are aware of these issues but are often limited by what can be achieved with current priority measures. Existing solutions like bus pre-signals and queue jump lanes require dedicated bus lanes on the approach to intersections, which is not feasible in many urban areas where land is expensive and space is constrained. Implementing such measures can lead to increased delays for general traffic and may not provide sufficient benefits to buses when intersections become congested.

Moreover, traditional bus priority measures like TSP allocate signal time from general traffic to buses, which can result in increased delays for other vehicles — even if the total person delay is reduced due to the higher occupancy of buses. This reallocation of signal time can be a significant barrier to the implementation of bus priority measures, as it may not be politically or practically feasible to prioritize buses at the expense of general traffic flow.

Therefore, there is a pressing need for innovative, space-efficient bus priority solutions that can improve bus operations at intersections — particularly for cross-traffic turns — without penalizing general traffic or requiring significant infrastructural changes. The challenge lies in designing a solution that minimizes the need for additional road space, enhances bus performance during cross-traffic turns, and potentially improves overall traffic efficiency.

This chapter addresses this problem by proposing the bus cross-traffic turn priority box, a novel bus priority measure designed to reduce bus delays during cross-traffic turns while improving delays for general traffic. Unlike traditional measures that require dedicated bus lanes or extensive infrastructure, the bus cross-traffic turn priority box allows buses to skip the cross-traffic turn queue by using existing through lanes, which typically have faster average speeds and lower delays compared to cross-traffic turning lanes. By offsetting only the cross-traffic turn lane, this design requires less road space and infrastructure but provides similar benefits for buses making cross-traffic turns.

This chapter develops a bus priority measure that overcomes the limitations of existing solutions by not requiring additional space and reduce delay to general traffic and effective in enhancing bus operational efficiency at intersections.

6.2.2 Description of bus cross-traffic turning priority box

The bus cross-traffic turn priority box (Table 6.1) is a novel priority measure designed specifically for buses making cross-traffic turns. It functions by enabling buses to bypass the queue in the cross-traffic turn lane using the through traffic lanes, which generally have faster speeds and less congestion. As a bus approaches an intersection where it needs to make a cross-traffic turn, it remains in the through lane and overtakes the queued turning vehicles. During the through traffic green phase, the bus moves into the designated priority box positioned ahead of the turning queue. The bus waits there until the cross-traffic turn signal turns green, at which point it proceeds to make the turn before other vehicles, significantly reducing bus delays. When there is no cross-traffic turning bus, the priority box can be utilized by other turning vehicles for pre-acceleration, allowing them to enter the intersection at higher speeds and thus improving efficiency. This system eliminates the need for dedicated bus lanes, reduces delays for general traffic, and enhances overall intersection efficiency.

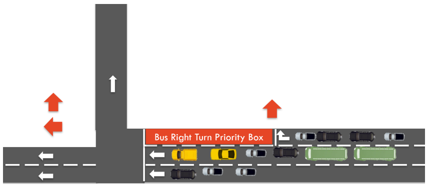
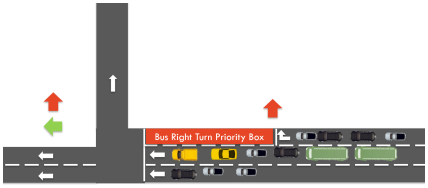
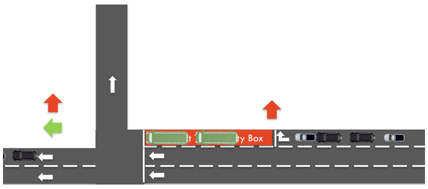
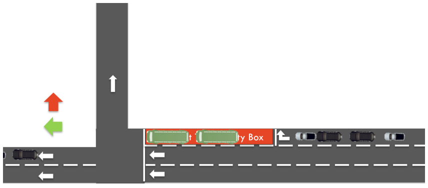
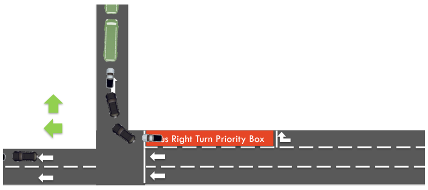
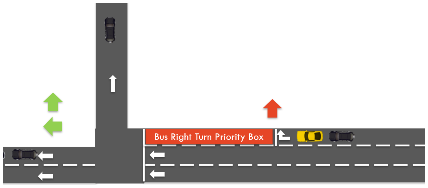
Implementing the bus cross-traffic turn priority box only requires the realignment of the cross-traffic turn lane to accommodate the priority box. This allows buses to use the faster-moving through lanes to bypass the cross-traffic turn queue, where the main delay for buses occurs. The length of the priority box is customized based on various factors such as demand, service rate, targeted pre-acceleration speed, and intersection geometry.

For the example intersections simulated below, the bus cross-traffic turn priority box is set to be 40 meters long. This length can accommodate two 14.5-meter-long buses, which is the typical bus length in Sydney, per signal cycle, corresponding to a service rate of approximately one bus per minute. Designing the length of the priority box is critical; an insufficient length could result in the bus queue spilling back into the through traffic lanes. Such spillover could significantly reduce intersection efficiency, particularly during peak hours.

6.2.3 Benefit for buses

The speed difference between through buses and buses making cross-traffic turns is generally expected to be significant (Figure 5.10). This discrepancy arises due to differences in conflict

TABLE 6.1: Step-by-step guide for implementing the bus cross-traffic turn priority box: Example intersection in Greater Sydney. The bus cross-traffic turn priority box is placed in the cross-traffic turning lane near the intersection. Buses bypass the cross-traffic turn queue by using the through lanes.

1: Offset the turning traffic to accommodate the bus cross-traffic turn priority box.  	2: The through traffic signal turns green, allowing buses to proceed forward. 
3: Buses use the through traffic lanes to bypass and wait in front of the cross-traffic turn queue.  	4: Eliminate the pre-cross-traffic turn signal. Other turning vehicles are allowed to use the unused part of the priority box for pre-acceleration. 
5: Buses turn first during the cross-traffic turn phase.  	6: Use the pre-cross-traffic turn signal to keep the bus cross-traffic turn priority box clear for the next cycle. 

*Buses arriving at the intersection during the cross-traffic turn green phase are expected to merge into the cross-traffic turn lane at the entrance of the bus cross-traffic turn priority box.
**According to state law, vehicles are required to yield to buses when buses indicate right.

points, resulting in variations in green time allocation — through traffic typically receives longer green times than cross-traffic turns. The primary delay reduction for buses comes from utilizing the faster through lanes to bypass the cross-traffic turn queue, thereby reducing the waiting time, as evidenced by the significantly lower average speeds shown in Figure 5.10. Additionally, since buses are generally able to pass through the intersection within a single signal cycle, this leads to increased reliability with less variance in travel time through the intersection.

6.2.4 Benefit for general traffic

Unlike other bus priority measures, the bus cross-traffic turn priority box not only provides travel time and reliability benefits for buses but also improves intersection efficiency for other road users. When there is no bus present within the priority box, general traffic can utilize it to pre-accelerate before the main signal, allowing vehicles to enter the intersection at higher speeds and significantly reducing start-up lost time for turning traffic. When buses are present at the beginning of the green phase, any remaining empty space within the priority box are used by general traffic, ensuring that no effective green time is wasted.

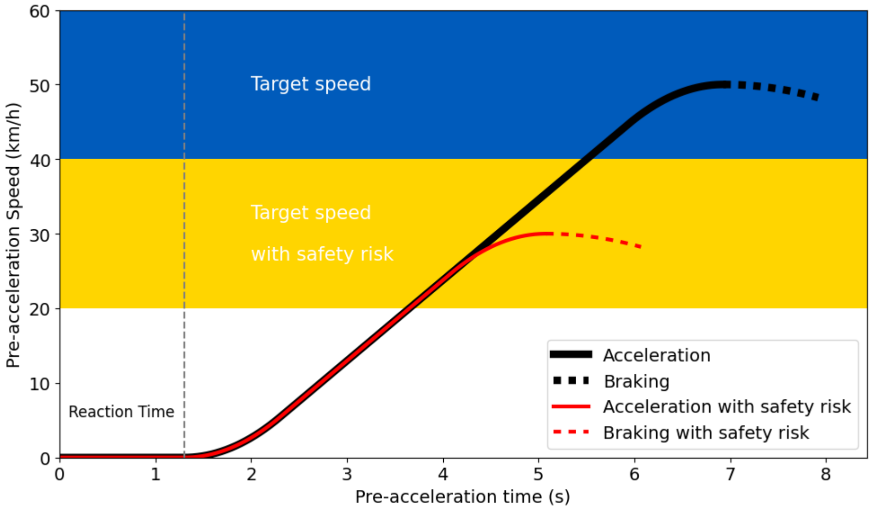


FIGURE 6.1: Suggested pre-acceleration speed profiles. The total pre-acceleration time refers to the time difference that the pre-cross-traffic turn arrow eliminates before the main cross-traffic turn signal turns green. The area under each speed–time curve represents the travel distance during pre-acceleration, which should not exceed the length of the bus cross-traffic turn priority box.

The primary benefit for general traffic arises from the pre-acceleration of turning vehicles before the main signal turns green. By granting a turning pre-signal a few seconds earlier, vehicles can begin accelerating from a standstill and approach the main signal at a higher speed, rather than starting from a complete stop. The total acceleration time includes both the driver's reaction time and the vehicle's acceleration time to reach the target speed (see Figure 6.1). To avoid abrupt braking, the travel distance during pre-acceleration should not exceed the length of the bus cross-traffic turn priority box.

