



Titre: Understanding Transit Travel Time Variations through Data
Title: Integration and Multi-Level Analysis

Auteur: Yuxuan Wang
Author:

Date: 2025

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Wang, Y. (2025). Understanding Transit Travel Time Variations through Data
Citation: Integration and Multi-Level Analysis [Thèse de doctorat, Polytechnique Montréal].
PolyPublie. <https://publications.polymtl.ca/72077/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/72077/>
PolyPublie URL:

**Directeurs de
recherche:** Catherine Morency
Advisors:

Programme: Génie civil
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

**Understanding Transit Travel Time Variations Through Data Integration
and Multi-Level Analysis**

YUXUAN WANG

Département des génies civil, géologique et des mines

Thèse présentée en vue de l'obtention du diplôme de *Philosophiæ Doctor*
Génie civil

Décembre 2025

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Cette thèse intitulée :

**Understanding Transit Travel Time Variations Through Data Integration
and Multi-Level Analysis**

présentée par **Yuxuan WANG**

en vue de l'obtention du diplôme de *Philosophiæ Doctor*
a été dûment acceptée par le jury d'examen constitué de :

Geneviève BOISJOLY, présidente

Catherine MORENCY, membre et directrice de recherche

Martin TRÉPANIÉ, membre et codirecteur de recherche

Francesco CIARI, membre

Peter FURTH, membre externe

DEDICATION

*To my professors and friends at Polytechnique who guided my steps,
To the bus driver friends who dance with the city's heartbeats,
To all who dream of re-imagining how we move together,
To anyone reading this thesis with curiosity and hope,
Let us keep moving,
with grace, with care, with determination,
towards a better future in motion. . .*

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my research directors, Professors Catherine Morency and Martin Trépanier. Your guidance, feedback, and continuous support have been fundamental to the development of this thesis. Thank you for providing not only academic mentorship, but also for facilitating access to data, funding, and collaborations with various partners. Without your trust and guidance, this research would not have been possible.

I would like to acknowledge the financial support from several key institutions that made this research journey possible. I thank the Natural Sciences and Engineering Research Council of Canada (NSERC), the Research Chair in Mobility, and the Research Chair in Transportation Transformation. Your investment in research and innovation in transportation has not only supported my academic pursuits, and I hope the results from my projects also contributed meaningfully to the advancement of this field.

I am also sincerely thankful to Société de transport de Montréal (STM) and Autorité régionale de transport métropolitain (ARTM) for providing the data that served as the foundation of this work. Your openness to collaboration were essential for me to explore the questions that shaped this research. The detailed scale, richness, and quality of the data allowed for the in-depth analysis I hoped to carry out, and the meaningful insights obtained from my projects would not have been possible otherwise.

Finally, I would like to extend a heartfelt thank you to the public transit drivers for your constant presence, hard work, and friendship along the way. Over the course of these four years, you carried me more than 40,000 kilometres and for over 2,270 hours on buses, metro, and trains. During these time, transit became more than just a mode of transportation, it became a workspace, a place for reflection, and often a much-needed place for joy. Rain or shine, snowstorm or heatwave, you showed up and kept the city moving. Your dedication and friendship carried me from early mornings to late nights. Our conversations also brought a human dimension to all the numbers in my research. This thesis would quite literally not have moved forward without you.

To all those who supported me along the way, in ways big and small, thank you.

RÉSUMÉ

Le transport en commun est un moyen efficace pour transporter de nombreuses passagers, ce qui le rend particulièrement pertinent face au défi du changement climatique. Il implique également tous les trois pôles du développement durable, l'environnement, l'économie, et les aspects sociaux. Il est nécessaire d'attirer davantage de personnes à utiliser les transports en commun afin de réduire les émissions de gaz à effet de serre qui contribuent aux changements climatiques et de fournir un bon service aux personnes qui ne veulent pas ou ne peuvent pas conduire une voiture dans une société centrée sur l'automobile.

Des temps de parcours précis et fiables sont essentiels pour que le transport en commun attirent les passagers. Pour les sociétés de transport, les temps de parcours influencent la planification, les coûts, l'efficacité, et la satisfaction des chauffeurs. Pour les passagers, les temps de parcours influencent les choix d'itinéraires, l'attractivité du service, et la satisfaction globale. Si les temps de parcours en transport en commun sont moins attrayants ou moins fiables, les gens éviteront de voyager en transport en commun et se tourneront vers d'autres modes. Nous devons donc obtenir une meilleure compréhension plus détaillée sur les temps de parcours.

Actuellement, de nombreuses distributions statistiques ont été utilisées pour décrire les temps de parcours en transport en commun, de nombreuses variables et méthodes ont été proposées pour modéliser les temps de parcours en transport en commun, et de nombreuses recherches ont tenté de quantifier les variations des temps de parcours en transport en commun. Cependant, il n'existe pas de niveaux ni de méthodes standards pour analyser les temps de parcours, puisque les définitions varient d'une société à l'autre et d'une recherche à l'autre. Il est nécessaire d'évaluer systématiquement les avantages et les désavantages de chaque niveau d'analyse et des différentes méthodes utilisées pour évaluer les temps de parcours avec des données plus granulaires, comme la localisation détaillée des véhicules et les transactions tarifaires. Donc, cette recherche se concentre sur l'évaluation des distributions statistiques, des moyens, et des variations des temps de parcours, ainsi que les impacts potentiels sur les passagers, afin d'améliorer notre compréhension des temps de parcours et d'aider les sociétés de transport à cibler des besoins spécifiques pour améliorer les temps, la fiabilité, et la performance du réseau.

La première contribution concerne la distribution statistique des temps de parcours. Des recherches récentes ont découvert que les temps de parcours peuvent suivre des distributions mixtes. Les temps de parcours suivant les distributions mixtes suggèrent que plusieurs con-

ditions de déplacement sont observées sur le même segment pendant la même période de observation, créant une incertitude supplémentaire lors de l'étude ou la modélisation de ces temps. Cette contribution tente d'identifier les segments sur l'ensemble du réseau, où les temps de parcours suivent des distributions mixtes à différents niveaux d'analyse, le niveau arrêt à arrêt, le niveau point de contrôle, et le niveau du parcours total. Le test de Hartigans est appliqué aux données archivées de localisation des véhicules de Montréal pour détecter la présence de distributions mixte. Nous identifions ensuite les facteurs environnementaux et opérationnels potentiels associés à ces distributions mixte observées en testant plusieurs modèles de classification.

Les résultats montrent que les distributions mixte peuvent être observées aux tous les trois niveaux d'analyse, et que la proportion de distributions de mélange varie dans le temps, à environ 35% en semaine et 13% la fin de semaine, ainsi que dans l'espace, avec des proportions plus élevées dans les zones centrales. Nous avons constaté que les variations de la demande, les feux de circulation, la fréquence de service, et la longueur des segments ont un impact important sur la présence de distributions mixtes. Ces résultats, ainsi que des données encore plus détaillées, permettront aux planificateurs de transport en commun à mieux diagnostiquer les causes des variations de conditions de déplacement et d'élaborer des stratégies pour réduire les conditions de déplacement plus lentes, améliorant la performance du système.

La deuxième contribution vise à évaluer les avantages et les désavantages de diverses variables dépendantes généralement utilisées pour modéliser les temps de parcours en transport en commun, et de divers niveaux d'analyse. La modélisation des temps directe ou indirecte via des vitesses ou des allures d'opération. Étant donné que le temps de parcours dépend de la distance et de la vitesse, il est possible d'utiliser les deux approches dans la plupart des logiciels de planification puisque les distances sont fixes. Dans cette contribution, nous essayons d'modéliser et d'évaluer les modèles de temps, de vitesse, et d'allure aux niveaux inter-arrêts, d'arrêt à arrêt, de point de contrôle, et de parcours total. Ensuite, nous testons ces modèles en utilisant deux scénarios typiques de planification du transport en commun, créant de nouveaux parcours, et prolongant les heures de service.

Les effets aléatoires indiquent qu'environ 50% de la variance totale est due à des différences non mesurées entre les segments individuels. Les résultats montrent que les modèles de temps, de vitesse, et d'allure fonctionnent de manière similaire pour les nouveaux scénarios d'heures de service, les modèles de temps étant globalement plus performants. Cependant, les modèles de vitesse ont tendance à être plus performants pour les nouveaux parcours pour lesquels aucune donnée historique n'a été observée. Les deux approches, directe et indirecte, fonctionnent bien aux niveaux plus agrégés, comme le niveau point de contrôle. Pour les

niveaux plus détaillés, les deux approches fonctionnent mieux au niveau inter-arrêts qu'au niveau arrêt à arrêt, soulignant la nécessité d'inclure des informations plus détaillées sur les feux de circulation et l'achalandage pour améliorer les modèles. Les erreurs des modèles de vitesse et d'allure sont légèrement plus asymétriques que le modèle de temps. Cependant, les erreurs relatives des modèles de temps sont plus importantes que celles des modèles de vitesse et d'allure sur des segments plus courts ou plus vites. Étant donné que chaque mesure d'erreur fournit des vues différentes des résultats de la modélisation, nous concluons que les planificateurs doivent choisir leurs mesures avec soin selon les applications et les niveaux d'analyse.

La troisième contribution tente d'expliquer la variance globale du temps de parcours en utilisant les données observées les plus détaillées. La littérature existante se concentre sur des aspects spécifiques des temps de parcours en bus, tels que les temps de parcours total, les temps d'arrêt, et les temps d'attente aux feux, mais rarement sur leurs interactions. Cette contribution propose de combiner les approches de recherche précédents en décomposant les temps de parcours observés en quatre éléments: l'heure de départ, une série de temps inter-arrêts, temps d'arrêt, et temps d'attente aux feux rouges, en utilisant trois mois de données archivées de localisation des véhicules et de transactions tarifaires. Ensuite, nous identifions les éléments de parcours affectant la variation globale du temps de parcours, ainsi que leur contribution respective aux variances globales. Plus précisément, des analyses de sensibilité globale basées sur la variance et des analyses "un à la fois" ont été utilisées pour répondre à ces questions.

Les analyses montrent que les temps de parcours globaux et les temps d'attente au feu rouge sont fortement influencés par les effets d'interaction avec les autres éléments du parcours. Les temps inter-arrêts et les temps d'arrêt sont principalement affectés par des variations individuelles importantes, vitesse entre arrêts et nombre de passagers, respectivement. Pour la plupart des lignes analysés, l'heure de départ et les effets d'interaction avec les feux de circulation contribuent à la majeure partie de la variance totale. L'heure de départ, même si elle est modifiée par quelques minutes, peut affecter considérablement les temps de parcours total et les temps d'attente aux feux rouges en raisons des durées des cycles variables dans la synchronisation des feux ainsi qu'à leurs interactions avec les autres éléments. Les résultats suggèrent que les sociétés de transport doivent considérer les effets en cascade, où de petites variations peuvent entraîner de grands changements dans les temps de parcours.

Le quatrième article s'appuie sur les résultats précédents et propose un modèle simple de temps de parcours qui prend en compte les composantes détaillées du voyage, étant donné que la littérature actuelle se concentre sur des niveaux d'analyse plus agrégés. Nous démontrons

que 15% du temps de parcours total est attribuable aux feux rouges, ce qui représente environ 20 000 heures-passagers de retard par jour. Nous démontrons que l'estimation des temps de parcours à l'aide de composantes détaillées peut potentiellement améliorer la précision face aux changements potentiels du réseau par une ou deux minutes. Le modèle proposé permettrait de reproduire de manière raisonnable les temps observés avec une bonne précision et pourrait potentiellement être utilisé pour améliorer les algorithmes d'optimisation des horaires.

La cinquième contribution explore les variations spatiales et temporelles potentielles de la fiabilité du temps de parcours du point de vue des passagers. Certains passagers peuvent être contraints d'effectuer des voyages entre des paires origine-destination peu desservis, avec des temps de parcours longs pour diverses raisons. Nous analysons la fiabilité potentielle en simulant des voyages passagers sur l'ensemble du réseau à l'aide d'un moteur d'itinéraire et des données archivées de localisation des véhicules. Les temps de parcours planifiés sont comparés aux temps réels qui auraient été vécus par les passagers afin de mesurer la fiabilité potentielle, en utilisant des métriques discrètes (ponctualité) et continues (temps tampons). Enfin, nous examinons les variations spatiales et temporelles de cette fiabilité potentielle, depuis une origine donnée et à l'échelle du territoire.

Nous avons constaté que les seuils de fiabilité influencent fortement les résultats, et que des variations spatio-temporelles existent. Certaines lignes présentent une fiabilité constante tout au long de la journée (bonne ou mauvaise), tandis que d'autres varient selon l'heure. De même, le nombre d'itinéraire fiables par zone d'origine est généralement similaire à l'échelle régionale, environ 60% de tous les itinéraires par zone, mais certaines zones connaissent des variations plus importantes que d'autres au fil de la journée. Cela implique que les impacts sont inégaux aux différents segments de population aux différentes périodes.

Compte tenu des résultats de recherche ci-dessus, des données plus détaillées à petite échelle permettent de découvrir des tendances et des problèmes supplémentaires liés aux temps de parcours en transport en commun. Différentes lignes peuvent rencontrer des problèmes distincts entraînant des temps de trajet longs ou peu fiables. Les sociétés de transport doivent continuellement améliorer la qualité et la granularité de leurs ensembles de données pour approfondir les analyses de temps de parcours. Les analystes pourraient s'appuyer sur ces efforts pour inclure des données détaillées afin d'identifier les causes des conditions de déplacement plus lentes, et « pousser » les temps de parcours dans ces conditions plus lentes vers des conditions plus rapides. Des modèles plus détaillés sur les variations d'achalandage, les feux de circulation flexibles et les comportements des conducteurs sont nécessaires pour améliorer les modèles et les estimations des temps de parcours. Les travaux futurs pourraient

également élargir nos recherches en utilisant des données historiques supplémentaires pour évaluer les mesures préférentielles pour les autobus, telles que les feux prioritaires et les voies réservées, afin de déterminer la meilleure façon d'améliorer les temps de parcours et la fiabilité des temps de parcours selon les contextes locaux. Enfin, les recherches sont nécessaires pour intégrer divers comportements et attentes des passagers dans les analyses de la fiabilité des temps de parcours, afin d'améliorer l'expérience globale des usagers du transport en commun.

ABSTRACT

Public transit is efficient in transporting many people, making it especially relevant under the challenge of global climate change. It is also a cross-cutting issue related to sustainable development involving all three pillars, environment, economy, and social aspects. We need to attract more people to use transit to reduce the greenhouse gas emissions contributing to climate change as well as providing good service to people who do not wish or cannot drive in a car-centric society.

Good, accurate, and reliable travel times are essential for transit systems to attract passengers. For agencies, travel times influence scheduling, costs, efficiency, and operator satisfactions. For passengers, travel times relate to route choices, service attractiveness and satisfaction. If transit travel times are less attractive or reliable, people will avoid travelling on transit and change to other modes. Thus, we need to gain more detailed insights into transit travel times.

Currently, many statistical distributions has been used to describe transit travel times, many variables and methods have been proposed to model transit travel times, and many research has attempted to quantify transit travel time variations. However, there are no standard levels or methods to analyze transit travel times, since the definitions vary from agency to agency and from research to research. There is a need to systematically evaluate the pros and cons of each analysis levels and the various methods used to evaluate transit travel times using more granular datasets, such as detailed vehicle locations and fare transactions. Thus, this research focuses on assessing the statistical distributions, the means, the variation of transit travel times, as well as the potential impacts on passengers in order to improve our understanding of transit travel times and to help agencies target specific needs to improve travel times and their reliability.

The first contribution focuses on the statistical distribution of transit travel times. Recent research has discovered that transit travel times may follow mixture distributions. Travel times following mixture distributions would suggest various underlying travel conditions being observed for the same period, which would create an additional layer of uncertainty when studying or modeling transit travel times. This contribution tries to identify segments systemwide, where transit travel times follow mixture distributions at different analysis levels, namely stop pair level, timepoint pair level, and service pattern level. Hartigans' Dip Test is applied to archived transit vehicle location data from Montréal to explore the presence of mixture distributions. We then identify potential environmental and operational factors re-

lated to the mixture distributions observed. We test several classification models to identify these potential factors.

The results show that mixture distributions can be observed at all three analysis levels, and the proportion of mixture distributions varies temporally at around 35% on weekdays and 13% on weekends, as well as spatially with higher proportions in the central areas. We found that demand variations, traffic lights, service frequency, and segment lengths have a larger impact on whether the given segment follows mixture distribution or not. These findings, along with even more detailed data, will help transit planners to later pinpoint the issues causing the various travel conditions and travel time variations on each segment, then create strategies to reduce the slower travel conditions, thus improving the reliability of our transit system.

The second contribution aims to evaluate the pros and cons of various dependent variables typically used to model transit travel times at various analysis levels, namely modelling travel times directly or indirectly using operating speeds or pace. Since travel time is a function of distance and speed, it is possible to use both as inputs in most scheduling software since the distances are fixed in fixed-route services. However, most literature focuses on travel times, and travel speeds are typically used to plan infrastructures or evaluate operated services. There is a lack of comparison between modelling travel times directly and indirectly through normalized measures, namely speed and pace, at various analysis levels. In this contribution, we try to fit and evaluate travel time, speed, and pace models at inter-stop, stop-to-stop, timepoint-to-timepoint, and service pattern levels. Then, we test these models using two typical scenarios in transit planning, creating new routes and expanding service hours.

The random effects show indicates that around 50% of the total variance is due to intrinsic unmeasured differences between individual segments. The results show that travel time, speed, and pace models perform similarly for new service hour scenarios, with the time models performing better overall. However, speed models tend to perform better for new route scenarios where no historical data has been observed. Both direct and normalized approaches perform well at more aggregated levels, such as the timepoint-to-timepoint level. For more detailed levels, both approaches perform better at the inter-stop level than the stop-to-stop level, emphasizing the need to include more detailed traffic signal and ridership information to improve the models. Errors calculated from the speed and pace models are slightly more skewed than the direct travel time model. However, the relative errors from travel time models are larger than the normalized models on shorter segments or faster segments. Since each error measure provides different views of the modelling results, we conclude that planners need to choose their measures carefully according to specific model

applications and analysis levels.

The third contribution tries to explain the overall travel time variance using the most granular observed data. Existing literature focuses on specific aspects of transit travel times, such as inter-stop times, dwell times, and red light waiting times, but less on how these aspects interact with each other. This contribution proposes to combine previous research efforts by further decomposing observed trip travel times into four types of elements, departure time, a series of inter-stop times, dwell times, and red light waiting times using 3 months of archived vehicle location data and fare transaction data. Then, we identify important trip elements affecting the overall travel time variation, as well as how much variation can be attributed to each trip element. More specifically, variance-based and one-at-a-time sensitivity analyses were used to answer these questions.

The analyses show that the overall travel times and red light waiting times are more affected by interaction effects between trip elements, whereas the overall inter-stop times and dwell times are mainly affected by large individual variations of inter-stop speeds and ridership respectively. For most routes analyzed, departure time and the interaction effects with traffic signals contribute to most of the total variance. Trip departure time, even when changed by a few minutes, can greatly affect the overall trip travel times and red light waiting times for most routes analyzed. This is mainly due to the varying cycle lengths in our fixed signal timing plans as well as their interactions with inter-stop and ridership variations. The results suggest that planners must consider potential chain reactions where small variations in one trip element can lead to significant changes in the overall travel times.

The fourth paper builds on the previous results and proposes a simple travel time model that accounts for the detailed trip components, given that the current literature focuses on higher aggregated analysis levels. We demonstrate that 15% of total travel times are attributed to red lights, contributing to roughly 20,000 passenger-hours of delays per day. The travel time estimations using detailed travel time components can potentially improve accuracy by a minute and improve the responsiveness to potential changes in the network. The proposed model could reasonably reproduce observed travel times with good precision and could potentially be used to expand and improve scheduling optimization algorithms.

Our fifth contribution aims to explore the spatial and temporal pattern of potential passenger travel time reliability. Some potential passengers may be forced to make a trip on transit between origin-destination pairs have little or no actual demand with long travel times due to various reasons. We analyze the potential travel time reliability by simulating passenger trips across the entire service area using a routing engine and archived vehicle location data. Then we evaluate the planned travel times against travel times that would have been delivered

to passengers to determine the potential reliability across the region using discrete metrics similar to on-time performance and continuous metrics like buffer times. Finally, we try to discover potential spatial and temporal variations of these potential transit reliability for a given origin as well as across the entire region.

We found that various reliability thresholds affect the reliability outcomes significantly and there are some spatial temporal variations in the potential transit reliability. Some route reliability measures remains constant all day, either good or bad, but some will change throughout the day. Similarly, the number of all reliable routes for each origin zones are similar across the region, at around 60% of all routes per zone, but the numbers would remain more consistent for some zones than the other zones during the day. As a result, the impacts on different population segments are different for various time periods.

Given the above research results, more detailed small-scale data can help discover additional patterns and issues related to transit travel times and that different routes may face different problems causing long or unreliable travel times. Agencies should continuously improve the quality and granularity of their datasets to help improve travel time analyses. Analysts could build on these efforts to include detailed data to identify causes of consistently slower travel conditions, such as long red times, and "nudge" the travel times under slower conditions towards faster conditions. More detailed models on ridership variations, flexible traffic signals, and driver behaviours are needed to improve the travel time models and estimations. Future works could also expand our research by using additional historical data to further evaluate bus preferential measures, such as transit signal priority and bus lanes, to determine the best way to improve travel times and improve travel time reliability, given the specific local contexts of a given route. Finally, more research is needed to incorporate various passenger behaviours and expectations into travel time reliability analyses to help improve the overall passenger experience in transit.

TABLE OF CONTENTS

DEDICATION	iii
ACKNOWLEDGEMENTS	iv
RÉSUMÉ	v
ABSTRACT	x
LIST OF TABLES	xx
LIST OF FIGURES	xxi
LIST OF SYMBOLS AND ACRONYMS	xxiv
LIST OF APPENDICES	xxvi
CHAPTER 1 INTRODUCTION	1
1.1 Why do We Care About Public Transit	1
1.2 Challenges Faced by Transit Agencies Around the World	2
1.3 General Aims and Benefits of this Research	4
1.4 Document Structure	6
CHAPTER 2 LITERATURE REVIEW	7
2.1 Transit Travel Time Performance and Reliability Definitions	7
2.2 Importance of Improving Transit Performance and Reliability	9
2.3 The Use of Automated Data Sources	11
2.4 Factors Affecting Travel Times and Reliabilities	12
2.5 Strategies to Improve Travel Times and Reliabilities	13
2.6 Vehicle Travel Times	15
2.6.1 Frameworks	16
2.6.2 Descriptive Studies	17
2.6.3 Factors Affecting Travel Times	19
2.6.4 Travel Time Variations	20
2.6.5 Travel Time Predictions	21
2.6.6 Statistical Distributions	21
2.6.7 Related Variables for Modelling Travel Time	22
2.7 Detailed Transit Travel Time Components	23

2.7.1	Dwell Times	24
2.7.2	Red Light Waiting Times	27
2.8	Transit Scheduling Methods	30
2.9	Passenger Preferences and Behaviours	33
2.10	Passenger-Centric Travel Time Measures	35
2.10.1	Wait Time	35
2.10.2	Transfer Time	38
2.10.3	Overall Passenger Travel Time	41
2.11	Summary of Literature Review and Research Gaps	46
2.11.1	General Gap 1: Systemwide Multi-level Comparisons	47
2.11.2	General Gap 2: Spatial and Temporal Variations	47
2.11.3	Detailed Gap 1: Statistical Distributions of Travel time	48
2.11.4	Detailed Gap 2: Various Average Travel Time Models	49
2.11.5	Detailed Gap 3: Understanding Travel Time Variations	50
2.11.6	Detailed Gap 4: Using detailed trip components in travel time estimation	51
2.11.7	Detailed Gap 5: Understanding Potential Passenger Travel Time Variations	51
CHAPTER 3 DATA AND GENERAL METHODOLOGY		53
3.1	Research Questions	53
3.1.1	General Question 1: How do different analysis levels compare in assessing systemwide transit performance, and what are their implications for transit planning and scheduling?	53
3.1.2	General Question 2: How do spatial and temporal variations affect transit travel times and reliability across a transit system?	54
3.1.3	Specific Question 1: Which segments have travel times that follow mixture statistical distributions, and which operational or environmental factors could explain mixture statistical distributions?	54
3.1.4	Specific Question 2: What are the pros and cons of different related dependent variables when modelling average travel times?	54
3.1.5	Specific Question 3: How much travel time variance can be attributed to specific trip components?	55
3.1.6	Specific Question 4: How can detailed trip components and their interactions be incorporated into travel time models to improve the accuracy of performance prediction under different planning scenarios?	56
3.1.7	Specific Question 5: What is the potential travel time reliability if passengers were to travel on public transport across the entire service area?	56

3.2	Data Used in This Study	57
3.2.1	Data in Montréal	57
3.2.2	Data Sources	57
3.2.3	Data Processing and Database Design	60
3.2.4	Developed Information Tools	68
3.3	General Framework, Methodology, and Expected Results	70
3.3.1	Overview of Research Framework	70
3.3.2	Contribution 1: Improve understanding of travel time distributions	71
3.3.3	Contribution 2: Improve understanding of average travel time modelling	74
3.3.4	Contribution 3: Improve understanding of travel time variance	75
3.3.5	Contribution 4: Improve understanding of travel time modelling using detailed trip components	76
3.3.6	Contribution 5: Improve understanding of the equity of experienced reliability of potential passengers	77
CHAPTER 4 ARTICLE 1: SYSTEMWIDE VARIATIONS AND FACTORS AF- FECTING MIXTURE TRANSIT TRAVEL TIME DISTRIBUTIONS		79
4.1	Introduction	81
4.2	Literature Review	82
4.2.1	Analysis scales and factors of transit travel time reliability	83
4.2.2	Studies on transit travel time distributions	84
4.3	Research Framework and Methodology	84
4.3.1	Data Sources	85
4.3.2	Hartigans' Dip Test	86
4.3.3	Overview of Classification Methods	88
4.4	Identify Segments with Mixture Distributions	89
4.4.1	Examples of Mixture Distributions in Transit Travel Time	89
4.4.2	Temporal Statistics	93
4.4.3	Spatial Distribution	96
4.5	Variables Related to Mixture Transit Travel Time Distributions	98
4.5.1	Variables Used in Classification Models	99
4.5.2	Classification Model Results	103
4.5.3	Identified Factors Related to Mixture Transit Travel Time Distribution	104
4.6	Conclusion	110
CHAPTER 5 ARTICLE 2: MODELLING BUS TRAVEL TIME, SPEED, AND PACE AT VARIOUS ANALYSIS LEVELS		113

5.1	Introduction	115
5.2	Literature Review	117
5.3	Research Framework and Methodology	120
5.3.1	Data	122
5.3.2	Analysis Levels	123
5.3.3	Modelling Methods	124
5.3.4	Input Variables	125
5.3.5	Model Evaluation Criteria	128
5.4	Results	129
5.4.1	Model Coefficients	129
5.4.2	Model Comparisons	135
5.5	Conclusion	143
CHAPTER 6 ARTICLE 3: DECOMPOSITION AND SENSITIVITY ANALYSIS OF BUS TRAVEL TIMES		147
6.1	Introduction	149
6.2	Literature Review	151
6.3	Methodology and Data	154
6.3.1	Defining the Trip Elements	156
6.3.2	Estimating Traffic Signal Settings	156
6.3.3	Getting Departure Times and Inter-Stop Times	157
6.3.4	Estimating Dwell Times	158
6.3.5	Example of decomposing a stop-to-stop travel time observation	159
6.3.6	Sensitivity Analyses	160
6.4	Case Study: Montréal	162
6.4.1	Overall Variance-Based Sensitivity Results for 20 Selected Routes	163
6.4.2	Detailed Variance-Based Sensitivity Results for Three Example Routes	167
6.4.3	One-At-A-Time Sensitivity Results for Westbound Route 27	177
6.5	Conclusion and Future Research	184
CHAPTER 7 ARTICLE 4: IMPROVING BUS TRAVEL TIME ESTIMATIONS BY INCORPORATING DETAILED TRAVEL TIME COMPONENTS AND INTER- ACTION EFFECTS		189
7.1	Introduction	191
7.2	Literature Review	192
7.3	Methodology and Data	195
7.3.1	Decomposed Travel Times	195

7.3.2	Short Summary on Variance-based Global Sensitivity Analysis Method	197
7.3.3	Proposed Travel Time Model	198
7.4	Case Study: Montréal, Canada	198
7.4.1	System-Wide Descriptive Statistics for Red Times	199
7.4.2	Potential Travel Time Estimation Scenarios for Westbound Route 27	202
7.5	Conclusion and Next Steps	213
CHAPTER 8 ARTICLE 5: SPATIAL-TEMPORAL EQUITY ASSESSMENT OF PUBLIC TRANSPORT TRAVEL TIME RELIABILITY		215
8.1	Introduction	217
8.2	Literature Review	218
8.3	Methodology and Data	220
8.4	Case Study: Montréal, Canada	224
8.4.1	Impact of Various Thresholds on Reliability Metrics	224
8.4.2	Spatial-Temporal Changes of Reliability	227
8.4.3	Number of Potential Destinations with Reliable Routes for All Origins	232
8.4.4	Potential Equity Implications for the Number of Destinations with Reliable Routes	236
8.5	Conclusion and Next Steps	239
CHAPTER 9 DISCUSSION		242
9.1	Overview of the Projects	242
9.2	Detailed Discussion of Research Works	245
9.2.1	General Question 1: How do different analysis levels compare in assessing systemwide transit performance, and what are their implications for transit planning and scheduling?	246
9.2.2	General Question 2: How do spatial and temporal variations affect transit travel times and reliability across a transit system?	247
9.2.3	Specific Question 1: Which segments have travel times that follow mixture statistical distributions, and which operational or environmental factors could explain mixture statistical distributions?	248
9.2.4	Specific Question 2: What are the pros and cons of different related dependent variables when modelling average travel times?	249
9.2.5	Specific Question 3: How much travel time variance can be attributed to specific trip components?	251
9.2.6	Specific Question 4: How can detailed trip components and their interactions be incorporated into travel time models to improve the accuracy of performance prediction under different planning scenarios?	252

9.2.7	Specific Question 5: What is the potential travel time reliability if passengers were to travel on public transport across the entire service area? . . .	253
9.3	Summary of Limitations and Directions for Future Research	254
9.4	Detailed Discussions on the Methodological Implications	259
9.5	General Discussions on the Policy Implications	262
CHAPTER 10 CONCLUSION		268
REFERENCES		271
APPENDICES		287

LIST OF TABLES

Table 4.1	Descriptive Statistics for the Variables	102
Table 4.2	Confusion Matrix for Classification Methods at Different Analysis Levels	104
Table 4.3	Logistics Regression Coefficients	105
Table 4.4	Average Variable Effect in Logistic Regression	109
Table 5.1	Summary of Dependent Variables and Corresponding Distances at each Analysis Level	124
Table 5.2	Random Effects for Linear Speed and Time Models	130
Table 5.3	Fixed Effects for Linear Speed and Time Models	132
Table 5.4	Aggregated Model Error Measures	136
Table 6.1	Departure Time Sensitivity Results for Westbound Route 27	170
Table 6.2	Inter-stop Time Averages and Variations at Different Time of the Day and Day of the Week for Northbound Route 30	173
Table 6.3	Inter-stop Time Sensitivity to Saint-Denis and Jean-Talon for North- bound Route 30	174
Table 6.4	Ridership Variation on Route 97 Direction East at Station Mont-Royal	175
Table 6.5	Dwell Time Sensitivity at Station Mont-Royal for Eastbound Route 97	176
Table 7.1	Estimated Travel Times Given Congestion and Ridership Scenarios .	205
Table 7.2	Estimated and Observed Median Travel Times for Westbound Route 27 Departing Around 6:30	207
Table 8.1	Summary of independent and dependent variables	237
Table 8.2	Spatial regression model results for the number of destinations with reliable routes (Jan-Mar 2024 data)	237
Table A.1	Correlation Matrix for Paper 1 Stop Level	287
Table A.2	Correlation Matrix for Paper 1 Timepoint Level	288
Table A.3	Correlation Matrix for Paper 1 Route Level	288
Table A.4	Correlations Matrix for Paper 2 Stop Level	289
Table A.5	Correlation Matrix for Paper 2 Timepoint Level	290
Table A.6	Correlation Matrix for Paper 2 Route Level	291
Table A.7	Correlation Matrix for Paper 5	291

LIST OF FIGURES

Figure 2.1	Example: Splitting a Stop to Stop Travel Time Observation from Ride-Check Data on Route 51	24
Figure 3.1	Example of mixture travel time distribution observed for trips departed at 6:30 on westbound route 27 in Montréal.	55
Figure 3.2	Screenshot of the Ride-Check Application	61
Figure 3.3	Data Processing Overview	62
Figure 3.4	General Research Framework	72
Figure 3.5	Illustration of Analysis Levels	73
Figure 4.1	Research Framework	86
Figure 4.2	Map for Route 51	90
Figure 4.3	Example of Mixture Transit Travel Time Distribution at Service Pattern Level (Route 51, Eastbound, Weekdays 11 AM to 12 PM, STM data between May 1, 2021 and Oct 24, 2022)	91
Figure 4.4	Example of Mixture Transit Travel Time Distribution at Timepoint Level (Route 51, Eastbound, from Decelles / Jean-Brillant to Station Édouard-Montpetit, Weekdays Between 5 PM and 6 PM, STM data between May 1, 2021 and Oct 24, 2022)	92
Figure 4.5	Example of Mixture Transit Travel Time Distribution at Stop Level (Route 51, Eastbound, , from Decelles / Jean-Brillant to Decelles / Édouard-Montpetit, Weekdays Between 5 PM and 6 PM, STM data between May 1, 2021 and Oct 24, 2022)	93
Figure 4.6	Temporal variations of Mixture Transit Travel Time Distribution at Service Pattern Level (STM data between May 1, 2021 and Oct 24, 2022)	95
Figure 4.7	Temporal variations of Mixture Transit Travel Time Distribution at Route Timepoint Level (STM data between May 1, 2021 and Oct 24, 2022)	95
Figure 4.8	Temporal variations of Mixture Transit Travel Time Distribution at Stop Pairs Level (STM data between May 1, 2021 and Oct 24, 2022)	96
Figure 4.9	Locations of Stop Level Mixture Transit Travel Time for Weekdays at 7 AM	97
Figure 4.10	Percent Timepoint Pair Segment Length Following Mixture Distributions in each Neighbourhood for Weekdays at 7 AM	98

Figure 5.1	Research Framework	121
Figure 5.2	Error Distributions for Time, Speed, and Pace Models	139
Figure 5.3	Errors Aggregated by Observed Average Segment Speed	140
Figure 5.4	Errors Aggregated by Segment Length	142
Figure 6.1	Example of mixture travel time distribution observed for trips departed at 6:30 on westbound route 27 in Montréal.	152
Figure 6.2	Research Framework	155
Figure 6.3	Example: Splitting a Stop to Stop Travel Time Observation	159
Figure 6.4	Most important factors contributing to overall red light waiting time variations for 20 routes at various time-of-day periods	164
Figure 6.5	Most important factors contributing to overall trip time variations for 20 routes at various time-of-day periods	166
Figure 6.6	Example: Total-Order Sensitivity Indices for Westbound Route 27	168
Figure 6.7	Example: Total-Order Sensitivity Indices for Northbound Route 30	172
Figure 6.8	Example: Total-Order Sensitivity Indices for Northbound Route 30	174
Figure 6.9	One-at-a-time analysis for dwell time or inter-stop time changes	178
Figure 6.10	One-at-a-time analysis for traffic signal changes	180
Figure 6.11	One-at-a-time analysis for departure time changes	182
Figure 7.1	Example: Splitting a Stop to Stop Travel Time Observation Reproduced from Wang et al. (2025)	197
Figure 7.2	Percent of Total Travel Time Attributed to Red Times for Weekday services from Jan. 2024 to Mar. 2024	199
Figure 7.3	Red Time Variations as a Percentage of Total Travel Time Variation for Weekday Services from Jan. 2024 to Mar. 2024	200
Figure 7.4	Estimated Bus Trajectory for Three Ridership Scenarios on Westbound 27 Departing at 6:30 in Time-Distance Graph	203
Figure 7.5	Travel Times Observed and Simulated for Trips Departing at 6:30 on Westbound Route 27 in Montréal	204
Figure 7.6	Travel Time Estimation Improvements Compared to Averages for Westbound 27 Trips Throughout the Day	209
Figure 7.7	Departure Time Adjustment to Improve Estimated Travel Times	212
Figure 7.8	Potential Travel Time Gained from Departure Time Adjustments	213
Figure 8.1	Research Framework	221
Figure 8.2	Reliability for routes from Polytechnique Montréal (purple star) to other destinations departing at 17:00 with various thresholds, by indicator (Jan-Mar 2024 data)	225

Figure 8.3	Reliability for the overall route travel times from Polytechnique (purple star) to other destinations for three departure times, by indicator (Jan-Mar 2024 data)	228
Figure 8.4	Route reliability breakdown from Polytechnique Montréal to other destinations at different times of the day, by indicator (Jan-Mar 2024 data)	229
Figure 8.5	Number of Destinations with Routes Considered Reliable Departing from Polytechnique Montréal, by indicator (Jan-Mar 2024 data) . . .	231
Figure 8.6	Number of Reliable Destinations for each Zone at Different Times, by indicator (Jan-Mar 2024 data)	233
Figure 8.7	Bus Service Performance by Route for 12:00 Departures (Jan-Mar 2024 data)	234
Figure 9.1	High Level Summaries and Relationships Between the Papers	243

LIST OF SYMBOLS AND ACRONYMS

AFC	Automated Fare Collection
APC	Automated Passenger Counter
AQTr	Association québécoise des transports
ARTM	Autorité régionale de transport métropolitain
AVL	Automatic Vehicle Location
BART	San Francisco Bay Area Rapid Transit
COTA	Central Ohio Transit Authority
COVID	Coronavirus Disease
CTA	Chicago Transit Authority
CTRF	Canadian Transportation Research Forum
GPS	Global Positioning System
GTFS	General Transit Feed Specification
GTFS-RT	GTFS Real-time
MBTA	Massachusetts Bay Transportation Authority
MTS	San Diego Metropolitan Transit System
NCHRP	National Cooperative Highway Research Program
NSERC	Natural Sciences and Engineering Research Council of Canada
NYCT	New York City Transit Authority
OD	Origin-Destination
RTL	Réseau de transport de Longueuil
SEPTA	Southeastern Pennsylvania Transportation Authority
STM	Société de transport de Montréal

STL	Société de transport de Laval
TCQSM	Transit Capacity and Quality of Service Manual
TCRP	Transit Cooperative Research Program
TfL	Transport for London
TOD	Transit-Oriented Developments
TRB	Transportation Research Board
TSP	Transit Signal Priority
TTC	Toronto Transit Commission
WCTR	World Conference on Transport Research
WMATA	Washington Metropolitan Area Transit Authority

LIST OF APPENDICES

Appendix A	Correlation Matrix for Model Inputs	287
------------	---	-----

CHAPTER 1 INTRODUCTION

This chapter introduces the importance of public transit in addressing environmental, economic, and social challenges, particularly in the context of sustainable development. At the same time, it highlights the growing challenges faced by transit agencies worldwide, including funding shortfalls, aging infrastructure, and shifting travel patterns. The chapter concludes by framing the research objectives: to improve the understanding of transit travel times and develop data-driven tools to enhance service reliability and efficiency.

1.1 Why do We Care About Public Transit

Public transit is efficient in transporting many people, making it especially relevant under the challenge of global climate change. It is also a cross-cutting issue related to sustainable development. It relates to all three pillars of sustainable development: environment, economy, and social aspects.

First, public transit has a lot of environmental benefits compared to private car use. It has a lower greenhouse gas emission per capita. In Quebec, transportation accounts for 43% of all greenhouse gas emissions (Gouvernement du Québec, 2021). By moving many people at once, transit can reduce the number of cars on the road, especially given the dominance of driving alone. Hence, efficient and effective public transit services can reduce congestion by reducing car traffic. One bus can replace 40 cars, thus reducing emissions dramatically. Electric buses, trams, or trains produce zero tailpipe emissions and lower life cycle emissions, especially when produced using renewable energy. It will also reduce other air and water pollutants (Litman, 2015). Transit-Oriented Developments (TOD) can also support walkability, which further reduces the need for long car commutes and urban sprawl.

Public transit also provides economic benefits. It saves costs for the individual passengers compared to owning and operating a car, since passengers don't need to pay for fuel, insurance, maintenance, and parking. The reduced congestion could also save billions of dollars by reducing delays and increase productivity (Wang et al., 2021). Transit can also increase property values and development around stations, which could also attract more residents, businesses, and retail stores. These could add up and help create jobs and improve economic mobility, which will boost the local economy.

Public transit also relates to social issues. Transit enables access to jobs, school, healthcare, and stores especially for people who do not wish or cannot drive. It helps promote equity

and inclusion by reducing transportation barriers and improving independence. Students and young adults rely heavily on transit to access schools, internships, and extracurricular activities. Wheelchair users and people with disabilities who can't drive also heavily rely on public transit. Public transit is also safer than driving, which creates more livable cities.

Hence, public transit helps improve the connection, environment, health, equity, inclusiveness, and the economy of the communities. We need to attract more passengers to use transit.

1.2 Challenges Faced by Transit Agencies Around the World

Transit agencies around the world are facing various financial, political, and operational challenges. These challenges impact their ability to deliver fast, reliable, equitable, and sustainable service to the passengers. Some of these issues are reaching breaking points for many agencies.

One of the most pressing issues for the agencies around the world is the funding shortfalls. Public transit relies heavily on fare revenue and government funding. The pandemic changed how people travel, causing reduced ridership and fare revenue, leading to long-term financial strain. Federal, state, or provincial support for public transit are also stagnating. Since transit agencies typically serve various jurisdictions, the changing leadership and political gridlock in the government also cause uncertainties in transit funding. These uncertainties, in turn, affect the operational budgets and service deliveries, where transit agencies could not better adapt their services to the changing demands from passengers. The uncertainties also affect capital investments, which are often delayed due to budget gaps, as well as the operating budget, which may reduce service, causing maintenance issues, then lead to a death spiral for transit services and passengers (Siddiq et al., 2023).

Operational budget shortfalls are a significant factor contributing to travel time and service delivery issues in transit systems, which directly affects the passengers, who are the reason why transit systems exist. When transit agencies face financial constraints, they often have to make decisions that impact the quality of services, such as reducing the service frequency. With fewer trips, passengers may experience longer wait times between vehicles, adding to the passenger travel times and crowding in vehicles. Longer wait times and extra crowding also increase the perceived travel times for passengers, further decreasing the travel experience. Less favourable perceptions, driven by the discomfort and unpredictability of public transit, can deter passengers from using the system altogether, further diminishing ridership (Carrel et al., 2013a).