When there are potential safety risks — such as heavy pedestrian activity or frequent red-light violations at the end of the preceding signal phase — it is advisable to set a lower pre-acceleration target speed, as illustrated by the red line in Figure 6.1. For safety, it may be beneficial to allow a slightly longer pre-acceleration time. This forces drivers to decelerate slightly before entering the intersection, as indicated by the braking (dashed) sections of the two curves in Figure 6.1. This approach can help shorten stopping distances and reduce collision risk. However, the trade-off between safety and performance warrants further investigation in future studies. Nevertheless, any degree of pre-acceleration leads to greater operational efficiency compared to having no pre-acceleration at all.

While the bus cross-traffic turn priority box offers significant efficiency benefits, higher approach speeds at intersections create a safety concern. It is recommended to limit the pre-acceleration speed of general traffic to 10–20 km/h below the speed limit. In situations with elevated safety concerns — such as high pedestrian volumes or frequent red-light violations — the desired pre-acceleration speed should be further reduced (Figure 6.1).

6.2.5 Microscopic simulation

Microscopic simulations are conducted using Aimsun microscopic traffic simulation to evaluate the effectiveness of the bus cross-traffic turn priority box. Two intersections were simulated: the first is a standardized three-way intersection, used to assess the priority box's performance under various demand scenarios, thereby providing robust evidence of its effectiveness. The second simulation applies the proposed solution to a real-world intersection, utilizing real-time data to validate its practical benefits. To ensure the robustness of the results and mitigate uncertainty due to stochastic traffic arrivals, each simulation scenario is replicated 100 times. The average values across these replications are taken as the final results.

The first simulation features a stylized three-way intersection, consisting of a two-through-lane arterial road and a one-through-lane collector road (Figure 6.2). All approach segments in the simulation are 500 meters long, exceeding the maximum queue lengths to ensure all delay is fully captured. Seven different demand scenarios were tested over a one-hour period to analyze the efficiency of the bus cross-traffic turn priority box under varying levels of traffic demand.

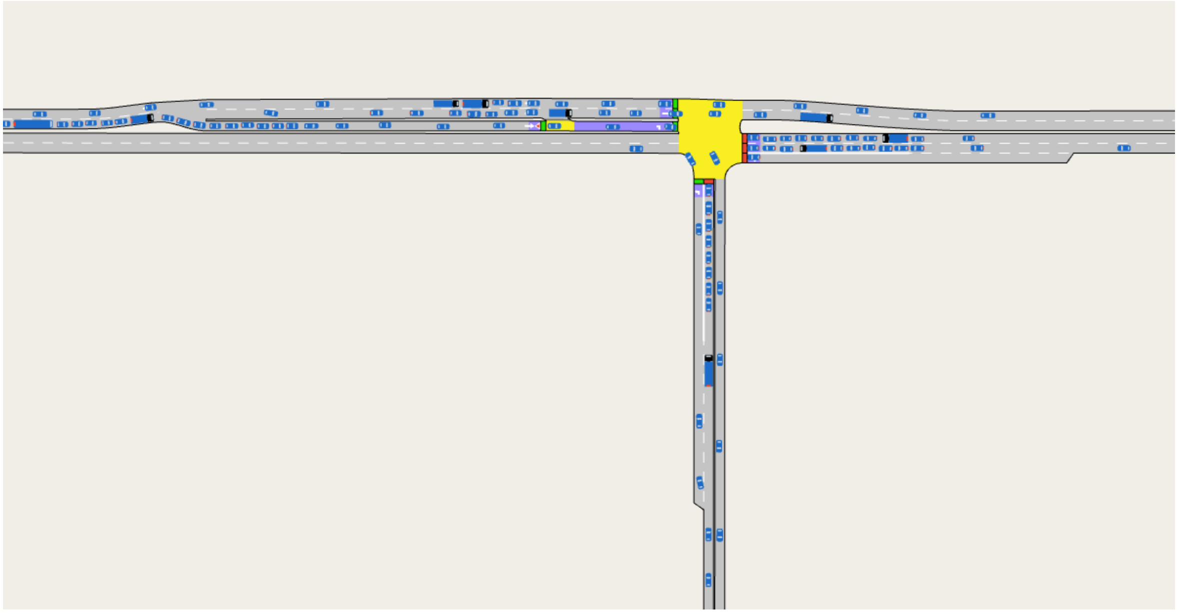
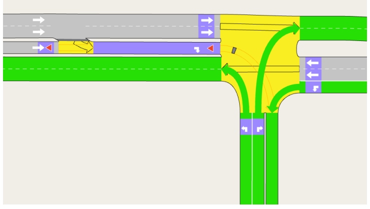
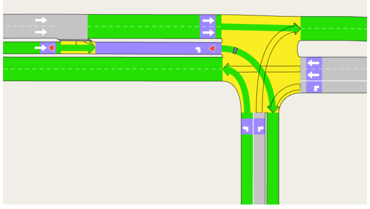
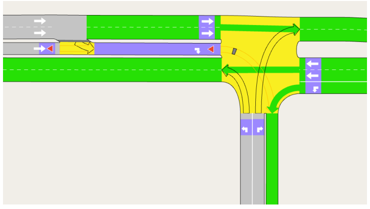


FIGURE 6.2: Simulation layout of the standardized three-way intersection. The bus cross-traffic turn priority box is located on the western approach. Eastbound and westbound through traffic are designated as the primary flow directions at this intersection.

The signal timing at this intersection is actuated (Table 6.2), simulating the adaptive control nature of the Sydney Coordinated Adaptive Traffic System (SCATS). The three-phase plan remains consistent between the base model without the bus cross-traffic turn priority box and the alternative scenario with the priority box. To emulate SCATS's adaptive features, the actuated traffic signal control applies a 3-second extension when detectors (shown in purple) detect a vehicle entering the intersection.

All phases at this intersection have a yellow time of 3 seconds and an all-red time of 4 seconds. The primary difference lies in the activation of the pre-cross-traffic turn signal on the western approach. In the alternative scenario, the pre-cross-traffic turn red arrow is deactivated 7 seconds before the start of Phase Φ_2 and activated at the end of Phase Φ_2. At the end of Φ_2, the main cross-traffic turn signal is extended by a maximum of 6 seconds to ensure clearance within the bus cross-traffic turn priority box. This extension allows vehicles within the priority box to proceed through, creating an empty priority box for pre-acceleration at the beginning of the next cycle. Consequently, this setup prevents vehicles from becoming stuck within the priority box and minimizes lost time at the intersection based on the microscopic results.

TABLE 6.2: Signal phase plan for the standardized three-way intersection. Actuated signal control dynamically adjusts each phase's green time from 0 seconds (skipping the phase) to the maximum allocated green time. The phase plan remains the same between the base design (without the bus cross-traffic turn priority box) and the alternative design (with the bus cross-traffic turn priority box).

 Φ_1 : 0-20s Northbound left turn and right turn, Westbound left turn	 Φ_2 : 0-20s Northbound left turn, Eastbound through and right turn	 Φ_3 : 0-60s Eastbound through, Westbound through and left turn

A total of seven traffic demand scenarios are employed for this standardized intersection. To simulate peak traffic effects, the hourly traffic demand presented in Table 6.3 was allocated as follows: 20% to the first 15 minutes, 30% from 15-30 minutes, and 50% for the final 30 minutes of the simulation. The high demand scenario reaches the intersection's capacity while preventing infinite queues that could result in a large number of vehicles failing to enter the simulated network. Different demand levels are tested, with medium and low demands set at 75% and 50% of the high demand scenario, respectively. Additionally, other traffic demand scenarios are derived by modifying the medium demand baseline. The high bus cross-traffic turn scenario simulates an extreme condition where one bus turns every minute, testing the potential effects of bus spill back from the bus cross-traffic turn priority box. The high cross-traffic turn scenario simulates a situation where the cross-traffic turn volume is heavy, with demand reaching half of the through traffic in the same direction. Two tidal flow demand scenarios are also tested for westbound and eastbound traffic, respectively. In these scenarios, the tidal flow demand is set to 10% higher than the high demand scenario, while the other directions maintain the medium traffic demand.

Table 6.3: Hourly traffic demand for seven scenarios tested at the standardized three-way intersection. The scenarios include high cross-traffic turn demands for buses alone and across all three modes, as well as tidal flows for westbound and eastbound traffic. These variations aim to capture a wide range of demand conditions, providing robust evidence of the efficiency improvements resulting from the implementation of the bus cross-traffic turn priority box.

Demand Scenario	**Direction**		**Car**	**Truck**	**Bus**
Low	Westbound	Through	800	40	20
		Left turn	100	5	2
	Eastbound	Through	800	40	20
		Right turn	200	10	5
	Northbound	Left turn	200	10	5
		Right turn	100	5	2
Medium	Westbound	Through	1200	60	30
		Left turn	150	8	4
	Eastbound	Through	1200	60	30
		Right turn	300	15	8
	Northbound	Left turn	300	15	8
		Right turn	150	8	4
High	Westbound	Through	1600	80	40
		Left turn	200	10	5
	Eastbound	Through	1600	80	40
		Right turn	400	20	10
	Northbound	Left turn	400	20	10
		Right turn	200	10	5

Demand Scenario	Direction		Car	Truck	Bus
High bus cross-traffic turn	Westbound	Through	1200	60	30
		Left turn	150	8	4
	Eastbound	Through	1200	60	30
		Right turn	300	15	60
	Northbound	Left turn	300	15	60
		Right turn	150	8	4
High cross-traffic turn	Westbound	Through	1200	60	30
		Left turn	150	8	4
	Eastbound	Through	1200	60	30
		Right turn	600	30	15
	Northbound	Left turn	600	30	15
		Right turn	150	8	4
High westbound	Westbound	Through	1760	88	44
		Left turn	220	12	6
	Eastbound	Through	1200	60	30
		Right turn	300	15	8
	Northbound	Left turn	300	15	8
		Right turn	220	12	6
High eastbound	Westbound	Through	1200	60	30
		Left turn	150	8	4
	Eastbound	Through	1760	88	44
		Right turn	440	22	11
	Northbound	Left turn	440	22	11
		Right turn	150	8	4

The second simulated intersection is selected to demonstrate the real-world implications of the proposed bus cross-traffic turn priority box. This simulation is based on real-world conditions at the intersection of Blaxland Road and Victoria Road in Ryde, Greater Sydney (Figure 6.3),

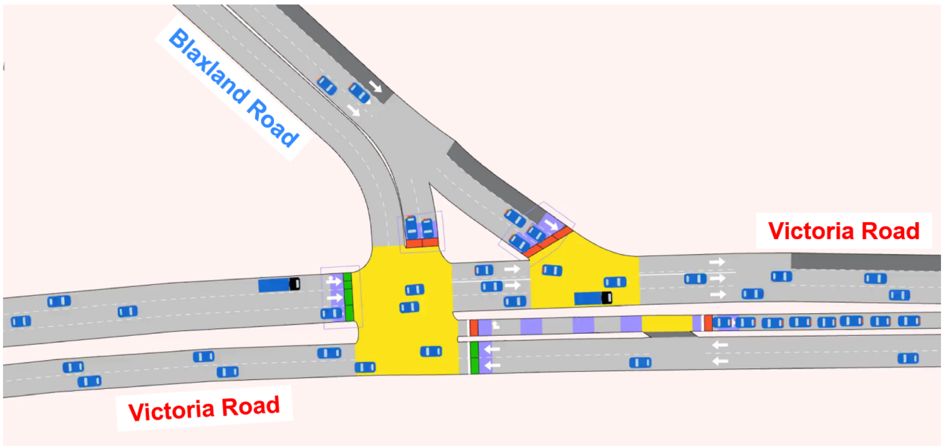
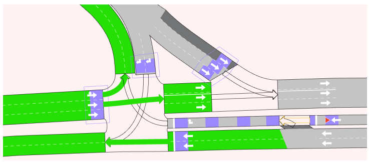
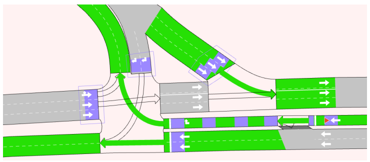
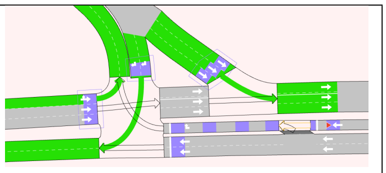


FIGURE 6.3: Simulation layout of the intersection at Blaxland Road and Victoria Road, Ryde, Greater Sydney, Australia. The bus cross-traffic turn priority box is implemented on the eastern approach of Victoria Road. Significant delays are experienced by westbound cross-traffic (right) turning traffic from Victoria Road into Blaxland Road.

during the morning peak hours (7–9 am). As shown previously in Figure 5.2, this intersection is among those identified with significant delays attributable to bus cross-traffic turns and is therefore used as one of the illustrative case studies. The simulated area includes the full stretch between the preceding and following intersections in all directions, ensuring that even the longest observed queues are fully captured. This intersection experiences substantial congestion caused by cross-traffic turns affecting both buses and general traffic during peak periods, making it an ideal location for testing the priority box intervention. The simulation inputs are based on average signal timings and traffic volumes observed on weekdays in July 2023, providing a realistic representation of typical operating conditions

The signal timing is the average phase durations for each hour from SCATS, archived from real-time traffic signal operations (Table 6.4). All phases at this intersection have a yellow time of 3 seconds and an all-red time of 3 seconds. To simulate the adaptive features of SCATS, the actuated traffic signal control is used with a 3-second extension time when detectors (shown in purple) detect a vehicle entering the intersection. The variance in phase durations is controlled to maintain the average phase times for each phase as outlined in Table 6.4. The three-phase plan remains consistent between the base scenario without the bus cross-traffic turn priority box and the alternative scenario with the priority box. The only difference lies in the activation of the pre-cross-traffic turn signal on the eastern approach. In the alternative scenario, the pre-cross-traffic turn red arrow is deactivated 6 seconds before the start of Phase Φ_2, and the main cross-traffic turn signal is allowed to extend by up to 6 seconds — rather than the usual 5-second extension in other signal phases — to ensure the priority box is cleared before the next cycle begins.

TABLE 6.4: Signal phase plan for the intersection of Blaxland Road and Victoria Road, Ryde, Greater Sydney, Australia. Since this intersection is actuated and controlled by the SCATS system, the signal times are allowed for a 5-second variation in signal phase durations based on the detection of arriving vehicles. The phase plan remains consistent between the base scenario (without the bus cross-traffic turn priority box) and the alternative scenario (with the priority box).

 Φ_1 : 65s Eastbound through and left turn, Westbound through	 Φ_2 : 18s Westbound through and right turn, Southbound left turn	 Φ_3 : 11s Eastbound left turn, Southbound right and left turn

The demand for the intersection of Blaxland Road and Victoria Road is updated every 15 minutes within the studied time period (Table 6.5). The main traffic flow is eastbound towards the city, with an average of 2,258 cars and 157 trucks per hour. Despite the westbound right-turn (cross-traffic turn) flow being relatively low — 359 cars, 8 trucks, and 12 buses per hour — the substantial volume of opposing eastbound traffic leads to significant delays for the westbound right-turn movement. Buses only operate between the eastern leg (Victoria Road) and the northern leg (Blaxland Road) due to the presence of a shopping center to the north on Blaxland Road. Consequently, there is a notable bus volume making westbound right turns and southbound left turns, with 12 and 27 buses per hour, respectively. This emphasizes the importance of optimizing this intersection for bus operations to enhance service reliability and reduce delays for both buses and general traffic.

TABLE 6.5: Hourly average traffic demand at the Intersection of Blaxland Road and Victoria Road, Ryde, Greater Sydney, Australia. The eastbound through movements, which are citybound, experience the highest traffic volumes. This creates significant delays for westbound right turns due to conflicts with the eastbound through traffic.

Direction		Car	Truck	Bus
Westbound	Through	1405	51	0
	Right turn	359	8	12
Eastbound	Through	2258	157	0
	Left turn	23	2	0
Southbound	Left turn	818	5	27
	Right turn	118	5	0

6.3 Results

The outcomes of the seven tested demand scenarios for the three-way standardized intersection are summarized in Figure 6.4 and Table 6.6. Overall, implementing the bus cross-traffic turn priority box yielded notable delay reductions for all three vehicle types, especially buses. Figure 6.4 provides a delay time comparison with and without the priority box under various demand scenarios, while Table 6.6 presents the percentage delay reduction for cars, trucks, and buses, along with person-delay savings that assume car/truck occupancy of 1.1 and bus occupancy of 20.

A key feature underpinning these improvements is the pre-acceleration effect. By allowing vehicles to begin moving before the cross-traffic turn phase starts, the priority box reduces start-up lost time, effectively increasing the usable green interval for cross-traffic turns. This enables more vehicles to pass within a given cycle, which can translate into up to a 19% delay reduction for cars and trucks under certain scenarios. However, under high demand in all directions, capacity constraints mean that cars and trucks see only limited improvements. Even in this challenging scenario, buses still benefit substantially by bypassing queues in the cross-traffic turn lane. Meanwhile, in scenarios exhibiting tidal flow — for example, when either the eastbound or westbound approach is notably higher than the other — spare green time can be productively reallocated to other directions, thereby producing a larger overall delay reduction for cars and trucks.

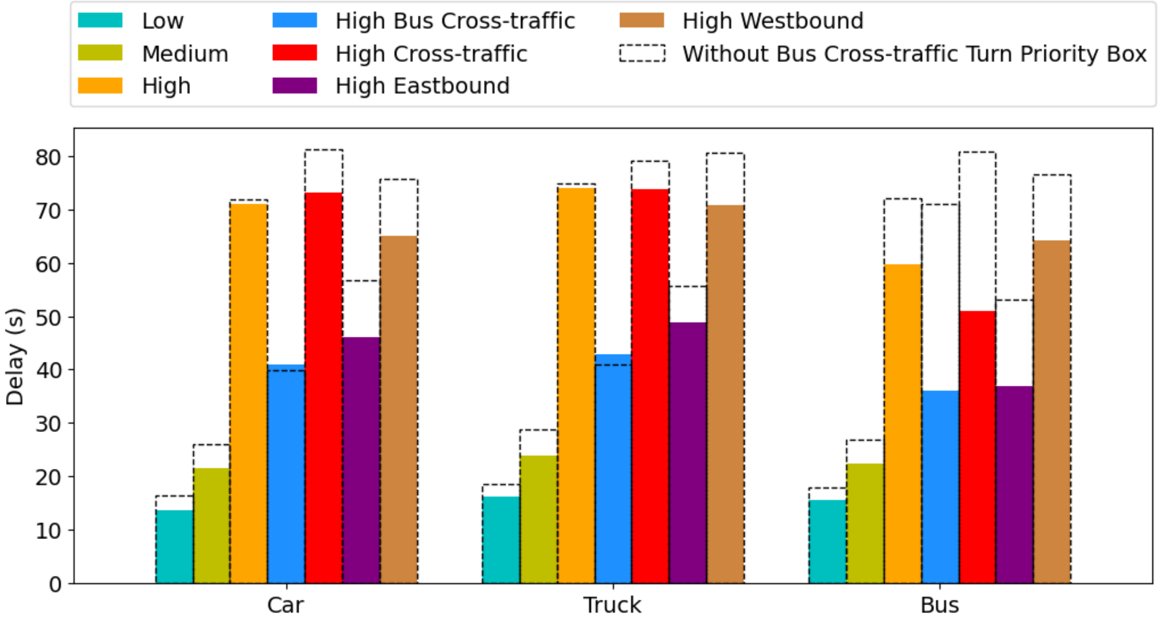


FIGURE 6.4: Delay time comparison with and without the bus cross-traffic turn priority box at the three-way standardized intersection under various demand scenarios. Delay reductions are observed for most scenarios, especially for buses. The pre-acceleration feature also benefits cars and trucks in most instances.