The existing travel time gap between private vehicles and public transit is still significant. Public transit trips often take 1.4 to 3.6 times longer than car trips, although transit trips are often shorter in distance (?). Again, the disparity makes public transit less attractive for time-sensitive commuters, which in turn affects passenger mode choice and undermines the case for further investment. Another dimension of the challenge is the variability and unreliability of transit travel times, which also contribute to passenger frustration beyond just long travel times. Variability in travel times, which could be due to various reasons like congestion, extra passengers, traffic lights, or missing transfers, is more damaging for the travel experiences for passengers compared to long but consistent travel times (Carrel et al., 2013a). Passengers want reliable travel times, since they care about how consistent the arrival at their destinations is, especially for time-sensitive trips to school or work.

Transit agencies also have issues with the aging infrastructure and fleet especially in North America. Many transit systems, especially in older cities (e.g. Montreal, Toronto, New York, Boston), struggle with decades-old infrastructure and deferred maintenance, leading to delays, breakdowns, and declining service quality. The repairs, modernizations, and expansions are very costly, and will become more costly as time goes on (Wang et al., 2021). In 2024, the U.S. Department of Transportation estimated that the maintenance backlog would continue to grow from approximately \$140 billion in 2022 in the upcoming years (U.S. Department of Transportation, 2024).

Transit infrastructures are also increasingly under stress due to climate change bringing more extreme weather, such as floods and heatwaves. Transit agencies must also adapt their vehicles and systems for resilience, while also aiming to decarbonize their fleets. For example, in 2025, the New York City Transit Authority (NYCT) is still dealing with the infrastructure damages caused by Hurricane Sandy from 2012, 13 years ago. These resiliency needs also require a lot of capital investments. However, due to the uncertainty and financial strains caused by the pandemic, many agencies also delayed or cancelled their capital investments (Siddiq et al., 2023)

Since the pandemic, transit agencies have been paying more attention to cleanliness and disease prevention. Many agencies have invested in more frequent cleaning, air filtration systems, hand sanitizer dispensers, and clear public health messaging to help improve public perception (Kamga and Eickemeyer, 2021). However, due to the changes in passenger behaviours and travel patterns, ridership in most transit systems is still below the pre-pandemic level (Ziedan et al., 2023).

There's a growing shortage of bus and rail operators, mechanics, and other essential staff. Recruitment and retention are hard due to demanding schedules (split shifts), relatively low

wages, and the rising cost of living in cities. Some agencies increased the wages, benefits, and implemented other initiatives to attract workers, pushing the labour cost up (Siddiq et al., 2023; Godfrey et al., 2024).

Concerns about safety are also increasingly challenging for the agencies. These issues relate to how safe people feel using transit, especially at night or in less-busy areas. There is an increase in harassment in transit vehicles or stations. Homeless populations also use transit vehicles and stations as temporary housing. In addition, many people with mental health issues also take refuge in transit stations. Agencies rely more on social workers or crisis response teams rather than police alone (Loukaitou-Sideris et al., 2023). These trends can be observed in many North American cities, leading to complex issues around safety, dignity, and public perception.

Another major challenge is the perception. North American cities have been shaped around the private automobile for decades. Sprawling low-density suburbs, wide roads, a lack of sidewalks, and the sea of parking make transit physically difficult to access and less convenient. Thus, public transit is often seen as a "last resort" instead of a viable option, due to the unnecessarily long wait times, long travel times compared to cars, the decreasing reliability, along with the aforementioned safety and cleanliness issues (Weitz, 2008).

1.3 General Aims and Benefits of this Research

As seen from previous sections, there are multiple viewpoints and challenges when thinking about public transit. There are two main focuses, one for agencies and one for passengers (National Academies of Sciences, Engineering, and Medicine and others, 2013). Travel times relate to the resources needed to operate the services planned, and could relate to most of the benefits and challenges. Faster and more reliable travel times can reduce the number of vehicles needed to operate the service while maintaining the same service level, relieving the operating budget shortfall and maintenance pressure. More competitive travel times can also help reduce delays and improve passenger and operator satisfaction, which makes transit a more viable option and attracts more passengers (Danaher et al., 2020).

The more detailed research questions and ideas will be developed in Chapter 3 after the literature review in Chapter 2. However, the broad research objective of this research is to:

- Improve our understanding of transit travel times
- Explore and integrate the existing automated data sources in more detail

- Develop tools and models to help pinpoint specific issues and strategies to improve transit travel times

Having good and reliable travel times is essential for planners to improve the efficiency of our transit systems and attract additional passengers. By addressing these general aims, this research would help transit agencies to better respond to demand, plan for the changing traffic conditions, and improve service attractiveness to passengers while saving operating costs.

Better travel times help increase competitiveness with cars (Chakrabarti, 2017). Travel time is one of the biggest factors influencing whether people choose public transit over driving. If transit can compete with or beat the travel time of cars (especially during peak hours), it becomes a far more attractive option. Commuters often choose a faster trip even if it costs more. If a transit trip takes twice as long compared to driving alone, many people won't even consider transit unless they have no choice. There is also a distortion of perceived travel times in public transit, where long wait times, crowded buses, and congested travel conditions increase the perceived travel times (van Exel and Rietveld, 2010).

Better travel time reduces vehicle and operator needs, which affects operating costs. When vehicles complete their routes faster, each vehicle can turn around and make more trips during the day. Good travel times also mean agencies can serve the same number of passengers using fewer vehicles or expand service elsewhere using the same vehicles while accounting for delays and schedule padding. Fewer vehicles needed can, in turn, reduce the fuel and maintenance needs, saving operating costs. Less random delays means more reliable recovery times and breaks for the drivers, which could improve driver satisfaction and retention (Danaher et al., 2020).

Even if transit time is slightly slower, reliable and predictable travel times can still make transit a viable option. Passengers will trust the system more when the vehicles arrive and depart when they are supposed to, which will increase their satisfaction and perception of the transit system. Reliable travel time could also reduce the time passengers need to budget to arrive at their destination on time, especially for time-sensitive trips to work, school, and medical appointments, which reduces their time budget and mental friction when choosing transit. Good passenger experiences are more likely to translate to better perceived experiences for later transit trips (Le and Carrel, 2021). Happy passengers, efficient operations, and better perception help create a positive feedback loop to improve and build support for transit services.

In addition, using existing automated data sources is becoming more popular, since it is

a cost-effective way for transit agencies to improve efficiency, service quality, and passenger satisfaction. Transit agencies already generate a massive amount of data, such as Automated Passenger Counter (APC), which tracks the number of boardings and alightings, Automatic Vehicle Location (AVL), which tracks the vehicle movements, Automated Fare Collection (AFC), which tracks the passenger fare transactions. These data could then be used to generate performance reports, which could turn them into insights to improve transit performance. The data could help improve operational efficiency, planning, and provide passenger information.

As mentioned before, many transit agencies face ongoing financial challenges, from declining fare revenues to rising operating costs after the pandemic. Analyzing these existing datasets can be a cost-effective way to help agencies make more strategic decisions, without significant spending on new systems and infrastructures.

Again, given the financial constraint, transit agencies have to maximize the amount of service they can provide without spending too much. Analyzing the existing data could help agencies pinpoint the strategies to best improve the performance given the financial constraints, such as implementing a new bus lane or signal priority to speed up buses. Agencies can use data to inform passengers, local partners, and government agencies of their challenges to help continuously improve transit services. Agencies could potentially better justify their service adjustments, be transparent, build trust with the public, and advocate for more funding. Data analysis lets agencies do more without significant investments and make strategic choices under financial constraints.

1.4 Document Structure

The structure of this thesis is the following. Chapter 2 will review the existing literature on transit service reliability. Chapter 3 will provide an overview of my research framework and the transit-related data sources. Chapters 4, 5, 6, 7, and 8 will present the five papers submitted for publication and conferences. Chapter 9 will present an overall discussion of the relations between each article, the main finding, their limitations, and a general discussion on the future methodology implications and research directions. Finally, Chapter 10 will provide a summary of my project and general policy discussions to improve transit services.

CHAPTER 2 LITERATURE REVIEW

This chapter reviews the existing literature on transit travel time performance and reliability, examining how these concepts are defined, measured, and perceived by different stakeholders. It explores the operational and passenger-centric perspectives on reliability, the role of automated data sources in performance monitoring, and the internal and external factors influencing travel times. The review also identifies key strategies for improving reliability and highlights significant research gaps, particularly the need for detailed, multi-level, systemwide analyses and better integration of passenger experiences into planning frameworks.

2.1 Transit Travel Time Performance and Reliability Definitions

Transit performance can reflect multiple points of view, namely transit passengers, transit operators, transit agencies, motorists, and community members (National Academies of Sciences, Engineering, and Medicine and others, 2013). Passengers are interested in travel times, service availability, delivery, and safety. Agencies are interested in maintenance and economics. The community is interested in the capacity, economic, employment, and environmental impacts. Motorists care about roadway capacity and their travel times. The literature pointed out a lack of standardized definitions for transit reliability.

In practice, the Transit Cooperative Research Program (TCRP) Report 215 (Danaher et al., 2020) surveyed many transit agencies in North America and reported that most agencies monitor travel times and define travel time reliability in terms of one or more specific performance measures. Most agencies surveyed use on-time performance as their service reliability definition, and only a few agencies use other measures like missed trips, travel time, and vehicle reliability. Meanwhile, 17% of the surveyed agencies do not have a formal definition of service reliability. The report also criticized some agencies for using the average travel time. Average travel times are important for planning but not necessarily reflect passenger experiences since passengers make their plans based on worst-case scenarios.

Agencies would focus on analyzing different measures at different levels. At the system level, they mainly focus on on-time performance and various factors related to non-operations. Most agencies focus on on-time performance, missed trips, headway adherence, recovery time, and travel time variations for route and trip levels. For stop level, they mainly focus on on-time performance. To analyze the performance, agencies tend to make basic assessments based on on-time performance, travel time, travel time variability, and wait time variability,

then focus on more detailed measures that address the underlying cause of these unreliability issues (Danaher et al., 2020).

Transit Capacity and Quality of Service Manual (TCQSM) (National Academies of Sciences, Engineering, and Medicine and others, 2013) is a comprehensive report on transit capacity and service quality. It emphasized considering passenger experiences when evaluating service reliability and pointed out that agencies typically focus on operational measures, not passenger measures. It defines transit performance measures as (National Academies of Sciences, Engineering, and Medicine and others, 2013):

- "a quantitative or qualitative factor used to evaluate a particular aspect of transit service;
- quality of service as the overall measure of the perceived performance of transit service from the passengers' point of view;
- and transit service measures as a quantitative performance measure that best describes a particular aspect of transit service and represents the passenger's point of view".

In the research literature, transit travel time reliability also involves multiple viewpoints and definitions. Ma et al. (2013) and Kimpel et al. (2000) pointed out that transit reliability is defined and perceived differently across stakeholders. There is no common consensus on what aspect or perspectives should be included in the analysis of transit reliability since various measures are used and proposed throughout the year.

Abkowitz et al. (1978) proposed an often-cited general definition of transit reliability, which is the variability affecting passengers' and agencies' decision-making. Zhao et al. (2013) interpret this definition as focusing on service delivery and quality perceived by passengers and agencies, which is consistent with the emphasis of TCQSM.

Some literature also uses specific measures to define transit performance, like the agencies. The common definitions are mostly related to schedule and travel time adherence, travel time variations, and waiting time variability. Turnquist et al. (1980) defined transit reliability as the ability to adhere to the planned schedule or headway with consistent travel times. Strathman et al. (1999) defined transit reliability as minimizing service delays and travel time variations. The terms consistent and variations used by these definitions imply that service performance should be measured over a period of time.

El-Geneidy et al. (2007) summarized service performance as:

- easily accessible by passengers at both origin and destination;

- arrives predictably;
- short waiting times;
- short running times;
- and low variance in running time.

However, there are more factors to consider when evaluating transit travel time performances. As previous literature pointed out, crowding or vehicle breakdowns affect the passenger perceptions of transit travel times and reliability more significantly (Wardman and Whelan, 2011). The travel time performance definition also seems to differ between individuals based on the consistency of their travel experiences (Carrel et al., 2013a). Someone often travelling at peak hours might have learned to expect longer travel times than off-peak hours. If the longer travel time is consistent or predictable, they might not consider the prolonged travel time as an issue. Other people frequently travelling during off-peak hours are used to the uncongested travel times. Suppose they make an occasional trip during peak hours. In that case, they might consider the congested time as bad and unreliable, even if the congested time is consistent and predictable for frequent travellers during peak hours.

2.2 Importance of Improving Transit Performance and Reliability

Good and reliable travel time is critical for both passenger experiences and agency operations. Improving travel times and their reliability benefits both passengers and agencies, and it is a win-win situation in general (Carrasco, 2012).

For operations, travel time and its reliability impact scheduling, which in turn impacts the operating costs. If the route is slower, more vehicles are required to provide a given service level or frequency (National Academies of Sciences, Engineering, and Medicine and others, 2013). Larger travel time savings can reduce vehicle requirements for the route, which saves operating costs for new services or service enhancements. Unreliable travel times increase the layover times needed at terminals and schedule padding to ensure vehicles can depart on time for future trips, which are part of the total cycle time, which will increase the agency's operating cost. Unreliability would also negatively affect operators' health due to the lack of breaks, which could then affect operation safety and operator retention issues (Danaher et al., 2020). An increase in transit service reliability should help schedulers remove excessive schedule padding and layover times, which would increase service productivity and resource utilization (Bowman and Turnquist, 1981). King County Metro (2021) also

states that improving transit reliability can help move more people using the same or fewer resources.

From the passengers' perspective, passengers want to arrive at their destination on time, reasonably fast, and have a reasonable wait time for the transit vehicle. Thus, travel times influence passengers' mode choice (Carrel and Walker, 2017), especially when competing with cars. Typically, researchers analyze the ratio between transit travel time and car travel time to determine the competitiveness of transit travel times. Small travel time savings can benefit the existing passengers. Walker (2012) pointed out the importance of considering people's movement, not vehicle movement, when calculating travel time savings. Even small travel time savings can have a big impact on people's movements. 1,000 bus passengers saving 5 minutes ($1000 \times 5 = 5000$), outweighs 2000 car passengers who might be delayed by 2 minutes ($2000 \times 2 = 4000$).

Abkowitz et al. (1978) pointed out that reliability is among the most crucial factors influencing passenger mode choice. Unreliable services increase the sources of uncertainty, which would then increase passenger anxiety and reduce the attractiveness and comfort of transit services (Bates et al., 2001). If the service becomes too unreliable, they will switch to different modes, and it will take major improvements for them to come back (Carrel and Walker, 2017). Unreliable services could cause passengers to experience longer wait times, delays, crowding, or bus bunching. If passengers believe the vehicle might arrive early, they must get to the stop earlier to avoid missing their trip. If the passenger believes the vehicle might not get them to the destination on time, they might choose earlier departures to ensure on-time arrival. In this case, passengers would lose time that could have been otherwise spent productively. It also impacts passengers monetarily, since they might need to find last-minute alternative transportation. The cost of travel time variation may be greater than the cost of regular travel time (Chen et al., 2003; Perk et al., 2008). Some research indicates that services with a small travel time deviation are more important than short-headway services (Balcombe et al., 2004; Daskalakis and Stathopoulos, 2008). An increase in travel time reliability has also been linked to increases in transit demand for particular routes. A model from the transit agency in San Diego, San Diego Metropolitan Transit System (MTS), found that service reliability-related variables have a more significant impact than demographic and economic variables in predicting ridership (Transportation Research Board and National Academies of Sciences, Engineering, and Medicine, 2006). Thus, the passengers' future mode choice heavily depends on good and reliable travel times (Carrel et al., 2013a).

2.3 The Use of Automated Data Sources

Existing technologies, like AFC APC and AVL can give agencies more information about their performance and passenger experiences (Hendren et al., 2015).

Agencies typically use AVL data for operation control purposes. These systems can help agencies identify any schedule deviations in real-time, which will help traffic controllers quickly implement strategies to get buses back on schedule (National Academies of Sciences, Engineering, and Medicine, 2006). It will also help agencies identify problematic trips or routes as well as potential driver issues. Although AVL provides comprehensive real-time data, it sometimes lacks the polling frequency to record precise arrival and departure times (Arias et al., 2021).

The GTFS Real-time (GTFS-RT) data provides bus tracking information in a standardized format. However, analytic tools are still scarce Aemmer et al. (2022). Agencies publishing real-time data are likely to have AVL analytic systems in place through their software vendors or in-house developments. However, these systems, maintained by each agency or software vendor, provide different measures and scales, thus making it hard to generalize the comparisons between transit systems (Aemmer et al., 2022). GTFS-RT feed produced by different software systems might also behave differently, since the official requirements are loose and flexible.

It is also important to adopt APC or AFC systems to evaluate the system from the passenger perspective, since a passenger trip might involve more than one transit trip. There is not always a one-to-one relationship between service delivery and passenger experiences. With the same vehicle delay, the overall impact on passengers may differ based on the number of passengers on board (Carrel et al., 2013b).

These automated measurements can improve service planning by routinely monitoring changes in service quality, evaluating operation performance, identifying causes of service issues, and predicting travel behaviours in response to changes in transit services. Automated data sources reduce the cost of the data collection process, albeit they sometimes require additional inference and processing (Uniman et al., 2010). A methodology that utilizes the existing data in a standardized format to improve transit performance is more appealing to agencies under severe budget constraints. With a clearer understanding of issues in the system, transit planners can better identify appropriate strategies to improve transit travel times and reliability for passengers, especially where passengers are experiencing delays or missed transfers (Arias et al., 2021).

2.4 Factors Affecting Travel Times and Reliabilities

There are several internal and external factors affecting transit travel time performance. The agencies can partially control most factors (Carrasco, 2012). Some factors occur daily, but some factors happen rarely. Some factors vary day-to-day, but some vary within the day.

Liu and Sinha (2007) categorized factors affecting bus reliability into the following groups:

- Traffic characteristics (e.g. congestion levels)
- Route characteristics (e.g. route lengths, traffic signals, number of lanes, on street parking, etc.)
- Passenger characteristics (e.g. passenger demands, passenger route choices, passenger arrival patterns, etc.)
- Vehicle characteristics (e.g. fleet maintenance, ticket system, etc.)

The commonly cited internal factors are finance and purchasing of vehicles, which affect vehicle quality and age, vehicle and track maintenance, which affect vehicle performance, vehicle availability, operator availability, transit infrastructures, route characteristics, service operation, and scheduling. The commonly cited external factors are the natural environment, traffic congestion, and passenger activities (National Academies of Sciences, Engineering, and Medicine and others, 2013). The detailed research models and results of these factors will be discussed in detail in later sections.

Buses are flexible by sharing existing infrastructures with other modes. However, the drawback is that buses are more subject to slower travel times and reliability issues caused by external factors such as traffic congestion, as well as shared curb use for boarding and alighting (Aemmer et al., 2022). Traffic characteristics can vary day to day and within the day. Schedule padding, used to account for traffic congestion, can affect passenger experiences (Wessel and Widener, 2017). If it is not enough, the trip would often run late. If it is adequate, the trip will run on time. If it is too much, the trip will run slower than expected or depart earlier than expected.

Dwell time variability relates to the passenger characteristics, which influence transit travel time reliability. Dwell time variations can be the result of varying passenger demand, fare payment methods, vehicle characteristics, schedule or headway deviation, crowding, lift or bike rack usage, or driver interactions with passengers. The variability will then affect the vehicle headways, which impacts transit reliability, reduces vehicle capacity, and increases

overall travel times (National Academies of Sciences, Engineering, and Medicine and others, 2013).

The passenger demand varies throughout the day, typically with two peaks during rush hours. The demand is typically measured by hours, but it might also vary within the hours. Hence, TCQSM suggests agencies consider the peak hour factors to express demand variations within a 15-minute window. People in densely populated areas are more likely to use transit, and more people are within walking distance of transit stops. Demographics, such as gender, age, employment, car ownership, and transportation demand management strategies, also influence transit demand (National Academies of Sciences, Engineering, and Medicine and others, 2013). However, there is still debate in the literature on whether passenger demand variations cause unreliability or unreliability causes passenger demand variations.

Vehicle capacity and reliability also depend on the operating environment, like mixed-traffic, exclusive, or grade-separated. Other traffic, right-of-way, signals, traffic delays, and speed restrictions will affect transit reliability. For semi-exclusive bus lanes and mixed traffic buses, other traffic becomes a more significant source of unreliability, from unauthorized use of transit infrastructure, right-turning vehicles, congestion, and traffic variability (National Academies of Sciences, Engineering, and Medicine and others, 2013). Deferred maintenance might lead to breakdowns during service operation, which affects service frequency and passenger demand variations.

2.5 Strategies to Improve Travel Times and Reliabilities

The strategies to improve travel time performance issues depend on the cause of the problem. Transit preferential treatments vary widely, and the solutions depend on several factors, including budget constraints, roadway configuration, and support from policymakers and the public (Arias et al., 2021). The related costs might also vary. Additional cycle time might require additional vehicles. Infrastructure improvements, such as bus lanes or signal priority, require capital investments (National Academies of Sciences, Engineering, and Medicine and others, 2013).

Agencies can revise their operator training programs to improve operators' experience and to equip operators with skills to recover from potential delays. The training program can be improved based on travel time monitoring, feedback from other departments, and field supervision (El-Geneidy et al., 2011).

For scheduling, transit travel times are a function of speed limits set by law or social constraints, random delays, and the amount of expected random delays. The speed limit is

the maximum speed a vehicle can operate, including physical and psychological constraints. Random delays are not predictable, possibly due to external factors like traffic or ridership variations. To mitigate the random delay issues, transit agencies typically use schedule padding to include the expected random delays on a given segment. It is essentially some extra time to allow vehicles to arrive on time. Wessel and Widener (2017) define it as the difference between the best-case travel time and the average travel times.

Operational controls also play a key role in improving travel times experienced by passengers. There are several response strategies, such as holding, short turns, adding vehicles, and modifying route configuration. The impacts of operations control interventions on passengers include (Carrel et al., 2013b):

- increased or decreased waiting time at the passenger trip origin;
- additional transfers and waiting times during the journey in the case of a short turn or diversion;
- increased in-vehicle travel time experienced by passengers on board due to holding at stations, travelling at reduced speed, or making unexpected stops.

However, there are only a few attempts to quantify the impact of different control strategies on agencies and passengers. (Carrel et al., 2013b)

Providing transit preferential treatments on congested segments or intersections can help address specific travel time issues and reduce delays. Thus, a reduction in delay yields a faster and more reliable service. Walker (2012) suggests transit agencies focus on reducing delays, which would also help decrease the travel times needed in the schedules. The two most commonly mentioned bus priority measures are bus lanes, which relate to the lack of space in mixed traffic conditions, and signal priorities, which reduce times caused by signal delays.

Bus lanes are a cost-effective way to improve transit travel time and reliability, and quantifying the benefits would help justify the projects to obtain public funding. Agencies with complex networks are likely to have unknown locations where transit priority infrastructures are needed (Arias et al., 2021). A dedicated bus lane might be suitable to reduce travel times and improve reliability on urban streets with high bus traffic and certain segments prone to excessive delays (King County Metro, 2021). Agencies could convert general-purpose lanes or parking lanes to create bus lanes. Projects that require roadway widening or median reconstruction will have higher costs (King County Metro, 2021). Although bus lanes can reduce the flow of vehicular traffic, multiple studies based on simulation have shown that

they can still increase person throughput because buses have greater occupancy than private cars (Walker, 2012).

A queue jump lane is a relatively short lane for buses to bypass traffic queues before a signalized intersection (National Academies of Sciences, Engineering, and Medicine and others, 2013). Queue jump lanes require less right-of-way than longer-distance bus lanes, while still being able to reduce some delays while travelling in mixed traffic (Arias et al., 2021). However, studies for queue-jump lanes typically assume vehicles can turn right while the signal is red. Cities like Montréal have a no-turn-on-red policy, right-turn cars might still block the bus departure even with the presence of a queue jump lane (Diab and El-Geneidy, 2013).

Transit signal priority gives priority at an intersection to transit, which would reduce signal delays. The passive Transit Signal Priority (TSP) involves manually adjusting signal timings so that the signal offsets match the time needed for transit vehicles to travel through the street segment, plus making stops for passengers. The active TSP mostly adjusts the signal timing (green extension or red truncation) when a bus is approaching, while some TSP continually adjusts traffic signals based on transit vehicle locations (King County Metro, 2021). Transit signal priority can improve the speed and reliability of transit vehicles. However, it requires coordination between government agencies (Arias et al., 2021).

Several papers also recommended changing route designs to obtain better reliability. The changes might include splitting routes (Abkowitz and Engelstein, 1984) and stop consolidation (El-Geneidy et al., 2011). However, there may be associated negative impacts on some passengers due to extra transfers, wait times, and walk times.

2.6 Vehicle Travel Times

This section focuses on the vehicle travel times, discussing the definition, related data, monitoring, importance, analysis standards, as well as the detailed studies on various aspects of vehicle travel times. Since most literature focuses on travel time analysis that includes dwell times and signal delays, we will specifically discuss these detailed components in the next section.

The vehicle travel time refers to the total time taken for a vehicle to complete an entire trip or to traverse specific segments of that trip (National Academies of Sciences, Engineering, and Medicine and others, 2013). Data on vehicle travel times are typically provided by AVL systems and measured at a specific time of day over a period from several days to a few months. From an agency survey, most agencies would include this measure in their performance analysis, since vehicle travel time is a key scheduling input and an important

performance measure for monitoring scheduled travel time accuracy (Danaher et al., 2020). For scheduling, TCQSM (National Academies of Sciences, Engineering, and Medicine and others, 2013) suggests agencies set vehicle travel times between timepoints equal to the mean observed running time. Some researchers also argued for using the 85th percentile running time to achieve high reliability, but with more holding. This comes from the assumption that the travel time distribution is normal, where the 85th percentile would equal the mean plus one standard deviation (National Academies of Sciences, Engineering, and Medicine, 2006). Analysts also commonly focus on travel time deviation (National Academies of Sciences, Engineering, and Medicine and others, 2013), i.e. delay, calculated as actual travel time minus scheduled travel time, which measures how well a bus is moving along each segment. A positive value of travel time delay means that a bus is having difficulty traversing the segment. The analysis also considers the variation using metrics like standard deviation, coefficient of variation, or high percentiles. High variability indicates that the travel time is less consistent, even if the bus departs at the same time every day.

Vehicle travel times are important to passengers to the extent that they affect trip planning and in-vehicle travel times, but it is not helpful to communicate with the passengers, since the measure doesn't necessarily correspond to the passengers' overall travel times or variations (Kimpel et al., 2000). It is not effective in identifying the causes of large travel time variations either, since it combines all three sources of potential delays, i.e. congestion, high ridership, and high signal delays.

In the subsections, we will discuss more detailed literature on vehicle travel times, focusing on the frameworks, descriptive statistics, and factors affecting travel times.

2.6.1 Frameworks

There are several frameworks for analyzing vehicle travel times proposed by researchers, each contributing unique components and perspectives to the broader understanding of transit performance.

Muller and Furth (2001) introduced a comprehensive theoretical system designed to evaluate transit travel times across multiple analysis levels. Their framework integrates a vehicle location system, an onboard computer to monitor vehicle status and door operations, trip reconstruction software housed within a database, and analytical tools for assessing service quality. This system would allow transit agencies to monitor operational quality using performance metrics, such as schedule adherence, headway consistency, travel speed, and transfer synchronization. This, in turn, allows agencies to improve their route and segment-level schedules,

diagnose operational inefficiencies, such as identifying slow zones, and better understand the relationship between passenger demand and their service delivery, since overcrowding and bunching can be identified more easily.

Later research generally followed similar processes, built upon this framework, often shifting focus toward the evaluation of specific travel time metrics. For example, researchers focused on evaluating various measures to evaluate travel times. Duddu et al. (2019) effectiveness of the percentage-based and fixed range-based travel time measures. Their methodology involved categorizing travel time data by time of day and service date, accounting for delay propagation across stops by comparing scheduled and actual travel times. Their findings suggest that percentage-based measures offer greater reliability in capturing performance variations.

Similarly, additional analysis components or data have been added to the framework to improve our understanding of specific issues related to travel times and their variations. Using detailed AVL and AFC data, Ma et al. (2013) expanded the framework to include factor analysis for identifying root causes of performance issues, cluster analysis for grouping similar service patterns, and distribution modelling to characterize diverse travel conditions. They also introduced a series of performance indicators aimed at improving communication of results, thereby supporting both agency-level service planning and passenger trip decision-making.

2.6.2 Descriptive Studies

Following the framework descriptions, several researchers have demonstrated descriptive statistics and visualization tools to further analyze transit travel time performance using data collected from onboard computer systems.

Some work focused on summarizing operational characteristics at both route and stop levels. Bertini and El-Geneidy (2003) used the archived AVL and APC data from TriMet to compute descriptive statistics related to transit service operation and service delivery. Their analysis included travel times, dwell times, transit availability, time-distance graphs, average speed, and schedule adherence at the route level. These measures can be used for more detailed analysis if extra attention is required on specific segments. They further examined stop-level performance measures, such as passenger boardings and on-time performance, since these directly affect passenger experience.

Similarly, also analyzing the stop-level data, Ma and Wang (2014) developed a platform for analyzing and visualizing Beijing's transit network using AVL data. Their work focused on

vehicle travel times and operating speeds, as well as the distribution of these metrics on various stop-to-stop segments on the entire network, allowing them to identify slow zones.

Beyond the overall performance summaries, several studies examined the temporal variability and reliability of travel times. Carrasco (2012) analyzed Zurich’s highly reliable transit system, finding that planned and median travel times align closely except during peak hours, when variability increases. Travel speeds were notably lower in the city center due to more traffic congestion, while corridors with transit-only lanes improved travel times. Delays, up to 90 seconds, were more frequent during peak periods, likely due to schedule padding, and the standard deviation of delays is greater during off-peak hours.

Currie and Mesbah (2011) visualized the difference in tram travel times using AVL data from March 2001 and March 2004 at the time point level to discover temporal changes in different years. Focusing on the morning peak (7–9 a.m.), they observed declining average tram speeds, attributed to rising car traffic. They made travel time isochrone maps illustrating travel times to the CBD, and delay maps revealing early arrivals in suburban routes. Their coefficient of variation analysis showed most stops within 0.1–0.2, with greater variability near the CBD.

While the previous studies focused on performance for routes or stops, others aggregated performance at the roadway segment level and classified different types of delays. Aemmer et al. (2022) introduced methods to extract transit performance metrics such as pace (inverse of speed) for street segments. They matched delay locations to street segments and noted that delay distributions were right-skewed. Their analysis also distinguished between systematic and stochastic delays, finding that stochastic delays were often negative, indicating buses sometimes ran faster than scheduled. Similar classification efforts in Montreal (Boudabous et al., 2024) aimed to separate predictable delays from random variations, supporting schedule optimization and delay forecasting.

Some researchers used alternative data sources, such as AFC data, to estimate travel times and performance measures, given its rich information on passenger boarding and fare transactions and better alignment with passenger experiences. For example, Trépanier et al. (2009) calculated performance metrics segmented by time of day, day of week, and route. Their findings revealed that Sundays had the highest average operating speeds, while weekday morning peaks were slower due to congestion. Express trips exhibited the fastest speeds. Regarding punctuality, most buses arrived late, with greater delays during morning peaks.

2.6.3 Factors Affecting Travel Times

Building on descriptive analyses, researchers have also sought to identify the underlying factors that influence transit running times and contribute to performance issues. These studies investigated physical and operational characteristics using various data sources and statistical models.

Early research from Abkowitz and Engelstein (1984) found that the mean travel time is affected by route length, passenger activity, and the number of signalized intersections. These findings have also been shown multiple times by other researchers in later studies such as El-Geneidy et al. (2007). Expanding on these results, a linear regression model from Berkow et al. (2009) further shows that distance, number of stops, direction, time of the day, dwell time, passenger activities, weather, and seasons have significant effects on route travel times.

Beyond identifying general determinants of travel times, some studies have focused on diagnosing specific causes of poor schedule adherence. For instance, Mandelzys and Hellinga (2010) used AVL and APC data to classify timepoint level performances according to TCQSM level-of-service standards, revealing that two-thirds failed to meet the standard. Then, they focused on the timepoints with frequent early or late arrivals and examined the causes. Their analysis showed that late arrivals were primarily due to longer-than-expected travel times in preceding segments, with longer dwell times being slightly less important. Early arrivals were linked to shorter travel times and early departures. From this study, the authors recommended that analysts start the analysis at the route level, which will help organize the data, before diving deeper into specific issues with stop-level travel times or dwell times, since most agencies do not develop their schedule at the stop level.

Since planners account for operational variations using schedule paddings, Wessel and Widener (2017) aimed to understand the reasons why buses run slowly, which could be due to random delays, waiting for the schedule to catch up, or both. They hypothesize schedule padding as a direct function of travel time variability. Using General Transit Feed Specification (GTFS) and AVL data from Toronto Transit Commission (TTC) in 2015, they estimated the schedule padding by comparing the scheduled travel times and the best case travel times, i.e. the 10th percentile travel time, to remove extreme values and data anomalies. At the same time, the legal speed limit is often ignored, taking the 10th percentile allows the estimation of psychological or social speed limits. The results indicate that 30% of the total scheduled service hours are related to schedule padding. Spatial analysis further revealed patterns of padding across the network, identifying areas where schedules are overly tight, excessively padded (e.g., downtown), or relatively balanced.

2.6.4 Travel Time Variations

In addition to analyzing and explaining the times delivered, it is important to understand running time variations as well, since the agencies need to schedule their layovers to recover from potential delays. Since layovers or recovery times are non-revenue hours, excess variations in travel times increase costs. From the passenger perspective, variability translates into unpredictable waiting and in-vehicle travel times, reducing perceived service reliability. Despite its importance, travel time variation models are less common.

First, there are conflicts between service reliability and the overall travel times (El-Geneidy et al., 2007). To achieve high reliability, agencies need to have more schedule padding to absorb potential delays, which increases the probability of arriving early or holding, which will increase the overall running times and operating costs. Reducing schedule padding or layover times can shorten travel times, but compromises reliability, creating a fundamental scheduling dilemma for agencies.

To quantify variability, some researchers have adopted the coefficient of variation as a metric, calculated as the standard deviation over the average. Currie and Mesbah (2011) recommend this measure because it is dimensionless, accounts for the section lengths, and can be applied to different scales.

Mazloumi et al. (2010) explored the day-to-day travel time variations using a linear regression to identify the factors affecting travel time reliability at the timepoint level. They find that land use, route length, number of signals, stops, and departure delays significantly influence variability, with land use having the strongest effect, likely as a proxy for demand and congestion. Longer length between timepoints is also a significant factor, so they recommended shortening the timepoint sections. Signalized intersections are also a key factor, highlighting the need for implementing queue jump lanes and transit signal priority. Interestingly, buses running early often have longer overall running times due to holdings at timepoints.

Some studies have evaluated the effectiveness of transit priority strategies in reducing variability. Diab and El-Geneidy (2013) assessed a series of reliability improvement measures on Montreal's St-Michel corridor (routes 67 and 467), including transit signal priority, bus lanes, limited-stop service, and smart card fare collection. Using AVL and APC data, they compared travel time models before and after the implementation of these treatments. Results showed mixed outcomes: smart card adoption unexpectedly increased travel time and its variability, while transit signal priority had little measurable effect due to no-turn-on-right policies and right-turning cars blocking the bus departures. Other strategies produced inconsistent impacts, highlighting the complexity in designing transit priority treatments.

2.6.5 Travel Time Predictions

Given the challenges of controlling variability, many researchers have turned to predictive modelling to support passenger information systems and operational planning. Some of the works used only AVL data (Gurmu and Fan, 2014; Chen et al., 2023), while others attempted to incorporate additional datasets to improve the prediction models, such as real-time traffic data (Ma et al., 2019; Han et al., 2020). More recently, researchers have also included various artificial intelligence methods to improve transit travel time models, such as using computer vision to incorporate roadside features (Abdelhalim and Zhao, 2024) and using neural networks to integrate vehicle positions with weather data (Alam et al., 2021).

Despite these advances, a common limitation is the difficulty of isolating the effects of individual factors at play, i.e. congestion, red lights, and passenger activities, possibly due to data constraints. This highlights the need for more comprehensive analytical frameworks capable of isolating these influences and capturing detailed sources of travel time variations.

2.6.6 Statistical Distributions

To capture both the average and the variance of travel times, some researchers examined the overall statistical distributions of travel times, rather than relying solely on a given measure. For the overall distributional characteristics of travel times, Kieu (2015) modelled route-level travel times using several statistical distributions. The results show transit travel times differ significantly from private vehicle travel times. Among the tested models, skewed distributions, particularly the log-normal, provided the best fit, reflecting the asymmetric nature of transit travel times.

Beyond skewness, several studies have observed multimodal mixture distributions in route-level or stop-level running times (Wang, 2020; Ma et al., 2014). For example, Wang (2020) demonstrated that these patterns can be effectively captured using a Gaussian Mixture Model, expressed as

$$p(\theta) = \sum_{i=1}^K \Phi_i N(\mu_i, COV)$$

where K equals the number of components, ϕ is the probability of the corresponding distribution, μ is the mean of the corresponding distribution, and COV is the covariance matrix between all the components:

However, prior literature has largely focused on modelling rather than explaining the causes of multimodality. Wang (2020) and Ma et al. (2014) suggested that mixed travel time patterns likely reflect underlying traffic conditions or operational modes, such as free-flow versus

congested states or due to traffic signals, resulting in distinct clusters of travel times. This complexity poses challenges for scheduling, as a single scheduled time may not adequately represent multiple operating scenarios.

Other studies have approached variability from a categorization and clustering perspective. Ma et al. (2013) applied cluster analysis to classify travel time patterns and reliability indicators. They find that holidays have a similar travel time pattern compared to weekends, while Thursdays and Fridays exhibit slightly longer times than other weekdays, possibly due to increased leisure activities at the end of the day. The authors emphasized that reliability indicators should reflect both actual service performance and perceived performance, linking them to the value of time to better inform investment and planning decisions.

2.6.7 Related Variables for Modelling Travel Time

Travel speed, pace, and time are related variables for fixed-route services, where the travel time equals the travel distance divided by the travel speed, and also equals the distance multiplied by the travel pace. By normalizing performance using speed or pace, distance is removed from the equation, enabling comparisons across segments of varying lengths. This makes speed and pace valuable indicators for system-wide diagnostics, schedule design, and infrastructure planning. Since many findings are similar to the ones from the aforementioned travel time studies, I will summarize key contributions focusing on speed and pace.

Speed and pace have been widely used to identify problematic segments and evaluate operational strategies. Several studies developed classification models to detect slow roadway segments using speed (Cortés et al., 2011; Du et al., 2017) or pace (Boudabous et al., 2024; Aemmer et al., 2022), providing agencies with insights on where to improve performance. Speed has also served as an important metric for assessing the effectiveness of transit priority measures, such as dedicated bus lanes (Russo et al., 2022) and transit signal priority (Wu et al., 2020).

Similar to travel time studies, there are also studies for transit speed modelling. Examples include modelling bus speeds with traffic characteristics using regression and neural network models (Du et al., 2017), as well as integrating detailed car speed data to estimate bus speed distributions (Zhang et al., 2020). Consistent with travel time studies, these models identify factors such as bus lanes, roadway classification, geographic context, peak direction, and service type as significant determinants of speed. Buses operating on main roads, in suburban areas, during off-peak hours, with wider stop spacing, or in bus lanes generally achieve higher speeds (Zhang et al., 2014; Kopsacheilis et al., 2023).

Despite the one-to-one relationship between time, speed, and pace, further research is needed to compare these measures across different analysis levels and operational contexts. Additional complexities, such as stop spacing, ridership variability, and priority treatments, might influence the suitability of each measure. In addition, the existing models mentioned above have used many error measures to evaluate their model performances, such as absolute and relative measures, typically aggregated into one number. As noted by Kolassa (2020), different measures evaluate the results "from different angles" (Kolassa, 2020), and few studies have compared between these measures. As transit services have various segment sizes, it is also necessary to compare the errors by segment for potential biases, since a short local segment is not directly comparable with a long highway express segment, for example.

To summarize, there are various analysis levels used in the previous studies to measure and model transit travel times. Most agencies and researchers focused on higher, more aggregated levels. Although most North American agencies schedule service at the timepoint level, the GTFS standard requires arrival and departure times for every stop to support passenger information systems. Intermediate stop times are often interpolated from timepoint data (Wessel et al., 2017), creating discrepancies between scheduling practices and what passengers see. Since riders board and alight at various stops, not just timepoints, this discrepancy highlights the need to further investigate and improve the scheduling practices and analysis frameworks, as also emphasized by Wessel et al. (2017); Graves et al. (2019). In addition, there is still a lack of comparison between these analysis levels. Hence, a multi-level analysis is needed to better isolate the effects of a specific issue to improve travel time models.

2.7 Detailed Transit Travel Time Components

In this section, we will discuss detailed literature on two detailed travel time components defined by the TCQSM, and the papers in this research will be mainly based on these definitions with some small adjustments, given data availability and analysis levels.

The basic idea for TCQSM is to split the trip travel times into three categories, inter-stop times, dwell times, and red light waiting times. To reconstruct the arrival times at each stop, we also need to include departure times in our analyses. Inter-stop times, or travel times between stops, are related to congestion. Dwell times are related to passenger activities. Red light waiting times are related to signal delays. Hence, by separating these elements, we can better understand the issues along the route.

However, due to data limitations, most literature had trouble separating these three categories of times, and focused on higher analysis levels reviewed in the previous section, such as the

timepoint level. Hence, most existing literature studied high-level vehicle travel times, which include all three categories. Given the lack of literature focusing on more detailed vehicle travel times, in this section, we will focus on the two detailed components, dwell times and red light waiting times.

Figure 2.1 shows an example time distance diagram for one stop-to-stop travel time observation. The stop-to-stop travel time is split first into the inter-stop time, where the bus travels to the next stop at the given departure time at the first stop. Then, as the bus stops at the second stop, the stopped times are split into dwell times according to the dwell time model, as well as red light times according to the traffic signal model mentioned above. The decomposition is performed for every stop-to-stop observation.

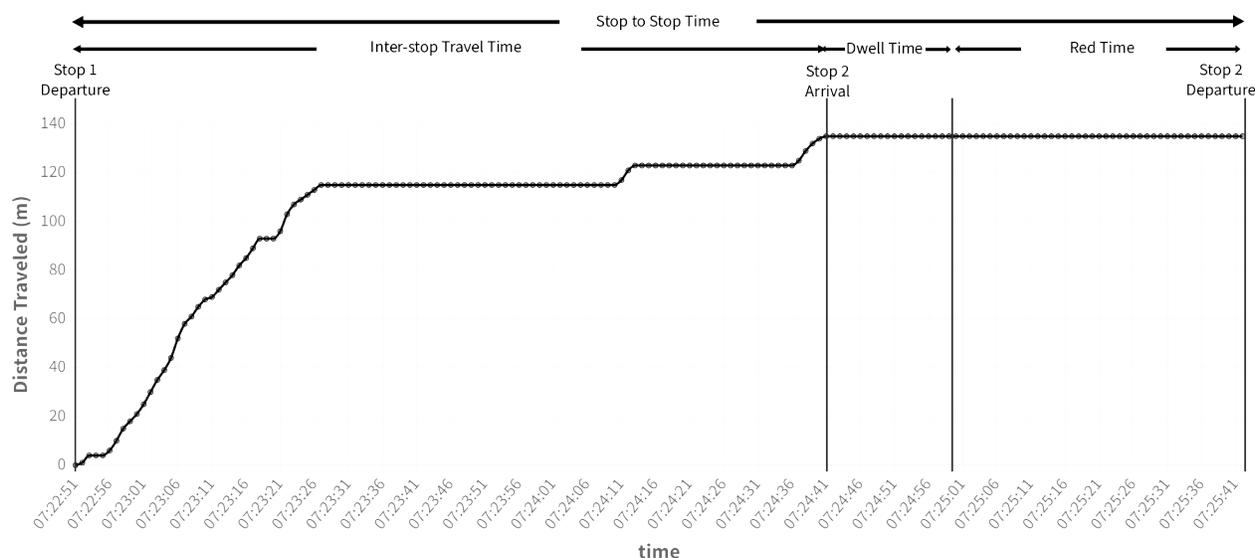


Figure 2.1 Example: Splitting a Stop to Stop Travel Time Observation from Ride-Check Data on Route 51

In the following subsections, we will discuss detailed studies on dwell times and signal delays, with their detailed definition, data requirements, usage, and the literature related to them.

2.7.1 Dwell Times

Dwell times are the time a bus spends at bus stops for passenger boardings and alightings. It is typically the time between the door opening and closing. Dwell times are part of the overall vehicle travel time from terminal to terminal. Therefore, changes in dwell times and variations would also affect the overall travel time. Again, the TCQSM, does not consider the time a vehicle remains stationary at the stops after passenger boardings and alightings

as dwell time, e.g., the red light waiting times after the boardings and alightings have been completed. Data on dwell time are typically collected through APC, AVL, and AFC systems. Researchers also utilize manual data collection and video feeds to infer dwell times.

In an agency survey, dwell time analyses are not common due to the challenges of extracting and matching data from different sources. However, agencies analyze ridership counts as a secondary measure to pinpoint the causes of travel time issues if they observe high overall vehicle travel times or high travel time variations (Danaher et al., 2020). The variability between stops could be due to different numbers of passengers boarding and alighting or due to vehicle crowding, which makes boarding and alighting difficult. Given the challenges, there is relatively less research attention on dwell times.

Focusing on describing ridership variations in the system, Berkow et al. (2009) created visualization tools using archived data from TriMet. They visualized average boardings and variations at all stops within the system using a three-dimensional map. For each route, they visualized the total boardings, allowing them to observe some seasonal effects, where the summer season might have lower boardings. They also segmented the data by day of the week, and they found that Saturday and Sunday have different ridership patterns, and the ridership variation is small. For each individual day, they found that transit ridership on weekdays is nearly constant, and the weather does not have a big effect on transit ridership. They also mapped out the total count of lift usage, which helped them identify stops with frequent lift usage, which would increase dwell times.

Typically, there are three main approaches to model bus dwell times, regression models (Dueker et al., 2004; Bertini and El-Geneidy, 2004), time-series models (Shalaby and Farhan, 2003), and machine learning models (Xin and Chen, 2016; Ding et al., 2015). Since dwell times are part of the overall travel times, the factors are similar to but not the same as vehicle travel time factors earlier. Each categories of factors affecting dwell times and related studies are discussed in the following paragraphs.

For bus characteristics, the determinants are mainly related to the vehicle types, such as articulated vs regular buses, number of doors, door sizes, and capacities. Researchers have found that low-floor buses have faster dwell times (Dueker et al., 2004), especially for passengers with reduced mobility (Levine and Torng, 1994). Articulated buses with more doors could reduce dwell times by allowing multiple passengers to board and alight at the same time, and by reducing crowding from the larger capacity (Al-Jumailey, 2011), but these buses are less manoeuvrable when merging with traffic, which could negate the dwell time savings (El-Geneidy and Vijayakumar, 2011).

Massachusetts Bay Transportation Authority (MBTA) evaluated the variance of bus dwell

times due to fare payments (Massachusetts Bay Transportation Authority, 2017). They calculated the time between consecutive payments at a stop to estimate the transaction duration. As a result, they were able to calculate the average transaction times of each fare media, as well as the distribution of the transaction times. The results show that smart card and ticket validation are faster than cash and top-up transactions. Smart card and ticket validations also have a smaller variance. For routes where cash and top-up payments are common, the larger variances could negatively impact transit reliability and headway regularities. The cost of implementing faster fare collection technologies and off-board fare collections could reduce dwell times (Jara-Díaz and Tirachini, 2013).

For passenger characteristics, Dueker et al. (2004) used Trimet AVL and APC data to evaluate the factors affecting bus dwell time. They found that boarding passengers add more dwell time than alighting passengers, but the additional passenger volume at the stop will take less time. Currie et al. (2013) evaluated the impact of crowding on vehicle dwell times. Using data from the Melbourne tram network, they showed that crowding negatively affects vehicle dwell times significantly and that crowding effects are more important than some other factors, such as entrance steps. In-vehicle crowding is also determined to be worse than platform crowding. Due to delays and the resulting headway variations, bus bunching could also increase dwell times due to the extra passengers that need to be picked up (Chen et al., 2013). In addition, passengers with reduced mobility in need of the kneeling and the ramp functions will increase dwell times significantly, and it is needed to model the dwell times with ramp usage separately (Dueker et al., 2004).

For temporal variations, peak hours tend to have the most efficient dwell times due to the more directional and regular passenger demand, and crosstown routes have shorter dwell times than local routes (Dueker et al., 2004). Routes that serve schools will also have higher riderships and dwell times on school days (Isukapati et al., 2017).

For operational and environmental characteristics, higher passenger demand and higher service frequency have been found to cause longer dwell times. More stops also correlate to less reliable service, which correlates to variations in dwell times (Chen et al., 2013). Bus lanes that do not have passing lanes also limit the bus flow, where buses that come after are stuck, which contributes to additional times needed at the stop (Kathuria et al., 2018). Diab and El-Geneidy (2015) compared the effect of farside and nearside stops on dwell times. Using 6 months of AVL data from STM and regression models, they find that nearside stops are on average 5 seconds slower than the dwell times on the far side of the intersection. This could be due to the no-turn-on-red policy on Montreal Island, which would cause right-turn cars to block the bus from departing. However, the stop placement does not seem to affect

dwell time variation. In addition, stops shelters also tend to have longer dwell times, since passengers need to walk from the shelter to the bus (Diab and El-Geneidy, 2015).

2.7.2 Red Light Waiting Times

This subsection focuses on the signal delays, or the red light waiting times, which play a significant role in the interaction between transit vehicles and traffic signals. Signal delays are the time a bus spends stopped at intersections without any passenger boardings or alightings, as defined by the TCQSM (National Academies of Sciences, Engineering, and Medicine and others, 2013), which can significantly affect the overall travel time of the vehicle. It is typically calculated using AVL and APC data or estimated in simulations.

There is less attention paid to the red light waiting times by transit agencies due to data limitations. However, red light waiting times could be a significant part of the overall vehicle travel time from terminal to terminal if there is no effective transit signal priority. Therefore, red light waiting times could also affect the overall running time and variations.

Most traffic signal-related studies in the transit context are related to transit signal priorities for transit vehicles, which is considered a promising way to improve bus travel times and reliability developed since the 1960s (Westinghouse Air Brake Company and United States Dept. of Housing, 1968). There are two main types of signal priority, passive priority and active priority.

Passive priority is to implement bus priority by predetermining the signal timing plans according to bus travel times to benefit buses. Hence, passive priority is easy to implement, and there are a few strategies to implement passive priorities (Lin et al., 2015). Traffic engineers can adjust signal timings. Longer cycle lengths can help optimize vehicle progressions, but red times will also be long. Shorter cycles will reduce red times, and the bus signal delays. Traffic engineers could also optimize signal timings along a corridor to reduce the people delays, not vehicle delays, going through the intersection, since buses carry many people at the same time (Lin et al., 2013).

Active priority involves sensors and communication infrastructure between buses and the traffic signals to allow signal priorities in real-time. There are three types of active signal priorities, unconditional priority, conditional priority, and model-based priority.

The unconditional priority will give immediate priority to the buses when they arrive at the intersection, such as extended greens, shortened yellows, or additional bus green times (Dion et al., 2005). When there are multiple requests, various rules are used to determine signal strategies, such as the first-come-first-served policy (Dion et al., 2005).

The conditional priority will give the bus priority when the bus is actually ready to depart to reduce impacts on other traffic. For example, some cities will only give priority to buses when the bus is late, some will limit the number of priority requests for consecutive signal cycles (Smith et al., 2005).

The model-based priority will calculate the best bus priority to minimize a target value, e.g. total bus delay or total person delay. The models require information on bus locations, operation conditions, and traffic conditions, thus, it is more difficult to implement (Vasudevan, 2005).

Typically, the signal priority strategies are evaluated using several indicators, travel time savings, increase in bus speed, reduction in travel time variances, total person delay reductions, car-delay variations, and queue length in non-priority approaches (Lin et al., 2015).

To evaluate the effectiveness of signal priority, there are typically three approaches, mathematical models, simulations, and field tests. Overall, many studies show positive impacts of transit priority signals on reducing transit travel times (Dion et al., 2005; D'Souza et al., 2010). However, some other studies failed to show significant travel time gains from signal priority (Diab and El-Geneidy, 2013). Hence, there is no consensus on the impacts of transit signal priorities in the literature.

For mathematical models, queuing theory is a popular choice to model the effectiveness of signal priority, vehicle delays with and without priority. They typically show localized signal delay reductions under a few assumptions, random traffic conditions, no over-saturated street segments, impact from upstream and on downstream segments are not significant, and no bus operation controls (Abdy and Hellinga, 2011). Unfortunately, these assumptions are not necessarily true for most principal or secondary corridors in central areas.

Since evaluating various traffic environments, operational characteristics, and built environments is not always possible, micro-simulation tests are also widely used to evaluate signal timing settings. Researchers have tested various signal priority strategies, priority thresholds, priority conditions, operational conditions, and various lane configurations. They found signal priority is effective when there is less cross-street traffic, high bus traffic, far-side stops to allow signals to react, good base signal coordination, queue jump lanes, or little turning traffic (Ngan et al., 2004; Zhou and Gan, 2005; Furth et al., 2025).

Overall, one of the common themes previous researchers identified is unconditional signal priority, which yields the best travel times for buses but has a significant impact on crossing traffic. They recommended conditional priority when buses run behind schedule (McLeod and Hounsell, 2003; Altun and Furth, 2009; Al-Deek et al., 2017). Hence, some recommend

agencies to schedule more aggressive travel times to take more advantage of signal priority (Altun and Furth, 2009), but passengers are more likely to arrive late. However, other researchers found less impact to crossing traffic in more suburban environments (Sheffield et al., 2021).

Another common theme from the previous researchers identified is the importance of underlying signal timing. Both passive priority, adjusting planned signal coordination for buses, and active priority, adjusting signal timing in real-time, are important. However, signal priorities are more effective when the underlying signal timing is more supportive for bus travel (Furth et al., 2025). Having passive priority, making underlying signal timing transit-friendly, along with active signal priority, can halve the delays compared to only using active priority without changing the underlying signal timing.

Field tests are costly but more accurate, it is also important to continuously evaluate the effectiveness of the signal priority strategies after the implementation to determine the actual benefits (Lin et al., 2015). For example, Transport for London implemented the iBus AVL system, which allows the agency to determine the bus locations in real-time. This system allowed buses to request signal priority at various virtual detection points, either extending the green or truncating the red, which reduces bus delays by a few seconds per intersection. After leaving the intersection, the priority requests are cancelled to reduce impacts on traffic from other approaches (D'Souza et al., 2010).

However, using the iBus system in Montreal, Diab and El-Geneidy (2013) only showed 1.4 seconds of improvements per timepoint and no impact on delays and travel time variations. They speculated that the no-turn-on-red policy, combined with near-side stops, hinders bus departures. Similarly, using trip-level data, the signal priority results in Portland, Oregon are inconsistent across routes, time periods, and various performance measures (Kimpel et al., 2005).

Hence, many researchers argued that evaluating signal priorities requires regular performance monitoring and calibration to identify problems, willingness to experiment by adjusting operational practices and signal timing plans, obtain detailed baseline observations for comparisons, select routes based on operational problems, not scheduling problems, and taking into account small-sized cities compared to large metropolitan areas (Kimpel et al., 2005; Vlachou et al., 2010).

To conclude, due to various signal timing configurations, conditions applied, models used, and priority strategies implemented, the results of signal priority vary a lot. There is a general lack of effort to quantify the overall signal delays. Some of the studies focused on delays at a given intersection, without considering the downstream impacts. The bus delay

could be reduced by 10 seconds at a given intersection, but if the next intersection is red anyway, signal priority at this intersection does not significantly impact the overall travel times. Hence, it is necessary to evaluate the chain reaction from the start to the end of the line.

For some cities, especially in downtown areas, the majority of the signalized intersections still use a predetermined coordinated fixed timing plan, which creates heavy time dependency between intersections. Researchers have suggested that good arrival time predictions are important in these cities for transit priority signals to be effective at reducing travel times (Li et al., 2012). Better inter-stop and dwell time estimations are needed to improve the arrival time predictions. Scheduling strategies can also be adapted to take advantage of transit priority signals (Altun and Furth, 2009).

However, these previous studies tend to focus on one specific intersection or a few consecutive intersections on a given corridor, which would have similar base timings. Similarly, some studies only included the number of priority signals to model their effects, which essentially assumes these signals behave similarly. Since buses can make turns and travel through multiple corridors, the signal synchronization and cycle lengths may all be different in reality. Smaller intersections typically have shorter cycle lengths, while large intersections have longer cycle lengths. There is still a need to better isolate and quantify signal delays in bus travel time observations, as well as to consider how the travel time and red light waiting times are affected by varying signal cycle lengths and synchronization patterns.

2.8 Transit Scheduling Methods

Travel time is an important measure for scheduling. In this section, we will discuss the status of current scheduling practices.

Coleman et al. (2018) provided a summary of a typical scheduling process in New York City Transit. In general, the route performances are reviewed at various intervals for different types of routes and schedules. Service changes generally happen at pre-defined times every year to facilitate operator sign-ups and schedule adjustments. The schedule revision process generally involves the level of ridership, which relates to the service frequency, and travel times between timepoints in the North American context. If passenger levels exceed a predefined agency standard, service frequency is adjusted. Travel times are also analyzed using an agency standard, and then adjusted both between the two terminals as well as between various timepoints using measures such as the mean, median, or a given percentile of observed travel times (Furth, 2000).

There are typically three types of headways (Ceder, 1987). Equal headway simply means headways are evenly spaced during each time period. Balanced headways are not evenly spaced, but they are set so that observed passenger loads would be similar on all buses. Smooth headways are simply the average of equal and balanced headways. Then, exact headways can be set based on passenger loads.

There are typically four headway setting methods based on point checks and ride checks to ensure enough capacity is allocated for the demand, i.e. ridership obtained at one point along the route and data obtained from a complete bus trip (Ceder, 1987; Vuchic, 2017). The first method is based on data gathered at the maximum load point during the whole day. The second method is to determine service frequency based on the hourly or period max load point. The third method sets the frequency so that the average load along the route is at or under the desired capacity. The fourth method adds a constraint to the third method that limits the length of the route where the load exceeds a certain overcrowding limit. In addition, special requests can be applied to the headways, such as clocked-faced headways.

Once the headways are set, the average round-trip times, including holding and layover times, are used as additional inputs for setting trip departure and arrival times. Levinson (1991) synthesizes the agency practices on bus running times. He points out that good schedules should allow enough running time to operate the route, account for congestion and signal delays, and provide enough layover time at the terminal so that late buses can start the next trip on time. He also found that if drivers leave the terminal late, it is unlikely that they can catch up on the lost time. Levinson (1991) also concludes that scheduled run times should be set at a value slightly less than the mean or median run time to ensure that the majority of operators do not have to kill time to maintain their schedule adherence.

However, in a later synthesis, Furth (2000) argued that whether scheduled running times are based on average running times, ideal running times, or high percentile running times (e.g. 75 to 95 percentile) is a matter of agency policy. Ideal running times help prevent buses from running early, which is less desirable for passengers than trips running late. However, under ideal schedules, trips will tend to run late, and will need greater scheduled layovers at terminals. The high percentile running times help prevent trips from being late, but increase the chances of running early unless agencies implement a strict policy to reduce early departures. The high percentile running times also mean less layover time is needed in the schedule. The average running time policy is a compromise between those extremes.

Furth and Muller (2007) argues that while average running time is sufficient to determine scheduled running time, scheduled layovers should be determined based on running time variability. As a result, most agencies choose to use a high percentile, typically the 85th

percentile, or the average running time plus one standard deviation as their half-cycle time, i.e. the terminal-to-terminal travel time plus layover time at the terminal. This allows agencies to set standards so that only a small percentage of vehicles will not be able to start the next trip on time.

With the implementation of AVL systems, people mistakenly assume these systems provide travel time analyses to agencies by default, simply because they track the vehicles (Muller and Furth, 2001). Precise travel time scheduling under urban congestion requires larger data provided by trip time analysis systems. Muller and Furth (2001) proposed a scheduling strategy that sets travel time to be the 85th percentile from every timepoint to the terminal. The idea is that if drivers end up early and hold at a time point, the remainder of the trip should also satisfy the on-time requirement with high probability. Then, the individual segment travel times are set using repeated subtractions.

Using the archived AVL data, agencies could also retroactively improve the accuracy of transit agencies' schedules. Using the observed AVL data, Wessel et al. (2017) created a bus timetable representing service delivered, which can then be used in assessing transit performance, reliability, and accessibility issues. They argue that transit schedules are often done at timepoints, whereas arrival and departure times for stops in between timepoints are not yet well defined and are typically interpolated using timepoint times in practice. Passengers, on the other hand, require arrival and departure times for every stop served by a certain trip. There is a discrepancy between the general scheduling practices and what's shown to the passengers through GTFS, which calls for further investigation into stop-level scheduling practices.

However, there are fewer articles in the scheduling literature for the dwell times and red times, and these two categories are typically included in the overall running times. One example is that, with the implementation of bus signal priority, planners should also adapt their scheduling practices to take advantage of the signal priorities. Altun and Furth (2009) developed simulation models to test the impact of schedule aggressiveness on bus running times, headway regularity, and crowding. They found that conditional signal priority, which only gives priority to late buses, combined with more aggressive schedules and strategic holding, can lead to a reduction in travel times. More recently, Gallo and Sacco (2022) proposed an optimization of bus dwell times to reduce the red times at the following intersection. However, their method only accounts for one intersection, which showed a reduction in red times. Yet, it is still necessary to consider the possible impacts for the future stops and intersections as well as the impact of various signal timing parameters.

2.9 Passenger Preferences and Behaviours

To further highlight the importance of this research, I will discuss the passengers' preferences and behaviours in this section, how travel times affect passengers and how passengers adapt to unreliable travel times.

Passengers have different preferences regarding transit. TCRP Project B-11 (Morpace International, 1999) used surveys from three different agencies and found that the top concerns are about fares and schedules, then comfort and safety. National Cooperative Highway Research Program (NCHRP) Report 616 (Barker et al., 2008) surveyed five transit systems. Passengers consistently identified service frequency as the top concern, while reliability, wait time, access, and service hours are also significant to passenger satisfaction. Some researchers tried to measure the generalized cost of transit trips, which is a way to associate a monetary value to passenger preferences and perceptions.

Passenger value of time differs by mode and by trip components. Wardman (2004) found that bus passengers have the lowest value of time, whereas rail and metro users have the highest value of time. Trip purpose also contributes to the difference in the value of times. Business trips have the highest value of time. Longer travel distances are also associated with a higher value of time since passengers would try to reduce their travel times. Access time, transfer time, and initial waiting times are also valued higher than in-vehicle time since these activities require more physical effort and are generally not productive for passengers.

Crowding and transit amenities also impact perceived waiting times. Balcombe et al. (2004) showed that platform crowding increases the passenger's perceived waiting time. Bus stop amenities and real-time travel information can reduce the perceived waiting times. For in-vehicle travel times, Wardman and Whelan (2011) found seated transit trips have the lowest perceived travel time, followed by standing trips, and finally, crowded trips. In other words, for the same travel time, passengers in crowded vehicles would perceive travel times longer than passengers in empty vehicles.

Passengers see travel in crowded conditions as worse than less crowded conditions, even when they have a seat. From the agency's point of view, as the passenger load increases, the cost-effectiveness increases, throughput increases, and dwell times increase as well (National Academies of Sciences, Engineering, and Medicine and others, 2013). Therefore, we need to balance passenger comfort and agency resources.

Some researchers also looked into the generalized cost for passengers due to unreliable transit services. Casello et al. (2009) used AVL data to measure schedule adherence at each stop. They defined the generalized cost as fare plus the value of access time, wait time, and

in-vehicle time. They then created three risk categories to model passenger preferences, risk-averse, risk moderate, and risk-neutral. Using the generalized cost and risk categories, their simulation shows that 78% of passengers will choose to arrive nearest to the necessary arrival time, and 17% of passengers would ride an earlier bus to avoid being late at their destination. The generalized cost for the least reliable stop is more than 15% compared to the most reliable stop.

Some literature tries to establish a link between passenger behaviours and transit reliability. From these papers, we can have a glimpse into how passengers adapt when they encounter transit reliability issues, and how these issues affect the passengers' future travel decisions. This would help us target areas that are important to passengers.

A smartphone-based survey was used to measure travel experiences as well as to measure objective travel times. The results show that wait time reliability and travel time reliability are important factors affecting passenger travel experiences (Carrel and Walker, 2017). In-vehicle delays on the metro were perceived as longer than delays on the bus, since passengers were stuck underground and usually cannot directly observe the cause of the delay. Crowding is also an important factor, since it could be related to bus bunching and affect perceived travel times. Scheduled in-vehicle travel time and early arrivals do not have a significant effect on user satisfaction. However, in-vehicle delay is an important driver of dissatisfaction. The transfer times were negatively correlated to satisfaction regardless of trip purpose, meaning longer transfer times correlate to dissatisfaction. Delays encountered at transfer stops have twice the influence on reducing transit use as delays at the origin stop (Carrel et al., 2013a). The respondents also ranked work trip reliability higher than non-work trip reliability, possibly due to people having to be at work at a given time (Wardman, 2004).

In a long-term passenger satisfaction survey, Le and Carrel (2021) found that the long-term retrospective satisfaction is lower than the daily assessments. The negative experiences are more memorable, which would stick with the passengers longer than the positive experiences. Passenger satisfaction will also lag to the following day. If transit passengers encounter dissatisfaction on the previous day, they will typically adjust and lower their expectations for the following day. Similarly, if the satisfaction is high on the previous day, they will raise their expectations on the following day.

After experiencing transit unreliability, the general adaptation strategy is to travel less with transit but more with other modes (Carrel et al., 2013a). However, even with unreliability, some people will not change regular transit routes, since they don't know the reliability of the new route, and some people don't have other route choices (Carrel et al., 2013a). Most regular passengers are aware of the specific travel time issues on their regular route, and

they have strategies to deal with them. In this case, people will typically allow extra time to account for the travel time variations. The extra time people allocate is generally around 10 to 20 minutes. Similarly, people are more forgiving of issues outside of the agency's control, such as police activities. However, the most influential issues that reduce transit use are mainly operation-related delays and denied boarding.

To summarize, if buses become more unreliable, passengers will avoid the service, causing a reduction in fare revenue, and leading to service cuts, which would lead to more passengers avoiding the service, causing a vicious cycle (Aemmer et al., 2022). Travellers tend to orient their life around a certain mode, e.g. by transit, and the decision is relatively stable. In other words, people who stopped using transit are less likely to return unless major improvements are made (Carrel and Walker, 2017).

2.10 Passenger-Centric Travel Time Measures

In this section, we will discuss how passenger travel times are evaluated. Even though our research mainly focuses on the operational issues, there are still potential methodologies or impacts on the passengers to keep in mind. Similar to vehicle travel times, most studies focus on the overall passenger travel times or specific components of passenger travel times. Since the in-vehicle travel times correspond to the vehicle travel times, we will focus on the overall passenger travel times and passenger wait times in this section.

2.10.1 Wait Time

The passenger wait times are the passengers' waiting time at the stop for the vehicle to arrive. Waiting time is a significant portion of passenger trips, and it is often cited as one of "the most important factors hindering the usage of bus transit" (Ma et al., 2013).

Passenger experience in transit is commonly assessed through two key dimensions: headway adherence (regularity of service) and schedule adherence (punctuality) (Barabino et al., 2017). Since passenger arrivals at the bus stops are not always directly observable, the passengers are generally assumed to arrive independently and randomly, i.e. following a Poisson distribution, for high-frequency lines. Therefore, the average wait time is half of the observed headway. However, as shown in some articles, some passengers would still follow the scheduled time for high-frequency routes (Luethi et al., 2007). For low-frequency lines, agencies would typically make some assumptions regarding passenger arrival rates (National Academies of Sciences, Engineering, and Medicine and others, 2013).

For low-frequency routes, agencies typically assume passengers care more about schedule

adherence and the excessive waiting time relative to the schedule is typically used. For high-frequency headway-based services, passengers typically care more about the headway adherence. Hence, the data needed for analyses are different based on service frequency. For high-frequency lines, the data needed are typically the headway variations at the given stop and ridership obtained from APC or Origin-Destination (OD) models. For low-frequency lines, the data needed are the ridership from APC, the assumed passenger arrival rate, and the ridership models.

The indicators for analyses are typically the following (Transport for London, 2022a).

- Scheduled Waiting Time (expected wait based on schedule)
- Actual Wait (observed wait times)
- Excess Waiting Time (difference between actual and scheduled wait)
- Budget Waiting Time (high percentile of observed wait)
- Wait Ratio (actual vs. scheduled wait)
- Probability Distribution of Wait Times (e.g., <10 min, 10–20 min)

Similar to operational travel time indicators, waiting time indicators can also be categorized into two groups, namely, mean-variance-based and extreme-value-based (National Academies of Sciences, Engineering, and Medicine, 2006).

At NYCT, they do not measure the wait time for every passenger but use the passengers with the longest wait time, and check if that is within a certain threshold. They found that excess waiting time is preferred because it reflects both the proportion of affected passengers and the severity of delays (Graves et al., 2019).

The current automated data sources, such as fare gates at the station entrance, can capture passenger arrivals at train stations. However, since most bus services require fare validations on board, they do not necessarily capture passenger arrival rates at bus stops. Researchers often estimate waiting times using GTFS, AVL, and AFC data, based on the aforementioned assumptions on passenger arrival rates. For example, Webb et al. (2020) calculated actual wait times from passenger tap-in times upon entering the station and actual vehicle arrivals, while Transport for London (TfL) used the difference between tap-in data on board and the scheduled bus arrivals to estimate excess waiting times.

Passenger waiting times are also modelled using different distributions, and the Gamma distribution fits the best (Webb et al., 2020). As for spatial and temporal variations, long

bus delays and morning peaks have longer passenger wait times. Wait time reliability would gradually decrease as the vehicle approaches the end of the route, which could be due to the lack of schedule control outside of terminals (Huo et al., 2014). Additional factors are also tested by researchers to capture the user experience. Regression models identify factors such as time of day, stop spacing, route length, and passenger load as key drivers of unreliability (Gittens and Shalaby, 2015).

Because waiting times are not always directly observable, some studies use simulation to evaluate reliability under scenarios like rising congestion, increased demand, or improved ticketing systems (Liu and Sinha, 2007; An et al., 2014). Using empirical data from one bus route, a simulation model for the travel time reliability, headway regularity, and passenger waiting time reliability can be developed. Results show perceived waiting time grows disproportionately with demand, while reducing boarding time significantly improves reliability.

Similarly, using archived vehicle locations and the previously mentioned passenger arrival behaviours for high and low frequency services, researchers have calculated the potential distribution of passenger wait times, which can then be used to calculate various wait time indicators (Furth and Muller, 2006). For example, by assuming passengers arrive uniformly at the stops for high-frequency services, researchers can calculate the excess wait times using the actual variation of headways. By assuming passengers consult the timetable for low-frequency routes, the actual arrival time of the vehicle can be used to calculate excess wait times at stops. A sensitivity analysis found passengers are more impacted by extreme values in the variations in headways or arrival times (Furth and Muller, 2006).

For passengers trying to arrive at their destination on time, they often budget extra time to account for uncertainty in bus arrivals, causing hidden waiting times (Furth and Muller, 2006). Huo et al. (2014) found extra waiting time contributes to 80% of the total extra travel time budget, and extra in-vehicle time contributes 20% to the extra travel time budget. Passenger behaviour models suggest that route knowledge, behaviour adaptation, and real-time information reduce perceived uncertainty and encourage adaptive choices, such as route or mode changes (Cats and Gkioulou, 2017).

Splitting long routes can improve route travel time reliability but introduces transfer penalties for passengers (Gittens and Shalaby, 2015). High passenger loads also degrade reliability, though different causes, whether inadequate frequency or bus bunching, require different interventions such as schedule adjustments for the former and travel time control for the latter (Huo et al., 2014; Gittens and Shalaby, 2015).

To summarize, the literature demonstrates that waiting time is a critical determinant of perceived reliability, influenced by both operational factors (e.g., stop spacing, route length,

passenger load) and passenger behaviour responses (e.g., route knowledge). While excess waiting time is widely recognized as a robust metric, challenges remain in accurately estimating waiting times, especially for intermediate stops and under varying demand conditions. Simulation studies provided insights but often rely on simplified assumptions on passenger arrival rates.

2.10.2 Transfer Time

Medium to large transit networks often consist of multiple modes and sometimes require passengers to transfer within or across different modes. It is widely shown that passengers dislike transfers. Passenger experiences are also based on the whole journey, which includes transfers. Therefore, it is important to examine transfer times closely.

The transfer time section is added by me in addition to the overall waiting times used by TCQSM and Danaher et al. (2020). As the TCQSM pointed out, the coefficient of variation of transfer times and transfer time deviations are possible indicators. However, day-to-day arrival time variations make the measurement difficult (National Academies of Sciences, Engineering, and Medicine and others, 2013), since there is a limited amount of literature available on transfer times.

Transfer time can contribute significantly to the overall reliability measures for passengers, as it is dependent on two transit trips, from a generating trip to a receiving trip. Missing transfers will result in additional delays for passengers, as the passengers need to wait for the next departure or find an alternative path.

Distinguishing different waiting times is important because passengers experience greater disutility at transfer stops (Wardman, 2004). The travellers have more control regarding arrival time at the first stop, but transfer times are often less flexible and unknown for passengers when travellers commit to a specific transfer point (Carrel et al., 2015). However, it is not possible to know if a passenger had a quick non-travel-related activity at the transfer points, or spent the time waiting.

Most of the literature focusing on passenger transfers has been done in systems with both tap-ins and tap-outs, which will tell the agencies the exact route and times related to the passenger trip. However, for systems without tap-out data, much of the literature relies on simulation and optimization to study passenger transfers.

Several studies optimize timed transfers to reduce waiting and missed connections. Ceder et al. (2013) propose a method to design networks with high transfer synchronization and simulate policies such as holding and skip-stop to assess passenger delays. Similarly, Parbo et al.

(2014) incorporates passenger route choice (via the Danish national route-choice model) into transfer optimization; the resulting timetable reduces weighted waiting time by an amount equivalent to 45 million Danish kroner per year. However, a key limitation is that these analyses are schedule-based, largely ignoring stochastic arrival/departure variation.

Another focus of the literature is the modelling of transfer times. Focusing on the distribution of vehicle arrivals and departures, Kieu (2015) simulated transfers between two bus routes using AVL-based observations. They showed that no planned transfer time minimizes the mean transfer time or the probability of missing a transfer. However, operational interventions, such as holding, scheduling adjustments, and synchronized transfer policies, can reduce mean transfer time by around 20% and cut the probability of missing a transfer by around 80%.

Using AFC data, Jang (2010) summarized transfer patterns between modes, such as bus-to-rail and bus-to-bus transfers, and the corresponding trip frequencies. The author found that most trips involve no transfers. For trips with transfers, passengers prefer to transfer near origins or destinations, and around 80% of transfers have less than 10 minutes of transfer time, which is a useful guide for agencies to target improvements.

Missing a transfer could greatly impact the overall transfer times and travel times for passengers, hence the literature also focuses on evaluating the impact of missed transfers. Mai et al. (2012) simulate two untimed transfer points using APC data derived passenger arrival distributions and AVL derived schedule deviation. Thus, in their simulation, passengers might miss both the generating and the receiving bus. Using these distributions, they found that the passenger OD travel times roughly follow the normal distribution. By simulating schedule changes, they found that shifting the receiving bus earlier, 144% more passengers will miss their transfer, but the mean travel time only increased by 1 minute, because transfer time is a small share of total trip time in the test, and some passengers benefit from shorter waits. Narrowing the scheduled transfer window reduces overall travel time despite more missed transfers, indicating that traditional metrics may not capture the impact of missed transfers fully.

Liu and Miller (2021) proposed different risk measures related to transfers, desynchronization degree (the number of trips gained or missed relative to the intended transfer), transfer time penalty (the time loss compared to the schedule at the receiving stop), and risk of missed transfer (share of missed transfers). Using data from Central Ohio, they showed temporal and spatial distributions of missed transfers, where major corridors have higher miss risk but lower penalties due to high service frequency, while peripheral roads exhibit lower risk but higher penalties. Holidays and summer show lower penalties due to less congestion and

traffic variations, and Wednesday to Friday services and peak hours services show a higher risk and penalties due to more varied traffic patterns.

Transit service quality and reliability have an impact on passenger route choices, and some researchers have included the impacts of transfer time in their studies through surveys. The general methodology is to infer the actual transit paths taken by the passengers, then match the paths to the schedule and AVL. The authors then generated a series of alternative paths to compare to the actual path and model mode choices by including different characteristics of the paths (Beduhn et al., 2015; Carrel et al., 2015).

Beduhn et al. (2015) examined the effect of transit service reliability on passenger route choices using an on-board survey. Their logit model shows that reliable transfers, like the probability of making a transfer and reliable departure times, significantly increase the utility of a specific route.

Similarly, using smartphone-based travel diaries matched to service delivery data via AVL, Carrel et al. (2015) infer origin and transfer waits, in-vehicle deviations, transfer-time variability, and overall travel-time variability by comparing observed trips to published timetables. Using decomposed travel times to understand the source of delays and variability, they find that the origin wait times do not strongly differ between frequent vs. infrequent service, likely due to real-time information and route knowledge adaptation, while around 70% of the waiting time was incurred at the transfer stop.

Crowdsourced traces also help map transfer realities. Traut and Steinfeld (2019) match crowdsourced passenger Global Positioning System (GPS) traces to GTFS data to estimate walking distances and transfer durations and identify popular transfer points (including those not officially recommended). They report most transfers involve less than 500 metres of walking, and the mean actual transfer times are shorter than scheduled. However, some popular unofficial transfer points not recommended by the agency have relatively higher crime rates, which affects the perceived passenger safety and passenger satisfaction. However, a caveat is self-selection bias, where riders may be more likely to report disrupted or unpleasant trips.

To summarize, in practice, most transfer analyses rely on AFC tap-in/tap-out or are simulated from demand. Extending such work to tap-in-only systems (common with flat fares in North America) is feasible with OD inference and modern routing engines, enabling comparable transfer analytics without exit data. Analyses should also recognize "optional" transfers (e.g., between buses on a shared corridor) that passengers may choose to skip due to crowding or limited-stop service patterns. In addition, we still need to understand why passengers do not transfer at a stop. It could be that the transfer is unreliable or too long, so passengers

learned to avoid certain stops. Finally, many studies remain highly aggregated, and there is a need for more decomposed analyses with finer spatial and temporal granularity to understand where and when reliability issues arise.

2.10.3 Overall Passenger Travel Time

The overall passenger travel time is the actual lengths of passenger trips, involving access time from the origin to the boarding stop, waiting time, in-vehicle travel time, potential transfer time, and access time from the alighting stop to the destination. Agencies sometimes would ignore the access times for the initial boarding and the final alighting, since their reliability is beyond the agency's control. So, most agencies would consider the passenger travel time as the initial access to the system until the passenger exits the system (Danaher et al., 2020).

It is not possible to offer one-seat rides to all origin and destination pairs, transfers are necessary for passengers. Each transfer adds to the total trip time and increases the possibility that missed connections will occur, which would further delay the passenger's final arrival (National Academies of Sciences, Engineering, and Medicine and others, 2013).

Passenger-based on-time performance can show the magnitude and the impact of system delays. A 30-minute delay during off-peak has a different impact compared to a 30-minute delay during peak. Passenger-based measures reflect the magnitude of delays (Buneman, 1984).

Hendren et al. (2015) argued that the focus should shift to customers, not vehicles, since vehicle-based travel time measures do not show how many people were affected and for how long. They drew similarities between moving transit vehicles and focused on volume capacity ratios in highway engineering. They are important from an infrastructure perspective but remove users from the equation.

Having customer-centric journey time metrics can also help improve public communication to riders and help measure how passengers are affected by service. (Graves et al., 2019) Passenger-based measures can improve transparency and give agencies metrics that better reflect their priorities towards passengers. Agencies could also use these metrics to understand how incidents and service adjustments affect passengers. (Halvorsen et al., 2019)

The data for passenger trips are typically obtained from fare card tap-ins and tap-outs. However, for agencies that use an honour-based fare system or a system without tap-outs, analysts would typically estimate these data using surveys or regional OD model runs. AVL data is also required to match passenger OD pairs to actual arrival and travel times.

The OD level data can help agencies estimate the boarding and alighting stops and the

ridership for the OD pairs. Then, agencies would use AVL to examine the travel time variability for each OD pair. High variability indicates passenger trip times are less consistent, forcing passengers to plan for extra times.

Passengers typically plan their trips based on worst-case scenarios. Analysts can also analyze the buffer times, which is the difference between a high percentile and the median. A high buffer time indicates passenger trip times are less consistent, forcing passengers to plan for extra time.

This analysis can provide detailed information regarding passenger experience. However, passenger-based measures don't translate very well to agency operations, and typically don't provide actionable information for the agencies to improve their service (Hendren et al., 2015). It is also not practical for use on a routine basis for agencies due to the difficulties and complexities in calculating the OD and passenger routes. TCRP report 215 (Danaher et al., 2020) suggests agencies use this measure for special or one-time studies.

Researchers have long emphasized the importance of measuring end-to-end passenger journey times as a core indicator of transit service quality. Many works used automated data sources to estimate passenger travel times and develop reliability metrics that better reflect passenger experience than traditional vehicle-based measures.

Many studies used AFC smart card data to infer passenger trip durations and speeds. For example, Jang (2010) estimated systemwide travel and transfer times using AFC data with tap-in and tap-out records showing passenger origins and destinations and computed travel speeds from timestamps and Euclidean distances. They visualized travel times to downtown in different modes, showing areas with slow services and subway trips outperform buses, especially for origins or destinations near stations. However, reliance on Euclidean distance underestimates actual network distance, penalizing neighbourhood coverage routes designed for access rather than speed.

Similarly, using the GTFS and AFC data from Hague, the Netherlands, Bagherian et al. (2016) summarized passenger travel patterns where they showed that 80% of the trips are direct without transfers, 17% have one transfer, and most transfers are less than 7 minutes. 20% stops accounted for 70% of the trips, stressing the need for passenger-level analysis. Temporally, services during AM peaks and weekends are more reliable, whereas services during Friday PM peaks are less reliable. Trams are less reliable than buses, though the study does not control for spatial context, where trams run in the centers vs. buses run in the suburbs, which may explain modal differences.

Li et al. (2021) added to the literature by estimating passenger door-to-door travel times

for the Chicago area using GTFS, AVL, AFC, OD model data from Chicago Transit Authority (CTA) and census. Using a routing engine to evaluate walking and in-vehicle times, they compared approaches with schedules only, with perfect foresight using AVL data, and behaviour-based OD travel times. Results show behaviour-based OD travel times exceed schedule and AVL estimates, with large discrepancies in suburban areas. However, assuming wait times as half-headway for infrequent routes may overstate delays where most riders would consult schedules, helping explain large differences relative to AVL-based times in some neighbourhoods.

To guide investment, Arias et al. (2021) identified corridors where bus lanes would yield the greatest passenger travel-time reductions. Using GTFS and stop-level ridership, they computed schedule padding and passenger-weighted savings. By comparing the scheduled time to Google’s car travel time estimates, they found GTFS underestimates peak travel times on 53.52% of segments on the network. The top 10% of segments, based on passenger-weighted travel time savings, contribute to 179,018 minutes of potential travel time savings, equivalent to 55% of all delays in the system and 124 days of delays for passengers on an average weekday. Mapping these hot spots helped the agency to target transit priority investments.

Other than calculating times directly, there are also various passenger-centric reliability measures proposed by researchers to capture the magnitude and distribution of impacts. Dating back to 1984, Buneman (1984) introduced two key measures, similar to operational metrics like on-time performance, using passengers’ fare transactions from AFC with tap-in and tap-out to match train movements in AVL data from San Francisco Bay Area Rapid Transit (BART). Many transit agencies and researchers have analyzed, adapted, and applied these measures.

1. Trip dependability, $\frac{\text{number of passenger trips on time} \times 100}{\text{total trips}}$, where an on-time trip is a trip with delays within a certain window, such as less than 5 min at NYCT and BART and 10 minutes at Washington Metropolitan Area Transit Authority (WMATA). Similar to on-time performance, this measure would be expressed as a percentage. For example, 99 percent would be very dependable service, whereas 50 percent would be very poor service.

2. Expected delay, $\frac{\text{total passenger minutes of delay (excess journey time)}}{\text{total trips}}$, which is the average passenger delay. Passenger delays can also be analyzed using a predefined percentile. The unit is in minutes. For example, 0.1 min shows very dependable service, whereas 5 min shows very poor service. This measure is sensitive to the duration of delays and the number of passengers affected.