TABLE 6.6: Percentage delay reductions for cars, trucks, and buses, as well as overall person delay, across various demand scenarios at the standardized three-way intersection and the intersection of Blaxland Road and Victoria Road. Cars and trucks show moderate delay reductions, while buses exhibit the highest reductions. Overall person-delay reductions are notably significant when accounting for high bus occupancy (20) compared to car/truck occupancy (1.1).

Demand Scenario	**Car**	**Truck**	**Bus**	**Person Delay**
Low	15.48	13.54	12.83	14.56
Medium	17.46	17.21	16.56	17.17
High	0.91	1.00	16.96	5.81
High Bus Cross-traffic	-2.66	-4.63	49.30	30.46
High Cross-traffic	9.94	6.87	36.71	17.96
High Eastbound	18.90	12.38	30.61	21.98
High Westbound	13.88	12.20	16.22	14.27
Intersection of Blaxland Road and Victoria Road	24.74	28.87	69.17	34.62

Several specific observations emerged from the scenario-based testing. When the eastbound demand is high, the cross-traffic turns are congested. The priority box shortens the cross-traffic turn phase, thus reducing total intersection delay. A similar mechanism applies under high westbound demand, where more efficient cross-traffic turns shorten red times for through traffic. An exception in delay reduction occurs in the high bus cross-traffic scenario, where the bus volume is so frequent (about one bus per minute) that buses occupy the priority box almost every cycle; as a result, cars and trucks gain negative delay outcomes since they lose the opportunity to pre-accelerate. Nonetheless, from a person-delay standpoint, the large capacity of buses ensures a net benefit to overall passenger throughput. In essence, while this scenario yields a small trade-off for cars and trucks delays, the reduction in bus delay indicates significant benefits once occupancy is factored into the analysis.

The intersection of Blaxland Road and Victoria Road in Ryde, Greater Sydney, is tested to validate the effectiveness of the priority box concept under real-world conditions. As illustrated in Figure 6.5 and detailed in the final row of Table 6.6, the implementation of the priority box resulted in substantial benefits for buses, with delays decreasing by over 69%. This significant improvement is expected, given that the primary bus delays at this intersection occur during cross-traffic turns from Victoria Road into Blaxland Road. In contrast, left turns from Blaxland Road into Victoria Road experience only minor delays.

Buses benefited from the priority box by bypassing the queue in the cross-traffic turn lane and positioning themselves at the front during the cross-traffic turn phase. This maneuver significantly reduced the waiting times typically encountered by buses, enabling them to complete cross-traffic turns within a single signal cycle. Consequently, this enhancement not only improved bus service reliability and punctuality but also optimized overall intersection efficiency.

In addition to the substantial benefits observed for buses, cars and trucks also experienced significant delay reductions at this intersection, as the westbound cross-traffic turn is the only congested movement. With the implementation of the bus cross-traffic turn priority box, the capacity for cross-traffic turns increased significantly, allowing more vehicles to be processed during each green phase — within the same green duration as without the priority box. This led to notable reductions in both delay and queue length. Importantly, since the green time for the westbound cross-traffic turn remains unchanged, there is no additional delay imposed on other movements. The high occupancy rates of buses contribute to an overall in-vehicle person delay reduction exceeding 34%. This highlights the substantial improvement in intersection efficiency achieved through the implementation of the bus cross-traffic turn priority box.

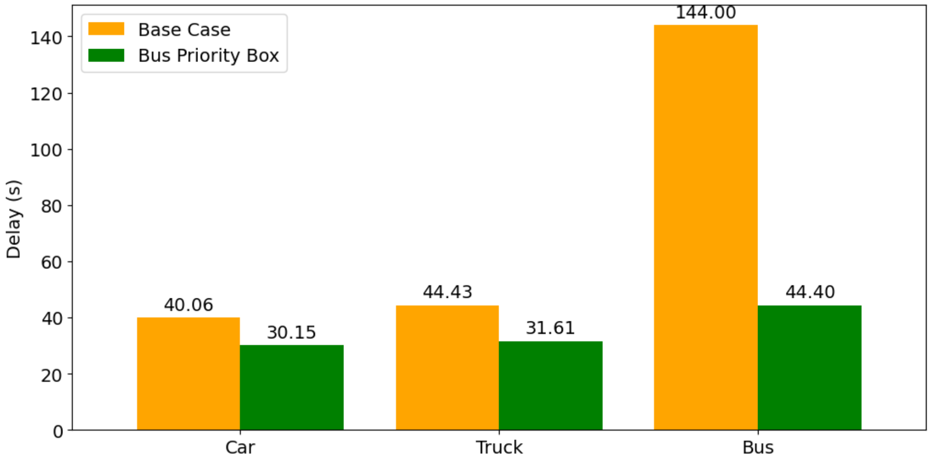


FIGURE 6.5: Delay time comparison with and without the bus cross-traffic turn priority box at the intersection of Blaxland Road and Victoria Road in Ryde, Greater Sydney. Significant delay reductions are observed for buses. Cars and trucks benefit to a lesser but still notable extent.

Taken together, these results underscore the effectiveness of the bus cross-traffic turn priority box at both a standardized three-way intersection with various demand and a real-world intersection in Sydney. Buses has the largest delay reduction, especially in scenarios where cross-traffic turn demand is congested or when bus volumes are high. Cars and trucks frequently see delay reductions as well, except under extremely high bus cross-traffic circumstances that limit pre-acceleration opportunities. Critically, considering the bus occupancy amplifies the person-delay benefits, indicating that prioritizing buses can achieve a substantial impact on overall passenger throughput.

A major concern with implementing the bus cross-traffic turn priority box is the potential for bus queue spill back at the entrance of the priority box. Figure 6.6 illustrates a scenario from the simulation where spill back occurs. This spill back may result from an insufficient length of the bus cross-traffic turn priority box or excessively high bus cross-traffic demand.

In the simulations conducted for this paper, the cross-traffic turn bays are generally filled when bus spillback occurs. This results in the obstruction of the inner through lanes by the

FIGURE 6.6: Spilled bus queue at the bus cross-traffic turn priority box in the three-way standardized intersection. The spilled bus queue blocks the inner through lane. No significant delay increases are observed for through traffic because the cross-traffic turn bay is generally full under this scenario, causing upstream blockage of the inner through lane.

accumulating cross-traffic queue upstream of the bus cross-traffic turn priority box entrance. At the two tested intersections, the through movement in the same direction as the critical cross-traffic movement is typically not critical, as it receives green time over two signal phases. This means that, although the spilled cross-traffic turning buses may block the inner through lane near the intersection, their impact on other through traffic is not detrimental to the delay for cars.

In practical terms, the priority box mechanism effectively addresses the longstanding issue of delayed cross-traffic turns without the need for additional infrastructure or dedicated lanes. By optimizing the utilization of existing road space and adjusting signal phases, the priority box minimizes delays for buses while enhancing traffic flow for other vehicles. As a result, the bus cross-traffic turn priority box emerges as a robust solution that not only improves public transport reliability and punctuality but also supports more efficient general traffic operations at congested urban intersections.

6.4 Conclusion

This chapter evaluates the effectiveness of the bus cross-traffic turn priority box in improving bus performance at intersections, with results showing significant delay reductions for buses, and some improvements for general traffic. The priority box allows buses to bypass the cross-traffic turn queue using through lanes, reducing bus delays. Testing across various demand scenarios, including high bus volumes and congested cross-traffic turns, demonstrates that the priority box significantly benefits buses.

In addition to the improvements for buses, cars and trucks also experience reductions in delay due to the pre-acceleration feature, which reduces the start-up loss time for cross-traffic turns. However, in extreme cases with high bus volumes, the opportunities for general traffic to pre-accelerate are limited, which may result in a minor increase in delays for cars and trucks. Nevertheless, when accounting for bus occupancy, the overall person-delay savings remain significant.

Compared with the traditional priority measures, the priority box is a space-efficient and practical solution for urban intersections, offering substantial benefits without requiring extensive infrastructure changes. It optimizes the use of existing road space, improves bus service reliability, and enhances traffic flow, making it a promising solution for addressing intersection congestion in rapidly growing cities. Further discussions on the results are presented in Chapter 7.

CHAPTER 7

Discussion and Conclusion

7.1 GTFS-Realtime Trip Updates data

Chapter 3 presents a comprehensive data processing pipeline for GTFS-Realtime Trip Updates data, bridging the gap between the vast volume of raw information generated by PTAs and the actionable insights necessary for effective transit planning and operations. Chapter 3 also demonstrates the development and technical validation of GTFS-Realtime Trip Updates data in Greater Sydney, which includes bus arrival and departure time estimates. The data outputs provide a reliable and flexible foundation for high-resolution bus performance analysis, offering valuable insights into operational efficiency at the stop level. By enabling detailed stop-to-stop performance evaluations, the dataset creates opportunities to explore critical but previously overlooked aspects of bus operations, including traffic-related delays, scheduling inconsistencies, and the spatial dynamics of urban transit systems.