In addition to measuring the passenger delays, Zhao et al. (2013) also considered early arrivals

using negative delays using the data from London Overground. They found that excessive journey times concentrated on heavily loaded lines or often-delayed lines. Further disaggregated the results by the direction of travel due to the difference in inbound vs outbound demands at the same time of the day. More recently, Hendren et al. (2015) compared the two approaches (i.e. on-time performance and passenger delay) using tap-in and tap-out data from WMATA to calculate the OD-level passenger travel times and variability. They showed individual travel times reflect the passenger experience much more accurately, and passenger-based metrics exposed system-wide ripple effects during disruptions.

In North America, most agencies use a flat fare structure without tap-out. In this setting, NYCT (Graves et al., 2019; Halvorsen et al., 2019) adapted the passenger delay measures using an OD-level ridership model inferred from the AFC data across many days. Then, the estimated OD-level data were matched to scheduled and actual bus and train movement data. They then compared the passenger-based metrics to the agency vehicle-based metrics. They showed these measures correlate with on-time performance but are more sensitive to major incidents, as they account for how many riders are affected.

Since passengers have different expectations and tolerances, they perceive travel times differently in various travel conditions. Standing and crowded changes how passengers perceive time, and are typically considered to be longer than travel times with a seat by passengers. Incorporating seat availability and crowding multipliers to the observed travel times (e.g., Wardman and Whelan (2011)), Jenelius (2018) show that perceived journey times are longer than the nominal travel time, especially during the afternoon peak when the ridership is high. The peak of perceived journey time also happens later than the peak of nominal travel times, which shows that crowding remains high even after the congestion has cleared.

Rather than analyzing the absolute delays, some researchers proposed to analyze the worst-case scenario since passengers plan extra time to ensure arriving at their destination on time, especially for time-sensitive trips to work or school. Buffer time is proposed to measure the extra time a passenger has to plan, in addition to the usual travel time, to arrive at the destination on time with a certain level of tolerance.

Researchers typically compare a high percentile, e.g. the 95th percentile travel time, to the median travel time, which would represent the extra time passengers have to plan to have at most 2 disruptions in a month (Danaher et al., 2020). The proposed 95th percentile value strikes a balance between passenger relevance and realistic expectations of the service and is not susceptible to any biases due to unusual individual passenger behaviour (e.g., waiting for a friend inside a station), which were found generally to occur beyond the 99.5th percentile for the case of the London Underground (Uniman et al., 2010). The median travel time

represents the typical duration of a journey and is preferred to the mean because it is not sensitive to outliers. A high buffer time indicates passenger trip times are less consistent, forcing passengers to plan for extra time.

Since it is a relatively new measure, the literature typically focuses on measuring the buffer time from different perspectives. Uniman et al. (2010) measured the experienced service reliability for passengers using the London Underground. Similarly, Carrel et al. (2013b) proposed a framework to evaluate the impact of operation control strategies on passenger buffer times. As an example, they found incident-related disruptions had larger effects in February than in November for the Victoria line. There is a 73% increase in the amount of time passengers need to budget in February. In November, passengers need to budget 42% more compared to the typical baseline. By breaking down journey times into components such as in-vehicle travel times and waiting times. Unreliability contributed 16% of total perceived journey time, in which 7% was caused by service interruptions. If non-recurring incidents could be completely eliminated, reliability can improve by 40% and perceived travel time can improve by 7%.

To generalize the buffer time measure for multimodal trips, Dixit et al. (2019) compared buffer times for different modes and routes, as well as the variability of different components of buffer times, i.e. wait times, in-vehicle times, and transfer times. They showed that trips using only the metro system tend to be the most reliable, followed by direct trips on a bus or a tram. Trams serve more congested inner-city areas, making service less reliable. Buses serve more suburban areas, which have better reliability, but the frequency is low, making transfers unreliable. Trips with a transfer increase buffer times. Transit hubs outside of the city center tend to be more reliable due to less congestion.

Since the buffer time measure focuses on a period of time, it does not directly correlate to the specific factors affecting transit reliability. Ma et al. (2014) tried to adapt the buffer time metrics for agency operational uses by using AVL data to calculate the line-level demand-weighted buffer times. They observed multimodal distributions in the travel time data similar to the vehicle-based literature and argued that agencies should consider passenger buffer times under different travel conditions. Planners can evaluate corresponding travel conditions to better align improvement strategies and to provide better travel information for passengers.

To summarize, most of the literature focuses on systems with tap-in and tap-out data, and there are limited studies applied to tap-in-only systems. For the paper that focused on tap-in only systems, they all created an overall averaged ridership model for the network using an OD pattern generated from one chosen day. The resulting OD pattern will not necessarily capture some nonrecurring trips or less-used trips. It will also omit day-to-day variations in

travel patterns. There is a need for more detailed analysis to observe any temporal or spatial changes and to model detailed factors affecting these measures.

2.11 Summary of Literature Review and Research Gaps

In general, there are many existing works looking into transit travel times, given its importance. However, due to the amount of information involved, most literature is limited in terms of analysis scales, as most of them focus on a few routes and a predetermined analysis level. There is no standard in analyzing transit travel times and travel time variability since the definitions vary from agency to agency, literature to literature, and passenger to passenger. Thus, there is a lack of systemwide analysis, multi-level comparison, and spatial-temporal comparison in the literature.

Agencies focus on operational travel times and reliability since they affect costs, operator and vehicle scheduling, fleet and facilities management, as well as capital investments. Most literature focuses on punctuality, travel time, and waiting time, and their corresponding variability for agency-based measures. These measures give agencies actionable items to improve their performance. Planners often need to add travel times or layover times to account for the increase or variation in travel times, which increases the operating costs. Missing a layover due to unrealistic or unreliable planned travel times could increase operators' dissatisfaction, causing operator retention issues and propagating delays to trips downstream (Danaher et al., 2020). Thus, improving travel times and reducing their variability benefits the agencies by reducing the excessive costs and focusing on improving passenger experiences.

Passengers care about the service delivered to them, but not necessarily the behind-the-scenes arrangements. A passenger trip might involve more than one transit vehicle or mode as well. Some agencies and researchers have also started to focus on passenger experiences and perceptions. From the passengers' perspective, long or unreliable travel times are less attractive, therefore affecting the passengers' satisfaction and mode choices, causing rider retention issues and reducing ridership. Once the passengers switch to other modes, it becomes a lot harder to get them to come back to transit unless major improvements are made (Carrel et al., 2013a). Some surveys pointed out that passengers value service reliability more than service frequency and travel times (Balcombe et al., 2004; Daskalakis and Stathopoulos, 2008).

As the technology progresses, automated passive datasets like AVL, APC, and AFC become more granular, and we can observe travel times with more details. The detailed observations on the segments can be summarized as an empirical probability density function. Hence, we

can improve our understanding of transit travel times by better modelling the empirical travel time distributions, the means, the variances, as well as the potential impacts on passengers. A summary of these gaps can be found below.

2.11.1 General Gap 1: Systemwide Multi-level Comparisons

For the analysis levels, the literature and agencies typically focus on the following levels for travel time variations, namely the route, timepoint, and stop levels (Danaher et al., 2020). For route and trip levels, agencies typically use travel times and travel time variations in scheduling, improving the on-time performance, meeting layover times laid out in the labour agreements, and assigning vehicle and operator resources (Danaher et al., 2020).

In practice, even though transit schedules are typically revised multiple times a year, travel times are revised less frequently at the timepoint level. Stop level departure and arrival times, which passengers need to plan their trips, are typically interpolated in North America, which may or may not reflect the reality (Coleman et al., 2018). Schedule controls, such as holdings, are also typically done at the timepoints. Whereas, in Europe, some agencies would consider every stop as a timepoint to avoid large schedule deviations (Muller and Furth, 2001).

Unfortunately, given the amount of information involved, most of the literature does not consider more detailed analysis levels than the stop-to-stop level. In addition, most literature focuses on a few routes and one chosen analysis level. Thus, there is still a need to examine the pros and cons of each analysis level using systemwide observations.

Travel time can then be further divided into the sum of a series of dwell times, red times, and inter-stop travel times. The current literature has mostly focused on travel times and dwell times as well as their variations for specific lines or corridors. However, travel time models might not be the same across the entire system. There is a need to cluster different routes or corridors so that we can discover routes with similar issues. This way, planners can target all of them at the same time to improve performance. It would also show which factors affect the entire system, and which factors have more local effects on different segments.

2.11.2 General Gap 2: Spatial and Temporal Variations

Similar to the previous observation, most studies do not consider the spatial differences, probably because they do not have systemwide data. Most of them show examples of one line or a few chosen segments, but there are line-to-line or segment-to-segment variations. Segments downtown might have different performances compared to segments in car-oriented suburbs. There is limited literature that examines travel time conditions and the spatial

differences.

Only a few studies looked into temporal changes other than the time of the day. The previous literature generally aggregates the entire dataset rather than some subsets. Other temporal changes could affect transit performance as well. The literature mainly focused on weekdays, since they have the most ridership. Weekends are likely to have different performances compared to weekdays due to different travel patterns and traffic conditions. Schedule changes could also affect transit performance. During the year, the agencies will update their schedules multiple times to address some service issues. However, there is still room to compare these different temporal variations. For example, when the schools are on vacation during the summer, the routes that pass near schools might have a faster travel time due to less demand and traffic. Similarly, during holidays, there is also less demand for work trips, which contributes to seasonal variations.

2.11.3 Detailed Gap 1: Statistical Distributions of Travel time

Transit travel time distributions can provide key statistics for analyzing service reliability, exploring causes of service unreliability, and adjusting service schedules (Mazloumi et al., 2010). Previous literature has attempted to fit transit travel times to symmetrical distributions Abkowitz and Engelstein (1984). However, Mazloumi et al. (2010) then observed that travel times could also follow skewed distributions, such as lognormal or gamma distributions. More recent literature (Ma et al., 2014; Kieu et al., 2015; Wang, 2020) has also observed mixture distributions from more detailed AVL data.

Since mixture distribution represents several unimodal distributions weighted with different probabilities, the possible explanation provided by Ma et al. (2014) is that there exist many travel time patterns related to underlying traffic conditions. For bus travel time, the multimodal distribution can be related to different operation characteristics, such as free-flow operation patterns in the early morning and congested operation patterns during peak hours. However, there is no discussion on what factors could relate to these distributions, and there is no guidance on what agencies should do in case of observing these multimodal distributions.

The gap is also closely tied to the general gaps mentioned above, as the current research on travel time distributions often focuses on one chosen analysis level of a given route. It is still needed to conduct more system-wide analysis to enable comparisons between routes at various analysis levels to identify potential problems. Again, since the current literature tend to focus on one route or one corridor, it is also needed to analyze the spatial and temporal contexts of travel time distributions. Analyzing the distributions across different spatial and temporal

contexts would help identify factors that affect the travel time distributions. Without this knowledge, transit agencies are left without actionable insights to manage service reliability effectively, especially under varying traffic conditions.

2.11.4 Detailed Gap 2: Various Average Travel Time Models

Many works have tried to predict transit travel or arrival times, and most of the literature focuses on time-related measures. Scholars have proposed methods for predicting transit travel times for planning or passenger information. Some of the works used only AVL data (Gurmu and Fan, 2014; Chen et al., 2023). Some incorporated additional information, such as real-time traffic data (Ma et al., 2019; Han et al., 2020), roadside features (Abdelhalim and Zhao, 2024), and weather data (Alam et al., 2021).

However, travel speed, pace, and time are related variables, where the travel time equals the travel distance divided by the travel speed and also equals the distance multiplied by the travel pace. By using a normalized unit of measure, such as speed or pace, we can remove the distance from the equation, and we can potentially find similarities and differences between various segments in the system. Therefore, operating speed and pace are other commonly used indicators for transit performance evaluations, and both can be used as inputs for various planning software.

There is still a need to compare the various dependent variables, time, speed, and pace, to examine their advantages and disadvantages given various planning scenarios. There have been some efforts to compare time and speed measures in the transportation field, especially from Bauer and Tulic (2018), which posed a similar question for taxi travel times. However, there are some additional considerations for public transit planning, such as stops, ridership variations, and transit priority measures. Thus, we ask the question of whether we could compare these two approaches for transit planning.

In addition, the works mentioned above have used many error measures to evaluate their model performances, such as absolute measures and relative measures. Yet, these measures are typically aggregated into one number. Different measures evaluate the results "from different angles" (Kolassa, 2020), and there are not many comparisons between the measures. As transit services have various segment sizes, it is also necessary to compare the errors by segment for potential biases, since a short local segment is not directly comparable with a long highway express segment, for example.

The gap in comparing average travel time models at different analysis levels also links directly to the general gaps, as different models may behave differently at different analysis levels

and the models are generally estimated based on a few selected routes. Hence, it is still important to evaluate the strengths and weaknesses of these models across different analysis levels. To analyze the models at the system-wide level, spatial and temporal variations need to be accounted for, which relates to the second general gap. Different modelling approaches might be more or less effective depending on the spatial and temporal factors. For example, a model that works well in wide suburban streets might not have the same performance in congested downtown corridors. A more comprehensive comparison across different contexts is needed to better understand travel time models.

2.11.5 Detailed Gap 3: Understanding Travel Time Variations

Currently, even the smallest scale so far, namely the stop-to-stop scale, consists of the travel time between the two stops, the dwell time at one stop, and potentially multiple traffic signal waiting times. A much finer analysis scale with the ability to better isolate various trip elements and handle interaction effects among these elements is needed to attribute the travel time variation to a specific issue or a combination of issues, which relates to the first general gap on multi-level analysis.

The literature mostly focused on the travel time impacts from one specific element of the transit system, namely the variation in travel times, dwell times, and signal priorities. There is still less attention on combining these various elements to examine how these travel time elements affect and interact with each other.

In addition, since transit agencies have limited resources, there is still a need to help planners prioritize their resources. Thus, it is important to determine which trip element is more important on a given route to reduce the overall travel time variations. By quantifying the importance of each trip element, planners can select a good and effective strategy to improve the reliability of a specific route.

Since congestion, ridership, and traffic signal programming have strong spatial and temporal patterns, transit travel time patterns might change depending on the time of day (e.g., rush hour vs. midday) or location (e.g., downtown vs. suburb). By combining these variations with multi-level analysis, transit agencies can better identify areas with the most severe or consistent delays and tailor interventions accordingly, which is also identified in the second general gap.

2.11.6 Detailed Gap 4: Using detailed trip components in travel time estimation

The existing literature has mostly focused on evaluating the overall travel time or one specific component of the overall travel time. Following the previous research gap, it is important to examine how detailed travel time components and their interactions can be used in a travel time modelling context with various planning scenarios. A travel time model that accounts for these detailed trip components is still needed to improve travel time and travel time variation estimations given different changes in the network. This would help planners better react to changes in the operational context and address potential reliability issues more proactively.

The lack of detailed models incorporating trip components into travel time estimations is strongly related to both general gaps. Detailed models that integrate components like dwell times and signal delays can help evaluate changes at specific locations on overall travel time, across the analysis levels. Individual trip components are often influenced by both spatial and temporal factors, such as ridership and traffic conditions, which are crucial to understand when analyzing the travel times as a whole. A detailed approach is needed to help include these detailed spatial and temporal variations in travel time models, which will improve the models' flexibility given various changes in the system.

2.11.7 Detailed Gap 5: Understanding Potential Passenger Travel Time Variations

To understand the passenger experiences, most literature focuses on AVL and AFC data with tap-ins and tap-outs. This would allow them to have the boarding and alighting stops for the passenger, as well as the potential vehicles the passengers boarded, which would help them calculate their journey times. Thus, passenger-based studies are mainly related to rail services. However, most bus services and metro systems in North America do not require passengers to tap-out. This would create a challenge in estimating passenger journey times. Some previous literature addressed this problem by using an OD matrix generated by AFC data once a month. Due to this challenge, most of these passenger-centric measures are also published once a month. These analyses do not include weekend or holiday services, since the travel patterns are different. London Underground even stopped publishing their passenger excess journey time metric during the Coronavirus Disease (COVID) disruptions due to the difference between their travel demand assumptions and the actual passenger travel patterns (Transport for London, 2022b).

Focusing on observed passenger AFC data or estimated demand could potentially hide the

travel experiences for passengers travelling between OD pairs with lower demand. Some passengers could change their transport mode choice due to service reliability issues, hence lowering the demand. Some passengers without private car access would still depend on transit to make a trip on a lower demand OD. Hence, it is still important to evaluate and compare the potential reliability issues on these lower-demand OD pairs to evaluate potential equity issues, thus relating to the first general gap on system-wide analysis. In addition, there is a need to better understand the spatial-temporal variations of passenger travel time reliability, which would indicate the specific time periods or areas with unreliable routes or services, so that the agencies could target their resources to improve reliability for passengers, which connects to the second general gap on spatial and temporal analysis.

There is also a need to connect agency-based measures with passenger-based measures since most passenger-based measures do not inform agencies on how to improve their services, possibly due to the limitations related to generating an OD matrix with trip times. There are some strategies agencies could implement, such as timed transfer to improve wait times and revise schedules to improve in-vehicle times. However, given the variation in route headways, travel times, and available resources, they may not necessarily be possible. There is still a need to better understand what the specific problems are for a potential passenger trip to be unreliable, whether it be the wait times or the in-vehicle times. Understanding passenger experiences in more detail could help agencies link the passenger-based measures to their operational measures, which would help identify strategies to improve service.

CHAPTER 3 DATA AND GENERAL METHODOLOGY

This chapter presents the general methodological framework used to improve the understanding of transit travel times and their variability. It outlines the general and specific research questions, the multi-level analysis approach, and the integration of various data sources used, including GTFS, GTFS-RT, AVL, AFC, OpenStreetMap, and ride-check observations. The chapter details how these datasets are processed and matched to enable fine-grained analysis across different spatial and temporal scales. It also introduces the five research papers that address distinct aspects of travel time performance.

3.1 Research Questions

As mentioned in the literature review, understanding transit travel times and their variability is essential for improving service reliability, passenger satisfaction, and operational efficiency. While many studies have explored aspects of transit performance, they often focus on limited routes or specific analysis levels, leaving gaps in systemwide understanding. This research addresses these gaps by posing two overall and five targeted research questions, each contributing to a comprehensive framework for analyzing transit travel times using automated data sources.

3.1.1 General Question 1: How do different analysis levels compare in assessing systemwide transit performance, and what are their implications for transit planning and scheduling?

As shown previously, a major limitation in current transit research and agency practice is the predominant focus on a single level of analysis, typically at the timepoint level or stop-to-stop level, when evaluating travel times and system performance. While these high-level analyses are essential for high-level planning and scheduling, they may mask the fine-grained variability in service reliability and performance. A more detailed analysis may reveal more details on specific issues for passengers and planners. Moreover, in North America, most stop-level times communicated to the passengers are interpolated between timepoints, potentially distorting passenger-facing schedules. Comparing different analysis levels could illustrate the different perspectives and insights they could offer to planners. This research question also seeks to evaluate the pros and cons of each analysis level from a systemwide perspective, identifying when and where each level is most useful in what scenario.

3.1.2 General Question 2: How do spatial and temporal variations affect transit travel times and reliability across a transit system?

Current literature often overlooks the spatial and temporal diversity within transit systems, possibly due to data limits. Most studies examine specific routes or limited geographic areas, which would not show variations across regions with different land use and traffic conditions. Similarly, while many papers address time-of-day effects, e.g. peak and off-peak, few investigate day-of-week or seasonal variations in depth. Without this knowledge, agencies may overgeneralize solutions that only work in specific routes, areas, or times, leading to ineffective service changes. Hence, this research attempts to include spatial and temporal variations of transit travel times.

3.1.3 Specific Question 1: Which segments have travel times that follow mixture statistical distributions, and which operational or environmental factors could explain mixture statistical distributions?

Many earlier studies in transit research assumed that travel time distributions are symmetrical and unimodal, typically Gaussian. However, more recent findings show that real-world travel times often follow skewed or even multimodal distributions with multiple peaks in the empirical probability density function (Example Figure 3.1). These more complex patterns reflect different operational states, e.g. free-flow vs. congested traffic, and complicate planning efforts. Multimodal distributions suggest that multiple distinct travel time regimes exist within the same segment or time period, potentially requiring different operational responses. Hence, we ask the following questions: where and when do mixture distributions occur, and what are the environmental and operational factors that relate to the mixture distribution. This question also relates to the two general questions, are there any spatial-temporal differences in the presence of mixture distributions among the various analysis levels.

3.1.4 Specific Question 2: What are the pros and cons of different related dependent variables when modelling average travel times?

While transit agencies and researchers often focus on predicting travel time directly, alternatives like operating speed or pace (inverse of speed) offer normalized approaches that can remove distance as a factor since the distances do not change for fixed-route services. However, each of these variables may have different implications at different analysis levels. What's missing from the literature is a systematic comparison of these variables across different spatial-temporal contexts and model performance metrics. Hence, we ask the questions,

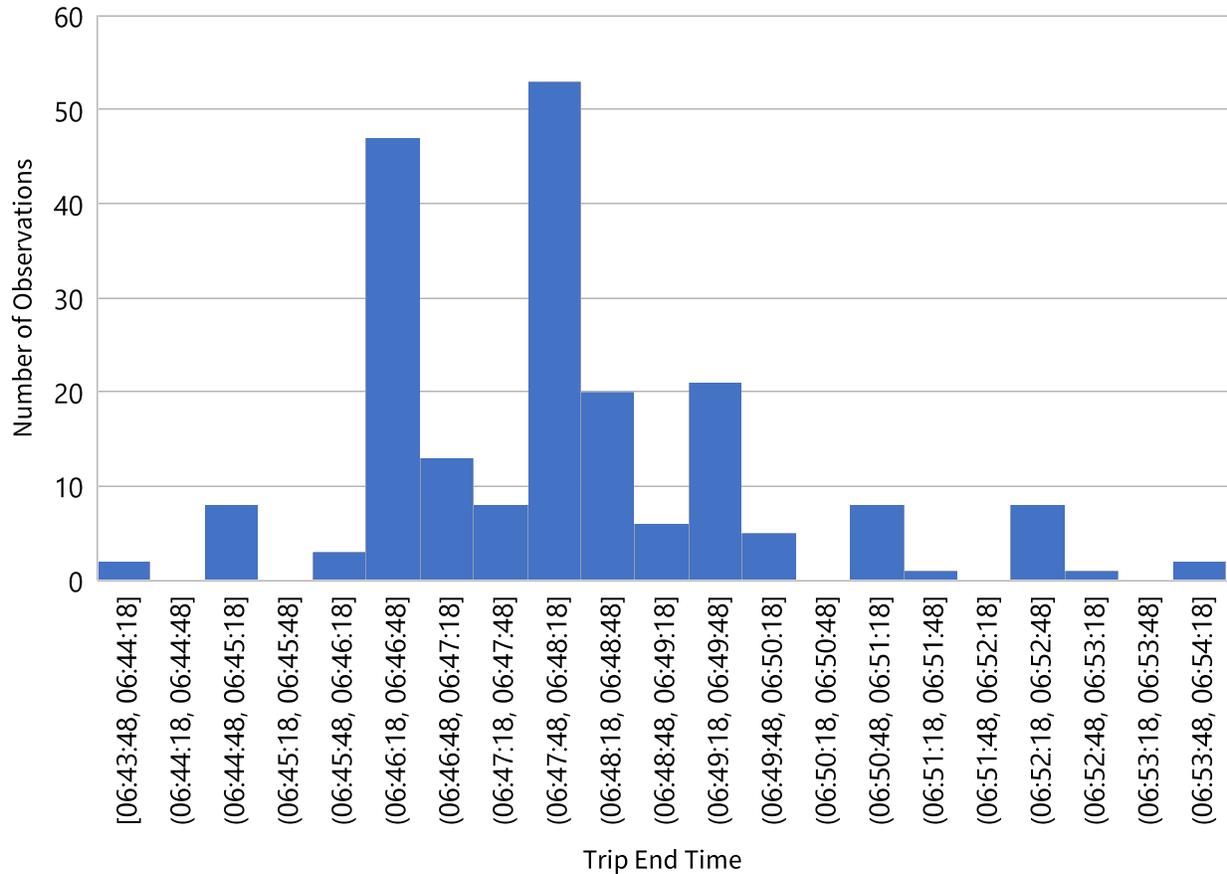


Figure 3.1 Example of mixture travel time distribution observed for trips departed at 6:30 on westbound route 27 in Montréal.

using the same spatial, temporal, and operational characteristics as inputs, does modelling travel time directly yield similar results compared to modelling normalized variables like speed or pace, does the model performance differ for each analysis level relating to the first general question, what are the spatial temporal characteristics for each model (general question 2), what are the advantages and disadvantages of using different evaluation criteria?

3.1.5 Specific Question 3: How much travel time variance can be attributed to specific trip components?

Transit travel time is composed of various detailed components, dwell times at stops, waiting at traffic lights, and inter-stop travel times, each corresponds to different sources of variability. However, most previous studies either aggregate travel time as a whole or examine only one component in isolation. This limits the ability to pinpoint what's actually causing travel time variations and makes it difficult for agencies to prioritize planning and operational

changes. Understanding which component contributes most to variability on different routes is important for effective resource allocation. Hence, we ask the questions: where do travel time variations come from? Relating to the different analysis levels in the first general question, How much overall travel time variation can be attributed to each trip component, i.e. departure time, inter-stop speed, ridership change, and traffic signal timing change? Does the contribution of each trip component change temporally and spatially (general question 2)?

3.1.6 Specific Question 4: How can detailed trip components and their interactions be incorporated into travel time models to improve the accuracy of performance prediction under different planning scenarios?

Current planning practice often relies on average travel times, which lacks the ability to test potential specific changes in different trip components. By incorporating detailed trip components, i.e. inter-stop travel times, dwell times, and red-light waiting times, travel time models may better reflect real-world decision processes and operational constraints. Hence, we try to see how to combine these details into a simple travel time model, what the model performance is for various planning scenarios compared to the average times, and what the temporal variations are in the model performances (general question 2).

3.1.7 Specific Question 5: What is the potential travel time reliability if passengers were to travel on public transport across the entire service area?

Understanding travel time reliability from passengers' perspective presents both data and methodological challenges, as it introduces additional analysis levels for the analysts. Most passenger-level analyses heavily rely on AFC, which is limited for systems where tap-outs are missing. As a result, agencies typically publish coarse or monthly metrics that may not reflect variability equally for passengers travelling on low-demand OD pairs. These OD pairs may face more reliability issues, but their importance is often overlooked due to lower ridership. This raises potential equity concerns, particularly for transit-dependent populations who rely on these services regardless of quality. Hence, we ask the following questions: what is the potential travel time reliability if passengers were to travel on public transport across the entire service area relating to the general question 1, what are the potential spatial and temporal distribution of travel time reliability in the region relating to the second general question, and what are the potential impacts on different population segments.

3.2 Data Used in This Study

There are several datasets used in the analyses. One of the main datasets is the GTFS published by STM. Next, the GTFS-RT provides updates on the real-time information from the AVL system. Then, the AFC transaction data provided by ARTM describes the time and location of each fare validation, thus giving a general outlook of passenger behaviours. Finally, combining these data sources with other data sources allows more detailed modelling of transit reliability.

The potential additional data sources could be OpenStreetMap extracts, existing transit preferential infrastructures, weather, and traffic data. Other potentially useful datasets include household travel surveys and census data, since they can provide a lot more detailed sociodemographic information. However, they are typically conducted every few years, and they might not update as frequently as the datasets mentioned above.

The following subsections will provide a summary of the available data and data processing steps.

3.2.1 Data in Montréal

This project will focus on the island of Montreal, where STM provides most of the transit services as well as the data used in my studies. STM provides metro and bus services on the island of Montreal, with some additional services in Laval and Longueuil. It operates 4 metro lines, 68 stations, and 71 km of revenue metro tracks. There are 219 lines, 1846 buses, and 439 km of bus lanes for its bus services as of spring 2025. Overall, the agency provides around 20 to 25 million trips per month.

The systems relevant to this project are the following. STM publishes its schedules in GTFS format, and we can then match the schedule to the real-time information provided by the agency. STM also implemented a real-time information system for its bus network called iBUS. It allows passengers to track the bus locations and view updated arrival times. Fare transactions can be done via OPUS cards, magnetic stripe cards, and cash. The AFC system will record each fare transaction for analysis.

3.2.2 Data Sources

GTFS

GTFS is a common file format to communicate planned transit services from transit agencies to third-party developers. It uses a flat-file format to describe planned services, such as the

planned routes, arrival times, and fare structures of different services. The files are similar to a relational database, except that it is stored in a series of comma-delimited text files.

A GTFS feed defines and describes the agency's common information that is unlikely to change frequently, such as their lines and schedules. The required information by the GTFS standard is agency, stop, routes, trips, stop times for each trip, stops, and service dates information. This project will extensively use all of the required information as well as optional information provided, such as the shape table for each service pattern and the service frequency table for metro timetables.

Since these files are updated infrequently, they are very easy to archive. So far, I have obtained all STM's GTFS feed since November 2013.

GTFS-RT

GTFS-RT is an extension to the GTFS standard. It allows agencies to provide updates on the current status of the network that might deviate from the planned service, such as arrival time updates, vehicle GPS locations, as well as service alerts to passengers. STM uses all three components of the GTFS-RT standard, namely trip updates, vehicle positions, and alerts.

Trip updates provide real-time updates on a trip's status. It can either show future updates, such as estimated arrival time for future stops or show past events, such as stop times for previous stops which allows passengers to determine whether the vehicle has passed. The data from STM provides a lot more detailed information compared to what is required by GTFS-RT contract. The data show the actual passing time if there is no passenger activity and the dwell time if there are door-opening activities. Stop cancellations due to roadworks and detours are also reflected here, although we will remove them from our studies.

Vehicle position shows real-time information about a vehicle's geographical location in latitudes and longitudes. The data from STM updates when there is a stop arrival event (when the driver opens the door at the stop), departure event (when the bus departs from the stop), or every 20 seconds if the bus is in transit between two stops.

Service Alerts give passengers information about service disruptions and future planned disruptions. STM mainly uses this feed for advertising stop relocation and detours.

GTFS-RT data collection started May 1st, 2021 for STM. The feed is archived every 10 seconds according to the developer agreement. For the potential future expansion of the analysis, the collection started on June 15, 2021, for Société de transport de Laval (STL) and EXO services. The data is archived every 30 seconds according to developer agreements.

EXO also publishes GTFS-RT data for Réseau de transport de Longueuil (RTL), therefore there is no separate data collection for RTL.

In the GTFS-RT standard, there is an experimental entity, Trip Modifications, which identifies a group of scheduled trips in the static GTFS trips affected by detours. When there is a detour, STM reprograms the driver consoles, iBus, to better assist drivers in their wayfinding. However, STM does not publish this feed using the GTFS standard, yet the reprogrammed routes and temporary stops are shown on the website, and reflected in the GTFS-RT feeds. Due to the lack of associated Trip Modifications or updated GTFS static information, we will remove the detours from our studies.

AFC

An AFC system is the automated version of fare collection systems. It typically involves a fare media, such as a smart card or a magnetic stripe card, which stores passenger ticketing data. In Montreal, buses are equipped with fareboxes and AFC readers for passengers to pay for their rides. In the metro, the validation is typically done at the turnstiles. Typically, the AFC transaction records will contain the validation time, route, trip direction, the fare product, card number, and the farebox ID, which contains the vehicle number or the station number.

Given the vehicle number, route-direction, and real-time vehicle locations, we could infer the corresponding GTFS trip. By adding passenger validation time, we could then match the boarding stops for each fare validation.

To deduce the destination of each passenger trip, we use the methodology developed by He and Trépanier (2015). To summarize, the algorithm tries to find the destination by assuming the passenger alights at a stop close to the next boarding location (e.g. transfers and trip chains) and the passenger behaves similarly according to the historical observations.

The estimated OD data will then be used to aggregate the number of boardings and alightings for each bus trip at each stop. The data will also help identify passenger experiences, such as travel times and delays, for analysis.

OpenStreetMap

OpenStreetMap (OpenStreetMap contributors, 2023) is an open map database contributed by volunteers around the world. It started with the road traces only, but later added other data types as well, such as land uses. There are three geometric data types, node (a point of interest), way (a line or an area marked by nodes), and relation (the relationship between

data points). Each geographical object is stored as one row in the data, and each object contains various tags (i.e. key-value pairs) that describe it. In the analyses here, we will focus on the roadway attributes, traffic signals, and land use attributes, since they contain more precise location and more detailed attributes compared to the city's open data.

Census

The only census attribute used in my analyses is the population count from the 2021 census. Here, I used the dissemination block data, which is the smallest geographic area for which population data are published. The dissemination blocks are also bounded by the road network, which is helpful when matching for bus routes. As for the socioeconomic variables, we will use the smallest geographic area with available data, the census tract data, to match with transit routes.

Ride-check Application Data

Due to various limitations in the dataset and to check the validity of the data, a ride-check application was developed to collect ground-truth data. Figure 3.2 shows a screenshot of the app.

On the top of the screen, users could select a specific trip we are checking. The selection will query for the most recent trip information according to the planned GTFS or the modified trip data from STM. The route, stops, planned times, planned speeds, and the estimated traffic signal states are displayed on the map view. Then, using the GPS location, we can calculate the corresponding delay and the current stop status, which are displayed at the top of the screen.

At the bottom, the buttons correspond to the events that users would like to check, i.e. the dwell and traffic signal information. Additional messages are available, such as cars blocking bus stops affecting travel times. However, these additional data are not used for the analyses. Finally, these calculated data and ride events are uploaded to the database for modelling and comparison with the data provided by STM.

3.2.3 Data Processing and Database Design

Each aforementioned data source is stored in separate databases. The raw inputs remain unmodified to account for future references or additional modifications. Then, additional layers of abstraction allow data integration among these data sources. This section describes

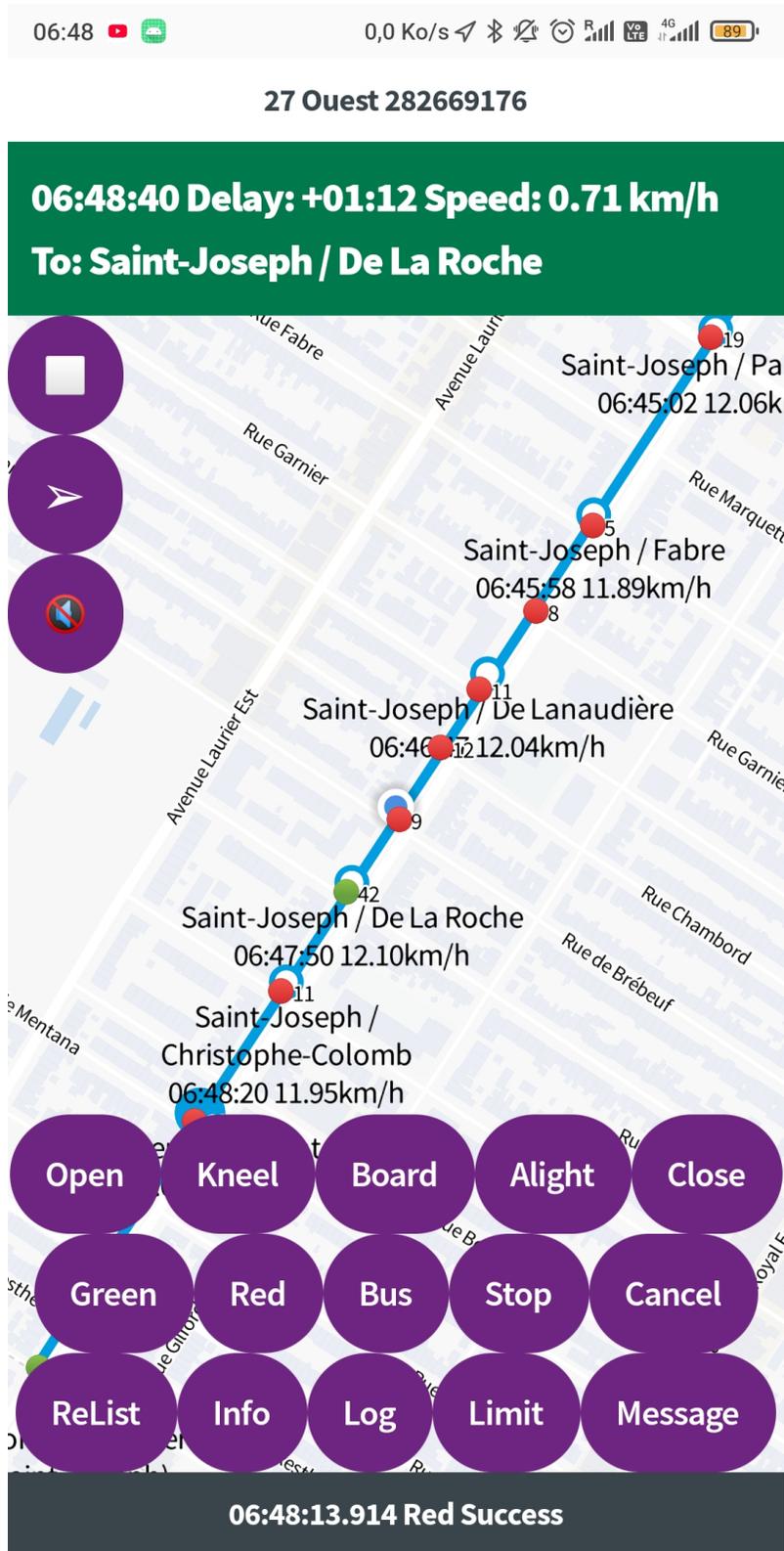


Figure 3.2 Screenshot of the Ride-Check Application

each database and how they are connected, and the overview of data processing steps is provided in Figure 3.3.

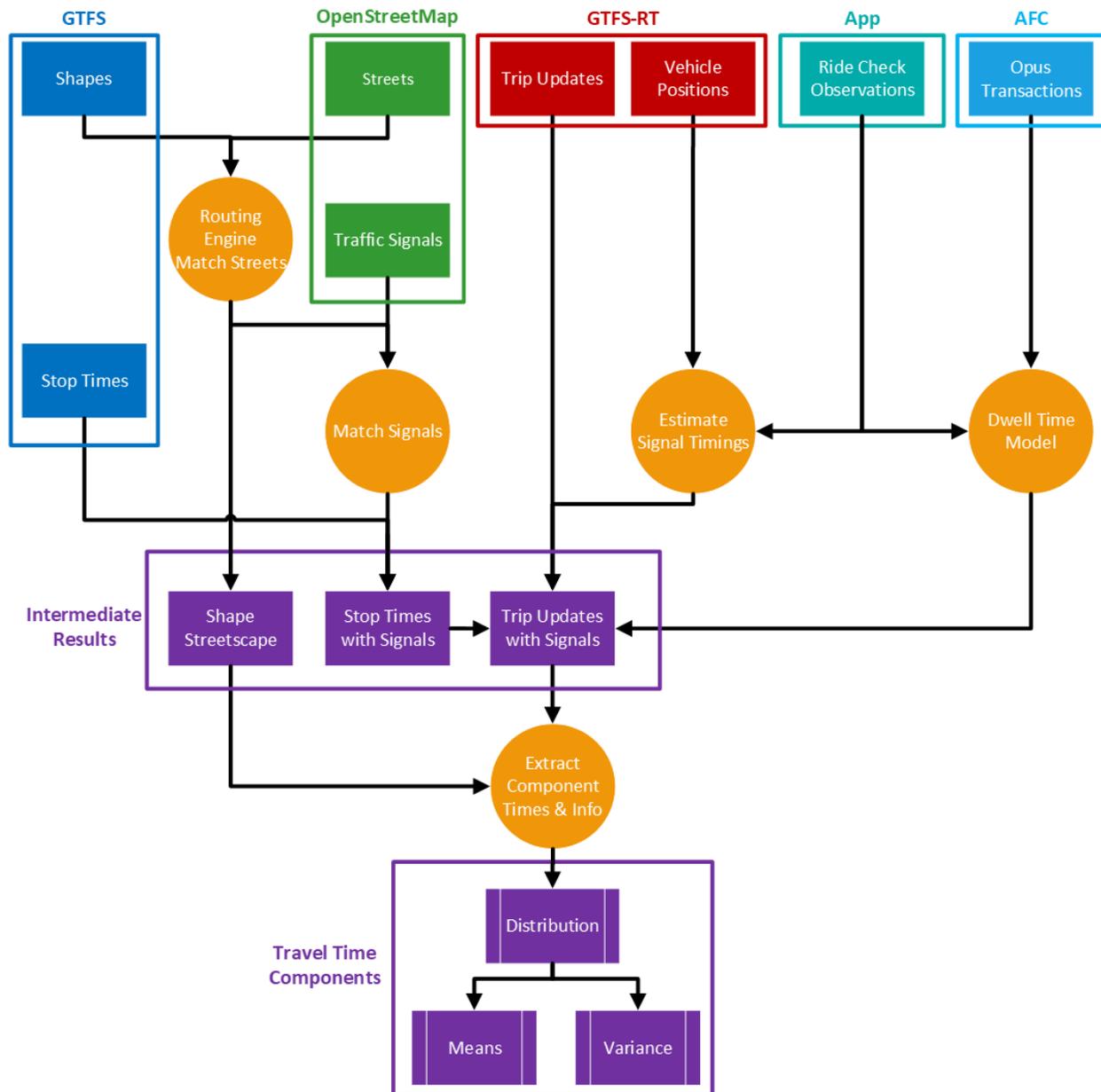


Figure 3.3 Data Processing Overview

GTFS

The GTFS data is created to store the raw input data provided by the agency, as well as the additional layers of abstraction needed for the analyses.

Raw Inputs

The agency table is required by GTFS, which contains the agency or agencies whose services are described in the feed.

Calendar and calendar_dates tables describe when and which service will operate, which is also required by GTFS standards. The calendar dataset allows agencies to distinguish weekday and weekend services by specifying a weekly schedule. Calendar Dates, on the other hand, define specific dates when a service will or will not occur. This is typically used to communicate holidays and other services that do not fit into a weekly schedule.

The routes table defines transit routes, which are a group of trips that are advertised as one single service. A route can have various service patterns, such as inbound and outbound directions and potential short turns.

The trips table lists the trips on each route. Each trip is defined as a trip from a given origin terminal to a given destination terminal at a specific time of the day.

The stops table is a collection of all the stops where transit vehicles pick up and drop off passengers or the entrances of station buildings.

The stop times table corresponds to each trip, and specifies a time when a transit vehicle is scheduled to arrive or depart at a specific stop. Note that, most transit schedules are created at timepoint levels, but GTFS requires the stop level arrival or departure times. This can cause a mismatch between the agency and passenger expectations.

The shapes table details the geographical paths a transit vehicle will follow for a collection of trips on the same service pattern.

The frequency table defines headway-based services, which are used to provide the timetable for the metro. However, by the June service change in 2022, this file was discontinued by STM, and the metro timetables will be communicated using the fixed schedule format, the same format as the bus service, in the stop times table.

Layer of Abstraction: Route Type

Since bus services are very flexible, there may be many service patterns for a given route-direction, such as short turns, branches, and deviations. To ensure comparability between trips, an additional layer of abstraction is created, route type, which only contains one service pattern with the same planned route and stops. Since service schedules are changed more frequently than service patterns, this abstraction will reduce the amount of repetitive calculations.

Trips are then associated with their corresponding route types. The additional calculated

data, such as street and population attributes, are also associated with each route type, which can then be applied to all trips of a given route type. An additional table associates the route types with the matched streets using the route shape, and another table includes the sequence of stops and traffic signals of the route type.

To match the planned route to streets, the routing engine uses the shapes, sequence of stops, and OpenStreetMap data as input. The routing results provide the OpenStreetMap IDs of the streets used, which can then be used to match detailed street information.

The detailed street information allows matching the traffic signal locations and the immediate surrounding land uses that the buses will encounter using OpenStreetMap data. The traffic signal locations are then added to the stop lists so that we can further decompose and analyze the travel times in greater detail.

The results are stored by route type, the added layer of abstraction. Three tables were attached to include detailed routing information with streetscape characteristics for each type, stop or traffic signal information for each type, and the surrounding land use for each type. These tables will then help us to analyze the attributes at various analysis levels.

GTFS-RT

The GTFS-RT data is created to store the raw input data provided by the agency, as well as the additional layers of abstraction needed for the analyses.

Raw Inputs

The trip updates provided by STM include the list of stops associated with a trip. A given trip can either be planned, unplanned (i.e. replacement bus), or cancelled. The list also includes all scheduled stops from the planned GTFS feeds and any temporary stops not in the planned service at the end of the list, which is not the standard practice given the GTFS-RT contract. Each stop has an arrival and departure time associated with it, and the status of each stop (e.g. cancelled due to detour).

The vehicle positions provide a given vehicle's location at a given instant, the time, latitude, and longitude. In addition, each position report is associated with a trip (either planned or unplanned), the vehicle load at the given instant using a categorical variable, the vehicle's relationship to the next stop. This feed is updated whenever there is a stop arrival or departure event, or every 20 seconds when the vehicle is travelling between two stops.

These two data sources are then stored as is into two tables for historical reference. Processed data are then stored in different tables for analysis.

Additional Processing

Since the real-time feeds include detours programmed by STM and STM does not publish detours in the GTFS-RT Trip Modification format, they don't necessarily match exactly to the planned GTFS. Therefore, the stop lists are compared to the planned list and the vehicle positions are matched to the planned route. If there are deviations found, the data are marked as noise and removed from the study. More detailed detour effects can be left for future research.

In addition, since the goal is to provide more detailed analyses between stops or traffic signals that are not near any stop, more matching is needed to add the traffic signal locations to the real-time feeds.

First, the updated stop lists, including the traffic signal locations, are calculated from the previous section to reconstruct the trip updates and the vehicle positions. Then, based on the vehicle positions and the original trip updates, the exact stop times observed are inserted into the new table. Next, the vehicle locations, which include details between two stops, allow better retrospective estimations of the arrival and departure times at each traffic signal. Thus, each stop or traffic signal will have arrival and departure times associated with it. Finally, the arrival and departure times at each signal will be used to infer signal settings, which will be discussed in detail later in the paper.

The newly expanded stop times lists enable the calculation of the travel times between stops or signals, dwell times, delays, and signal states where applicable. Then, these data can be aggregated to remove noise, and then be used for analyses at any level we defined, namely, route, timepoint, stop, and inter-stop levels.

AFC

The AFC data is created to store the raw input data provided by the agency, as well as the additional layers of abstraction needed for the analyses.

Raw Inputs

In Montreal, buses are equipped with fareboxes and AFC readers for passengers to pay for their rides. In the metro, the validation is typically done at the turnstiles. Typically, the AFC transaction records will contain the validation time, route, trip direction, the fare product, card ID, and the farebox ID, which contains the vehicle number or the station number.

Additional Processing

Given that vehicle location data are more precise, and can be used to estimate the boarding

and alighting stops. Using the vehicle number, route-direction, and real-time vehicle locations, it is possible to infer the corresponding GTFS trip. By adding passenger validation time, we could then match the boarding stops for each fare validation.

He and Trépanier (2015) developed a methodology to deduce the destination of each unlinked passenger trip. To summarize, the algorithm tries to find the destination by assuming the passenger alights at a stop close to the next boarding location (e.g. transfers and trip chains) and the passenger behaves similarly according to the historical observations.

The estimated OD stops will then be used to calculate unlinked passenger delays. Then, the sequence of a passenger trip can be used to estimate the trip chain.

Finally, these data will then be used to aggregate the number of boardings and alightings for each bus trip at each stop. The data will also help identify passenger experiences, such as travel times and delays for other analysis.

OpenStreetMap

The OpenStreetMap data is created to store the raw input data provided by the agency, as well as the additional layers of abstraction needed for the analyses.

Raw Inputs

The few attributes that we are interested in are the traffic signals, land use, and roadway information.

The traffic signals are stored as points along a roadway. They are typically mapped on the stop lines at each intersection and tagged with the direction affected. To facilitate the analyses, the attributes are expanded to include the estimated traffic signal programmes by including the day type, time-of-the-day, the offset, green length, and cycle length for each programme.

The roadways and their attributes are stored as a line. They are typically mapped at the center line of the road. The tags of interest contain the road classifications, one-way streets, number of lanes, speed limits, and bus lane hours where applicable. This information has been cross-checked with Montréal's open data to ensure validity.

The land use attributes are stored as a polygon, bounded by the area of the given land use. For each land use polygon, the tag will show the exact land use attributes, such as residential, commercial, industrial, etc. Similarly, the information has been verified with the city's open data.

Additional Processing

To use for routing purposes, the OpenStreetMap street network has to be modified slightly. The raw data does not require street lines to be split at the intersections, as long as they have the same attributes. Since buses can turn off a street in the middle of a continuous line, it is necessary to identify the intersections using the point connected to multiple streets. Then, the street line is split at the intersections for routing purposes. Then, the updated street line IDs ensure that each segment of the same street can be identified in the routing results and matched to the original dataset.

Census

The Census data is created to store the raw input data provided by the agency, as well as the additional layers of abstraction needed for the analyses.

Raw Inputs

The two most important data from the census are the census block shapes and the data files for each census attribute. Since the focus is on the population data, there are only two tables to store.

Additional Processing

Since the shape files use a different coordinate system compared to the GTFS standard, we will convert the shape files into the same coordinate system. Then, using the spatial buffer, we can match the census blocks near all bus stops. The blocks associated with each stop are considered the catchment area, and are then aggregated to calculate the population density associated with the routes or stops.

Ride-Check Data

The ride-check data is created to store the raw input data provided by the agency, as well as the additional layers of abstraction needed for the analyses.

Raw Inputs

The raw inputs of the ride-check data include the GPS tracks and trip events recorded during ride-checking on the bus.

The GPS location is recorded every second to allow detailed comparison to the less frequent AVL data. Each location record contains the recorded time, which is synchronized to Canada's official time provided by the National Research Council within around 50 millisecond accuracy. Each location record also includes the elevation, bearing, and speed information.

The ride event table also contains the timestamp for each recorded event, the trip information, the location where the event occurred, the event type, and the associated stop status of the event.

Additional Processing

Since the goal is to verify the low-frequency AVL data and the estimated traffic signal data using detailed high-frequency ground-truth observations, the GPS locations, stop arrival and departure data are matched to the GTFS-RT feeds. The passenger boarding and alighting numbers are matched to the AFC data to check the estimated vehicle load and OD flow. The traffic signal observations are compared to the estimated data to verify the errors. Finally, any other changes in speed limits, traffic signal configurations, and land uses were recorded and updated in OpenStreetMap.

3.2.4 Developed Information Tools

To support the aforementioned data integration and processing steps across different operating systems, a custom data processing pipeline in Java is developed with very limited dependency on other libraries. This system is designed to be platform-independent and can be used on Windows and Linux environments. It handles tasks such as parsing, validating, and upgrading GTFS and GTFS-RT feeds to include more details such as traffic signals. It also improves the estimated vehicle arrival and departure times at stops and traffic signals by matching vehicle positions to planned routes and storing them in the database for analysis. The Java-based architecture ensures compatibility across devices, allows specific optimizations given the data structure, and enables parallelism, making it easier to deploy in different systems while analyzing these large transit datasets.

To support both the research and practical applications of this project, a suite of online information tools has been developed that integrates multiple transit data sources, GTFS, GTFS-RT, and AFC, into a unified, interactive online platform. Although not directly used in the research results, these tools are designed to make planned and delivered transit data searchable and easy to understand when developing data processing algorithms, implementing analysis methods, and demonstrating an example in articles. They facilitate visualizing the planned and delivered transit services via tables, graphs, and maps, which enables users to explore transit operations, performance, and reliability in both real-time and retrospectively.

The online tools are organized into three main categories: Schedule Search Tools, Real-Time Tools, and Analysis Tools. The Schedule Search Tools include features such as the Route Schedule, which displays scheduled trips for a given route and day in a list format;

the Trip Schedule, which shows scheduled times at regular stops and interpolated times at traffic signals as operational stops for individual trips; and Block Assignments, which group trips into operational vehicle blocks to help understand relationships between trips. These tools help users understand planned service patterns and operational structures, as well as search through archived historical service plans. In addition, limited editing functions were implemented to modify the planned GTFS data, since the agency does not publish long-term detours or stop relocations using GTFS standards. These editing tools, although not comprehensive, allow users to modify planned routes on a map based on OpenStreetMap data, insert new stops, and adjust existing timetables. Again, since the agency does not publish these service changes, these edits are for the ride-check app (Figure 3.2) and validating real-time data only and are not included in the analyses.

The Real-Time Tools provide dynamic insights into current and historical vehicle movements. The Current Vehicle Positions map displays all tracking STM vehicles with estimated delay levels, while the Trip Tracking History and Vehicle Tracking History tables allow users to trace which vehicles operated specific trips or vice versa. The Trip Playback feature combines tabular, graph, and map-based views to show detailed arrival and departure times at regular and operational stops as well as vehicle trajectories for any recorded trip on a selected day. Finally, maps and graphs are generated automatically to validate and discover potential differences in the real-time observations using the high-resolution ride-check observations. These tools are particularly useful for analyzing anomalies when debugging and validating analysis results.

The Analyses section shows various existing performance indicators used by many agencies, such as the On-Time Performance Tracker, which visualizes stop-level punctuality (similar to the official definition at STM from 1 minute early to 3 minutes late) across all bus services since May 2021. Similarly, various maps show segments or routes with consistent delays or slow travel conditions. Custom graph and map creation tools are also available when summarizing research results by simply querying the result tables in the database. This tool enables longitudinal performance monitoring and supports data-driven evaluations of service reliability, allowing users to identify potential issues related to specific times and segments, discover potential trends in transit travel time performance, as well as create graphs and maps programmatically.

Technically, the online tools are implemented using a full-stack web development approach and hosted on a Linux-based server. The frontend is built with HTML, CSS, and JavaScript, providing a responsive and interactive user interface that supports dynamic visualizations for transit data. The backend is developed in PHP, which handles database connections, user

queries, and dynamic content generation.

3.3 General Framework, Methodology, and Expected Results

The projects aim to improve our understanding of transit travel times by addressing the questions identified in the literature above. More specifically, this project will be analyzing transit travel times at 4 different analysis scales, i.e. inter-stop, stop to stop, timepoint to timepoint, and service pattern levels. In this section, I will demonstrate the general research framework, introduce the methodologies, and present the expected results to answer the research questions.

3.3.1 Overview of Research Framework

The overall framework of this research is shown in Figure 3.4. In this figure, the data used in this project are coloured blue. The process to integrate various data sources is marked in green and explained in more detail in the data section. The five contributions are identified at the bottom in yellow, i.e. analyzing the overall statistical distribution, the mean, the variance of transit travel times, improving travel time estimations, and the potential impact on passenger experiences.

To reiterate, the goal is to use as many existing automated data sources as possible, since manual data collection or additional equipment would cost more money for the agency. Hence, there are five main existing data sources, GTFS, GTFS-RT, AFC, OpenStreetMap, and the census data. However, to verify the automatically collected data, as well as to mitigate the missing dwell time information from these data sources, ride-check observation data are added to the project.

Then, the data sources are matched following the descriptions in the data section later. In general, a routing engine matches the planned routes to the street network from OpenStreetMap. This would allow the expansion of the planned data to include traffic signal locations, streetscape, land use, and population information. Using the combined stop and traffic signal locations, the planned route and the corresponding data are split into the 4 analysis levels, again, inter-stop, stop to stop, timepoint to timepoint, and service pattern levels. For the analyses on passenger travel times, the journey times are generally categorized into overall wait times and overall in-vehicle times since a passenger journey can involve multiple routes and multiple wait times and in-vehicle times. These analysis levels are illustrated in Figure 3.5. Then, vehicle locations are used to match the planned times with the actual times observed for each level. Finally, the integrated data can be used for the analyses.

Finally, the integrated data allow more detailed analyses of transit travel times. Here, the research aims to improve our understanding of the statistical distribution, the mean, and the variance of transit travel times, which corresponds to the first, second, and third papers submitted. Paper 4 tries to integrate the knowledge gained from previous analyses into a travel time model, aiming to improve the travel time estimations for planning purposes. Paper 5 attempts to evaluate the potential passenger impacts from travel time variations. The more detailed methodology descriptions of each paper can be found in the subsections below, and the complete papers can be found in the chapters that follow.

3.3.2 Contribution 1: Improve understanding of travel time distributions

The first paper focuses on this question and was selected for presentation at World Conference on Transport Research (WCTR) 2023 and was published in Transportation Research Procedia on January 9, 2025.

To answer the first general question on system-wide multi-level analysis, the second general question on spatial-temporal variations, and the first specific question on travel time distributions, the first paper looks at the overall distribution of bus travel times at all four analysis levels. Most previous studies on mixture travel times only focus on a few routes or one transit corridor, and they did not provide additional information on the factors related to the presence of these mixture distributions. Travel times following mixture distributions create more difficulties for agencies to provide consistent service, since multimodal distributions imply several different underlying travel time distributions at once. Thus, it is important to better understand the segments with travel times following mixture distributions.

We aim to provide these insights on the underlying travel conditions for agencies to consider when improving their existing services. Identifying potential factors that relate to these mixture travel time distributions can help agencies to focus their resources when implementing or evaluating transit preferential measures like dedicated lanes and signal priorities. Hence, the first paper identifies the segments that follows mixture distributions, the spatial and temporal variations of the segments following mixture distributions, as well as the operational and environmental factors that could contribute to these distributions.

More specifically, there are two steps for this research. The first step involves using the Hartigans' Dip Test (Hartigan and Hartigan, 1985) to identify mixture transit travel time distributions from the integrated dataset at the 4 different analysis levels. We then aggregate the results to provide descriptive statistics related to mixture travel time distributions. The second part of the research is to identify environmental and operational factors that could relate to the presence of mixture transit travel time distributions. To achieve this, we used

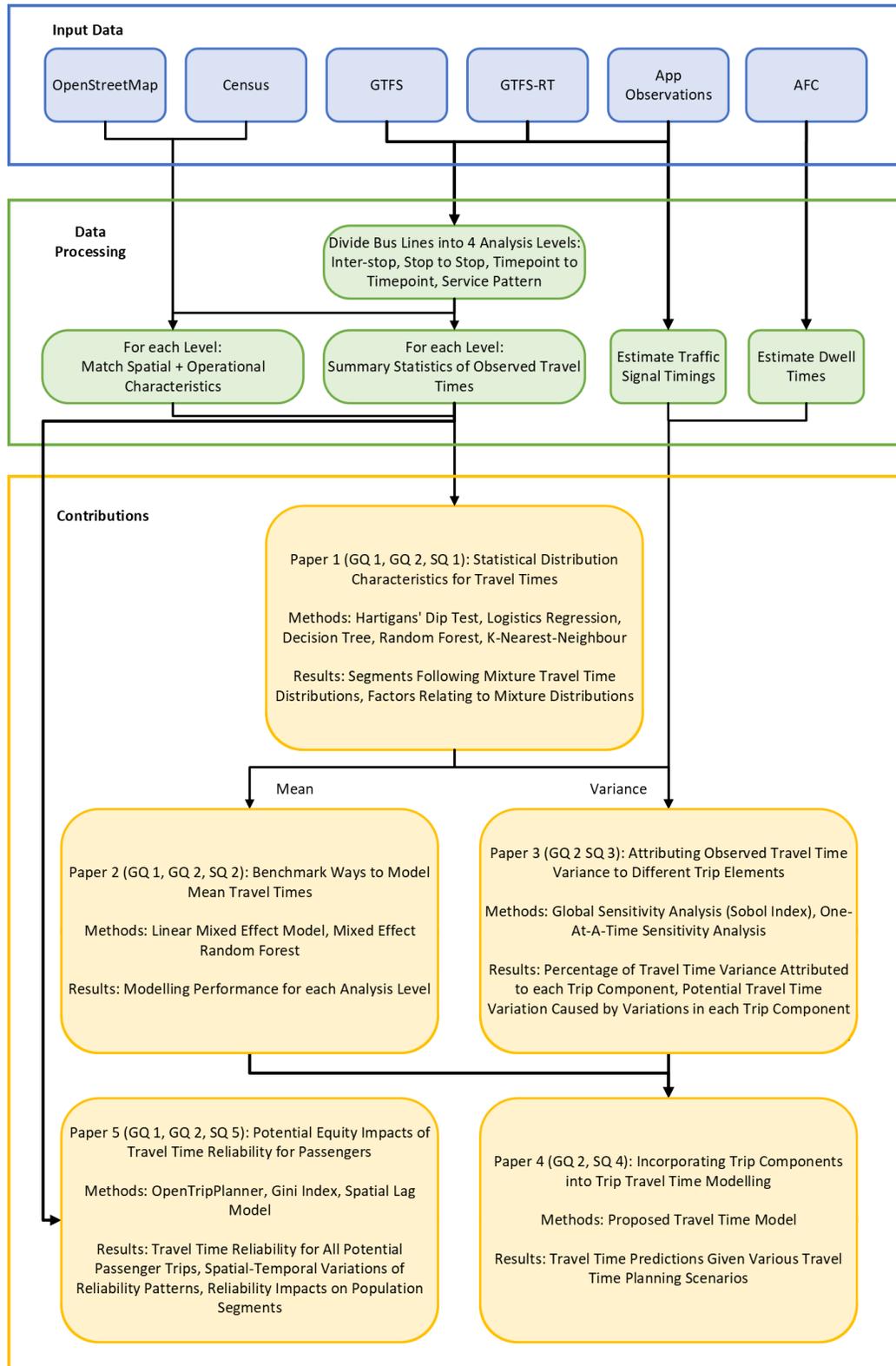


Figure 3.4 General Research Framework

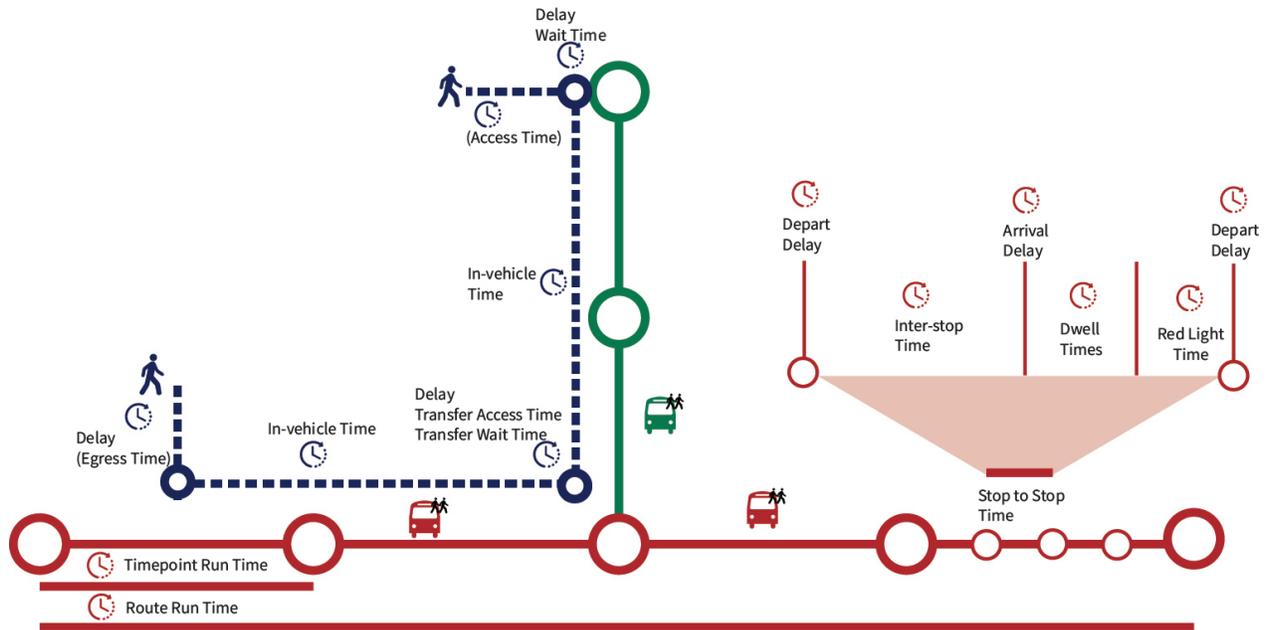


Figure 3.5 Illustration of Analysis Levels

various classification models to model whether the transit travel times on a given segment follow the mixture distribution.

The descriptive results would show the presence of mixture transit travel time distributions at three analysis levels, stop-to-stop, timepoint-to-timepoint, and route levels. It is also expected to show the spatial variations through maps and temporal variations (time of the day and time of the week) through graphs, which would help provide more detailed information on when and where mixture distributions occur to focus the attention of transit service planners. The classification model results would show the most effective approaches when modelling the distribution patterns. The model coefficients would also help reveal the influence of different environmental and operational variables on the statistical distribution of transit travel times, such as passenger demand variations, time of the day, service frequencies, traffic lights, segment length, and land use variables.

The result would also relate to the first general research question by modelling systemwide multi-level observations, and it is expected to demonstrate the mixture travel time distribution patterns across stop, timepoint, and route levels. For the second general research question, the result is expected to identify mixture distributions concentrated in specific spatial zones and temporal periods, such as corridors with traffic signals or during peak hours.

3.3.3 Contribution 2: Improve understanding of average travel time modelling

The second paper corresponds to this specific question and was selected for presentations at Canadian Transportation Research Forum (CTRF) 2023 and Association québécoise des transports (AQTr) 2023, for a poster session at Transportation Research Board (TRB) 2024, and the revised paper was submitted for publication to Public Transport on March 15, 2025.

Aiming to better understand the average transit travel time models and to help transit planners better choose their future modelling approaches, the second paper attempts to answer the first general question on system-wide multi-level analysis, the second general question on spatial-temporal variations. Previous literature mostly focused on travel time predictions for specific routes, related variables speeds and paces were typically used for infrastructure planning, and various error measures were used to evaluate the models.

We propose to benchmark average travel time models using different related dependent variables and various analysis levels, which relates to the second specific question on modelling mean travel times. Using the integrated data, we first summarize the observed travel times, operating speeds, operating pace, and the corresponding environmental and operational characteristics at the 4 analysis levels. Then, these data are used as inputs for modelling.

Next, we will benchmark the model performance using two common planning scenarios. One is to expand or modify the services into a new route. In this case, agencies may not have historical data available at all. To account for this scenario, we reserve 10% of the segments from our dataset for testing. Another common scenario is to expand the service hours on an existing line. Therefore, agencies may not have historical data for a specific time of the day on a given segment. Thus, we reserve another 10% of the remaining data with various time-of-the-day values for testing. Finally, we compare and evaluate the model results using a few commonly used error measures, including relative and absolute measures, to discover potential biases in the model results.

The model coefficients would show the influence of each temporal, spatial, and operational variable on the average travel time, speeds, and paces, as well as how much variance can be explained by unmeasured sources of variance that affect specific segments. The model performances would show which model is more effective under which planning application. For error measures, these aggregated and disaggregated error results would reveal potential biases of these various modelling approaches, which allow us to discuss the advantages and disadvantages of different models, analysis levels, and evaluation criteria, so that we can make recommendations to other researchers and planners when deciding future modelling approaches, as well as to transit agencies for their future planning and operations.

The results would also relate to the first general question by estimating models for different analysis levels using system-wide observations. The models would also include the spatial and temporal variables, and the model coefficients would show the impacts of different spatial and temporal variables, which relate to the second general question.

3.3.4 Contribution 3: Improve understanding of travel time variance

The third paper relates to this question and was selected for presentations at TransitData 2024, TRB 2025, CTRF 2025, and AQTr 2025, and the paper revision was accepted for publication on Transport Research Record on September 25, 2025.

The third paper aims to combine the knowledge gained from previous research efforts on the statistical distributions and the mean travel times, and focuses on the variance aspect (the third specific question) of travel times. It also relates to the first general question by adding more details to the existing analysis levels. Since most previous literature did not focus on attributing travel time variations to specific trip components, we propose to decompose the trip travel times into the sum of a sequence of times according to TCQSM, namely, inter-stop times, dwell times, and red light waiting times. To reconstruct the arrival times at each stop and determine the traffic signal states, we also need to include departure times in our analyses. Thus, there are four types of variables used in this project.

We then used the integrated dataset to obtain these detailed times. Due to the lack of door-closing times, we estimate a simple dwell time model at the stop level using ride-check observation data and apply the model to the fare transaction data. The traffic light settings are also estimated following the methodology proposed by Fayazi et al. (2015) due to data availability issues. For the departure times and inter-stop times, we can obtain them directly from the GTFS-RT data. Since there may be multiple traffic signals between two bus stops, we further add to the existing analysis levels by considering traffic signals as an operational stop, better isolating each component of travel times.

Then, the decomposed travel times are grouped by service patterns and three different time-of-the-day and used for sensitivity analyses, which relates to the second general question. Here, we focus on both the variance-based analysis (Saltelli et al., 2008), which is a global method that quantifies the proportion of variance that can be attributed to each variable, while having the ability to handle non-linear and non-additive variables. We will also include one-at-a-time sensitivity analysis to better illustrate the non-linearity between the input variables and the overall travel times. We will present different results and problems faced by different routes to illustrate the different needs for improvements.

The global sensitivity analysis provides two indices, first-order and total-order indices, which shows the proportion of variation that can be attributed to each trip component without and with interaction effects. These results will help rank the issues along each route analyzed and help planners prioritize their limited resources to address the most pressing issues given the financial constraints facing the agencies. The one-at-a-time analysis results could also help planners to quantify potential travel time variations resulting from various changes in the network, thus allowing planners to react more quickly to address future issues.

Due to data limitations, the results in the third paper would only include select routes from the system, but more complete systemwide results can be found in paper 4 which help contribute to the system-wide analysis (general question 1). The spatial and temporal analysis could only focus on times of the day due to data limitation, which would show the changes in sensitivity indices for each trip component throughout the day.

3.3.5 Contribution 4: Improve understanding of travel time modelling using detailed trip components

The fourth paper answers this question and is sent for review to be presented at WCTR 2026 on October 31, 2025, and hopefully to be published in the conference proceedings subsequently.

Building on the methodology from the previous papers, we aim to test the possibility of using the travel time components in the previous analyses in a travel time planning and estimation context, which relates to the added analysis level in the previous papers as well as the first general question on multi-level analysis. We consider how to combine these details into a simple travel time model and what the model performance is for various planning scenarios compared to the average times, since most literature focused on more aggregated levels of analysis.

For the travel time model, we proposed to reconstruct the bus trajectory by incorporating various decisions that a driver needs to make, which relates to the fourth specific question. First, the vehicle departs; add the inter-stop time to the next stop or signal; add the estimated dwell time if applicable; add the red time required if applicable; then travel to the next stop or signal. The decisions repeat until the end of the route, which would result in an estimated trip trajectory at each stop and an estimated overall travel time for the trip.

To test the model performance, we first compare the estimated travel times to the observed values to check if the model would generate similar values to the observations. We then estimate travel times for future service periods using data from the January 2024 service

period. Then, we compare the estimated values to the observed trip times to check the performance and the potential improvements of our travel time model.

The result would illustrate the system-wide impacts of red light waiting times on the overall travel times to highlight the needs to include them in travel time modelling, as there are fewer empirical evaluations in the literature. Finally, we evaluate the performance of our proposed simple scenario-based travel time model using detailed trip components by comparing the travel time predictions against the previously observed averages typically used in planning. The results would show the improvements in travel time estimations under various changes in planning scenarios, such as departure time, ridership, and traffic signal timing changes. The results would also show the potential scenarios where the model estimations are not as good as the averages.

3.3.6 Contribution 5: Improve understanding of the equity of experienced reliability of potential passengers

The fifth paper addresses this question and is sent for review to be presented at WCTR 2026 on October 31, 2025, and hopefully to be published in the conference proceedings subsequently.

To improve our understanding of the spatial and temporal patterns of transit reliability (general question 2), as well as attempt to recognize potential equity issues (specific question 5) related to system-wide service reliability (general question 1), we aim to remove the passenger demand from the equation by analyzing the transit reliability from a given location to all possible destination zones as if passengers were to make these trips regularly. This would help provide more detailed information to planners on potentially unreliable origin-destination pairs as well as when and where in the network the reliability issues occur.

Trip generations and attractions are mostly analyzed at neighbourhood levels, which may involve several different transit route combinations, thus, we need to limit analysis area to ensure passengers in the given area has similar route choices when analyzing the question on the potential passenger experiences. We propose to split the entire service area into hexagon grids that are roughly equal to a 10-minute walk within each zone, which allows us to simulate trips across the entire service area. Then, we calculate potential passenger trip plans from a given zone to all other zones, given the scheduled times using a routing engine, OpenTripPlanner. Using the fastest trip as a reference trip with extra multipliers added for waiting and transfer times, we simulated the passenger journey and the resulting travel times that would have been delivered to the passengers using 3 months of archived vehicle location data. Then, we calculated the reliability metrics for all weekday services within our study

period, which are then stored in an origin-destination matrix for analysis. To illustrate the results, we will show spatial and temporal variations for one origin and for all origins.

The results would include the travel time reliability for simulated system-wide passenger trips between all origin and destination zones for various departure times, which also contribute to the first and second general questions. The resulting reliability indices would include both discrete measures similar to on-time performance as well as continuous measures similar to buffer times, showing the likelihood of on-time arrival and the variability of overall travel times. Furthermore, passenger travel times would be categorized into wait times and in-vehicle times, where the reliability indices would further illustrate the source of variations in passenger travel times.

Finally, we aim to provide a short equity analysis, which aims to evaluate the fairness of societal outcomes of a given system. Transit service reliability issues can either be considered as an outcome of the system, i.e. the delayed trips for passengers, but it also impacts the passengers in their decisions whether to avoid unreliable transit by switching modes. Although there is no standard definition for fair societal outcomes in evaluating transit service reliability, we will take similar principles from Martens et al. (2012) based on equitable accessibility measures. We consider it as maximizing the reliability performance everywhere while meeting a minimum threshold for all and constraining the disparity in reliability performance. Hence, we will calculate the number of reliable routes for each zone using the minimum thresholds similar to those defined by the agency, and evaluate the evenness or spread of disparities in the reliability outcomes using the Gini Index (Ceriani and Verme, 2012) as well as the potential societal impacts on different population segments using spatial regression models and a few general demographic attributes.

CHAPTER 4 ARTICLE 1: SYSTEMWIDE VARIATIONS AND FACTORS AFFECTING MIXTURE TRANSIT TRAVEL TIME DISTRIBUTIONS

Yuxuan Wang^{acde}, Catherine Morency^{acde}, Martin Trépanier^{bce}

^a Department of Civil, Geological and Mining Engineering, Polytechnique Montréal, Montréal, Québec, Canada

^b Department of Mathematics and Industrial Engineering, Polytechnique Montréal, Montréal, Québec, Canada

^c Interuniversity Research Centre on Enterprise Networks, Logistics, and Transportation, Montréal, Québec, Canada

^d Canada Research Chair in Mobility, Montréal, Québec, Canada

^e Canada Research Chair in Transportation Transformation, Montréal, Québec, Canada

Published on 9 January 2025 in

Transportation Research Procedia, Volume 82, 2025, Pages 460-480

DOI: <https://doi.org/10.1016/j.trpro.2024.12.055>

Abstract

Understanding transit service reliability is essential for agencies to improve their operations and passenger experiences. Transit travel times that follow mixture distributions would create an additional layer of uncertainty when studying transit reliability. This paper tries to identify segments where transit travel times follow mixture distributions at different analysis levels, namely stop pair level, route timepoint level, and service pattern level. We then identify potential factors related to them. Hartigans' Dip Test is applied to archived transit vehicle location data from Montreal to explore the presence of mixture distributions. The results contain mixture distributions at three analysis levels, and the proportion of mixture distributions varies temporally and spatially. Then we test several classification models to identify the potential factors that affect transit travel time distributions, where we found demand variations, traffic lights, service frequency, and segment lengths have a larger effect on the results. The findings will help transit planners to later pinpoint the issues causing transit travel time variations on each segment, then create strategies to reduce the transit travel time variations thus improving the reliability of our transit system.

Keywords: Transit Reliability; Transit Travel Time; Transit Service Planning; Transit Operation

4.1 Introduction

Many transit agencies and planners have been trying to improve their service reliability, which is important for both agency operations and passenger experiences. From the agencies' perspective, unreliable services affect the operator and vehicle scheduling. Planners need to add additional schedule padding times or layover times to account for the unreliability, which increases the operating costs (Danaher et al., 2020). Missing a layover due to unreliability could increase operators' dissatisfaction causing operator retention issues and delays propagating further downstream (Danaher et al., 2020). From the passengers' perspective, unreliable services are less attractive, therefore affecting the passengers' satisfaction and mode choices (Carrel et al., 2013a). Some surveys pointed out that passengers value service reliability more than service frequency and travel times (Balcombe et al., 2004; Daskalakis and Stathopoulos, 2008; Chen et al., 2003; Perk et al., 2008).

There are many works looking into the variability of transit travel times, which is important in many aspects of transit planning and operations such as scheduling (Trépanier et al., 2009), vehicle arrival time prediction (Chen et al., 2023), transfer coordinations (Kieu et al., 2017), and passenger experience evaluation (Ma et al., 2014). Several pieces of literature have observed that transit travel time distributions might follow a mixture distribution (Kieu et al., 2015; Ma et al., 2014; Wang, 2020). A mixture travel time distribution can be considered as several different underlying travel conditions happening with different probabilities. In other words, we first select one travel condition from a collection of travel conditions with different probabilities, then we observe the travel time value from the selected travel condition random variable.

Travel times following mixture distributions create more difficulties for agencies to provide consistent service, since multimodal distributions imply several different underlying travel time distributions at once. They can also cause inconsistencies in passenger experiences or when predicting the arrival times for passenger information. Thus, it is important to better understand the segments with travel times following mixture distributions.

This paper aims to build upon the previous literature and provide additional analyses on mixture travel time distributions for agencies to consider when planning and improving their services. We first propose to identify the systemwide presence of mixture transit travel time distributions, which would help provide more detailed information on when and where mixture distributions occur to focus the attention of transit service planners.

Here, we use 1.5 years of archived transit vehicle location data from Montreal, Canada as a case study to test for mixture travel time distributions. We focus on three analysis

levels, namely, (1) stop pair level, (2) route timepoint level, and (3) service pattern level. Analyzing the travel time of a given service pattern, i.e. the travel time from one terminal to another, is important for agencies to schedule vehicle and operator duties. Since there may be multiple service patterns on the same route, such as short turns, and they may have different numbers of stops and ridership characteristics, we will analyze them separately. Most transit agencies in North America develop and evaluate their schedules at timepoint levels, and then interpolate the timepoint to timepoint travel times to produce the stop level arrival times. Thus, we aim to understand the distribution of travel times at both timepoint and stop levels, to help agencies improve their schedule accuracy and service consistency.

Additionally, we provide new insights for agencies to consider when improving their existing services, by identifying potential factors that relate to these mixture travel time distributions. We use additional built environment and operational attributes as inputs to fit several classification models. The classification models will help agencies to target their resources when implementing or evaluating transit preferential measures like dedicated lanes and signal priorities. By identifying the related factors, agencies can then "nudge" the higher travel time conditions to a lower condition, thus reducing the number of underlying travel conditions and improving travel times for passengers.

The structure of this paper is roughly summarized as the following. We provide additional research contexts in section 2. The research framework, data sources, and overview of our methodology are provided in section 3. Section 4 shows the descriptive statistics and results on mixture transit travel time distributions. Section 5 details the factors related to mixture travel time distributions. Finally, we conclude this article in section 6.

4.2 Literature Review

In this section, we quickly identify the definitions for transit travel time reliability, the commonly used analysis scales, as well as the previous studies specifically on transit travel time distribution.

A quick literature review shows that there is no standard definition of transit travel time reliability since it relates to multiple points of view, such as passenger and agency points of view (National Academies of Sciences, Engineering, and Medicine and others, 2013).

Although there is no standard definition of transit reliability, most agencies and literature define transit reliability as the variability that affects the agency's or passenger's decision-making (Abkowitz et al., 1978), ability to adhere to the planned service schedule or headway (Turnquist et al., 1980), and have consistent service deliveries (Strathman et al., 1999). The

terms consistent and variation used by these definitions imply that service performance should be measured over a period of time. Thus, the distribution of travel times can help create an overview of transit travel time reliability.

4.2.1 Analysis scales and factors of transit travel time reliability

To evaluate transit services over a period of time, we need to first identify the analysis time scales. These temporal scales are summarized by Noland and Polak (2002) as the following:

- Vehicle-to-vehicle variability is the travel time variation between different trips operating on the same segment at the same time of the day period. Traffic signals and operator preferences typically cause the variations.
- Period-to-period variability is the travel time variation between different trips operating on the same segment at different times of the day periods. The variations are typically caused by the variation of congestion levels, demand variations, weather conditions, as well as the potential occurrence of traffic incidents.
- Day-to-day variation is the travel time variation of the same trip made on different days. The variations are typically caused by the variations of traffic levels, demand, operator preferences, weather conditions, and the potential occurrence of traffic incidents.

As for spatial scales, the literature and agencies typically focus on the following levels for travel time variations, namely the route, trip, timepoint, and stop levels (Danaher et al., 2020). For route and trip levels, agencies typically use travel times and travel time variations in scheduling, improving the on-time performance, meeting layover times laid out in the labor agreements, and assigning vehicle and operator resources (Danaher et al., 2020).

In practice, even though transit schedules are typically revised multiple times a year, travel times are revised less frequently at the timepoint level and then interpolated to the stop level for passenger information in North America (Coleman et al., 2018). Schedule controls, such as holdings, are also typically done at the timepoints. Whereas, in Europe, some agencies would consider every stop as a timepoint to avoid large schedule deviations (Muller and Furth, 2001).

Since the goal of our study is to help agencies to improve the existing services as well as passenger experience, we will focus mainly on the vehicle-to-vehicle and day-to-day variation of the same stop pair segment, timepoint pair segment, and service pattern levels.

4.2.2 Studies on transit travel time distributions

To reiterate, transit travel time distributions can provide key statistics for analyzing service reliability, exploring causes of service unreliability, and adjusting service schedules (Mazloumi et al., 2010). The travel time distributions are also an important input for transit simulation models, such as transfer simulations (Kieu et al., 2017).

Previous literature has attempted to fit transit travel times to different parametric distributions. Abkowitz and Engelstein (1984) showed that travel times follow a symmetrical distribution. However, Mazloumi et al. (2010) then showed that travel times could also follow skewed distributions, such as lognormal or gamma distributions.

More recent literature (Ma et al., 2014; Kieu et al., 2015; Wang, 2020) has also observed mixture distributions. Kieu et al. (2015) fitted corridor level travel times using several parametric distribution models. Skewed distributions, such as log-normal distribution, fit the best, while some limited number of trips show bimodal mixture distribution. Similarly, Ma et al. (2014) fitted travel times on several route segments and showed that the multimodal distribution of bus running times can be estimated by using the mixture model.

Since mixture distribution represents several unimodal distributions weighted with different probabilities, the possible explanation provided by Ma et al. (2014) is that there exist mixed travel time patterns related to underlying traffic conditions. For bus travel time, the multimodal distribution can be related to different operation characteristics, such as free-flow operation patterns in the early morning and congested operation patterns during peak hours. However, they did not discuss more detailed causes.

One thing to note here is that, as data becomes more precise and the analysis scale becomes smaller, we can now observe more detailed variations in transit travel times. The more aggregated view provided by previous literature could hide the detailed variations that operators and passengers experience. There is still some research needed to determine the shapes of travel time distributions, and the cause for mixture distributions in more detail.

4.3 Research Framework and Methodology

As mentioned in the literature review section, most of the cited literature only focuses on a few routes or one transit corridor. In addition, previous research have observed the presence of mixture distributions in transit travel times. However, they did not provide additional information on the factors related to the presence of these mixture distributions.

We propose to address these limitations by producing systemwide summary statistics re-

garding the presence of mixture travel time distributions at different analysis levels to help agencies better understand transit reliability. Again, our goal is to help agencies focus their resources and improve their existing services and develop more precise schedules for passengers and operations in the future.

In this section, we provide a high-level overview of our analysis, data source, and methodology, so that other agencies or researchers could build on top of this article. Here, we use the data from Montreal as a case study to illustrate these analysis steps. The overall research framework and steps are shown in Figure 4.1 and explained here after.

More specifically, the first step of the research involves merging the GTFS and GTFS Real Time data to extract travel times at different levels, namely the stop pair level, route time-point pair level, and route service pattern level. Then, we use the Hartigans' Dip Test (Hartigan and Hartigan, 1985) to identify mixture transit travel time distributions at these different analysis scales. We then aggregate the results to provide descriptive statistics related to mixture travel time distributions.

The second part of the research is to identify factors that relate to the presence of mixture transit travel time distributions. We aim to provide more information regarding the shape of transit travel time distributions, which could be used by future research to further infer the cause of transit unreliability. To achieve this, we propose to develop a classification model on the shape of transit travel time distribution, using the available open datasets on Openstreetmap and the City's open data. More information on the datasets is detailed below.

4.3.1 Data Sources

This paper focuses on the bus system on the island of Montréal, Québec, Canada, operated by Société de transport de Montréal (STM). The bus system involves 224 bus lines, 2006 buses, approximately 17,000 bus trips on average weekdays, and 439 kilometers of bus lanes. STM publishes its schedules in GTFS format, which is a standard data format used to distribute transit timetable data to various passenger trip planning softwares. To improve passenger information, STM also implemented a real time information system iBus, which records the bus arrival and departure times. Using 1.5 years of archived data from STM, from May 1st, 2021 to October 24, 2022, we calculate stop pair, route timepoint pair, and route service pattern travel times for STM's bus services.