Validation against in-person observations demonstrates that the pipeline's forecasts closely align with observed bus arrivals and departures, maintaining small tolerances of 8 seconds for arrivals and -5 seconds for departures. This level of accuracy makes the pipeline well-suited for second-level and stop-to-stop granularity analyses, providing valuable insights into operational efficiency. Such precision is typically unattainable when relying on static schedules or manual observations, underscoring the data's potential for driving evidence-based improvements in public transport services.

The data processing pipeline and its associated dataset serve as powerful tools for both transit authorities and researchers. With high-resolution, consistently updated data, stakeholders can monitor bus delays, gauge service reliability, and develop targeted interventions such as bus lanes and transit signal priority. As showcased in Chapters 4 and 5, the pipeline enables the identification of segment-specific speed fluctuations, intersection-related delays, and impacts of priority measures, ultimately supporting more efficient bus operations. Moreover, its adaptability — tested across six metropolitan areas in Australasia — illustrates global scalability, paving the way for standardized, high-resolution datasets that can inform data-driven decision-making across diverse transit systems.

The pipeline's robust data outputs also highlight the potential for broader applications. Incorporating other data sources, such as roadworks, weather conditions, and traffic volumes, would further enhance its capacity to optimize transit systems. The flexible architecture of the pipeline allows for future integration with emerging data standards (e.g., GTFS-Realtime v2 or SIRI), thus expanding its applicability to a wide range of transit environments.

Ultimately, the processed GTFS-Realtime data offers a valuable framework for advancing research and practice in public transit performance. By enabling fine-grained analyses of bus operations, the pipeline supports the development of efficient, reliable, and sustainable transportation networks. Its utility for assessing transit interventions, measuring operational improvements, and guiding policy decisions underscores the transformative impact that real-time, high-resolution data can have on addressing urban mobility challenges in a data-driven era.

7.2 Bus lane performance evaluation

The data outputs in Chapter 3 provide the foundation for the analysis in Chapter 4, where the impact of route characteristics — such as bus lanes, traffic volume, bus cross-traffic turns, and traffic signals — on stop-to-stop marginal delay at a second-level granularity is examined. The research in Chapter 4 showcases the application of GTFS-Realtime data to assess the effectiveness of bus priority lanes at a microscopic level, utilizing detailed spatial information on bus priority infrastructure. By analyzing 20 months of data from the Sydney metropolitan area, the findings confirm that bus lanes can enhance bus performance primarily by reducing the variability in marginal delays, rather than significantly increasing average speeds. While their direct effect on efficiency may diminish under during periods of severe congestion, the significance of bus lanes lies in their capacity to create more predictable travel conditions for passengers and operators. In other words, even if travel times are not drastically shortened when congestion is most intense, bus lanes contribute to more stable and reliable journey times, mitigating the uncertainties that passengers and operators typically face during peak periods.

These results, although centered on Sydney, likely generalize to other urban contexts where traffic congestion and variable travel conditions challenge bus service reliability. Although statistically significant impacts on the average marginal delay are observed, it is expected that long-term adjustments — such as refined timetabling informed by ongoing data analysis — will counterbalance these fluctuations. Ultimately, the more notable advantage of bus lanes is their ability to reduce the variability in marginal delay, thereby improving overall reliability and predictability. Regression tree, pooled OLS, and between-effects models all produce intuitive and statistically significant coefficients for bus lanes, while within-effects models further underscore that bus lanes effectively counteract the unreliability associated with peak periods.

The findings in Chapter 4 also indicate that implementing suitable bus priority measures, such as transit signal priority and route designs that minimize cross-traffic turns, can be effective in enhancing schedule adherence and reliability. Traffic signals increase the average and variability of marginal delay due to the absence of active bus signal priority during the study period, although such measures have been deployed in the study area historically. Effective transit signal priority should integrate realtime information on bus delay, vehicle occupancy, and flexible signal phase adjustments. Cross-traffic turns, in particular, emerge as a key challenge for bus performance. When designing routes, it is generally advisable to avoid these turning movements where possible. However, at intersections where bus turns cannot be eliminated, adopting priority geometries — such as dedicated bus queue jump lanes and implementing adaptive signal timing can help maintain efficient bus operations. These measures should be carefully calibrated to balance improvements in bus service with the overall efficiency of general traffic. Building on these insights, Chapter 5 employs GTFS-Realtime data to further examine the impact of cross-traffic turns more thoroughly and Chapter 6 develops novel strategies for mitigating their adverse effects.

A common strategy to enhance bus performance is schedule padding, which involves increasing the timetabled journey time to lower the chances of buses arriving late (Aemmer et al. 2022). However, this can lead to early arrivals, causing longer waiting times for passengers than late buses. The models in this study illustrate how bus lanes can overcome this issue. In the average marginal delay model, Bus-taxi and bus-HOV lanes show negative parameters, suggesting these lanes help reduce delay in the study area. The data, which spans the COVID period, shows reduced passenger boarding and traffic volumes, leading many buses to arrive ahead of schedule. However, the standard deviation models of marginal delay reveal that bus lanes significantly improve bus reliability, justifying further development of the bus priority network irrespective of passenger demand levels or timetable adjustments. Additionally, the study area's map underscores the limited implementation of bus lanes in Sydney, highlighting the potential to expand these reliability benefits.

As an alternative to schedule padding, transit agencies could consider demand responsive service and adjusting timetables using real-time data to improve on-time performance. When the travel demand fluctuates, especially during less predictable peaks in demand such as weather, road incidents or train system failures, the operator could dynamically add trips with real-time adjusted timetables to accommodate the demand without impacting the existing services.

Chapter 4 only analyses the marginal delay on stop-to-stop link level, which ignores the effects of road network. Bin et al. suggest considering link interactions among the bus lanes (Yu et al. 2015). Long distance public transport riders can have different travel options using different corridors, and the conditions on nearby, intersection or parallel corridors might impact performance. By focuses on stop-to-stop link level and utilizing the departure delay at previous stops and arrival delay at the next stops, this study designed to isolate the stop effects including dwell time from the stop-to-stop variation.

7.3 Bus cross-traffic turn

Chapter 4 indicates that cross-traffic turns, along with high traffic volumes, traffic signals, and right turns, have been identified as major contributors to bus unreliability. Building on the insights from Chapter 4, Chapter 5 delves deeper into the microscopic level impacts of cross-traffic turns by drawing on the GTFS-Realtime Trip Updates data introduced in Chapter 3.

Analyzing second-level bus performance data, the statistical tests in Chapter 5 highlight significant differences in the mean and standard deviation of stop-to-stop marginal delays, as well as operational speeds, caused by cross-traffic turns. These findings confirm that cross-traffic turns not only increase schedule delays and reduce operational speeds but also disproportionately contribute to service unreliability. This highlights the substantial impact of cross-traffic turns on overall bus performance and reinforces the critical need for implementing bus cross-traffic turn priority measures. Notably, stop-to-stop links without cross-traffic turns consistently outperform those with such maneuvers, highlighting the importance of minimizing or eliminating cross-traffic turns when designing bus routes.

To enable the intersection level of analysis, GTFS-Realtime Vehicle Position data is employed for a microscopic study that aggregates bus AVL data. The processed Vehicle Positions data facilitates the creation of speed maps, showcasing 5-meter average speed intervals for bus trajectories. The average speed map can not only help in having a detailed understanding of local delay cause, but can also identify the location of start of queue. Within mixed traffic lanes when traffic data is unavailable, the GTFS-Realtime Vehicle Position data can be used to identify the length of the queue. The positive validation outcomes affirm the method's practicality, suggesting that GTFS-Realtime Trip Updates and Vehicle Positions data can effectively be combined in future research to enrich transit operation studies.

Among the various types of bus cross-traffic turns analyzed, cross-traffic turns at signalized intersections were found to have the most significant negative impact on bus operational efficiency. While sign-controlled intersections also contribute to delays, their impact is generally less severe. This suggests that avoiding cross-traffic turns at signalized intersections and performing early or late cross-traffic turns at sign-controlled intersections, which are typically less congested, can improve bus performance.

Yield/stop sign-controlled intersections and roundabouts have similar effects on bus operations, with roundabouts causing slightly more negative impacts due to their smaller typical turning radius. Although avoiding the problems identified with cross-traffic turns at signals, this can result in alternative delays for buses. This finding highlights the importance of design considerations in traffic management systems to accommodate public transit efficiently.

Past strategies for mitigating the adverse effects of cross-traffic turns have focused on route designs that avoid such turns, especially at congested intersections. Examples include performing early or late cross-traffic turns to bypass congestion, incorporating jughandle ramps, or in extreme cases, using three non-cross-traffic turns instead. Another approach is to minimize lane changes required for executing these turns during route planning by being aware of the placement of the bus stop immediately before the intersection.

When cross-traffic turns are unavoidable, particularly at intersections with significant bus cross-traffic turn demand and congestion, bus priority measures such as queue jump lanes and transit signal priority can be implemented. Queue jump lanes are generally effective but require substantial road space dedicated to bus cross-traffic turns, making them most suitable in areas with sufficient lateral space. However, in space-constrained areas, traditional bus priority measures may fail to improve bus performance and can cause disproportionate delays to general traffic. This, in turn, may exacerbate congestion, making it more challenging for buses to approach the intersection and potentially increasing delays for both buses and general traffic.

7.4 Bus cross-traffic turn priority box

Building on the findings from Chapter 5, Chapter 6 introduces a novel solution to bus cross-traffic turn delays: the bus cross-traffic turn priority box. This novel in-lane queue jump lane allows buses to bypass the cross-traffic turn queue by utilizing the faster through traffic lanes, significantly reducing bus delays.