More specifically, we aggregate all stop pair travel time data without considering the route differences, since their built environment and operational characteristics are the same. How-

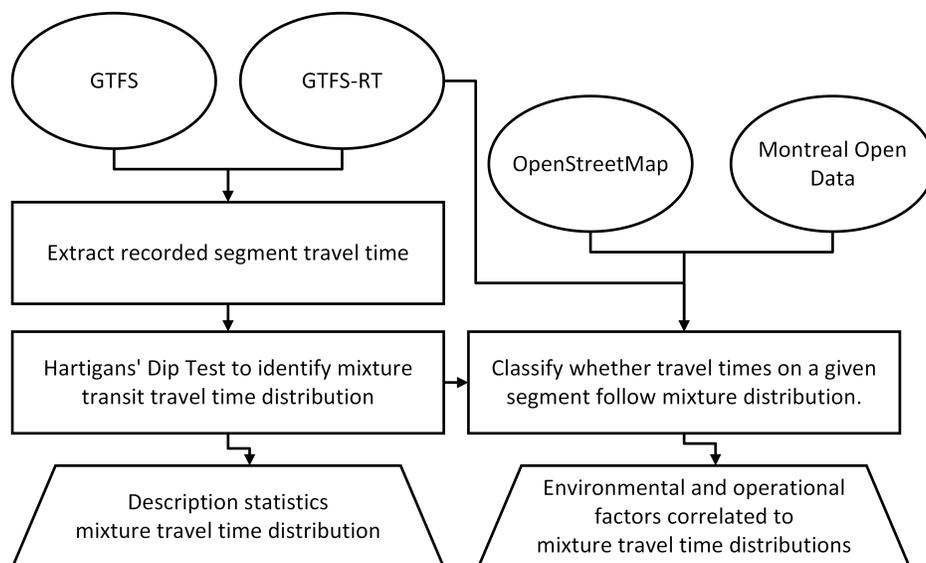


Figure 4.1 Research Framework

ever, for a given timepoint pair or a terminus pair, the built environment and operational characteristics could differ. There might be a local route operating alongside an express route, or routes may not operate on the same street. Thus, we split the dataset into different routes for higher level analyses.

For modeling, we add factors mostly related to the physical environments and operational characteristics of a given segment. This involves adding street information from OpenStreetMap, which is an open crowdsourced database of the physical attributes of the streets. We mainly use Openstreetmap to calculate the number of turns, speed limits, and street configurations for the classification models. However, due to the crowdsourced nature of Openstreetmap, some sectors have less available data. We use the data provided by the city as a supplement to OpenStreetMap. More detailed variable descriptions are available in section 5.

4.3.2 Hartigans' Dip Test

In this section, we provide a quick overview of the methodology for the first analysis in our framework, identifying the segments with mixture travel time distributions.

To model transit travel times, we can consider the transit travel times on a given segment as a random variable with a continuous distribution following a specific probability density function. However, the underlying distribution might not be directly observable. In this case, we can use samples of observed travel time data to estimate the underlying probability

density function.

Previous studies, such as Ma et al. (2014), uses parametric distributions. Using parametric distributions requires prior knowledge or assumptions regarding the shape of the distributions. In the parametric approach to estimating the probability density function, one has to assume that the data were sampled from a given distribution. However, as previous literature pointed out, there are many parametric distributions that could be used to fit our dataset (Mazloui et al., 2010). Fitting a mixture distribution would also require a specific number of underlying distributions as input. However, these required inputs are not necessarily known by planners or researchers beforehand.

We have also identified a few non-parametric tests for the presence of multimodality so that we don't need the number of underlying distributions as an input. Ameijeiras-Alonso et al. (2019) provided a comparison among several commonly used tests. They found older models, such as the Hartigans' dip test, tend to perform more conservatively than the more modern proposals. The previous research results using various parametric distributions have shown that the number of modes in observed travel times or travel speeds could be as high as five (Du et al., 2017). Since a higher number of underlying distributions relates to different travel conditions, it would make the planning process harder for planners to manage. Thus, we decided to act more conservatively for our methodology. Nevertheless, future research could build on top of our research and examine the cases with a higher number of underlying distributions.

In our study, we selected a commonly used non-parametric test, Hartigans' dip test, to check the multimodality proposed by Hartigan and Hartigan (1985). The idea is to find the "dip" in a given probability density function. If a distribution is a unimodal distribution, the probability density function would increase up until the mode, then decrease. Since there is no upper bound for probability density functions, we can translate probability density functions to cumulative density functions, which are bounded between 0 and 1. The cumulative distribution functions would be convex up until the mode then concave greater than the mode. If a distribution follows a mixture distribution, the probability density function would have a region where the density decreases and then increases. The dips in probability density functions would make the shape of cumulative density functions switch between convex and concave several times.

In short, the dip test constructs a unimodal cumulative density function that minimizes the difference from the empirical cumulative density distribution of the given dataset. Then, we find the dip statistic, which is the largest difference between the estimated constructed cumulative density distribution and the empirical cumulative density distribution. Intuitively,

A larger difference indicates that the data is more likely to have multiple modes.

More specifically, the test constructs the unimodal distribution function by analyzing all possible modal intervals $(x_{LowerBound}, x_{UpperBound})$. Then, for each interval, it calculates the greatest convex minorant in the interval $(-\infty, x_i)$ and the least concave majorant in the interval (x_i, ∞) . The best-fitted unimodal cumulative density function is chosen as the one with the least difference from the empirical cumulative density function. Using this best-fitted unimodal cumulative density function and its difference with the empirical cumulative density function, we can then calculate the probability of dips less than a given threshold, such as a p-value less than 0.05 in our case.

4.3.3 Overview of Classification Methods

Based on the observations from the descriptive statistics shown in later sections, we want to include a few simple and well-known models that cover a wide variety. Since the spatial and temporal variations tend to be non-linear, we hypothesize that non-linear methods would perform better to classify the segments. In addition, we also hypothesize that similar segments would behave similarly. Thus, we decided to include examples of both linear and non-linear models, as well as models based on similarity for comparison. This allows us to determine which method works better in classifying whether travel times on a given segment follow mixture distributions or not. The classification methods included here are Logistics Regression, Classification Tree, Random Forest, and K-Nearest-Neighbour. There are other methods in the aforementioned categories, and evaluating their performances can be left for future works. In this section, we provide a quick summary of the advantages and the prediction process of each method. The description of input variables is detailed in section 5.

1. We choose logistics regression due to its linear nature and its ability to allow us to interpret various coefficients (Hosmer Jr et al., 2013). Logistic regression predicts whether the outcome is true or false, instead of a continuous outcome typically used in linear regression. Logistics regression tries to fit a logistics distribution line between the data points of two categories. The logistic function is bounded between 0 and 1, and it tells us the probability of a positive outcome. If the probability of a positive outcome is greater than 50%, we then classify it as positive. We then select the best-fitted curve based on the maximum likelihood.
2. The reason for us to choose a decision tree to classify the shapes of transit travel time distributions is due to its non-parametric and non-linear nature (Quinlan, 1987). It

is an algorithm that recursively generates a set of binary split rules to best split the dataset into the given categories. The algorithm examines each predictor and possible split values by calculating the Gini's impurity index in our case to determine the best split point. To predict a given data point, the algorithm simply follows these split rules from the root of the tree to the leaves of the tree.

3. We recognize that by combining multiple decision trees, we could potentially achieve better classification results. This method is the random forests method proposed by Ho (1995). The drawback of this method, however, is that it does not allow easy interpretation due to it involving multiple decision trees. The basic idea is to generate multiple decision trees using various subsets of the data sample and various subsets of sample variables. To predict a new data point, the algorithm will follow the split rules of all the decision trees, then use the majority outcome class as the final output.
4. The K-Nearest-Neighbour is a another simple classification method, which predicts the outcome by looking at the known category of similar data points (Cover and Hart, 1967). The distance between each data point is typically calculated as the Euclidean distance. In addition, we also put a weight on each neighbour based on the distance, where the nearest neighbour has more weight than the others. Then, we select the nearest neighbours and use their weighted outcome class as the final output.

4.4 Identify Segments with Mixture Distributions

In this section, we demonstrate the clear presence and some descriptive statistics of mixture travel time distributions. Generally, the results agree with the previous research efforts, and we provide examples of mixture distributions at all three analysis levels. In addition, we also show some spatial and temporal descriptions of the mixture distributions in the following subsections.

4.4.1 Examples of Mixture Distributions in Transit Travel Time

Travel time distribution can be multimodal at multiple aggregation levels. We have observed mixture distributions in all three levels. To reiterate, our focus is on stop pair, route timepoint pair, as well as route service pattern levels. In addition, previous research has mainly focused on mixture distributions with two underlying components. However, we were able to observe mixture distributions with more components.

In this section, we use route 51 as an example. The 11.3-kilometer route mainly runs through



Figure 4.2 Map for Route 51

residential neighborhoods and many schools. It has 50 stops, 48 traffic lights along the way, and it operates on the main transportation corridors. We include a route map (Figure 4.2) to help the readers conceptualize the route. In our examples, we focus mainly on the section near the University of Montreal campus, which is located on the right half of the diagram.

Mixture distributions can be observed when analyzing a service pattern. Figure 4.3 is an example taken from all trips on weekdays eastbound Line 51 between 11 AM and 12 PM. On the route map (Figure 4.2), a given vehicle would travel from left to right. The sample contains 2900 observations. In the figure, we can observe two distinct modes in the distribution, at 2920 seconds, and 3195 seconds. The scheduled travel time, 3060 seconds, corresponds to 55th percentile of observed travel times.

The difference between the two modes is 275 seconds, roughly 4.5 minutes. This large difference again highlights the future research needs to pinpoint the causes for these mixture travel time distributions with more detailed data. There are various day-to-day and trip-to-trip factors that could explain the differences, such as operator preferences, road works, and the number of traffic signals encountered by each trip. Since our data collection started during the pandemic, there could also be period-to-period variations that cause the mixture distribution, such as ridership variation and congestion variation.

Next, we analyze the travel time data at timepoint scale, and we can still observe mixture distributions. Figure 4.4 is an example of a mixture distribution with two underlying distributions. The travel times samples are again aggregated from all trips on weekday eastbound Route 51 between 5 PM and 6 PM between timepoint Decelles / Jean-Brillant and timepoint Station Édouard-Montpetit. On the route map (Figure 4.2), a given vehicle would travel from left to right between the two timepoints near the center of the map.

This timepoint pair segment is roughly 1650 meters long, with 8 stops along the way. It runs through a university area and a residential area. There are five traffic signals between the first and the last stop on this segment.

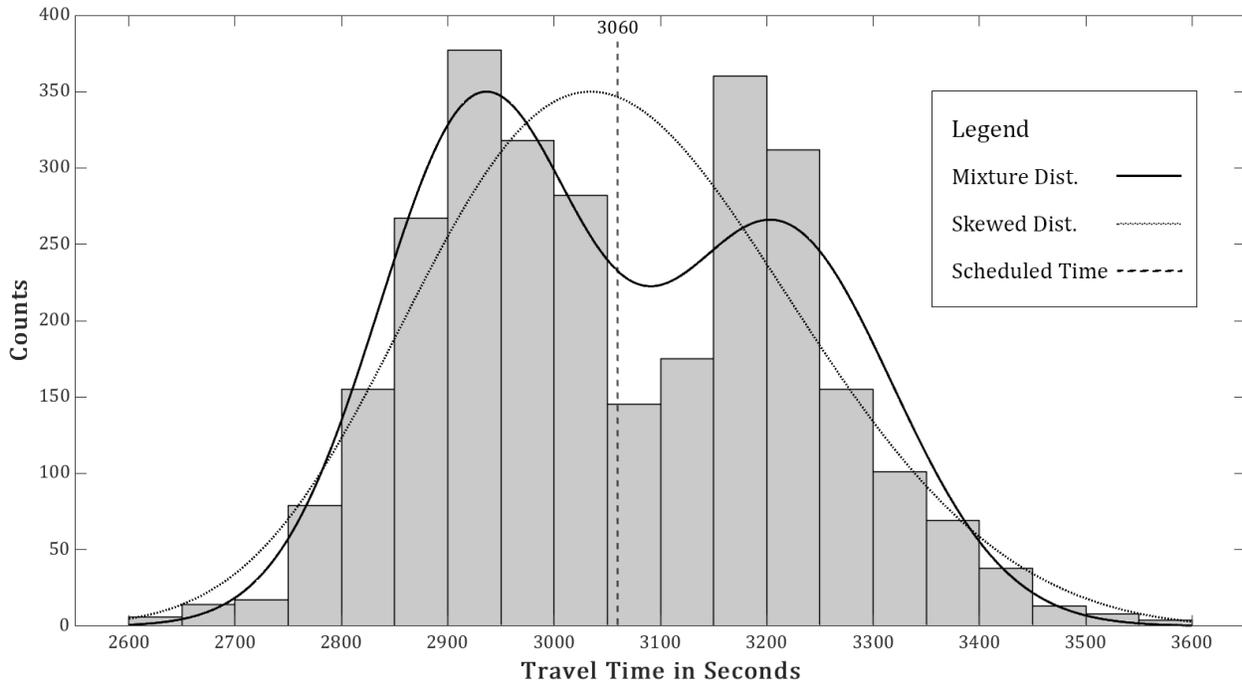


Figure 4.3 Example of Mixture Transit Travel Time Distribution at Service Pattern Level (Route 51, Eastbound, Weekdays 11 AM to 12 PM, STM data between May 1, 2021 and Oct 24, 2022)

The sample contains 2100 observations. In the figure, we can observe two distinct modes in the distribution, at 427 seconds, and 519 seconds. The scheduled travel time, 420 seconds or 7 minutes, corresponds to 28th percentile of observed travel times.

The difference between the two modes is 92 seconds. This is similar to one or two traffic light cycles, which suggests that some trips encounter fewer red lights than others. The potential explanations could be linked traffic lights, where some trips encounter a less ideal cycle. Another potential explanation is due to operator preferences, where some operators might rush through yellow lights and some others might not.

Figure 4.5 is an example of a mixture distribution at stop pair segment level with three modes. The travel times samples are aggregated from all trips on weekday eastbound Route 51 between 5 PM and 6 PM between stop Decelles / Jean-Brillant and stop Decelles / Édouard-Montpetit. On the route map (Figure 4.2), a given vehicle would travel up one stop, from the said timepoint.

This stop segment is roughly 250 meters long. It starts near a university building and runs through a residential area. Since the stops are on the near side, there is one traffic signal right after the first stop, one traffic signal in between two stops, and one traffic signal right

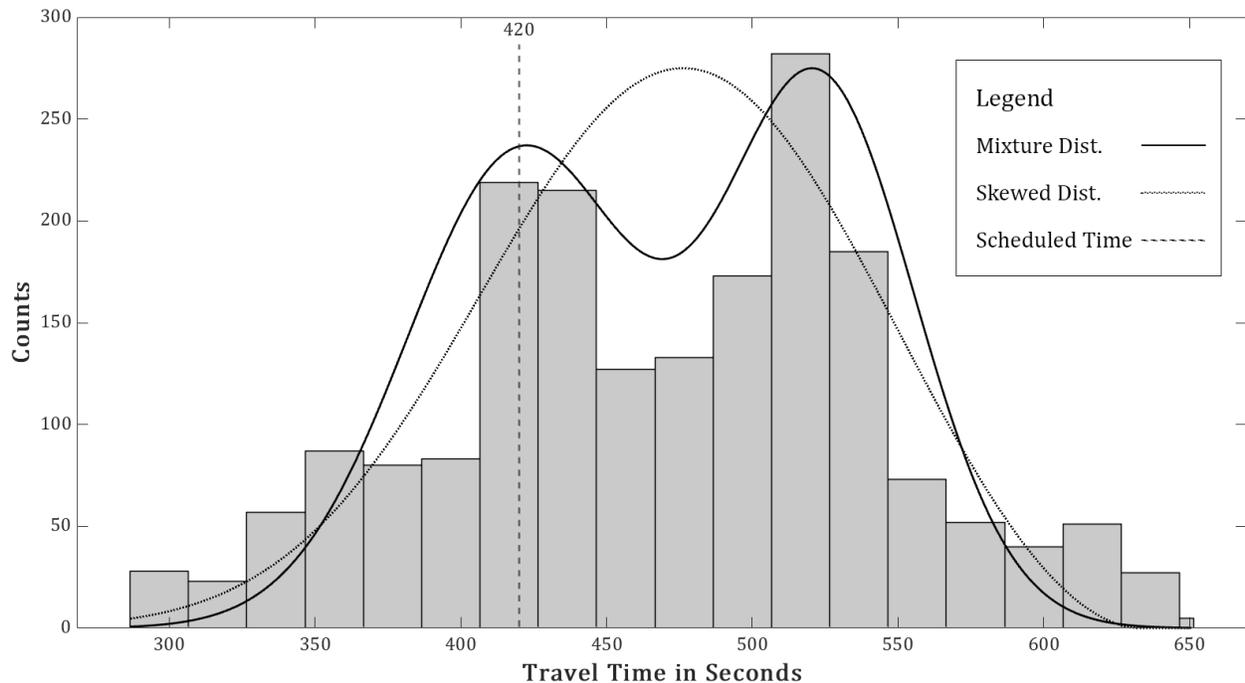


Figure 4.4 Example of Mixture Transit Travel Time Distribution at Timepoint Level (Route 51, Eastbound, from Decelles / Jean-Brillant to Station Édouard-Montpetit, Weekdays Between 5 PM and 6 PM, STM data between May 1, 2021 and Oct 24, 2022)

after the end of this segment.

The sample contains 2100 observations. From the figure, there are three modes, at 37 seconds, 91 seconds, and 161 seconds. The scheduled travel time, 67 seconds, corresponds to 21th percentile in the estimated distribution. In other words, only 21 percent of the trips could travel faster than the interpolated travel time.

The difference between the first two modes is 54 seconds, and the differences between the last two modes are 70 seconds. Unfortunately, without more detailed traffic light setting data or passenger data, it is hard to say what contributed to the differences.

Based on the differences and the surrounding environment, an educated guess is that buses could have issues clearing these signals, assuming passenger arrivals follow Poisson distribution due to the high service frequency with headways less than 10 minutes. The number of underlying travel time distributions could relate to the traffic signal states. We will explore more potential factors in section 5.

As we can observe from this section, there is a lot of potentials for agencies to reduce the travel time variance for passengers, increase service reliability, as well as allowing operators to have more layover times to rest and prepare for the next trip. If the travel time savings is

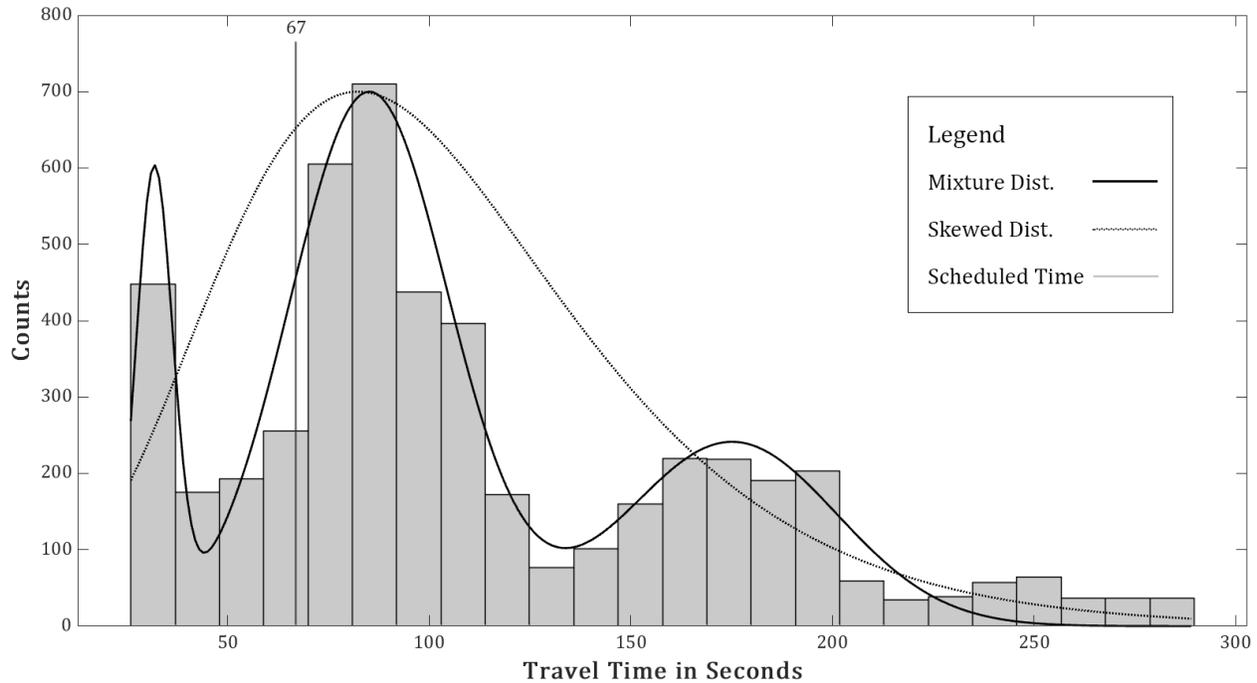


Figure 4.5 Example of Mixture Transit Travel Time Distribution at Stop Level (Route 51, Eastbound, , from Decelles / Jean-Brillant to Decelles / Édouard-Montpetit, Weekdays Between 5 PM and 6 PM, STM data between May 1, 2021 and Oct 24, 2022)

large enough, the agency could potentially save one unit from this line, thus saving operating costs. Again, there are further research needs to identify the causes for the slower travel condition and identify methods to "nudge" the travel times under the slower travel condition into the faster distribution components.

4.4.2 Temporal Statistics

To better understand the temporal variations of segments with travel times following mixture distributions, we aggregated the results by the time of the day and the day of the week attributes. For the time of the day, we aggregated the results by the hour. For the day of the week, we aggregated the results by Weekday, Saturday, and Sunday services, which corresponds to the published schedule. Due to the low service frequency and the small sample of night buses, we excluded them from the figure, since a small change could lead to a large variation in terms of percentages.

The service pattern level variations are shown in Figure 4.6. To reiterate, there could be multiple service patterns on each route that has different demand or operational characteristics, such as short-turns. Thus, we aggregated the data based on individual service patterns

which would help the agencies plan their services. Here, each data point corresponds to the percent of service pattern travel times that follow mixture distributions. For example, in the figure, the first red data point shows that there are 46% service pattern travel times that follow mixture distributions on weekdays between 4 and 5 AM.

In the figure, we can see a fluctuation in the percentage of segments that follow mixture distributions. Interestingly, the mixture distributions contribute to a less proportion during weekday rush hours. Although the number of service pattern travel times that follow mixture distributions is higher, there are more unimodal rush hour-only service patterns being added. This is somewhat counterintuitive since peak hours might have more variation in travel conditions. One hypothesis could be that the large variance in travel times makes the modes harder to separate. Another hypothesis could be that the additional service patterns added during rush hours are more consistent than the all-day service patterns.

The variation in mixture travel time distributions between route timepoint pairs is shown in Figure 4.7. Again, there could be multiple routes running between two timepoints, such as local and express routes. Since they may have different numbers of stops, we aggregate them separately. The first red data point shows that there are 37% route timepoint pair travel times that follow mixture distributions on weekdays between 4 and 5 AM.

Here, we can observe a more stable trend compared to the route level, possibly due to a larger sample size, since there are multiple timepoints on one route. However, there are differences compared to the service pattern travel times. Mixture distributions contribute to a larger proportion in the evening peak, but a smaller proportion during the morning peak.

Similarly, the stop pair level variations are shown in Figure 4.8. Since there is usually only one way to travel between two given stops, we are not separating the data by route here. The first red data point shows that there are 33% stop pair travel times that follow mixture distributions on weekdays between 4 and 5 AM.

We can observe that stop pair level mixture distributions are also quite stable. However, unlike, the timepoint level and route level observations, there are two small increases in the percentage of mixture distributions around peak hours and lunch hours. Similar differences regarding the peak hours at different analysis levels can also be observed for weekend services. The proportion of mixture distributions gradually increase until the afternoon peak for the stop pair level, whereas the other two levels stay relatively stable. It could be interesting to further investigate the observed differences during rush hours between these three analysis levels.

For all three levels, Saturday and Sunday travel times have roughly half the percentage that

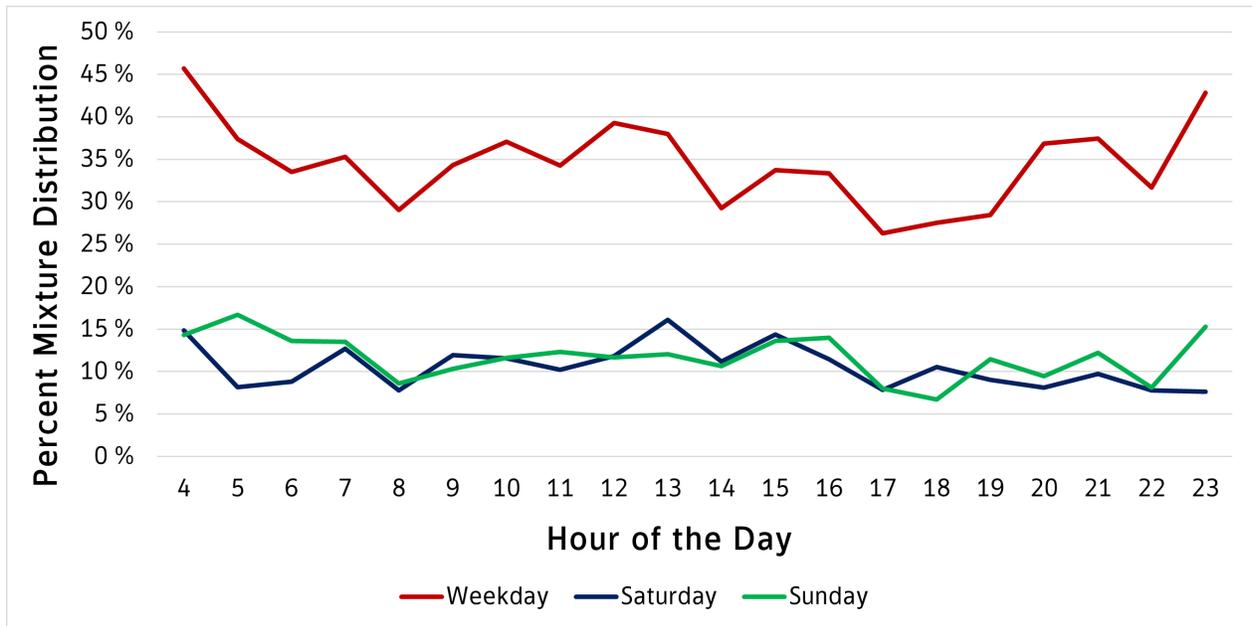


Figure 4.6 Temporal variations of Mixture Transit Travel Time Distribution at Service Pattern Level (STM data between May 1, 2021 and Oct 24, 2022)

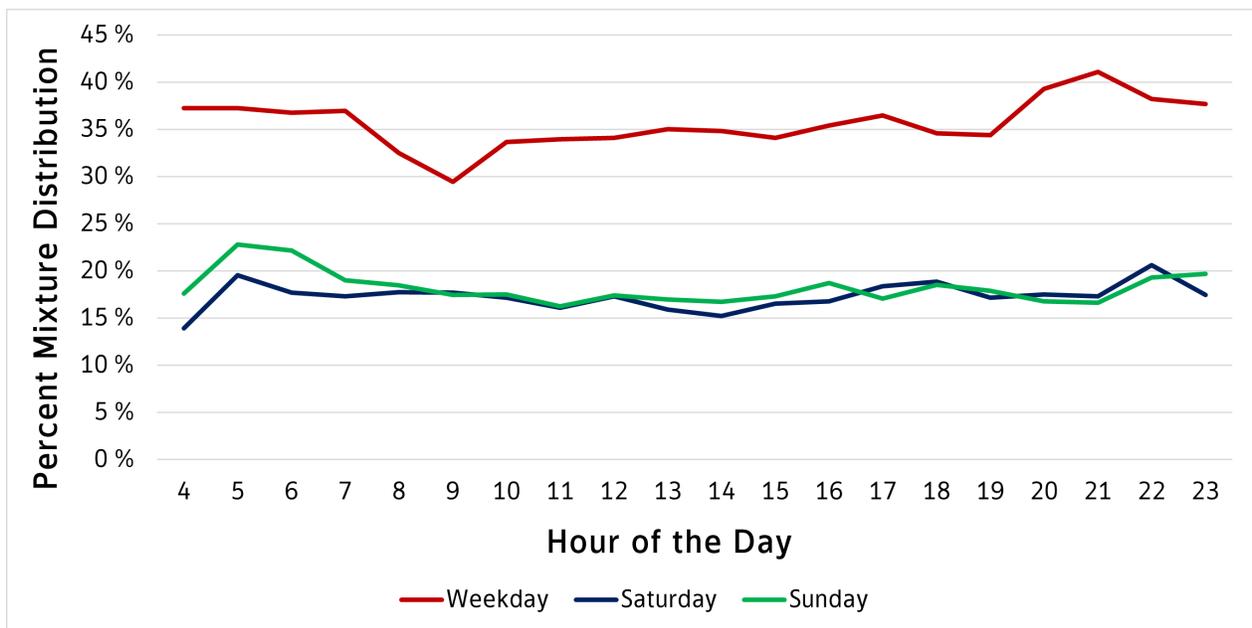


Figure 4.7 Temporal variations of Mixture Transit Travel Time Distribution at Route Time-point Level (STM data between May 1, 2021 and Oct 24, 2022)

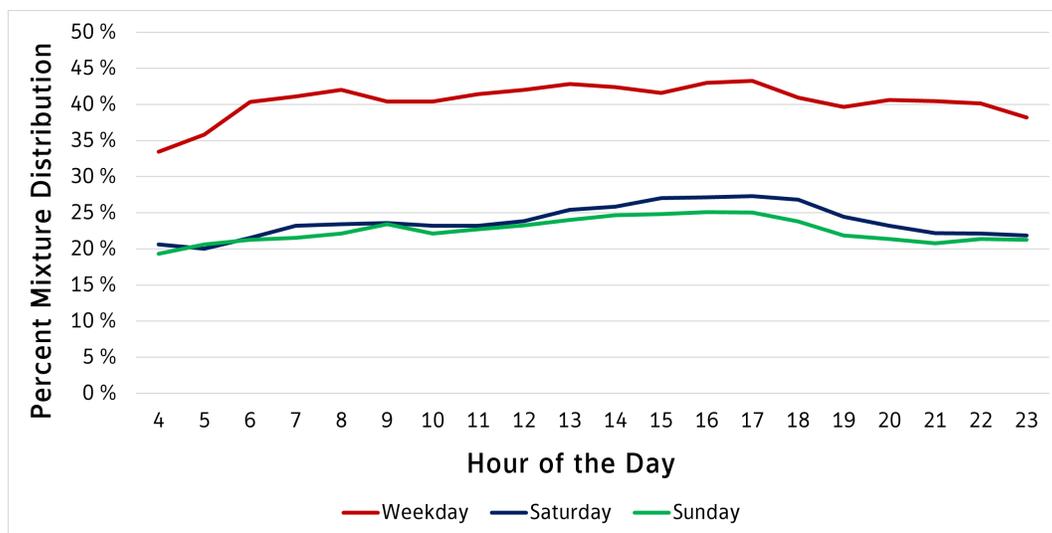


Figure 4.8 Temporal variations of Mixture Transit Travel Time Distribution at Stop Pairs Level (STM data between May 1, 2021 and Oct 24, 2022)

follows mixture distributions compared to weekday services. This could be due to the smaller data sample on each segment. Based on the types of services, weekday services have five times more data compared to weekend services by definition. Weekend services may only have 70 to 100 data points in a given segment. Therefore, there may not be enough data points to increase the power of our statistical tests. This behavior could also be due to the general reduction in travel demand and congestion during the weekends.

4.4.3 Spatial Distribution

After examining the proportion of segments following mixture travel times distributions, we mapped these segments on a map to determine where the mixture distributions exist in the system.

To demonstrate the concept, Figure 4.9 shows the travel time distribution of each scheduled segment on Weekdays at 7 AM, where green segments do not follow mixture distribution and red segments follow mixture distribution. Due to the limitation in picture sizes, it is hard to demonstrate the inbound and outbound segments here on the map. In reality, planners could zoom in using their geographic information systems to get a better view.

To help the readers of this article, we include a more aggregated view by neighbourhoods (Figure 4.10) to demonstrate the geographical variation. Here, we show the percentage of segments with mixture distributions in a given neighbourhood. More specifically, the red neighbourhoods correspond to more than 35% of the segments following mixture distributions,

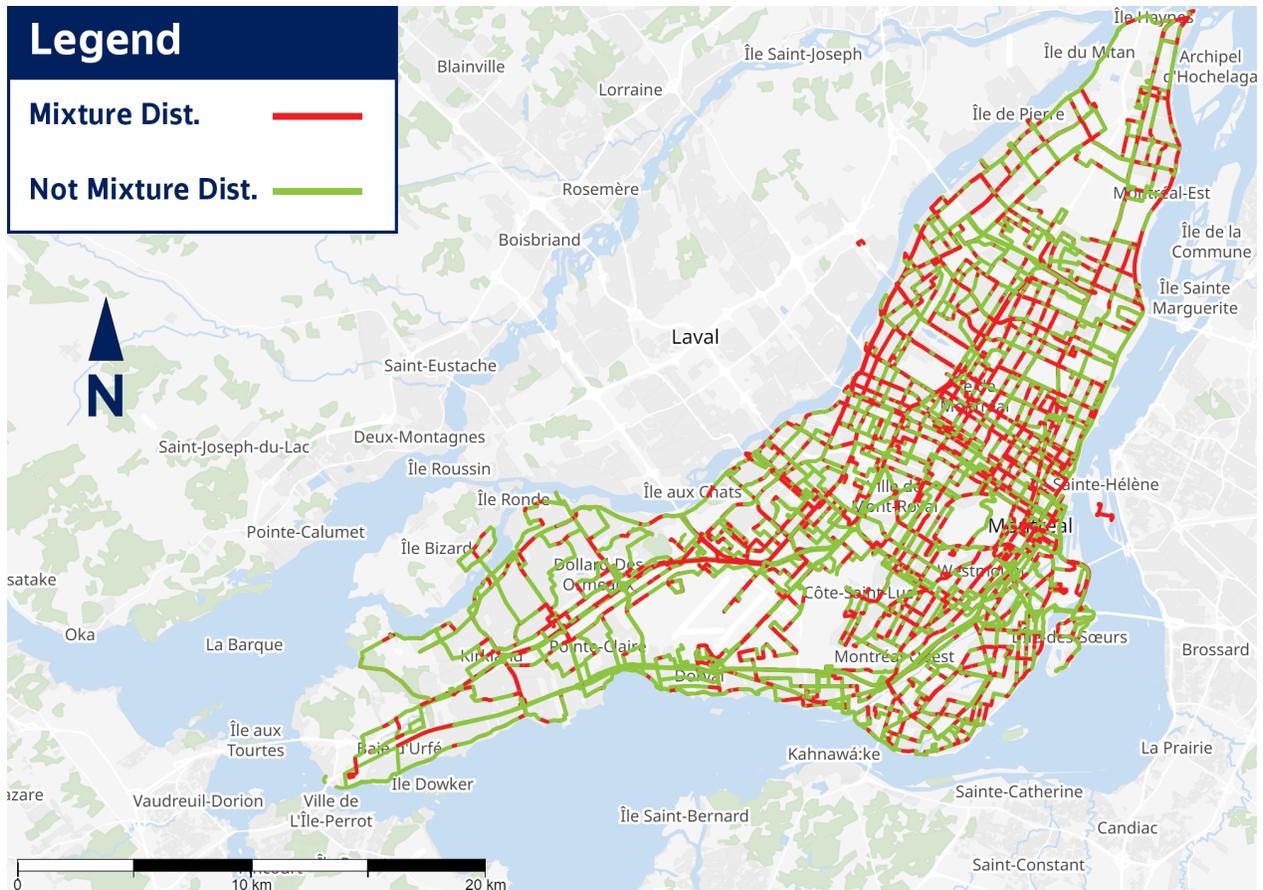


Figure 4.9 Locations of Stop Level Mixture Transit Travel Time for Weekdays at 7 AM

orange neighbourhoods correspond to 20% to 35%, and green neighbourhoods correspond to less than 20% percent.

From the two graphs, we can observe that the city's near east side (or geographically the northeastern side) has more segments that follow mixture distributions. Segments near major transportation corridors, also have more mixture distributions. These areas tend to have higher population densities with more transit demand, as shown by the density of transit segments as well. The streetscapes tend to be less car-friendly, with short blocks and traffic lights within short distances. This might create more variations in travel times.

More suburban areas, such as the west island, have fewer mixture distributions. This is also expected, since these areas tend to have less transit demand and more car-oriented streets. These segments tend to have wide streets, with longer block distances. Therefore, we hypothesize that population density, land use, and streetscapes could also influence the shape of travel time distributions.

The observations are similar for all three analysis levels. For simplicity, we are not repeating

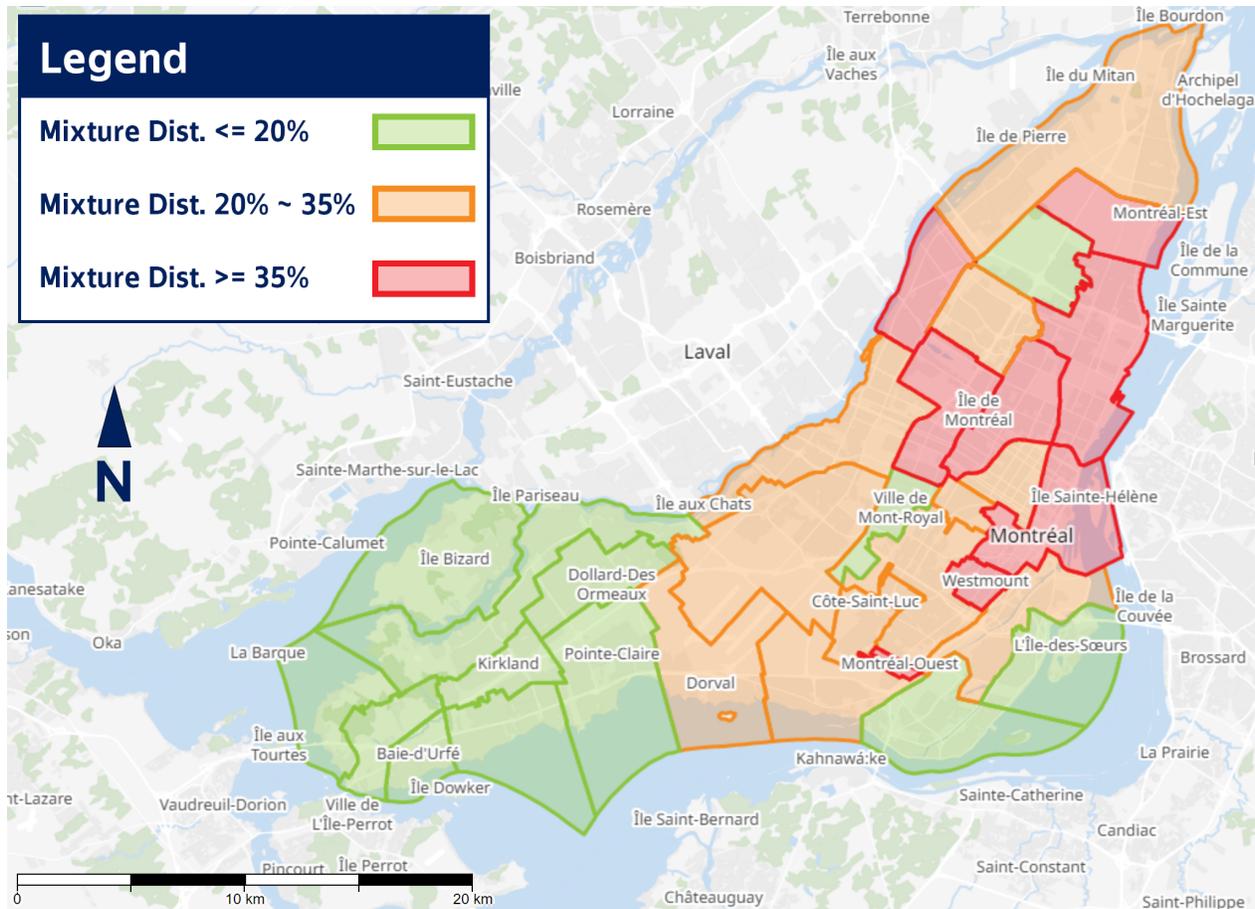


Figure 4.10 Percent Timepoint Pair Segment Length Following Mixture Distributions in each Neighbourhood for Weekdays at 7 AM

the same observations for all three levels in this section.

4.5 Variables Related to Mixture Transit Travel Time Distributions

After identifying these unimodally and multimodally distributed segments, we combine them with other open data sources, such as OpenStreetMap and the city's open data to identify some systemwide factors contributing to mixture travel time distributions. To reiterate, we hypothesize that the ridership variation, population density, traffic signals, and streetscapes could influence the shape of travel time distributions. We propose to develop a few classification models to identify the influence of these factors.

Although we have observed segments with three or more underlying distributions, the sample size with more than two underlying components is relatively small. Due to the unbalanced data size, the classification methods are less accurate for these categories. For example, the

models could simply ignore these small samples and still achieve high accuracy. Therefore, we set the outcome variable to either true, if a segment follows a mixture distribution, or false, otherwise.

4.5.1 Variables Used in Classification Models

Based on the literature review and the observation from the descriptive statistics, we select a few operational and built environment-related variables as inputs for our classification model to classify whether a segment follows mixture distribution or not at a given service type and time of the day. We briefly explain these variables in this section.

- Service type. This is a series of binary variables that identify the service type. We included three service types, weekdays, Saturdays, and Sundays. In the models, we consider weekdays as a base case, and include one binary variable "Saturday" and one binary variable "Sunday".
- Time of the day. This is a series of binary variables that identify the time of the day. Due to the non-linear nature of time variables, where the time would cycle back from hour 23 to hour 0 at midnight. For simplicity, we include a few binary variables for time of the day categories, namely "Early AM", "AM Peak", "PM Peak", "Evening", "Late Night". We consider midday hours as a base case.
- Service frequency. Since most STM schedules do not use clock-faced headways, we use the average scheduled service frequency during the study period. The unit for this variable is transit unit per hour.
- Number of stops. The number of stops between two timepoints on a given route or along a service pattern. This variable is only applicable to the route timepoint pairs and service pattern levels.
- Number of streets. This variable is calculated from OpenStreetMap, where we match the shape of the segment to the street network. We hypothesize that when transit vehicles make turns, transit vehicles could be affected by other traffic or traffic signals. Therefore, it could potentially affect the shape of travel time distributions.
- Number of Lights. The number of traffic lights along a given segment. This variable is obtained from OpenStreetMap and the city's open data. Since most of the STM stops are at the near side of the intersection, we also include the traffic light right after the end stop of a given segment if there is one. Our reasoning is that, due to heavy traffic,

buses might have to wait for a traffic light cycle before they can get to the stop at the end of the segment, thus affecting the travel time distributions.