The priority box functions similarly to a bicycle box, offsetting the cross-traffic turn queue upstream from the intersection. It is specifically designed to mitigate bus cross-traffic turn delays at critical intersections without requiring additional lateral space. This approach reduces bus delays while causing minimal disruption to general traffic, making it an effective alternative to traditional bus priority measures, which often require significant infrastructure changes or dedicated lanes.

Testing results in Chapter 6 across various demand scenarios, including high bus volumes and congested cross-traffic turns, confirms that the priority box benefits buses while maintaining or even enhancing the efficiency of general traffic in critical scenarios. Overall, the results show that the bus cross-traffic turn priority box is an effective solution for reducing delays for buses and improving intersection performance in congested urban environments. By addressing the challenges associated with cross-traffic turns and optimizing the use of limited road space, this measure offers a practical approach to improving public transport reliability without negatively impacting, and in most cases even benefiting, general traffic. This makes the priority box a promising alternative to traditional bus priority measures.

Buses can bypass queues in the cross-traffic turn lanes by using the faster through traffic lanes and positioning themselves at the front during the cross-traffic turn phase. This maneuver enables buses to complete cross-traffic turns within a single signal cycle, significantly enhancing service reliability and punctuality. However, the benefits of the priority box depend on a sufficient speed difference between through lanes and cross-traffic turn lanes. If the through lanes are slower than the cross-traffic turn lanes, buses cannot bypass the cross-traffic queue and must join the end of the queue, thereby negating the primary advantage of the bus priority box. Nonetheless, the pre-acceleration feature still provides benefits by allowing turning vehicles to begin accelerating earlier, maintaining some level of efficiency improvement.

The bus cross-traffic turn priority box does not conflict with current traffic regulations and is easy for road users to understand. The bus cross-traffic turn priority box is designed exclusively for protected cross-traffic turn intersections and is not applicable to permitted cross-traffic turn intersections.

Since the bus cross-traffic turn priority box relies on vehicle detectors within the designated area, the green phase for cross-traffic turns are extended as long as vehicles are present in the box. As a result, the need for a dedicated yellow cross-traffic turn phase may be eliminated, since the green arrow would only terminate once the box is clear. This mechanism has the potential to further enhance intersection efficiency by reducing lost time associated with phase transitions. While this effect was not simulated in the current study due to safety considerations, it represents a promising avenue for future research.

However, safety concerns arise from the higher approach speeds enabled by pre-acceleration. To address these concerns, it is recommended to limit pre-acceleration speeds to 10–20 km/h below the speed limit, especially in areas with high pedestrian volumes or frequent red-light violations. Another concern is the potential for right-turn queue spillback. The priority box is specifically designed for locations where turning bay length is not a major constraint but lateral road space is limited. In scenarios where the priority box length is insufficient, queues may spill back into the through lanes, potentially reducing intersection efficiency. Nevertheless, since the priority box improves turning efficiency, the overall length of turning queues is expected to be shortened in most cases, thereby minimizing the risk and impact of spillback even under peak traffic conditions.

The bus cross-traffic turn priority box offers a compelling advancement in traffic management by balancing the needs of public transportation and general traffic flow. Its implementation can lead to substantial improvements in intersection efficiency, reduced bus delays, and enhanced overall traffic performance, making it a valuable addition to urban transit strategies. The findings in Chapter 6 support the adoption of the bus cross-traffic turn priority box in similar urban settings where space constraints limit the feasibility of traditional bus priority measures. Further research and real-world trials are recommended to validate these results, explore long-term benefits, and address any potential safety considerations associated with higher approach speeds due to pre-acceleration. By doing so, cities can adopt more efficient and reliable transportation systems that cater to both public transit and general traffic needs without necessitating significant infrastructural investments.

7.5 Limitations and future work

While Chapter 3 addresses numerous barriers to using GTFS-Realtime data, several limitations and opportunities for future improvement persist. First, although the data processing pipeline presented in Chapter 3 is designed to be adaptable for most GTFS-Realtime data worldwide, it has only been tested in Australian and New Zealand cities. Its robustness in regions such as North America or Europe, where data formats may differ significantly, remains unverified. Future research should therefore evaluate the pipeline's performance in these areas and adapt it to support a broader range of GTFS-Realtime formats.

Additionally, the absence of passenger boarding information restricts the dataset's capacity to account for crowding and its effect on bus performance. As TfNSW's transition to the GTFS-Realtime v2 standard, refining the pipeline to include enhanced data fields could improve overall data quality, while incorporating other data standards (e.g., SIRI) may expand its applicability to more diverse transit systems worldwide. Moreover, although the GTFS-Realtime Trip Updates discussed in Chapter 3 were validated at a single bus stop, resource constraints have thus far prevented validation at larger scales, such as across entire metropolitan areas. Finally, since TfNSW remains the only source of the GTFS-Realtime Trip Updates data — without transparent information on how arrival and departure times are estimated — underlying system errors may exist. Despite the reasonable results observed in Chapters 4 and 5, further testing and validation of the GTFS-Realtime dataset are recommended.

In Chapter 4, the analysis relies on linear regression, a supervised method that requires predefined input variables. Although various regression models were tested, as well as regression trees, the selected model could still potentially be further optimized for better performance. Additionally, while machine learning models could potentially provide a better fit for the data, these models often lack interpretability, making them less suitable for understanding and explaining results in a human-readable format.

Due to limitations in computing resources, the regression analysis in Chapter 4 involved sampling, which may have resulted in some features being excluded from the models. Future studies should consider using more computationally efficient techniques or more powerful hardware to obtain more robust results.

The impacts of bus priority lanes on general traffic were not evaluated in Chapter 4 due to the lack of intersection-level traffic volume data. Although prioritizing buses may potentially delay general traffic, it is crucial to assess both the benefits to transit operations and any adverse effects on overall traffic flow. In future work, incorporating intersection-level traffic data will enable a more comprehensive examination of these trade-offs, ensuring that bus priority measures yield net positive outcomes, including reductions in person delays.

While research elsewhere suggests that restricting certain turning movements can improve the benefits of bus lanes (Levinson and St. Jacques 1998), we currently lack sufficient data on intersection-level turning volumes to directly assess these effects. In the Sydney context, bus lanes are typically implemented along major arterials where non-cross-traffic turning maneuvers are relatively limited. When buses are heavily affected by non cross-traffic turning conflicts, the buses can use through lanes to proceed. Consequently, non-cross-traffic turning movements are not generally a significant concern for bus operations in this setting. In future research, acquiring more detailed turning movement data could enable a deeper investigation of these interactions and inform strategies to further enhance bus priority measures in other urban contexts.

The stop-to-stop link comparisons in Chapter 5 use two-sample t-tests and K-S tests to compare the distribution of stop-to-stop links with and without cross-traffic turns. While the results show significant differences (low p-values), the large dataset size may have influenced these outcomes. Furthermore, some data distributions are skewed, which could affect the reliability of the test results. Alternative statistical techniques are recommended for future research.

The intersection comparison results in Chapter 5 introduce a novel method that spatially averages GPS data into 5-meter trajectory segments. While the average speed trajectory links provide promising results, only two intersections were analyzed, which may limit the generalizability of the findings. Although validation between the GTFS-realtime Vehicle Position and Trip Updates datasets shows reasonable alignment, both datasets are derived from the same GPS data source, which may introduce biases. Further investigation into this method is warranted.

The effectiveness of the bus cross-traffic priority box are tested using microscopic simulation in Chapter 6, but the simulation may still contain system errors. The behavior models within the simulation are simplified and may not fully capture real-world traffic dynamics, particularly for novel priority measures like the bus priority box. Additionally, results from only two scenarios may not be generalizable. Given that microscopic simulations rely on numerous input parameters, these inputs can significantly influence the reliability of the results. Future research should incorporate analytical methods that use formulas to further validate the effectiveness of the bus cross-traffic priority box.

No pedestrian behavior was modeled in the simulations presented in Chapter 6. Pedestrians influence traffic flow by requiring sufficient green time for crossing and by necessitating that turning vehicles yield to them. The impact of pedestrian movements on intersection capacity is not fully understood without pedestrian simulation. Additionally, jaywalking pedestrians could complicate cross-traffic turns. Future studies should include pedestrian modeling to better capture these dynamics and assess the potential impact on bus priority measures.

7.6 Conclusion

This study underscores the transformative potential of GTFS-Realtime data in advancing public transit performance analysis. By developing a data processing pipeline, the research bridges the gap between the vast quantities of raw data generated by PTAs and the actionable insights required for effective transit planning and operations. The pipeline enables granular evaluations of bus performance, providing the foundation for understanding and addressing critical factors such as traffic, scheduling inconsistencies, and spatial dynamics within urban transit systems.

Key findings underscore the critical role of targeted interventions, such as dedicated bus lanes, transit signal priority systems, and optimized route designs, in reducing delays and improving the reliability of bus services. Dedicated bus lanes not only provide a consistent travel path for buses but also significantly improve reliability by reducing their exposure to general traffic congestion, thereby enhancing operational efficiency.

Among all models tested, the most significant challenge in bus operation identified in this research is the impact of cross-traffic turns, a major source of bus delays and service unreliability. These maneuvers, often at signalized intersections, result in delays due to conflicting traffic flows and the large turning radius required for buses compared to smaller vehicles. Addressing this issue is vital to improving overall service quality and passenger satisfaction.

To tackle the challenges posed by bus cross-traffic turns, this study introduces the innovative bus cross-traffic turn priority box. This measure offers a practical and space-efficient solution by allowing buses to bypass cross-traffic turn queues via through lanes, significantly reducing delays and improving reliability at intersections. The priority box enhances intersection performance by enabling pre-acceleration for general traffic when buses are not present. The bus cross-traffic turn priority box contributes to more reliable, efficient, and user-friendly public transportation systems.

The analysis reveals that high-resolution, real-time data is essential for evaluating and implementing effective transit interventions. The processed dataset not only supports detailed performance assessments but also opens avenues for integrating additional data sources, such as traffic volumes and weather conditions, to further enhance transit system evaluations.

In conclusion, this research demonstrates how leveraging GTFS-Realtime data can improve public transit analysis, offering data-driven solutions to improve reliability, efficiency, and sustainability in urban transportation systems. By addressing key operational challenges and proposing innovative strategies, this work lays the groundwork for developing smarter, more resilient public transit networks that meet the demands of growing urban populations.

Bibliography

Abkowitz, Mark, Amir Eiger and Israel Engelstein (1986). ‘Optimal control of headway variation on transit routes’. In: *Journal of Advanced Transportation* 20.1, pp. 73–88.

Abusalim, Mahmoud (2020). ‘Accuracy and Effectiveness of GTFS Transit Feeds for Trip Planning in Public Transit Networks’. In.

Aemmer, Zack, Andisheh Ranjbari and Don MacKenzie (2022). ‘Measurement and classification of transit delays using GTFS-RT data’. In: *Public Transport*, pp. 1–23.

Arias, Daniel et al. (2021). ‘Using gtfs to calculate travel time savings potential of bus preferential treatments’. In: *Transportation Research Record* 2675.9, pp. 1643–1654.

Balcombe, Richard et al. (2004). ‘The demand for public transport: a practical guide’. In.

Baltagi, Badi Hani (2008). *Econometric analysis of panel data*. Vol. 4. Springer.

Barbeau, Sean (Oct. 2017). *What’s new in GTFS-realtime v2.0*. en. URL: `https://barbeau.medium.com/whats-new-in-gtfs-realtime-v2-0-cd45e6a861e9` (visited on 01/11/2021).

Barrett, Jeff (Apr. 2018). *Up to 73 Percent of Company Data Goes Unused for Analytics. Here’s How to Put It to Work*. en. URL: `https://www.inc.com/jeff-barrett/misusing-data-could-be-costing-your-business-heres-how.html` (visited on 01/04/2022).

Böcker, Lars, Martin Dijst and Jan Prillwitz (2013). ‘Impact of everyday weather on individual daily travel behaviours in perspective: a literature review’. In: *Transport reviews* 33.1, pp. 71–91.

Chang, Mark et al. (2008). ‘A review of HOV lane performance and policy options in the United States’. In.

Changnon, Stanley A (1996). ‘Effects of summer precipitation on urban transportation’. In: *Climatic Change* 32.4, pp. 481–494.

Chen, Xumei et al. (2009). ‘Analyzing urban bus service reliability at the stop, route, and network levels’. In: *Transportation research part A: policy and practice* 43.8, pp. 722–734.

Chin, Teck Kean and Tingsen (Tim) Xian (2024). *SCALUT/GTFS-Data-Pipeline-TfNSW-Bus: GTFS Data Pipeline for TfNSW Bus Datasets*. URL: `https://github.com/SCALUT/GTFS-Data-Pipeline-TfNSW-Bus` (visited on 29/05/2024).

Cződörová, Renáta, Marek Dočkalik and Jozef Gnap (2021). ‘Impact of COVID-19 on bus and urban public transport in SR’. In: *Transportation Research Procedia* 55, pp. 418–425. ISSN: 2352-1465. DOI: `https://doi.org/10.1016/j.trpro.2021.07.005`. URL: `https://www.sciencedirect.com/science/article/pii/S2352146521003999`.

DePrator, Anthony J, Owen Hitchcock and Vikash V Gayah (2017). ‘Improving urban street network efficiency by prohibiting conflicting left turns at signalized intersections’. In: *Transportation Research Record* 2622.1, pp. 58–69.

Dunne, L and G McArdle (2022). ‘A large scale method for extracting geographical features on bus routes from OpenStreetMap and assessment of their impact on bus speed and reliability’. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 48, pp. 37–44.

European Committee for Standardisation (2021). *Service Interface for Real Time Information (SIRI)*. en. URL: `https://siri-cen.eu` (visited on 01/11/2021).

Frick, Karen Trapenberg, Tanu Kumar and Alison Post (2020). ‘Background Paper: The General Transit Feed Specification (GTFS) Makes Trip-Planning Easier—Especially During a Pandemic—Yet its Use by California Agencies is Uneven’. In.

Gan, Albert et al. (2003). ‘Development of operational performance and decision models for arterial bus lanes’. In: *Transportation research record* 1858.1, pp. 18–30.

Google Developers (Nov. 2021a). *GTFS Realtime Protobuf*. en. `https://developers.google.com/transit/gtfs-realtime/gtfs-realtime-proto`. (Visited on 01/11/2021).

— (Nov. 2021b). *Protocol Buffers*. Google Developers. `https://developers.google.com/protocol-buffers`. (Visited on 01/11/2021).

Google Developers (2022a). *GTFS Realtime Overview*. URL: `https://developers.google.com/transit/gtfs-realtime`.

— (2022b). *GTFS Static Overview*. URL: `https://developers.google.com/transit/gtfs`.

Guler, S Ilgin and Monica Menendez (2014). 'Analytical formulation and empirical evaluation of pre-signals for bus priority'. In: *Transportation Research Part B: Methodological* 64, pp. 41–53.

— (2015). 'Pre-signals for bus priority: basic guidelines for implementation'. In: *Public Transport* 7, pp. 339–354.

Guo, Zhan, Nigel HM Wilson and Adam Rahbee (2007). 'Impact of weather on transit ridership in Chicago, Illinois'. In: *Transportation Research Record* 2034.1, pp. 3–10.

He, Qing, K Larry Head and Jun Ding (2014). 'Multi-modal traffic signal control with priority, signal actuation and coordination'. In: *Transportation research part C: emerging technologies* 46, pp. 65–82.

Hickman, Mark D (2001). 'An analytic stochastic model for the transit vehicle holding problem'. In: *Transportation Science* 35.3, pp. 215–237.

Hine, Julian and Judith Scott (2000). 'Seamless, accessible travel: users' views of the public transport journey and interchange'. In: *Transport policy* 7.3, pp. 217–226.

Hofmann, Markus and Margaret O'Mahony (2005). 'The impact of adverse weather conditions on urban bus performance measures'. In: *Proceedings. 2005 IEEE Intelligent Transportation Systems, 2005.* IEEE, pp. 84–89.

Jenelius, Erik and Matej Cebecauer (2020). 'Impacts of COVID-19 on public transport ridership in Sweden: Analysis of ticket validations, sales and passenger counts'. In: *Transportation Research Interdisciplinary Perspectives* 8, p. 100242.

Kaddoura, Ihab et al. (2015). 'Optimal public transport pricing: Towards an agent-based marginal social cost approach'. In: *Journal of Transport Economics and Policy (JTEP)* 49.2, pp. 200–218.

Kashfi, Syeed Anta, Brian Lee and Jonathan Bunker (2013). 'Impact of rain on daily bus ridership: a Brisbane case study'. In: *Australasian transport research forum 2013 proceedings*. Australasian Transport Research Forum, pp. 1–18.

Kim, Hyung Jin and (2003). ‘Performance of bus lanes in seoul: Some impacts and suggestions’. In: *IATSS research* 27.2, pp. 36–45.

Knapp, Keith K and Leslie N Jacobson (1995). ‘Safety-Related Characteristics of Arterial High-Occupancy Vehicle (HOV) Roadway and Lane Treatments’. In: *Report SWUTC95/721944*.

Kujala, Rainer et al. (2018). ‘A collection of public transport network data sets for 25 cities’. In: *Scientific data* 5.1, pp. 1–14.

Kwon, Jaimyoung and Pravin Varaiya (2008). ‘Effectiveness of California’s high occupancy vehicle (HOV) system’. In: *Transportation Research Part C: Emerging Technologies* 16.1, pp. 98–115.

Levinson, David M (1998). ‘Speed and delay on signalized arterials’. In: *Journal of Transportation Engineering* 124.3, pp. 258–263.

Levinson, Herbert S and Kevin R St. Jacques (1998). ‘Bus lane capacity revisited’. In: *Transportation Research Record* 1618.1, pp. 189–199.

Li, Zheng et al. (2022). ‘Simulation-based optimization of large-scale dedicated bus lanes allocation: Using efficient machine learning models as surrogates’. In: *Transportation Research Part C: Emerging Technologies* 143, p. 103827.

Ma, Wanjing, K Larry Head and Yiheng Feng (2014). ‘Integrated optimization of transit priority operation at isolated intersections: A person-capacity-based approach’. In: *Transportation Research Part C: Emerging Technologies* 40, pp. 49–62.

Manual, Highway Capacity (2000). ‘Highway capacity manual’. In: *Washington, DC* 2.1, p. 1.

Mathieu, Edouard et al. (Mar. 2020). ‘COVID-19: Stringency Index’. In: *Our World in Data*. URL: https://ourworldindata.org/covid-stringency-index (visited on 28/06/2023).

Mendes-Moreira, Joao et al. (2015). ‘Validating the coverage of bus schedules: A machine learning approach’. In: *Information Sciences* 293, pp. 299–313.

Messer, Carroll J and Daniel B Fambro (1977). *Effects of signal phasing and length of left-turn bay on capacity*. 644. Transportation Research Recorded.

MobilityData (Nov. 2021a). *General Transit Feed Specification*. URL: https://gtfs.org/ (visited on 01/11/2021).

— (Nov. 2021b). *Python GTFS-realtime Language Bindings*. GitHub. URL: https://github.com/MobilityData/gtfs-realtime-bindings (visited on 01/11/2021).

Montero-Lamas, Yaiza et al. (2023). 'A new big data approach to understanding general traffic impacts on bus passenger delays'. In: *Journal of Advanced Transportation* 2023.