- Number of lanes. This variable is average number of lanes of a given segment, calculated from OpenStreetMap.
- Speed limit. This variable is the average speed limit on a given segment, mainly obtained from OpenstreetMap. In case there is no speed limit information, we use the default speed limit according to the city ordinances. Please note that not everyone strictly follows the legal speed limit. Some suburban streetscapes could encourage drivers to drive faster than the legal limit.
- Bus lane. This variable is the percentage of bus lanes in operation on a given segment, calculated from OpenStreetMap and data provided by the agency. The variable is set only if the reserved lane is in operation, if the reserved bus lanes are not in operation, the variable is set to 0.
- Oneway. This variable indicates the percentage of oneway streets or streets with medians separating different directions on a given segment, calculated from OpenStreetMap.
- Population density. This variable is the average population density in the census tracts within 500 meters of the given segment, obtained from the census data from Statistics Canada.
- Lengths. This is a series of variables that show the length of different street classifications on a given segment. The data is obtained from the city's open data.
- Land use. This is a series of variables that show the proportion of land uses within 500 meters of the given segment. The data is obtained from the city's open data.
- Vehicle Load. This variable is calculated from the vehicle location data. We included both the average and the standard deviation of this variable. Passengers might take more time to board and alight the vehicle if the vehicle is crowded, thus might affect the travel times of the vehicle.
- Delays. This variable is calculated from the vehicle location data. We included both the average and the standard deviation of this variable. Since delays might affect how operators behave or react to certain traffic situations, we hypothesize that they could affect the shape of the travel time distributions as well.

- Percent Boarding. This variable is inferred from the vehicle location data. For our case, buses will only stop if there is someone waiting at the stop, or someone requests to get off. We hypothesize that the percentage of trips making a stop could affect the shape of the overall travel time distribution.

The descriptive statistics of each individual variable (Table 4.1) are included here for readers' reference.

Table 4.1 Descriptive Statistics for the Variables

Variable	Stop Pairs				Route Timepoint Pairs				Service Pattern							
	Min	Max	Avg	SD	Min	Max	Avg	SD	Min	Max	Avg	SD				
Is Mix Dist.	0.00	1.00	0.29	0.45	0.00	1.00	0.25	0.43	0.00	1.00	0.20	0.40				
Saturday	0.00	1.00	0.33	0.47	0.00	1.00	0.31	0.46	0.00	1.00	0.31	0.46				
Sunday	0.00	1.00	0.34	0.47	0.00	1.00	0.31	0.46	0.00	1.00	0.31	0.46				
Early AM	0.00	1.00	0.10	0.30	0.00	1.00	0.10	0.30	0.00	1.00	0.10	0.30				
AM Peak	0.00	1.00	0.21	0.41	0.00	1.00	0.20	0.40	0.00	1.00	0.21	0.40				
PM Peak	0.00	1.00	0.20	0.40	0.00	1.00	0.20	0.40	0.00	1.00	0.20	0.40				
Evening	0.00	1.00	0.14	0.35	0.00	1.00	0.13	0.34	0.00	1.00	0.13	0.34				
Late Night	0.00	1.00	0.08	0.27	0.00	1.00	0.08	0.27	0.00	1.00	0.08	0.26				
Service Frequency	0.02	22.67	2.34	1.69	0.02	17.09	2.11	1.36	0.04	16.61	2.15	1.47				
Number of Stops	-	-	-	-	1.00	40.00	5.14	3.32	1.00	116.00	35.12	17.66				
Number of Streets	0.00	14.00	1.34	0.73	0.00	16.00	2.47	1.64	1.00	41.00	11.00	6.03				
Number of Lights	0.00	36.00	1.18	1.27	0.00	36.00	4.43	3.55	0.00	121.00	28.66	15.88				
Number of Lanes	1.00	6.00	2.39	0.92	1.00	6.00	2.54	0.82	1.19	5.33	2.56	0.54				
Speed Limit	20.00	100.00	40.83	8.85	21.56	96.86	42.18	8.32	30.07	81.55	42.72	7.61				
Bus Lane	0.00	100.00	1.71	11.69	0.00	100.00	2.35	11.89	0.00	96.74	2.26	8.39				
Oneway	0.00	100.00	39.05	45.62	0.00	100.00	46.06	39.71	2.52	100.00	45.56	24.94				
Pop. Density	0.00	35.50	6.08	4.01	0.00	27.55	5.56	3.52	0.00	13.45	5.25	2.66				
Length - Residential	0.00	3.11	0.05	0.10	0.00	3.11	0.52	0.52	0.00	8.27	1.53	1.51				
Length - Collector	0.00	3.55	0.08	0.15	0.00	7.61	0.85	0.81	0.00	17.30	3.30	3.04				
Length - Secondary	0.00	7.53	0.12	0.23	0.00	12.23	1.03	0.89	0.00	19.34	5.08	4.02				
Length - Primary	0.00	5.75	0.03	0.12	0.00	5.75	0.95	0.86	0.00	12.94	2.60	3.31				
Length - Motorway	0.00	14.03	0.02	0.32	0.00	7.35	1.12	1.57	0.00	26.32	3.42	5.45				
Land Use - Mixed	0.00	100.00	10.35	28.31	0.00	100.00	11.94	25.28	0.00	90.36	10.62	13.50				
Land Use - Downtown	0.00	100.00	5.44	22.25	0.00	100.00	8.17	25.82	0.00	100.00	7.67	18.80				
Land Use - Residential	0.00	100.00	67.88	44.36	0.00	100.00	63.62	41.18	0.00	100.00	67.28	26.35				
Land Use - Industrial	0.00	100.00	12.91	32.16	0.00	100.00	11.31	27.37	0.00	100.00	9.58	16.99				
Average Load	1.28	32.82	8.19	3.86	0.00	21.30	4.69	2.73	0.00	21.36	5.23	2.78				
SD Load	0.00	21.13	4.38	2.71	0.45	30.77	7.40	3.88	1.44	26.15	7.65	3.10				
Average Delay	-1.22	8.24	1.97	1.31	-1.45	5.80	1.79	1.54	-1.00	5.74	1.64	1.09				
SD Delay	0.20	9.41	3.77	1.94	0.30	11.30	4.71	3.07	0.49	9.00	3.85	1.85				
Percent Boarding	0.00	100.00	39.06	28.75	0.00	93.85	30.67	20.65	0.00	80.30	31.36	15.52				
Number of Segments					10309				3176				552			

4.5.2 Classification Model Results

To reiterate, we select four classification models, K-Nearest-Neighbour, Decision Tree, Random Forest, and Logistic Regression due to their own unique properties. We apply all four models on the three different analysis levels, namely the stop pair level, route timepoint pair level, and service pattern level.

To evaluate each method, we use 80% of our dataset to fit these models, then use the rest 20% data unseen by the models as test sets. Table 4.2 shows the confusion matrices of each classification method at a given analysis level as well as their overall accuracies. The rows in the confusion matrix correspond to the actual categories, and the columns correspond to the category generated by the classification model.

Take K-nearest-neighbour method at stop pairs level as an example. We can observe that the overall accuracy is 0.79. More specifically, the true negative rate is 0.78, false positive rate is 0.22, false negative rate is 0.20, and true positive rate is 0.8.

From the table, we can observe that the random forest and K-nearest-neighbour methods generally perform better than the decision tree and logistic regression methods for the test sets. In addition, we can see that the correct responses from these models are fairly balanced.

We can infer two pieces of information from these models. One is that the relationship between the input variables and the outcome variable might be non-linear, since the logistic regression analysis performs worse than the others. Another observation is that k-nearest-neighbour performs well, which suggests that similar segments behave similarly. Thus, it suggests a potential new way for performance analysis or scheduling, where planners could potentially group similar segments together when analyzing travel times or adjusting transit schedules. When planning new services, agencies could also refer to the data on similar existing segments.

As for the analysis levels, the accuracy is the best for the route timepoint pair level, then the stop pair level, and the service pattern level perform the worst. There could be a few reasons for this behaviour. The sample size for the route level is very small, so the models may not have good classification power. In addition, a service pattern can also pass through many neighbourhoods and land uses, and the streets could all have different characteristics, making it difficult to produce a simple model to capture all these variations.

As for the stop pair levels, the stop segments are relatively short in Montreal. Common distances between two stops are around 300 meters. One route may even have two stops at both the near side and far side at the same intersection. Therefore, short stop distances could introduce more uncertainty for the stop-level analysis.

Table 4.2 Confusion Matrix for Classification Methods at Different Analysis Levels

		Stop Pairs		Route Timepoint Pairs			Service Pattern		
		False	True		False	True		False	True
K-NN	False	0.78	0.22	False	0.77	0.23	False	0.71	0.29
	True	0.20	0.80	True	0.16	0.84	True	0.28	0.72
	Accuracy		0.79	Accuracy		0.80	Accuracy		0.72
Decision Tree	False	0.73	0.26	False	0.76	0.24	False	0.66	0.34
	True	0.28	0.70	True	0.25	0.75	True	0.26	0.74
	Accuracy		0.71	Accuracy		0.75	Accuracy		0.70
Rand Forest	False	0.79	0.21	False	0.84	0.16	False	0.75	0.25
	True	0.20	0.80	True	0.18	0.82	True	0.27	0.73
	Accuracy		0.80	Accuracy		0.83	Accuracy		0.74
Logistic Reg.	False	0.66	0.34	False	0.71	0.29	False	0.70	0.30
	True	0.36	0.64	True	0.31	0.69	True	0.30	0.70
	Accuracy		0.65	Accuracy		0.70	Accuracy		0.70

When planning for a new service, there would be no actual passenger demand and delay data for reference. We tried to fit these models without delay and demand variables to see if the models could be useful for agencies when planning a new service. We found that the general observation still holds, except that the accuracy would decrease by roughly 2 to 3 percent in all cases. Thus, we will not be repeating the results in detail here.

4.5.3 Identified Factors Related to Mixture Transit Travel Time Distribution

Table 4.3 Logistics Regression Coefficients

Variable	Stop Pairs				Route Timepoint Pairs				Service Pattern			
	Coeff.	Err.	T-Stat	p	Coeff.	Err.	T-Stat	p	Coeff.	Err.	T-Stat	p
Intercept	-0.24	0.04	-5.66	0.00	1.14	0.09	13.32	0.00	-0.01	0.39	-0.03	0.97
Saturday	-0.84	0.01	-78.67	0.00	-1.11	0.02	-47.48	0.00	-1.56	0.07	-22.30	0.00
Sunday	-0.80	0.01	-73.08	0.00	-1.00	0.02	-42.44	0.00	-1.48	0.07	-21.65	0.00
Early AM	0.04	0.02	2.38	0.02	0.32	0.03	9.11	0.00	-0.08	0.11	-0.72	0.47
AM Peak	0.10	0.01	7.68	0.00	0.06	0.03	2.16	0.03	-0.18	0.08	-2.32	0.02
PM Peak	-0.02	0.01	-1.76	0.08	-0.05	0.03	-1.88	0.06	-0.43	0.08	-5.60	0.00
Evening	0.01	0.01	0.60	0.55	0.05	0.03	1.59	0.11	-0.12	0.09	-1.40	0.16
Late Night	0.05	0.02	3.22	0.00	0.19	0.04	5.14	0.00	-0.18	0.11	-1.66	0.10
Service Frequency	0.16	0.00	54.45	0.00	0.21	0.01	27.43	0.00	0.29	0.02	13.04	0.00
Number of Stops					-0.06	0.01	-10.82	0.00	0.01	0.00	3.55	0.00
Number of Streets	-0.07	0.01	-9.88	0.00	-0.17	0.01	-21.20	0.00	-0.01	0.01	-1.48	0.14
Number of Lights	0.17	0.01	33.20	0.00	0.07	0.00	15.96	0.00	0.00	0.00	0.27	0.78
Number of Lanes	0.00	0.01	-0.69	0.49	-0.01	0.01	-0.87	0.38	-0.10	0.06	-1.59	0.11
Speed Limit	0.00	0.00	-1.03	0.30	0.00	0.00	-2.55	0.01	0.02	0.01	3.49	0.00
Bus Lane	0.00	0.00	0.33	0.75	0.00	0.00	2.61	0.01	0.00	0.00	-0.04	0.97
Oneway	0.00	0.00	11.93	0.00	0.00	0.00	11.42	0.00	0.00	0.00	-1.48	0.14
Pop. Density	0.03	0.00	22.63	0.00	0.02	0.00	6.87	0.00	0.02	0.02	1.46	0.14
Length - Residential	-2.09	0.06	-34.50	0.00	-0.47	0.04	-13.47	0.00	0.00	0.01	-0.14	0.89
Length - Collector	-1.78	0.04	-42.37	0.00	-0.59	0.03	-22.53	0.00	-0.08	0.01	-6.86	0.00
Length - Secondary	-0.96	0.03	-28.27	0.00	-0.34	0.02	-14.48	0.00	-0.03	0.01	-2.66	0.01
Length - Primary	-0.53	0.04	-12.30	0.00	-0.03	0.03	-1.04	0.30	0.04	0.01	2.65	0.01
Length - Motorway	0.08	0.02	4.74	0.00	0.04	0.02	2.31	0.02	0.04	0.02	2.47	0.01
Land Use - Mixed	0.00	0.00	2.65	0.01	0.00	0.00	3.40	0.00	-0.01	0.00	-2.87	0.00
Land Use - Downtown	0.00	0.00	5.09	0.00	0.00	0.00	-1.62	0.10	-0.01	0.00	-3.66	0.00
Land Use - Residential	0.00	0.00	1.36	0.17	0.00	0.00	8.82	0.00	0.00	0.00	-0.56	0.58
Land Use - Industrial	0.00	0.00	-7.33	0.00	0.00	0.00	3.15	0.00	-0.01	0.00	-2.28	0.02
Average Load	-0.07	0.00	-29.31	0.00	0.22	0.01	32.43	0.00	0.02	0.03	0.72	0.47
SD Load	0.13	0.00	34.67	0.00	-0.15	0.01	-29.63	0.00	0.07	0.03	2.77	0.01
Average Delay	0.08	0.00	26.26	0.00	-0.11	0.01	-16.03	0.00	-0.04	0.02	-1.57	0.12
SD Delay	0.00	0.00	0.82	0.42	0.01	0.00	3.78	0.00	0.00	0.01	-0.07	0.94
Percent Boarding	0.00	0.00	22.99	0.00	-0.01	0.00	-22.36	0.00	-0.02	0.00	-7.97	0.00

In this section, we mainly focus on interpreting the logistic regression model coefficients due to their linearity. As a quick reminder, the logistic regression is fitted by log transforming the logistic function. Thus, the coefficients listed in this section are the log odds of having a mixture distribution given one unit of change. The coefficients that are statistically significant ($p < 0.05$) are marked with bold fonts.

As expected and demonstrated in earlier sections, the binary variables for Saturday and Sunday services have a negative sign in all three models compared to the base case of weekday services. For example, at stop pair level, the coefficient for Saturday service is -0.84 . This indicates that Saturday service has $e^{-0.84} \approx 0.43$ times the odds of being a mixture distribution compared to the weekday services all else being equal. In other words, the Saturday travel times are less likely to follow mixture distributions. Again, there could be two possible scenarios to explain this behaviour. One is that, during the weekend, the traffic and demand levels are lower. Another one is that, due to the smaller data sample, the statistical significances are not enough to reject the null hypothesis that the travel times do not follow mixture distribution.

As for the time of the day variables, the three models paint a different picture. Compare to midday services, morning travel times tend to have greater odds of following mixture distributions at stop and route timepoint pair levels. As for the afternoon peak hours, only the service pattern level coefficient is statistically significant with a negative sign. For evening services, no coefficients are significant, suggesting similar travel conditions compared to midday hours. However, for late night services, both stop and timepoint pair levels are significant with positive signs. This means that late night services are more likely to have underlying travel conditions at a smaller scale. This could be due to the additional late-night leisure activities at the end of the work week and during the weekend. Planners might consider adjusting transit schedules for the end of the work week or weekend only.

Service frequency is a significantly positive variable for all three levels. The variable could correlate to the variation in underlying travel conditions and operator preferences. For example, frequent service correlates to high passenger demand and congestion. More frequent service also correlates to the increased number of operators on the route, thus there could be various operator preferences. For example, some are more likely to rush through a yellow light, while some might not. In addition, service frequency could also relate to vehicles interacting with each other on the same corridor. For example, bus bunching may occur on high frequency services. On common corridors shared by many routes, different vehicles may also interact with each other. In these scenarios, the vehicles that follow may be affected by the travel times of the previous vehicle.

Interestingly, more stops on a given segment would decrease the odds of having mixture distributions at the route timepoint level but increase the odds for service pattern level. More research is needed to understand the reason behind this observation.

As for the number of streets, the coefficients are negative, which is also surprising. One would assume that as the vehicle makes more turns, there are more chances to encounter different travel conditions. Yet, the coefficient suggests that travel times are less likely to follow mixture distributions if there are more turns. However, having more turns would also correlate to having a longer segment, thus the longer segment lengths could hide more detailed underlying travel conditions.

The number of traffic lights is significant and positive for stop pair and route timepoint pair levels but is not significant for the service pattern level. This suggests that having additional traffic lights on a given segment would increase the odds of having mixture travel time distributions at a smaller scale. This is expected, since the travel times with and without waiting at a traffic signal are much easier to be distinguished at smaller scales.

The average number of lanes is not significant in all three models. The average speed limit is only significant in two models, yet the magnitudes for all three models are very small. For 1 kilometer per hour average speed limit increase, the odds of having mixture distribution is multiplied by roughly 1.02.

The percentage of bus lanes in operation and the percentage of oneway sections do not seem to have a big influence on the odds of having mixture distributions. Despite some coefficients being significant, the magnitudes are very close to 0.

The population density is also positive. This is expected, since population density correlates to more transit demand, traffic congestion, as well as more diverse land use activities. All of these factors could then contribute to the variation in travel conditions.

For the length of the segment, the signs are mostly negative, except for the segments that contain motorway sections. The magnitudes are also in decreasing order from stop pair to service pattern levels. This is again expected since the longer segments correlate to longer travel times, which could hide more detailed variations. The segments that include motorways tend to be longer express routes. In case of disruptions on the highway, there is less chance for buses to deviate, which could lead to different underlying traffic conditions.

As for the percentage of land use in surrounding areas, the coefficients are very small despite some of them being significant.

The demand-related variables are split between the three models. The average load coefficient is negative for stop pairs yet it is positive at higher levels. The load standard deviation is

positive for stop pair and service pattern levels, but negative for timepoint levels.

The average delay coefficients are also split between the three models. Additional delays correspond to higher odds of having mixture distribution at the stop pair level, yet correspond to lower odds at route timepoint pair and service pattern levels. The standard deviation of delays and percent of samples with boarding activities do not seem to have a big impact on the overall odds, due to their small coefficient magnitudes. More research is needed to explain the discrepancies observed for load and delay related variables.

Table 4.4 Average Variable Effect in Logistic Regression

Stop Pairs		Route Timepoint Pairs		Service Pattern	
Variable	Avg Effect	Variable	Avg Effect	Variable	Avg Effect
Average Load	-0.59	Intercept	1.14	Speed Limit	1.05
SD Load	0.57	SD Load	-1.11	Percent Boarding	-0.77
Service Frequency	0.38	Average Load	1.05	Service Frequency	0.62
Saturday	-0.28	Length - Collector	-0.50	SD Load	0.53
Sunday	-0.27	Service Frequency	0.45	Saturday	-0.48
Intercept	-0.24	Percent Boarding	-0.44	Sunday	-0.46
Number of Lights	0.20	Number of Streets	-0.42	Number of Stops	0.38
Pop. Density	0.18	Saturday	-0.35	Length - Collector	-0.25
Average Delay	0.17	Length - Secondary	-0.35	Number of Lanes	-0.24
Length - Collector	-0.15	Sunday	-0.31	Length - Secondary	-0.16
Length - Secondary	-0.11	Number of Lights	0.30	Length - Motorway	0.14
Number of Streets	-0.10	Number of Stops	-0.30	Pop. Density	0.12
Length - Residential	-0.09	Length - Residential	-0.24	Number of Streets	-0.11
Oneway	0.05	Land Use - Residential	0.21	Average Load	0.11
Land Use - Industrial	-0.03	Average Delay	-0.20	Land Use - Residential	-0.11
Speed Limit	-0.03	Speed Limit	-0.17	Oneway	-0.10
Land Use - Residential	0.03	Oneway	0.15	Land Use - Mixed	-0.10
AM Peak	0.02	Pop. Density	0.12	Length - Primary	0.10
Length - Primary	-0.01	SD Delay	0.06	Land Use - Downtown	-0.09
Land Use - Downtown	0.01	Length - Motorway	0.04	PM Peak	-0.09
Number of Lanes	-0.01	Early AM	0.03	Land Use - Industrial	-0.06
Land Use - Mixed	0.01	Number of Lanes	-0.03	Average Delay	-0.06
SD Delay	0.00	Length - Primary	-0.03	AM Peak	-0.04
PM Peak	0.00	Land Use - Mixed	0.02	Number of Lights	0.03
Late Night	0.00	Late Night	0.01	Evening	-0.02
Early AM	0.00	AM Peak	0.01	Late Night	-0.01
Length - Motorway	0.00	PM Peak	-0.01	Intercept	-0.01
Evening	0.00	Land Use - Industrial	0.01	Early AM	-0.01
Bus Lane	0.00	Evening	0.01	SD Delay	0.00
Percent Boarding	0.00	Land Use - Downtown	-0.01	Length - Residential	0.00
		Bus Lane	0.00	Bus Lane	0.00

Finally, we try to rank the variables by determining their average effects to the results. Table 4.4 shows the ranking of each variable's magnitudes where we applied the estimated coefficients to the average value of each input variable.

From the results, we can observe that variables related to vehicle load, average delay, service frequency, weekday service, number of traffic lights, and the lengths of collector and secondary streets consistently rank high in all three models. These higher ranking variables are related to passenger demand variations, operator preferences, vehicle-to-vehicle interactions, weekend services, traffic lights, and segment lengths. Other variables such as time of the day, reserved lanes, and land use variables tend to have less impact on the odds of having mixture distributions. The rankings should help agencies to focus their attention when analyzing segments with mixture travel time distributions.

4.6 Conclusion

Transit reliability is important for both transit agencies and passengers. To improve the agency's operation efficiency and passenger experience, previous literature pointed out the need to develop more precise models related to transit travel times. Some literature has observed mixture transit travel time distributions.

Aiming to provide more comprehensive evaluations of mixture travel time distributions and to provide more insights to the transit agencies, we conducted a systemwide study to illustrate the presence of these mixture distributions. We also identified potential spatial and temporal patterns related to the shape of transit travel time distributions. Then we fitted several classification models to classify if a given segment would follow mixture distribution at a given time. The fitted models allowed us to examine the importance of each predictor.

The results show the presence of mixture distributions in various analysis levels, stop pair level, route timepoint pair level, and service pattern level. Further analysis could help agencies to pinpoint the cause of such mixture distributions, thus could help agencies to "nudge" the slower travel conditions towards the faster ones. In addition, the percentage of route timepoint pairs and stop pairs that follow mixture distributions stay relatively stable throughout the day. However, the variation at the service pattern level is higher potentially due to the smaller sample size. The weekend services also tend to have fewer segments identified as mixture distribution. However, this could be also due to the smaller sample size. The geographical distribution shows that segments near traffic lights or on major transportation corridors tend to have more mixture distributions. Neighborhoods with more passenger demands, higher population density, and mixed-land use could have more segments identified

as mixture distribution. From the classification models, we showed that non-linear methods and similarity-based models work the best in classifying the shape of transit travel time distributions. The results highlight the need to consider the non-linearity and suggest similar segments tend to behave similarly. Agencies could potentially try to schedule similar segments at the same time. The logistic regression model highlighted the potential effects of passenger demand variations, weekday services, operator preferences, vehicle-to-vehicle interactions, traffic lights, and segment length variables. These results could help agencies to focus their attention and resources when trying to improve transit vehicle travel conditions.

The limitations of this research could be in the following areas. One is with regard to the methodology. In our analyses, we aggregate the entire dataset without considering the service changes that occurred during the study period. Additional research could examine potential seasonal variations in mixture distributions. Another limitation is that our classification models only included a binary outcome variable, due to the smaller sample size with a higher number of components. Future research could analyze the higher components in more detail. Finally, we only selected four classification models, from a wide range of categories. Future research could test additional classification models and compare their performances.

One potential expansion of this research is to use more detailed data on transit ridership, traffic counts, and traffic light settings to help infer the reason causing consistent slower travel conditions on each trip or segment. Using identified causes, agencies can target specific areas and potentially "nudge" the travel times under slower conditions towards faster conditions. Similarly, there are other occasional factors causing slower travel conditions beyond transit agencies' control, such as traffic incidents and weather. By incorporating these data, agencies could evaluate potential response strategies to increase the reliability and resilience of a given transit network. These expansions would help reduce overall travel times, improve vehicle travel conditions, and improve service consistency.

Future works could also expand our research by using additional historic data to further evaluate bus preferential measures, such as transit signal priority and bus lanes. Agencies could analyze the travel time distributions before and after to determine the effectiveness of these strategies using the underlying travel time distributions. This way, agencies could have a better understanding of how these strategies impact the services. Agencies could also incorporate these more precise travel time distributions into other analyses to better understand the passengers' experience, such as transfer simulations or buffer time estimations. Finally, transit agencies could use this information to provide more precise travel planning information to passengers, such as the likelihood to get to their destinations or the planned transfer vehicles for example.

Acknowledgements

The authors wish to thank the Société de transport de Montréal for providing the data. The study was funded by the Natural Science and Engineering Research Council of Canada (NSERC) and the Canada Research Chair in Transportation Transformation.

CHAPTER 5 ARTICLE 2: MODELLING BUS TRAVEL TIME, SPEED, AND PACE AT VARIOUS ANALYSIS LEVELS

Yuxuan Wang^{acde}, Catherine Morency^{acde}, Martin Trépanier^{bce}

^a Department of Civil, Geological and Mining Engineering, Polytechnique Montréal,
Montréal, Québec, Canada

^b Department of Mathematics and Industrial Engineering, Polytechnique Montréal,
Montréal, Québec, Canada

^c Interuniversity Research Centre on Enterprise Networks, Logistics, and Transportation,
Montréal, Québec, Canada

^d Canada Research Chair in Mobility, Montréal, Québec, Canada

^e Canada Research Chair in Transportation Transformation, Montréal, Québec, Canada

Revision submitted for review on 15 March 2025 for Public Transport

Abstract

Good travel time estimates are important for transit agencies and passengers. Since travel time is a function of distance and speed, it is possible to use both as inputs in most scheduling software since the distances are fixed in fixed-route services. However, most literature focuses on travel times, and travel speeds are typically used to plan infrastructures or evaluate operated services. There is a lack of comparison between modelling travel times directly and indirectly through normalized measures, namely speed and pace, at various analysis levels. In this paper, we try to fit and evaluate travel time, speed, and pace models at inter-stop, stop-to-stop, timepoint-to-timepoint, and service pattern levels. Then, we test these models using two typical scenarios in transit planning, creating new routes and expanding service hours.

The results show that travel time, speed, and pace models perform similarly for new service hour scenarios, with the time models performing better overall. However, speed models tend to perform better for new route scenarios where no historical data has been observed. Both direct and normalized approaches perform well at more aggregated levels, such as the timepoint-to-timepoint level. For lower levels, both approaches perform better at the inter-stop level than the stop-to-stop level, emphasizing the need to include more detailed traffic

signal and ridership information to improve the models. Errors calculated from the speed and pace models are slightly more skewed than the direct travel time model. However, the relative errors from travel time models are larger than the normalized models on shorter segments or faster segments. Since each error measure provides different views of the modelling results, we conclude that planners need to choose their measures carefully according to specific model applications and analysis levels.

Keywords: Transit Planning, Transit Operations, Transit Travel Times, Travel Time Modelling

5.1 Introduction

Reliable transit travel time estimates are important for transit agencies' operations and passenger satisfaction. For a transit agency, unreliable travel time estimates affect scheduling, where planners need to add schedule padding to improve reliability (Wessel and Widener, 2017) and thus increase operating costs. Unrealistic travel times may also cause vehicles and operators to miss their scheduled layovers, propagating delays to future trips, as well as causing operators' satisfaction and retention issues (Danaher et al., 2020). Similarly, for passengers, underestimated travel times may cause them to arrive at their destination late or miss their connecting trip. Overestimated travel times may lead passengers to perceive transit services as slow due to the additional holding or schedule paddings. These unreliable travel times also affect passengers' satisfaction and mode choice (Carrel et al., 2013a). Thus, it is important to improve transit travel time models.

One way to account for travel time variations is to review and adjust service schedules periodically. In practice, transit departure times are adjusted frequently according to ridership fluctuation, but transit travel times are less frequently adjusted (Coleman et al., 2018). Up until now, transit agencies and academics have mostly focused on travel time adherence and prediction (Trépanier et al., 2009). This is possibly due to the fact that the schedules communicate arrival and departure times to transit operators and passengers, which are directly related to travel times. Transit planners could simply diagnose the issues of a given segment based on direct observations and adjust travel times accordingly.

However, there are several scenarios where transit vehicles can spend longer travel times than planned. Delays can sometimes be attributed to slower travel conditions. For example, on a snowy day, buses are more likely to travel at a lower speed to ensure safety, even when there is no other traffic nearby. Similarly, areas with high traffic volume can also force transit vehicles to slow down. Another potential scenario is that a vehicle might get stuck, such as at stops with too many passengers or in front of traffic lights. Transit vehicles may spend a long time waiting for passengers to board and alight or for the traffic lights to turn green. In congested areas, vehicles may not only travel slowly, they may also spend more time waiting for several traffic light cycles. Overall, the delays can potentially be summarized into two categories, how fast the bus can operate between two stops, and how long the bus is expected to stop.

How fast a bus can operate between stops is related to the travel conditions on the route or the travel speed. How long the bus is expected to stop is more related to passenger activities, congestion level, and traffic lights, i.e. the time we are expected to stop. Thus, to

improve transit reliability and better adjust transit schedules, we need to better understand transit travel conditions. This knowledge would help transit planners develop a more precise schedule for operators and passengers. It would also help pinpoint the cause of transit travel time issues and evaluate potential transit priority measures to improve transit service quality in the future.

Travel time is a function of speed and distance. A long segment can have a short travel time when driving quickly, and a short segment can have a long travel time in congested conditions. A normalized measure, like speed and pace, also allows comparisons between various segments in the network since stop-to-stop distances are not necessarily the same for all segments. A local bus may make a stop every 300 meters every minute, and an express bus may have a non-stop segment of 15 kilometres taking 20 minutes, whereas speed or pace is more directly comparable. For example, the operating speed for the aforementioned local bus may be roughly 15 kilometres per hour, but the express bus may be at 75 kilometres per hour. Communicating using speed may also be more intuitive for transit operators to understand the expectations of the schedules and evaluate potential actions for buses to remain on time.

We are also inspired by the comparison between predicting travel time and speed approaches from Bauer and Tulic (2018) using floating taxi data. However, taxis tend to travel point-to-point without a predefined route, whereas transit vehicles need to make regular stops along a fixed route to pick up and drop off passengers. Previous studies have often focused on a few routes or a small geographical area, a more varied data sample is still needed for comparisons. Most of the existing works have also only focused on one modeling level, yet transit planning typically involves multiple analysis levels. For example, when planning for a new route or new infrastructure, planners typically assume an operating speed for the entire route, whereas when evaluating the existing schedules, planners usually analyze at timepoint levels (National Academies of Sciences, Engineering, and Medicine and others, 2013). Thus, a more detailed comparison of these levels is still needed. In addition, researchers have also focused on different sets of model evaluation criteria (absolute or relative errors), but there is less discussion on the advantages and disadvantages of these commonly used criteria.

To better understand transit travel time models and to help transit planners better choose their future modelling approaches, we pose the following questions: using the same information as inputs, does modelling travel time directly yield similar results compared to modelling normalized variables like speed or pace, does the model performance differ for each analysis level, what are the advantages and disadvantages of using different evaluation criteria?

To answer the aforementioned questions, we develop a framework to allow us to compute and compare commonly used measures in transit planning. First, we summarize the observed

travel times at various analysis levels, namely inter-stop, stop, timepoint, and service pattern levels, using two years of archived service delivery data from Montréal, Québec, Canada. Then, we compute the normalized measures, i.e. travel speed and pace, which allow us to remove the varying segment distances. Next, we will integrate some commonly used spatial and temporal variables as inputs for our models. Finally, we calculate and compare model performances and discuss the advantages and disadvantages of these various modelling approaches, analysis levels, and evaluation criteria, so that we can make recommendations to other researchers when deciding future modelling approaches as well as to transit agencies for their future planning and operations.

This paper is organized in the following ways. In section two, we will go through the related literature on travel times and travel speeds as well as their applications in planning processes. Next, in section three, we will describe our research framework and methodology. Then, in section four, we will show the modelling results and the evaluations. Finally, in section five, we will provide a quick summary of our research to conclude this paper.

5.2 Literature Review

Transit performance measures are commonly used by transit agencies in their planning and operations. Academics have also studied existing measures and proposed additional transit performance measures. Two commonly used measures for planning are travel time and travel speed, which can be easily obtained with the help of Automated Vehicle Location (AVL) systems.

Travel time is an important measure for scheduling. Coleman et al. (2018) provided a summary of a typical scheduling process. In general, the route performances are reviewed at various intervals for different types of routes and schedules. Service changes generally happen at pre-defined times every year to facilitate operator sign-ups and schedule adjustments. When revising schedules, the analyses generally involve the level of ridership and travel times between timepoints in the North American context. If passenger levels exceed a predefined agency standard, service frequency is adjusted. Travel times are also analyzed using an agency standard and then adjusted both between the two terminals as well as between various timepoints using measures such as the mean, median, or a given percentile of observed travel times (Furth, 2000).

With the help of AVL data, travel times are also used to diagnose schedule adherence issues. In general, agencies and scholars have proposed to classify each timepoint's on-time performance and the frequency of schedule adherence issues. Then, analysts can use more detailed

travel times and dwell times as diagnostic tools to infer the cause of the problems (Mandelzys et al., 2010). Most works found that late buses are mostly caused by longer-than-planned travel times or late departures from previous segments.

There are also attempts to account for the variations in transit travel times using AVL data. Wessel and Widener (2017) calculated the schedule padding using best-case transit travel times recorded. They found that 30% of total scheduled service hours are padded in their case study, and that downtown and rush hours tend to have more paddings. In case of better travel conditions, drivers need to wait for the schedule, thus, schedule control could contribute to slower travel times.

More recently, there are also attempts at classifying delays into systematic (recurrent foreseeable deviations) or stochastic delays (sporadic unforeseeable deviations) (Boudabous et al., 2024), which would then help the agencies to improve their schedules and forecast delays.

To help provide passenger information, many works have tried to predict transit travel or arrival times, and most of the literature focuses on time-related measures. Scholars have proposed methods for predicting transit travel times. Some of the works used only AVL data (Gurmu and Fan, 2014; Chen et al., 2023). There are also attempts to incorporate additional datasets to improve the prediction models, such as real-time traffic data (Ma et al., 2019; Han et al., 2020). More recently, researchers have also included various artificial intelligence methods to improve transit travel time models, such as using computer vision to incorporate road side features (Abdelhalim and Zhao, 2024) and using neural networks to integrate vehicle positions with weather data (Alam et al., 2021).

Using AVL data, it is also possible to evaluate the service delivered to passengers. Wessel et al. (2017) proposed a method to retroactively improve the accuracy of transit agencies' General Transit Feed Specification (GTFS) feed by using archived AVL data. Agencies have also started developing passenger-centric performance measures using their AVL and origin-destination data (Graves et al., 2019).

However, travel speed, pace, and time are related variables, where the travel time equals the travel distance divided by the travel speed and also equals the distance multiplied by the travel pace. By using a normalized unit of measure, such as speed or pace, we can remove the distance from the equation, and we can potentially find similarities and differences between various segments in the system. Therefore, evaluating speed and pace could potentially allow us to create schedules or target issues at the systemwide scale. Therefore, operating speed and pace are other commonly used indicators for transit performance evaluations.

Travel speeds and paces are also used in diagnostic processes. Various attempts have created

many classification models identifying slow roadway segments using speed or pace for agencies to improve performance (Cortés et al., 2011; Du et al., 2017). Speed has also been widely used in assessing the effectiveness of various transit priority strategies, such as bus lanes (Russo et al., 2022) and transit signal priorities (Wu et al., 2020).

More recently, researchers have also started to examine the possibility to model travel speeds by correlating transit speed with various traffic factors using descriptive and neural network models (Du et al., 2017) and by integrating more detailed car speeds to model bus speed distributions (Zhang et al., 2020). Existing literature has found bus lanes, road classifications, geographical area, peak direction, and service types affect bus speeds, and that buses on main roads, in outskirt neighbourhoods, during off-peak hours, with larger stop spacings, or in bus lanes tend to travel faster than on other segments (Zhang et al., 2014; Kopsacheilis et al., 2023).

Travel pace is defined as the inverse of the travel speed. There is relatively less research focusing on pace models specifically, but pace is still commonly used as an indicator for various delay classifications (Boudabous et al., 2024) or performance evaluation models (Aemmer et al., 2022).

The previous literature could all be helpful in identifying a slow segment, modelling the transit systems, or predicting vehicle arrival times. However, given the one-to-one relationship between time and speed, there is still a need to compare the time and speed measures to examine their advantages and disadvantages at various analysis levels. There have been some efforts to compare time and speed measures in the transportation field, especially from Bauer and Tulic (2018), which posed a similar question for taxi travel times. However, there are some additional considerations for public transit planning, such as stops, ridership variations, and transit priority measures. Thus, we ask the question if we could compare these two approaches for transit planning.

There are a few additional questions to answer. Even though most of the scheduling is done at the timepoint level in North America, GTFS standard requires arrival and departure times for every stop served by a certain trip for passenger information. Unfortunately, arrival and departure times for stops in between timepoints are not clearly defined (Wessel et al., 2017), and are typically interpolated using the timepoint inputs. Thus, there is a discrepancy between the general scheduling practices and what is shown to the passengers, since passengers do not necessarily board and alight at timepoints. This calls for further investigation into stop-level scheduling practices, also pointed out by other researchers (Wessel et al., 2017; Graves et al., 2019).

In addition, the works mentioned above have used many error measures to evaluate their

model performances, such as absolute measures and relative measures. Yet, these measures are typically aggregated into one number. Different measures evaluate the results "from different angles" (Kolassa, 2020), and there are not many comparisons between the measures. As transit services have various segment sizes, it is also necessary to compare the errors by segment for potential biases, since a short local segment is not directly comparable with a long highway express segment, for example.

Thus, in this paper, we aim to create a framework to compare direct travel time models and normalized models by using modelling speed and pace measures. Given the limitations of earlier studies, we also try to compare travel times and travel speeds at various analysis levels, namely, the inter-stop, stop-to-stop, timepoint-to-timepoint, and service pattern levels. Then, we evaluate these different models using various error measures. We hope to provide more nuances for future researchers and planners to consider when planning or modelling transit networks.

5.3 Research Framework and Methodology

In this section, we present an overview of our research framework. Then, we provide more details regarding the data and the methodology.

The overall research framework is summarized in Figure 5.1. We first use GTFS and GTFS Real-Time data provided by Société de Transport de Montréal in Canada as inputs to calculate the travel times, speeds, and paces. Those who wish to produce these statistics elsewhere could also use the archived data from similar data standards like Network Timetable Exchange (NeTEx) and Standard Interface for Real-time Information (SIRI) or other agency internal datasets.

Typically, agencies have their own policies on how to analyze travel times (Furth, 2000), whether using the average or a predefined percentile. The average running time policy is a compromise between having buses run early and having buses run late (Furth, 2000). Thus, we use the average travel times and speeds as the dependent variables of our models. Other researchers and planners could, nevertheless, test other statistics, such as percentile-based statistics, using the same research framework in the future.

Then, we integrate additional spatial characteristics using OpenStreetMap (OpenStreetMap contributors, 2023) and the open data provided by the city of Montréal. These spatial attributes, along with the calculated travel times, speeds, and paces, are then used as inputs for modelling. In addition, we apply the models to various analysis levels, inter-stop, stop-to-stop, timepoint-to-timepoint, and service pattern levels, to compare their planning

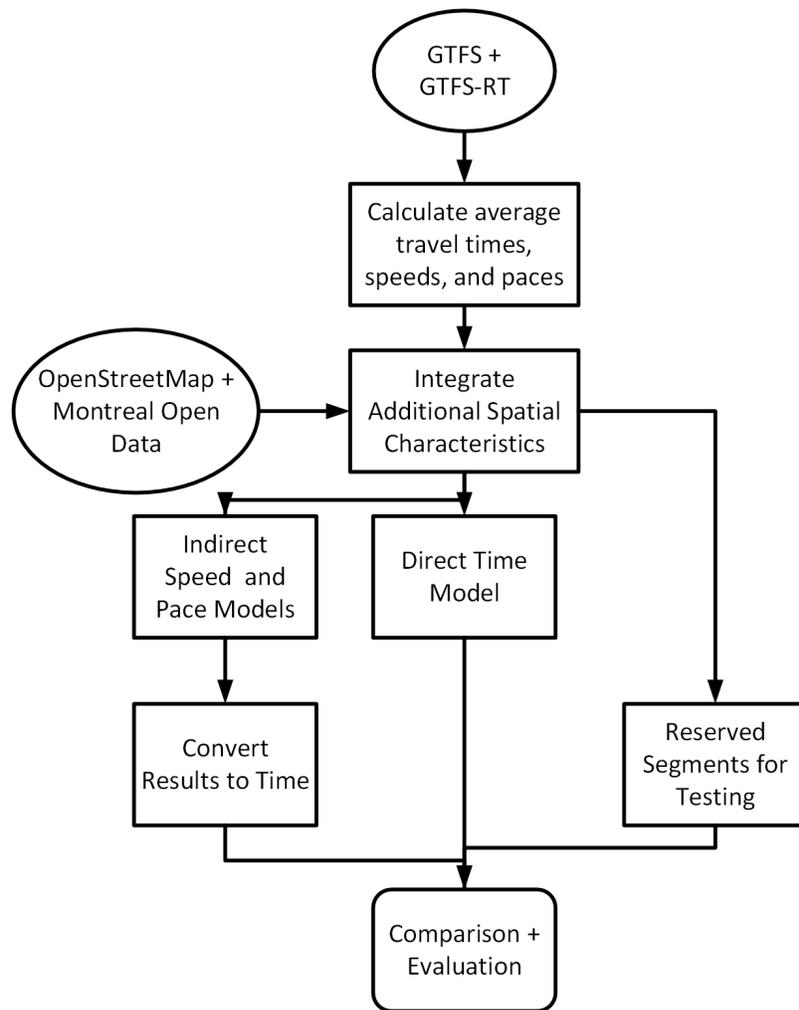


Figure 5.1 Research Framework

implications.