Nakanishi, Yuko J (1997). 'PART 1: bus: bus performance indicators: on-time performance and service regularity'. In: *Transportation Research Record* 1571.1, pp. 1–13.

National Association of City Transportation Officials (NACTO) (2016). *Transit Street Design Guide*. URL: https://nacto.org/publication/transit-street-design-guide/.

— (2018). *Global Street Design Guide*. en-US. URL: https://nacto.org/global-street-design-guide-gsdg/ (visited on 20/12/2021).

Newell, Gordon F (1959). 'The effect of left turns on the capacity of a traffic intersection.' In: *Quarterly of Applied Mathematics* 17.1, pp. 67–76.

Nowlin, Lewis and Kay Fitzpatrick (1997). 'Performance of queue jumper lanes'. In: *Traffic Congestion and Traffic Safety in the 21st Century: Challenges, Innovations, and OpportunitiesUrban Transportation Division, ASCE; Highway Division, ASCE; Federal Highway Administration, USDOT; and National Highway Traffic Safety Administration, USDOT.*

Petit, Antoine et al. (2021). 'Dedicated bus lane network design under demand diversion and dynamic traffic congestion: An aggregated network and continuous approximation model approach'. In: *Transportation Research Part C: Emerging Technologies* 128, p. 103187.

Prommaharaj, Postsavee et al. (Apr. 2020). 'Visualizing public transit system operation with GTFS data: A case study of Calgary, Canada'. In: *Heliyon* 6.4, e03729. ISSN: 2405-8440. DOI: 10.1016/j.heliyon.2020.e03729. URL: https://www.sciencedirect.com/science/article/pii/S2405844020305740 (visited on 15/05/2024).

python-holidays — holidays documentation (June 2023). URL: https://python-holidays.readthedocs.io/en/latest/ (visited on 28/06/2023).

Ren, Yicheng, Jing Zhao and Xizhao Zhou (2021). ‘Optimal design of scheduling for bus rapid transit by combining with passive signal priority control’. In: *International Journal of Sustainable Transportation* 15.5, pp. 407–418.

Shalaby, Amer S. (Sept. 1999). ‘Simulating Performance Impacts of Bus Lanes and Supporting Measures’. EN. In: *Journal of Transportation Engineering* 125.5. Publisher: American Society of Civil Engineers, pp. 390–397. ISSN: 0733-947X. DOI: `10.1061/(ASCE)0733-947X(1999)125:5(390)`. (Visited on 11/09/2021).

Sheth, Chintan, Konstantinos Triantis and Dušan Teodorović (2007). ‘Performance evaluation of bus routes: A provider and passenger perspective’. In: *Transportation Research Part E: Logistics and Transportation Review* 43.4, pp. 453–478.

Shu, Shijie, Jing Zhao, Yin Han et al. (2019). ‘Novel design method for bus approach lanes with bus guidance and priority controls for prioritizing through and left-turn buses’. In: *Journal of Advanced Transportation* 2019.

Singhal, Abhishek, Camille Kamga and Anil Yazici (2014). ‘Impact of weather on urban transit ridership’. In: *Transportation research part A: policy and practice* 69, pp. 379–391.

Smith, Harriet R, Brendon Hemily and Miomir Ivanovic (2005). ‘Transit signal priority (TSP): A planning and implementation handbook’. In.

Sterman, Brian P and Joseph L Schofer (1976). ‘Factors affecting reliability of urban bus services’. In: *Transportation Engineering Journal of ASCE* 102.1, pp. 147–159.

Strathman, James G and Janet R Hopper (1993). ‘Empirical analysis of bus transit on-time performance’. In: *Transportation Research Part A: Policy and Practice* 27.2, pp. 93–100.

Sun, Wenzhe, Jan-Dirk Schmöcker and Koji Fukuda (2021). ‘Estimating the route-level passenger demand profile from bus dwell times’. In: *Transportation Research Part C: Emerging Technologies* 130, p. 103273.

Surprenant-Legault, Julien and Ahmed M El-Geneidy (2011). ‘Introduction of reserved bus lane: Impact on bus running time and on-time performance’. In: *Transportation Research Record* 2218.1, pp. 10–18.

Suwardo, W, Madzlan Napiah and Ibrahim Kamaruddin (2009). ‘On-time performance and service regularity of stage buses in mixed traffic’. In: *International Journal of Business, Economics, Finance and Management Sciences 1: 3 2009* 1.3, pp. 176–183.

Tantiyanugulchai, Sutti and Robert L Bertini (2003). 'Analysis of a transit bus as a probe vehicle for arterial performance measurement'. In: *ITE Annual Meeting and Exhibit, Seattle, WA*. Citeseer.

Tao, Sui et al. (2014). 'Exploring Bus Rapid Transit passenger travel behaviour using big data'. In: *Applied Geography* 53, pp. 90–104. ISSN: 0143-6228. DOI: https://doi.org/10.1016/j.apgeog.2014.06.008. URL: https://www.sciencedirect.com/science/article/pii/S0143622814001246.

Tao, Sui et al. (2018). 'To travel or not to travel:'Weather'is the question. Modelling the effect of local weather conditions on bus ridership'. In: *Transportation research part C: emerging technologies* 86, pp. 147–167.

Transport for NSW (2021a). *Sydney Clearways Strategy*. en. URL: https://roads-waterways.transport.nsw.gov.au/projects/easing-sydneys-congestion/sydney-clearways-strategy.html (visited on 20/04/2022).

— (Nov. 2021b). *TfNSW Open Data Hub and Developer Portal*. URL: https://opendata.transport.nsw.gov.au/ (visited on 01/11/2021).

Transport for NSW (TfNSW) (2018). 'General Transit Feed Specification (GTFS) – Timetable and Realtime Feed for NSW Buses - Fileset Consumer Guide'. In: *Open Data Hub*. URL: https://opendata.transport.nsw.gov.au/sites/default/files/TfNSW_Realtime_Bus_Technical_Doc_v2.5.pdf.

Transportation Research Board and National Academies of Sciences, Engineering, and Medicine (2013). 'Transit Capacity and Quality of Service Manual, Third Edition'. English. In: Washington, DC: The National Academies Press. DOI: 10.17226/24766. URL: https://www.nap.edu/catalog/24766/transit-capacity-and-quality-of-service-manual-third-edition (visited on 05/10/2021).

Truong, Long Tien, Majid Sarvi and Graham Currie (2015). 'Exploring multiplier effects generated by bus lane combinations'. In: *Transportation Research Record* 2533.1, pp. 68–77.

Tsitsokas, Dimitrios, Anastasios Kouvelas and Nikolas Geroliminis (2021). 'Modeling and optimization of dedicated bus lanes space allocation in large networks with dynamic congestion'. In: *Transportation Research Part C: Emerging Technologies* 127, p. 103082.

Wang, Wei, John P. Attanucci and Nigel H. M. Wilson (Oct. 2011). ‘Bus Passenger Origin-Destination Estimation and Related Analyses Using Automated Data Collection Systems’. en. In: *Journal of Public Transportation* 14.4, pp. 131–150. ISSN: 1077-291X. DOI: 10.5038/2375-0901.14.4.7. URL: https://www.sciencedirect.com/science/article/pii/S1077291X22002156 (visited on 13/07/2023).

Waterson, BJ, B Rajbhandari and NB Hounsell (2003). ‘Simulating the impacts of strong bus priority measures’. In: *Journal of Transportation Engineering* 129.6, pp. 642–647.

Wong, James C (2013). ‘Use of the general transit feed specification (GTFS) in transit performance measurement’. PhD thesis. Georgia Institute of Technology.

Wu, Jianping and Nick Hounsell (1998). ‘Bus priority using pre-signals’. In: *Transportation Research Part A: Policy and Practice* 32.8, pp. 563–583.

Yan, Wenbo et al. (2021). ‘Impacts of School Reopening on Variations in Local Bus Performance in Sydney’. In: *Transportation Research Record* 2675.9, pp. 1277–1289.

Yan, Yadan, Zhiyuan Liu and Yiming Bie (2016). ‘Performance evaluation of bus routes using automatic vehicle location data’. In: *Journal of Transportation Engineering* 142.8, p. 04016029.

Yang, Yingxiang et al. (2013). ‘Potential of low-frequency automated vehicle location data for monitoring and control of bus performance’. In: *Transportation Research Record: Journal of the Transportation Research Board* 2351.1, pp. 54–64.

Yu, Bin et al. (2015). ‘A bi-level programming for bus lane network design’. In: *Transportation Research Part C: Emerging Technologies* 55. Engineering and Applied Sciences Optimization (OPT-i) - Professor Matthew G. Karlaftis Memorial Issue, pp. 310–327. ISSN: 0968-090X. DOI: https://doi.org/10.1016/j.trc.2015.02.014. URL: https://www.sciencedirect.com/science/article/pii/S0968090X15000698.

Zhao, Jing and Xizhao Zhou (2018). ‘Improving the operational efficiency of buses with dynamic use of exclusive bus lane at isolated intersections’. In: *IEEE Transactions on Intelligent Transportation Systems* 20.2, pp. 642–653.

Zhou, Dong, Qiang Li and Lixin Miao (2009). ‘Modeling on setting up bus lane in urban area with the constraint of travel time reliability’. In: *2009 International Conference on Management and Service Science*. IEEE, pp. 1–4.

Zhou, Guangwei and Albert Gan (2005). ‘Performance of transit signal priority with queue jumper lanes’. In: *Transportation Research Record* 1925.1, pp. 265–271.

— (2009). ‘Design of transit signal priority at signalized intersections with queue jumper lanes’. In: *Journal of Public Transportation* 12.4, pp. 117–132.