Next, we will test the models given two common planning scenarios. One is to expand or modify the services into a new route. In this case, agencies may not have historical data available at all. To account for this scenario, we reserve 10% of the segments from our dataset for testing. Another common scenario is to expand the service hours on an existing line. Therefore, agencies may not have historical data for a specific time-of-the-day on a given segment. Thus, we reserve another 10% of the remaining data with various time-of-the-day values for testing.

Finally, we compare and evaluate the model results using a few error measures, which are outlined in an upcoming section. We also discuss their planning implications in the results section.

5.3.1 Data

We use the bus system of Société de Transport de Montréal on the island of Montréal in Canada as a case study. To summarize the system, it has 222 bus lines in operation, 2012 buses in the fleet, and more than 17,000 published bus trips on average weekdays.

The GTFS file provides detailed information on the planned services, such as schedules and geographical information for the routes and stops. The GTFS Real-Time data provides the actual bus arrival and departure times at stops, as well as detailed bus location and speed information every 5 to 20 seconds. In this paper, we used the archived data from May 1st, 2021 to March 24th, 2023.

Since this project focuses on the mean travel times and travel times do not have an upper bound, outlier observations, such as from mechanical issues, major events, detours, or traffic incidents, might greatly affect the mean observation. Thus, we will remove these outliers from the analysis using Density-Based Spatial Clustering of Applications with Noise (DBSCAN) (Ester et al., 1996), which is a density-based algorithm to identify clusters and outliers in the data. For each segment, we calculate the density according to the travel time and delay observations. The outliers are identified from the lower-density areas, where the observations are less similar to the others. For example, travel times that are unusually short or long or departures that deviate significantly from the planned times will be removed. We chose to keep 80% of the data as inputs, and this parameter choice and model sensitivity can be a subject of future research.

5.3.2 Analysis Levels

Again, in this paper, we will focus on various analysis levels, the inter-stop, stop-to-stop, timepoint-to-timepoint, and service pattern levels. In this section, we will quickly define each level, as our dwell time definition is slightly different from the Transit Capacity and Quality of Service Manual (TCQSM) (National Academies of Sciences, Engineering, and Medicine and others, 2013) due to data limitations.

The dwell time is typically defined as the time a vehicle stops to allow passengers to board and alight at a given bus stop according to the TCQSM (National Academies of Sciences, Engineering, and Medicine and others, 2013). However, since most stops are on the nearside and we do not have more detailed door opening or closing data nor traffic light timing data, our study would thus combine both the time for passenger activities and the time waiting for green lights into our dwell time calculations. The estimations of these detailed data can be left for future research.

First, the inter-stop travel time includes the total time between the departure from the first stop and the arrival at the second stop, which would include any traffic light waiting times or congestion between the two stops. It does not include the dwell times and traffic light waiting times at the stops.

The stop-to-stop time is defined as the total time between the departure from the first stop to the departure of the second stop, which includes the dwell time at the second stop and the inter-stop travel time between the first and second stops.

The timepoint-to-timepoint time is defined as the total time between the departure at a timepoint to the next timepoint, which would include the sum of travel times of all stop-to-stop segments between the two timepoints.

Since a route may have different service patterns, such as short turns and branch lines, these service patterns would have different travel times. Thus, we will analyze the travel times for each service pattern to ensure that the travel times are comparable. The service pattern travel time is defined as the time between the departure from the first stop and the arrival at the final stop.

Finally, to reiterate, we hope to test the possibility of using normalized measures to model travel times by taking distance out of the equation, given the one-to-one correspondence between time, speed, and pace. Thus, we will calculate the corresponding speed and pace measures using the distance provided from the GTFIS and the times observed in the GTFIS-RT data for our models. The summary of each measure, namely time, speed, and pace, can be found in Table 5.1. From the table, we can observe that the segments have vastly different

lengths. The normalized measures, speed and pace, have less significant ranges between the minimum and the maximum compared to the travel times. The mean and medians are closer to each other for normalized measures, meaning they are less skewed compared to the travel times and the shape of the distribution is more symmetrical.

Table 5.1 Summary of Dependent Variables and Corresponding Distances at each Analysis Level

	Route	Timepoint	Stop	Inter-Stop
Min Time (sec)	324.00	27.00	10.00	10.00
Average Time (sec)	2076.64	273.56	60.97	46.47
Median Time (sec)	1949.00	259.00	44.00	30.00
Max Time (sec)	4345.00	1024.00	974.00	943.00
Min Speed (km/h)	9.39	7.68	3.74	6.77
Average Speed (km/h)	19.75	20.25	23.57	29.63
Median Speed (km/h)	18.39	19.23	22.89	29.28
Max Speed (km/h)	48.56	51.10	58.79	63.53
Min Pace (sec/km)	76.76	82.87	62.26	58.68
Average Pace (sec/km)	200.31	208.90	212.24	149.42
Median Pace (sec/km)	196.55	191.43	173.20	127.46
Max Pace (sec/km)	350.00	528.20	1288.73	785.49
Min Distance (km)	0.81	0.04	0.02	0.02
Average Distance (km)	12.22	1.67	0.35	0.35
Median Distance (km)	10.44	1.38	0.25	0.25
Max Distance (km)	38.92	16.84	15.25	15.25

5.3.3 Modelling Methods

Since our research deals with repeated measurements on a subject, in our case, a segment along a bus route, the resulting data points on each segment may be correlated. For example, if we have a linked traffic light, the traffic light may always be green for the given segment. The resulting impact of traffic lights on travel times is negligible. Thus, we need to adopt a mixed model to account for these unobserved differences between each segment (Yang et al., 2014).

In our research, the random effects, or the grouping factors, are crossed random effects between segments and time periods. For each segment, there are various time period measurements. Similarly, for each time period, there are many segments being measured. In this study, we will only allow random intercepts, which allows each subject to have a different intercept while keeping the slopes the same.

Since travel conditions, traffic lights, or interactions between vehicles could contribute to non-linear relationships between the dependent and independent variables, we decided to test both linear and non-linear models for comparisons.

The linear mixed model is similar to the regular linear regression model, but with an additional term to account for the grouping factors. Coefficients are estimated by solving the mixed model equations using maximum likelihood estimates (Laird and Ware, 1982). To predict a new data point not in any existing groups, the model uses the population level coefficients without considering any group-specific effects.

The non-linear method used here is the random forest method originally proposed by Ho (1995). Combining multiple regression trees was found to achieve better results than using one regression tree, albeit the model is less explainable due to it involving multiple trees. Hajjem et al. (2014) proposed an extension to account for the mixed effects. The basic idea is to generate multiple regression trees using various subsets of the data sample and various subsets of sample variables within a given group. To predict a new data point not in the existing groups, the algorithm follows the split rules according to the population level variations not specific to any pre-existing groups.

5.3.4 Input Variables

There are many factors affecting transit travel time and travel conditions. The Transit Capacity and Quality of Service Manual (National Academies of Sciences, Engineering, and Medicine and others, 2013) provides an excellent summary of these factors. Thus, we try to include these temporal, spatial, and operational variables. In this subsection, we describe the independent variables.

The temporal variables included in our studies are related to the daily, weekly, and seasonal changes in travel time or travel speed. They are defined as the following:

- Service Period. It is a categorical variable corresponding to each service change during the year. In Montreal, there are five service periods in a year, namely January, March, June, September, and November. Here, we use the January period as the base case.
- Time-of-the-Day. Due to the non-linear nature of traffic and the time periods, we simplified time as a categorical variable to correspond to various attributes for which we do not have detailed data. In Montreal, the majority of traffic signals use fixed-time coordinated plans with 3 different programs. For weekdays, the AM Peak program is generally in use between 6:00 and 10:00; the midday program is used roughly between

10:00 and 15:00; the PM Peak program is roughly in use between 15:00 and 19:00, and the evening program is used between 19:00 and 6:00 the next day. For weekends, the AM peak program is used between 10:00 and 19:00 for major corridors, and the evening program is used between 19:00 and 10:00 for major corridors and all day for smaller streets. Similarly, the bus lanes tend to operate during peak hours (roughly 6:30 to 9:30, then 15:30 to 18:30) in the peak direction only. Bus services are also divided into various time-of-the-day categories. Peak-hour only services tend to run between 6:00 and 10:00 then 15:00 to 19:00, and the night routes are in service from 23:00 to 6:00. Since these discrete changes in traffic and service patterns all happen close to 6:00, 10:00, 15:00, 19:00, and 23:00, we identified five time of the day categories for weekdays, namely morning peak (6 - 10), midday (10 - 14), evening peak (15 - 19), evening (19 - 23) periods, and late-night (23 - 6). For weekends, we used 4 time categories, i.e. morning (6 - 10), midday (10 - 19), evening (19 - 23), and late-night (23 - 6). Here, we use the Weekday PM peak as the base case, and these time categories are applied across all segments analyzed. Thus, we include both time of day and the day of the week in our categories.

The spatial variables included are related to street characteristics, land use characteristics, and the population density near a given segment. These data are obtained mainly from the city of Montréal and OpenStreetMaps (OpenStreetMap contributors, 2023). They are defined as:

- Number of Turns given a stop-to-stop segment. We hypothesize that turning would require buses to slow down to account for other traffic or pedestrians, thus increasing travel times.
- Number of Lanes, the average number of lanes on a given segment. We include this variable since it is related to the street classifications. Wider streets typically correlate to more traffic, which could act as a proxy for traffic data.
- Number of stop signs given a segment. Traffic is required to stop before the stop sign before continuing by law. Thus, stop signs would impact the overall travel time and speed.
- Number of traffic lights given a segment. Traffic is required to stop before the light when it is red. Thus, traffic lights would impact the overall travel time and speed. However, there are a variety of traffic lights in operation, and due to the lack of data, we can only include the total number in our model.

- Speed limit, the average legal speed limit of a given segment in kilometre/hour. This variable gives a rough approximation of how fast the vehicles travel on a segment. Legally, vehicles should travel at or below the speed limit. However, in practice, due to congestion, travel speeds on some segments may never reach the legal limit. Similarly, in less congested areas, people may drive well above the speed limit.
- Segment length, the length for each street classification category of a given segment in kilometres. In Montreal, the streets can be roughly classified into five categories, namely, local, collector, secondary, primary, and motorways. Here we separate the length of different street categories, due to the fact that travel conditions on local streets may be very different compared to a highway.
- Land use, a categorical variable on the land use surrounding a segment. For the models, we include five main land uses, namely, commercial, industry, downtown, green spaces, and residential as a base case.
- Population density, the average population density for the surrounding census tracts in thousand people per square kilometre. Due to the limitations of our ridership data, we include this variable as a proxy for ridership, since ridership is higher in high-density areas.
- Distance to downtown, the straight line distance to downtown in kilometres. We include this variable since traffic generally gets less congested further away from downtown areas. Thus, we can include neighbourhood differences in our model.

As for operational variables, we include them to account for operation-related variations. They are the following:

- Bus lane status, a categorical variable related to bus lane operations. Since bus lanes are typically located in congested areas to facilitate transit operation, the speed and travel times would be longer compared to less congested streets without bus lanes. In addition, the bus lanes are not necessarily in service all day, thus, we further divide the data to account for the differences when bus lanes are in service. Thus, we include three categories in our models, namely bus lane not in service, bus lane in service, and no bus lanes as the base case.
- Average frequency, the average number of buses passing through this segment during an hour. This is related to the ridership as well as the potential bus congestion. As the frequency increases, the risk of bus bunching increases. Some buses may be stuck behind another one, thus affecting the speed and travel times.

- Number of stops, the number of stops on a given segment for passengers to get on and off. This variable is only available for the timepoint-to-timepoint and service pattern levels.

5.3.5 Model Evaluation Criteria

To evaluate the models estimated using the variables and methods outlined above, we use four commonly used good-of-fit measures to evaluate the errors, namely coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE). Unfortunately, since the two models have different dependent variables with different scales and bounds, we cannot use a statistical test to directly measure the significance of their differences. Thus, we will convert the speed model results to time results using the segment distance, so that they are comparable in terms of their good-of-fit measures. In this subsection, we will provide a quick summary of these measures.

The coefficient of determination, or R-squared, is a measure to determine the proportion of variance in the dependent variable explained by the given independent variables. It is calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

where \hat{y}_i is the predicted value, y_i is the actual value, \bar{y} is the average of the dependent variable, and N is the sample size.

The root mean squared error is defined as the square root of the mean squared error (MSE). The least squares method of linear regression minimizes the MSE since it is always greater than or equal to zero. It is also an unbiased estimator since minimizing MSE is the equivalent of minimizing the variance. To better interpret the results, we take the square root of MSE (RMSE), which yields the same units as the actual values. However, the RMSE is scale-dependent, which means we cannot compare values if their scales are different. Mathematically, it is calculated as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

where \hat{y}_i is the predicted value, y_i is the actual value, and N is the sample size.

The mean absolute percentage error is a relative error measure commonly used to evaluate regression problems. It is the mean of prediction errors as a percentage of the actual values.

Since it is a percentage, it is not scale-dependent. However, due to the division, the actual data cannot contain actual zeros, since the results are undefined. It can be calculated as:

$$MAPE = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

where \hat{y}_i is the predicted value, y_i is the actual value, and N is the sample size.

5.4 Results

In this section, we first present the fitted linear model results, and then provide a comparison and an evaluation of the results for the time, speed, and pace models. Finally, we will provide a more detailed analysis of the errors to demonstrate potential biases for each method.

5.4.1 Model Coefficients

Since we use a mixed model, there are two sections to the coefficients, fixed effects and random effects. In this section, we will first demonstrate the random effect and then the fixed effect from the linear mixed model, which is easier to interpret given its linear assumption.

Random Effects for Linear Mixed Model

Table 5.2 shows the summary of random effects or the grouping factors. To reiterate, the random effect shows the unobserved individual differences between each segment, such as different congestion levels and traffic light synchronizations. For each segment, the random effect of a given segment shows the additional changes in travel speeds that are due to the differences of the segment itself (Yang et al., 2014).

In the table, we show the standard deviation of each segment in each analysis level. As we can observe, the speed models have similar random effects, around 5 km/h, for inter-stop and stop-to-stop levels. This means that the individual differences for each segment would contribute to around 5 km/h travel speed differences. For timepoint-to-timepoint and route levels, the standard deviation are slightly lower, around 4 km/h. For the pace model, the standard deviations are low for inter-stop and route levels, around 40 sec/km. The stop-to-stop level is the highest at around 79 sec/km. The observations from speed and pace models suggest that higher analysis levels may hide more detailed variations at smaller scales. For the time model, we can observe that as the analysis level goes up, the individual time differences get longer. This is as expected since the more aggregated analysis levels tend to correlate to

longer travel distances, which would contribute to larger variations in travel times.

We also included the adjusted intraclass correlation coefficient (ICC), which explains the proportion of the total variance in travel times or speeds that can be accounted for by simply grouping the observations on the same segment alone (Hox et al., 2017). Here, we can observe that all models have an ICC above 0.5 with most of them above 0.7, which indicates there are differences between individual segments, and shows the importance of using the mixed model to account for individual segment differences. It also highlights the importance of improving our models with more detailed data that are unobserved in our study, such as traffic variations, traffic light settings, and ridership variations. The smaller analysis scales tend to have larger ICC, except for the speed model at the route level and the pace model at inter-stop level. This means that the individual differences between segments become more important at smaller scales. This further highlights the fact that higher analysis levels may hide variations in smaller levels, and more research is needed for stop-stop level scheduling, as also pointed out by other researchers (Wessel et al., 2017; Graves et al., 2019).

Fixed Effects for Linear Mixed Model

Table 5.3 shows all the fixed effect coefficients estimated from the linear model. As a quick reminder to help readers interpret the coefficients, the units used in this paper for speeds are in km/h, the units for pace are in sec/km, and the units for times are in seconds. For the description of each variable, please refer to the earlier sections. In addition, we italicized variables with a p-value less than 0.05, since most of the values are statistically significant for smaller analysis scales.

Generally, most of the coefficients and signs are as expected. In addition, we can observe the opposing signs between speed and time variables. Given a fixed distance, if the speed is higher, then the time and pace values are lower. However, the higher-level analyses tend to

Table 5.2 Random Effects for Linear Speed and Time Models

	Inter-stop	Stop	Timepoint	Route
Speed SD	5.08	5.62	3.78	4.05
Speed Adj. ICC	0.81	0.75	0.79	0.85
Pace SD	43.77	78.51	62.87	39.79
Pace Adj. ICC	0.52	0.66	0.70	0.50
Time SD	20.89	23.07	58.46	320.10
Time Adj. ICC	0.72	0.70	0.74	0.53

have fewer significant variables. To better illustrate the differences between analysis levels, we still need to demonstrate the results to better indicate to planners and operators to what extent the results are significant, since they are relevant for planners to make decisions.

Table 5.3 Fixed Effects for Linear Speed and Time Models

	Speed				Pace				Time			
	Inter	Stop	TP	Route	Inter	Stop	TP	Route	Inter	Stop	TP	Route
Intercept	27.09	20.25	12.66	<i>2.27</i>	130.80	207.90	271.53	382.62	-8.97	3.82	80.70	1138.75
Period Mar	0.21	-0.13	<i>0.01</i>	<i>-0.05</i>	<i>0.29</i>	2.72	<i>0.71</i>	<i>1.37</i>	<i>0.09</i>	0.69	<i>0.57</i>	<i>13.61</i>
Period Nov	0.39	-0.04	0.06	<i>0.03</i>	-1.30	1.87	<i>0.11</i>	<i>0.86</i>	-0.41	0.27	<i>-0.13</i>	<i>6.56</i>
Period Sep	0.69	<i>0.03</i>	0.07	<i>0.01</i>	-1.51	2.81	<i>0.46</i>	<i>0.55</i>	-0.27	0.72	<i>0.31</i>	<i>4.88</i>
Period Jun	0.69	0.22	0.16	<i>0.11</i>	-2.22	<i>0.41</i>	-1.32	<i>-1.53</i>	-0.67	<i>-0.07</i>	-1.37	<i>-11.27</i>
Bus Lane On	-0.77	-0.92	-0.93	<i>-0.43</i>	4.62	7.67	<i>2.87</i>	<i>0.66</i>	6.22	7.81	19.00	<i>72.41</i>
Bus Lane Off	-0.58	-1.16	-1.05	<i>-0.38</i>	<i>2.52</i>	7.85	<i>5.66</i>	<i>0.94</i>	4.26	5.81	13.89	<i>72.43</i>
Late Night All	5.83	9.59	6.47	5.77	-24.95	-84.60	-65.66	-72.43	-7.57	-21.08	-91.08	-620.01
AM Peak Week	0.80	1.19	1.01	0.76	-6.44	-16.71	-15.58	-14.28	-2.34	-4.78	-20.97	-107.03
Midday Week	1.25	2.16	1.64	1.88	-9.83	-25.66	-21.76	-31.82	-3.36	-7.20	-31.26	-245.68
Eve. Week	2.49	4.79	3.88	3.93	-17.83	-56.28	-51.57	-56.91	-5.57	-14.44	-67.38	-474.39
AM Weekend	3.29	5.08	4.13	4.78	-19.82	-56.41	-55.55	-71.63	-6.24	-14.47	-70.74	-587.38
Midday Weekend	1.73	3.04	2.23	2.10	-9.74	-31.26	-30.42	-35.49	-3.16	-8.10	-39.91	-282.98
Eve. Weekend	2.75	5.28	4.19	4.12	-16.91	-56.03	-52.81	-61.57	-5.25	-14.28	-70.36	-512.52
Avg Freq	-0.11	-0.18	-0.02	<i>0.02</i>	<i>-0.19</i>	1.40	2.11	<i>0.44</i>	-0.11	0.13	0.82	8.93
Num Turns	-3.41	-2.71	-0.44	-0.15	28.79	27.97	3.54	<i>0.29</i>	14.68	15.72	9.18	22.96
Avg Lanes	0.70	0.28	-0.67	<i>-1.50</i>	-2.25	<i>2.27</i>	<i>4.17</i>	<i>7.41</i>	-2.30	-1.51	<i>-1.92</i>	<i>-4.25</i>
Avg Speed	0.06	0.05	0.14	0.45	-0.24	-0.27	-0.72	-3.42	<i>-0.03</i>	<i>-0.02</i>	<i>-0.19</i>	-19.01
Num Stop Signs	-1.77	-1.02	-0.41	<i>0.00</i>	6.68	<i>-2.10</i>	2.58	<i>0.18</i>	5.47	4.44	7.53	<i>5.11</i>
Num Signals	-1.76	-2.59	-0.54	<i>-0.03</i>	11.18	23.73	3.83	<i>0.36</i>	8.96	12.70	13.20	16.78
Length Local	4.73	8.06	3.98	0.88	-41.78	-90.80	-39.49	-8.43	89.20	81.17	82.06	<i>39.04</i>
Length Collector	6.36	9.62	3.23	0.75	-44.94	-100.30	-28.51	-6.12	92.35	88.25	87.02	60.16
Length Secondary	7.86	11.13	3.95	0.68	-54.27	-116.30	-36.85	-5.74	89.04	83.79	83.45	60.67
Length Primary	5.96	8.95	3.88	0.64	-42.18	-110.10	-38.09	-5.95	99.92	93.04	78.37	69.41
Length Highway	4.25	4.52	2.50	1.05	-22.43	-26.16	-12.45	<i>-2.89</i>	48.02	46.91	51.71	54.63
Green Space	1.05	1.50	<i>0.35</i>	<i>-0.51</i>	<i>-3.62</i>	-8.52	<i>-1.16</i>	<i>2.36</i>	<i>1.41</i>	<i>0.29</i>	<i>0.21</i>	<i>13.14</i>
Downtown	-1.54	<i>0.28</i>	<i>0.12</i>	<i>-1.04</i>	30.66	25.04	20.34	18.63	3.60	<i>1.93</i>	<i>7.78</i>	<i>32.42</i>
Retail	-1.31	-1.87	-1.38	<i>-0.17</i>	10.22	23.69	12.48	<i>-0.59</i>	3.71	7.37	22.65	<i>127.01</i>
Industry	2.62	3.47	0.61	<i>-0.12</i>	-8.40	-20.02	<i>-2.19</i>	<i>0.60</i>	2.74	<i>0.47</i>	<i>-3.32</i>	<i>19.46</i>
Pop. Density	-0.11	-0.23	-0.24	<i>-0.29</i>	0.84	3.72	3.07	<i>3.21</i>	<i>-0.09</i>	0.29	2.52	<i>-2.83</i>
Dist. Downtown	0.24	0.30	0.29	<i>0.12</i>	-0.91	-2.04	-2.95	<i>-1.84</i>	<i>-0.06</i>	-0.23	-2.32	<i>-11.48</i>
Num Stops			-0.28	-0.15			1.16	0.79			4.34	10.78

More specifically, for the service period variables, all of the speed coefficients are positive, and all of the time or pace coefficients are positive or insignificant for inter-stop models. This is expected since we chose the January or the winter schedule as the base case. During the winter, the speeds or times are also affected by adverse weather events like snow storms, resulting in worse travel conditions. However, the coefficients for higher levels, such as route levels, become insignificant. Despite the buses being able to travel faster in other periods, the faster travel condition does not reduce the overall travel time significantly. Thus, future researchers could consider including additional factors affecting the overall travel times to improve the model, such as more detailed traffic signal timings.

As for the time of days, we can observe that all the speed coefficients are positive and all the time or pace coefficients are negative. Once again, this is expected since we chose weekday afternoon peak as the base case, and it is typically the most congested period. We can also observe the typical traffic variation, where the speed gets slower for the morning peak, then stays a bit faster throughout the day, and gets slower again for the evening peak, and gets faster again for the evening. Since Saturdays and Sundays do not have peak hours, the travel conditions during weekday and weekend midday periods are similar, and the weekday evening coefficients are similar to weekend mornings and evenings.

An increase in the number of stops on a segment can also result in longer travel times and paces or slower travel speeds, since buses need to start and stop more often. However, we believe the stop level ridership could be a better indicator since buses are not obligated to stop if no passenger is getting on or off. Unfortunately, we will leave this to future studies to test due to our limited data sources.

For the bus lane operations, we can observe that both variables are negative for speeds and positive for times. This is expected since we chose streets without a bus lane as the base case. The results show that streets with bus lanes are more congested than those without bus lanes, which is why the agency implemented the bus lanes on these streets. In addition, if the bus lanes are in service, the negative impact on bus speeds is generally smaller for higher analysis levels, bringing these segments more in line with less congested segments without bus lanes.

As for the average frequency, the coefficients are mostly negative for speed and mostly positive for time and pace. Again, this is expected, since service frequency correlates to the ridership. More ridership means buses will spend more time dwelling at stops, thus reducing the speed and increasing the time and pace.

The number of turns also negatively affects the bus speeds. This is typically due to buses having to slow down to manage the turn as well as to yield to pedestrians and other vehicles.

The number of lane coefficients are all negative for the time models. They are also positive for speed models and negative for pace models at smaller scales. More lanes mean wider streets, which typically correlate to better manoeuvrability. However, more lanes can also correlate to more traffic volume and more traffic signals, which could contribute to slower speeds and paces at higher levels.

Similarly, the speed limits are also positive for bus speeds and negative for time and paces. However, it is mostly not significant for the time models. Stop signs and traffic lights also negatively affect the bus speeds since buses are obligated to stop before them.

As the distance between stops increases, vehicles typically have more time to accelerate to a higher speed. Thus, all the speed variables are positive and all the pace coefficients are negative. As for the time models, since the distance units are in kilometres and time units are in seconds, the coefficients can be interpreted as the pace to travel one kilometre on a given street. Hence, the smaller the coefficient, the faster a bus travels through a kilometre. Take the inter-stop time as an example, travelling one kilometre on a local street is roughly 90 seconds, which corresponds to 40 kilometres per hour. Notice that the coefficients are fixed for the time models, which might be too restrictive since not all local streets behave the same. It may become problematic if we try to create a schedule for a new area without historical data.

As for the land use variables, we used residential land use as a base case. Vehicles can travel faster near parks and industrial areas since they typically correlate to longer street block distances. In downtown and commercial areas, vehicles typically travel slower, which is expected due to the higher traffic, ridership, and pedestrian counts in these areas. Thus, it is important for agencies to improve service in these areas to improve passenger experiences. Most variables are also not significant in time models, suggesting that time models may be too restrictive for the spatial variations. Interestingly, most land use variables are not significant at route levels, which suggests, again, that faster travel conditions do not necessarily translate to overall time reduction.

Vehicles also travel slower in densely populated areas, which makes sense since it correlates to higher ridership, traffic, and pedestrian counts. Vehicles can travel faster in suburban areas further away from the city center since it generally correlates to a decrease in ridership and pedestrian volume. These two variables are, again, not significant for higher analysis levels.

To summarize, the coefficients behave as we expected. However, time models consider many spatial variables, such as land use, to be not significant. It may be too restrictive and underestimate the spatial variations in case we try to plan for a new route for a new neighbourhood. Thus, we need to test their performances more closely in the next subsections.

5.4.2 Model Comparisons

In this subsection, we compare the advantages and disadvantages of the above models. First, we use a few common aggregated measures in previous literature to summarize the performance of the models. Then, we provide a more disaggregated view to compare and evaluate the models to demonstrate the potential issues of using aggregated measures.

To reiterate, we created two testing scenarios to evaluate the models by holding back some data from the overall dataset. One scenario is service expansion with new routes where there are no existing data. We reserved 10% of all segments in each analysis level for testing. Another scenario is expanding the service hours on an existing segment. We reserved 10% time-of-the-day data from the segments left for testing. We test both modelling methods, the linear mixed model and mixed effect random forest. For each method, we tested the models using both scenarios. Since the time model and the normalized speed and pace models are not directly comparable, for the normalized models, we then converted the results to travel times for comparisons in this section.

Aggregated Measures

Table 5.4 shows the results of the commonly used aggregated error measures of each model using the two testing scenarios as well as the existing data used to fit the model. The better-performing models in each category are marked with bold fonts. Readers can refer to earlier sections for the definitions of these measures.

In general, the random forest method performs slightly better than the linear method. The R^2 values are generally higher and the RMSE and MAPE measures are lower for the random forest models. The differences in MAPE are generally around one to three percent. The differences in RMSE between the two methods are generally around a few seconds. However, the RMSE differences are larger for service pattern levels due to their longer distances, and a small percentage error can correspond to a relatively larger absolute error. Overall, given the segment-specific intercepts and the additional variables available to us, the differences between the two modelling methods are not too large.

We can also observe that the prediction results for the new service hours scenario are similar to the modelling results from existing services used to fit the models. This result suggests that planners can make a good estimate if they wish to expand service hours on existing segments based on historical observations. The modelling results from the existing services and the new hours scenario perform better than the new routes scenario where no historical observations are available, which is to be expected. Future research could potentially improve

Table 5.4 Aggregated Model Error Measures

		Existing Services		New Hours		New Route		
		Linear	Forest	Linear	Forest	Linear	Forest	
Inter-Stop	Time Direct	R^2	0.900	0.920	0.931	0.933	0.459	0.805
		RMSE	16.139	14.055	12.856	12.777	38.196	17.246
		MAE	3.608	2.985	3.699	3.238	10.564	7.835
		MAPE	0.083	0.067	0.085	0.072	0.293	0.201
	Time from Speed	R^2	0.854	0.859	0.903	0.913	0.794	0.827
		RMSE	19.338	19.199	16.005	14.505	18.190	16.228
		MAE	4.298	3.886	4.318	4.010	7.625	6.975
		MAPE	0.099	0.089	0.099	0.093	0.172	0.154
	Time from Pace	R^2	0.861	0.893	0.909	0.925	0.683	0.794
		RMSE	18.834	16.672	14.707	13.459	25.046	17.717
		MAE	4.549	3.808	4.526	4.066	8.733	7.643
		MAPE	0.111	0.096	0.112	0.101	0.217	0.180
Stop to Stop	Time Direct	R^2	0.903	0.910	0.919	0.928	0.463	0.736
		RMSE	17.145	16.097	15.185	14.394	40.615	23.505
		MAE	6.793	5.149	6.977	5.743	15.182	13.448
		MAPE	0.137	0.099	0.141	0.110	0.313	0.288
	Time from Speed	R^2	0.556	0.859	0.653	0.902	0.487	0.790
		RMSE	33.987	20.670	27.789	16.792	29.749	22.902
		MAE	8.448	6.894	8.425	7.235	15.068	13.349
		MAPE	0.147	0.123	0.147	0.129	0.241	0.231
	Time from Pace	R^2	0.825	0.896	0.855	0.917	0.464	0.684
		RMSE	23.236	17.734	19.564	15.426	35.211	25.054
		MAE	8.436	6.137	8.530	6.703	15.399	14.331
		MAPE	0.165	0.126	0.167	0.135	0.314	0.292
Timepoint to Timepoint	Time Direct	R^2	0.957	0.978	0.956	0.973	0.848	0.867
		RMSE	34.136	24.765	34.953	27.242	60.234	56.452
		MAE	22.018	15.158	22.786	16.979	42.435	41.876
		MAPE	0.127	0.064	0.130	0.071	0.214	0.246
	Time from Speed	R^2	0.951	0.966	0.949	0.960	0.846	0.875
		RMSE	36.820	30.422	37.314	33.111	60.703	54.770
		MAE	20.729	16.594	21.272	18.381	42.136	38.804
		MAPE	0.086	0.070	0.085	0.075	0.171	0.159
	Time from Pace	R^2	0.941	0.975	0.939	0.970	0.791	0.855
		RMSE	40.274	26.311	41.063	28.548	70.783	58.810
		MAE	24.131	15.910	24.841	17.690	45.251	42.651
		MAPE	0.101	0.070	0.101	0.076	0.193	0.189
Route	Time Direct	R^2	0.874	0.917	0.864	0.868	0.829	0.786
		RMSE	295.737	240.053	305.717	301.758	337.412	377.801
		MAE	153.583	114.945	151.483	129.140	243.657	288.020
		MAPE	0.081	0.057	0.081	0.064	0.120	0.153
	Time from Speed	R^2	0.870	0.904	0.861	0.866	0.722	0.843
		RMSE	300.496	257.673	309.200	304.213	430.785	323.698
		MAE	122.478	102.826	122.196	113.633	287.925	222.926
		MAPE	0.056	0.047	0.057	0.051	0.128	0.104
	Time from Pace	R^2	0.813	0.933	0.822	0.877	0.786	0.814
		RMSE	360.062	215.130	350.678	290.761	378.238	352.399
		MAE	182.558	96.575	182.282	114.164	232.938	250.299
		MAPE	0.088	0.047	0.091	0.055	0.105	0.124

the results by integrating more detailed traffic data and ridership estimations into the model. The direct time models perform better for existing segments in new service hours, whereas the normalized speed models perform better for reserved segments in the new routes scenarios. In our models, we included both speed-related, such as streetscapes and land use, as well as time-related variables, such as the number of traffic signals, which are related to a fixed time coordinated plan regardless of speed. The random intercepts included in the time models could help alleviate some limitations due to the lack of detailed signal timing plans or ridership counts. We believe the speed models might be more intuitive for the new route scenarios when there is no observed time-related information available, such as traffic light timing plans. This suggests that planners could potentially use speeds from existing similar segments as a starting point when planning for a new route, which is in line with current practices.

The results from the pace model perform in between the time and speed models. They do not "win" in any category except for the route-level. They generally perform better than the speed model for the new service hours scenario. However, they generally perform worse in the new route scenario. To reiterate from the literature review, pace is less used as an input for planners, as it is more used in the performance evaluation context. More research and examination are still needed to improve and apply the pace models.

Despite missing some detailed time information, the inter-stop models, which does not include the dwell times and signal waiting times, perform better than the stop-to-stop models. To improve stop-to-stop models, future work can consider adding more detailed traffic signal and ridership information by creating a hybrid model and combining both time and speed models.

For higher levels, namely timepoint-to-timepoint and the route levels, the differences between the direct time model and the normalized speed and pace models become smaller. The evaluation for new route scenarios is also relatively better compared to the two smaller scales. This once again highlights the fact that higher analysis levels may hide more detailed time variations in smaller levels. Thus, we add further evidence for the need to examine bus travel time modelling at these smaller scales.

Another interesting observation at these higher levels is that the different measures are more likely to indicate different "winners" in the same category, even for the existing data used to fit the models. For example, in the new routes scenarios at the timepoint-to-timepoint level, the R^2 and RMSE measures would indicate that the direct time model is better, but the MAE and MAPE measures would indicate that the result from normalized speed models performs better. This indicates some potential biases that different error measures might

reward. For example, the MAPE measure prefers to forecast lower values (Goodwin and Lawton, 1999), which is once again related to the original question of this paper. Due to the length differences, we may prefer smaller errors on shorter segments, and we may tolerate slightly larger errors on longer segments. A 20-second error may be great for a segment of 15 kilometres. It may not be as desirable for a short segment of 150 meters. For scheduling, planners may prefer to add some schedule padding to improve on-time performance. However, for arrival time predictions for passenger information, agencies may prefer to underestimate travel times to ensure vehicles do not leave passengers behind given a travel time prediction. Thus, transit planners need to decide if we would prefer certain biases when we model our transit systems, since different error measures evaluate the results "from different angles" (Kolassa, 2020). To illustrate these different biases, we need to analyze the errors in more detail in the next subsection.

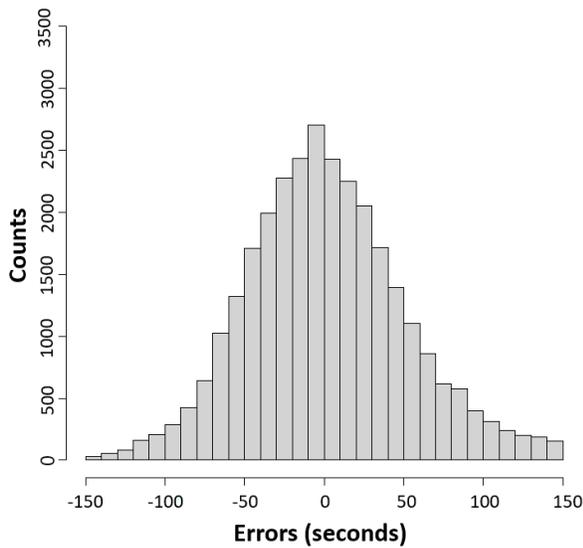
Disaggregated Measures

In this section, we will demonstrate some additional biases in these models that might influence our model choices. In the previous subsection, we observed a few cases where the error measures indicated different "winners" in higher-level analyses. In this subsection, we will use the direct and normalized linear mixed models at the timepoint-to-timepoint level for the new routes scenario as an example for simplicity since the observations are similar for the other models.

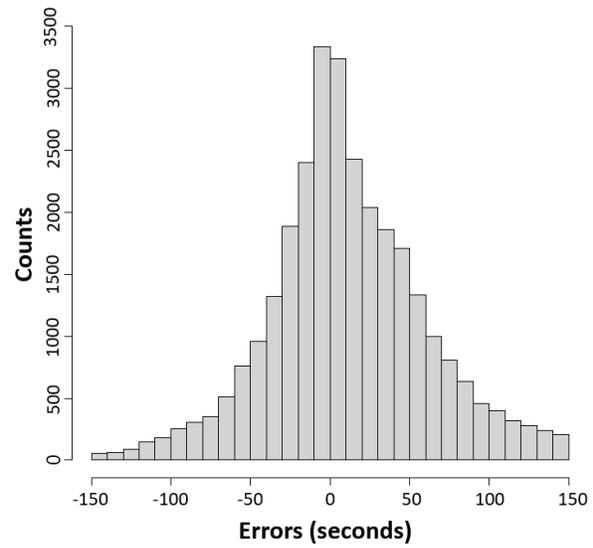
Figure 5.2 shows the error distributions for the direct time results and normalized model results from the speed and pace models. To help readers see the differences, we use the same x and y scale for these figures.

The errors for the direct time model (Figure 5.2a) are closer to the normal distribution. This is expected since the models try to directly minimize the MSE, which yields unbiased estimates. Given the long travel times on longer segments, the direct model may place more emphasis on long segments than on shorter segments. However, the errors from the two normalized models are more centered around 0.

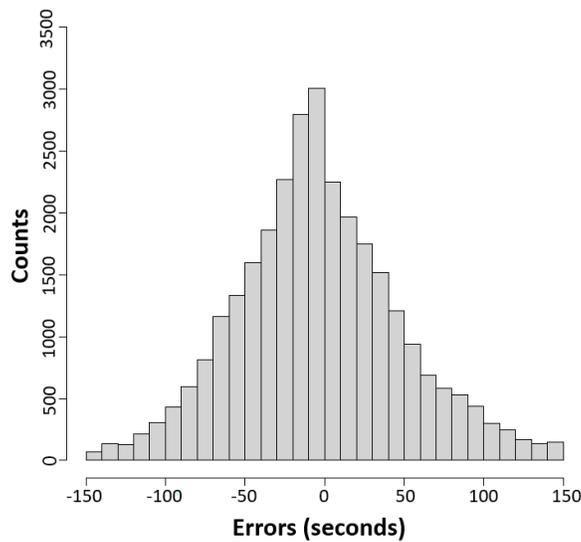
The speed results (Figure 5.2b) are skewed towards the right, whereas the pace results (Figure 5.2c) are skewed slightly to the left. In other words, the speeds models tend to underestimate the average travel times, but the pace models tend to overestimate the average travel times. In addition, speeds are bounded values between zero and the top speed of the vehicles, whereas time and pace values do not have an upper bound. Thus, the congested observations may have more impact on the average time measures.



(a) Errors from Direct Time Model



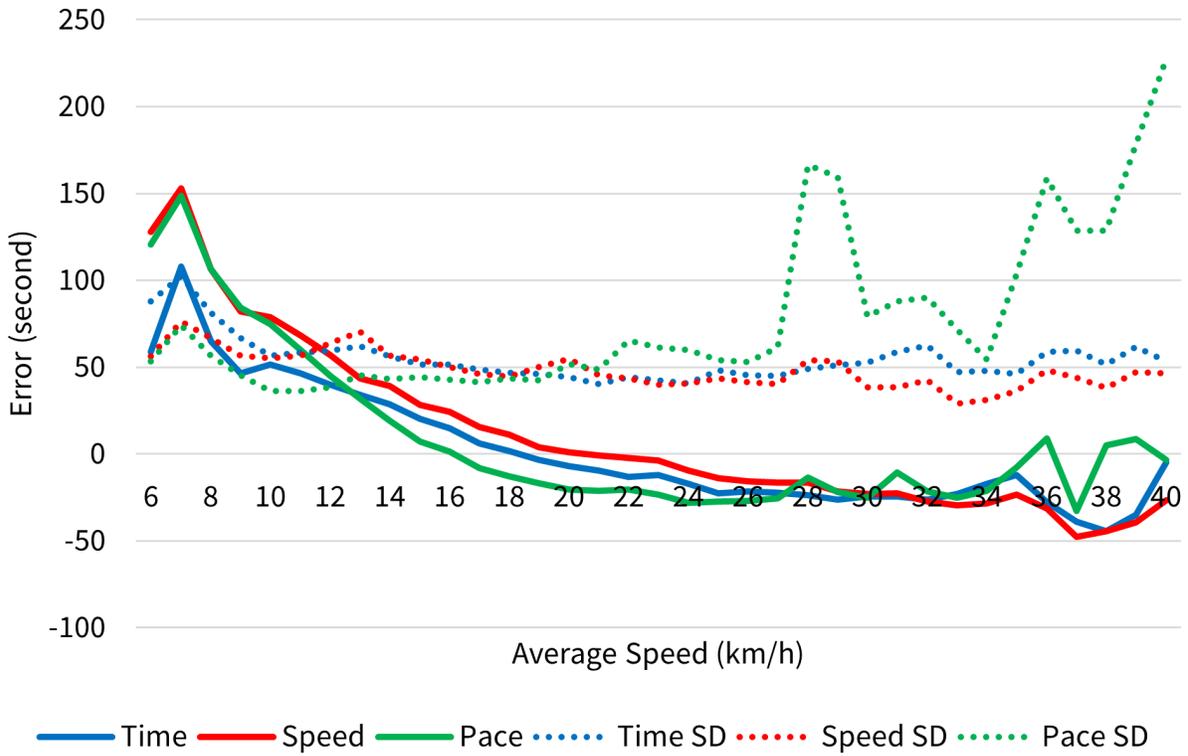
(b) Errors from Normalized Speed Model



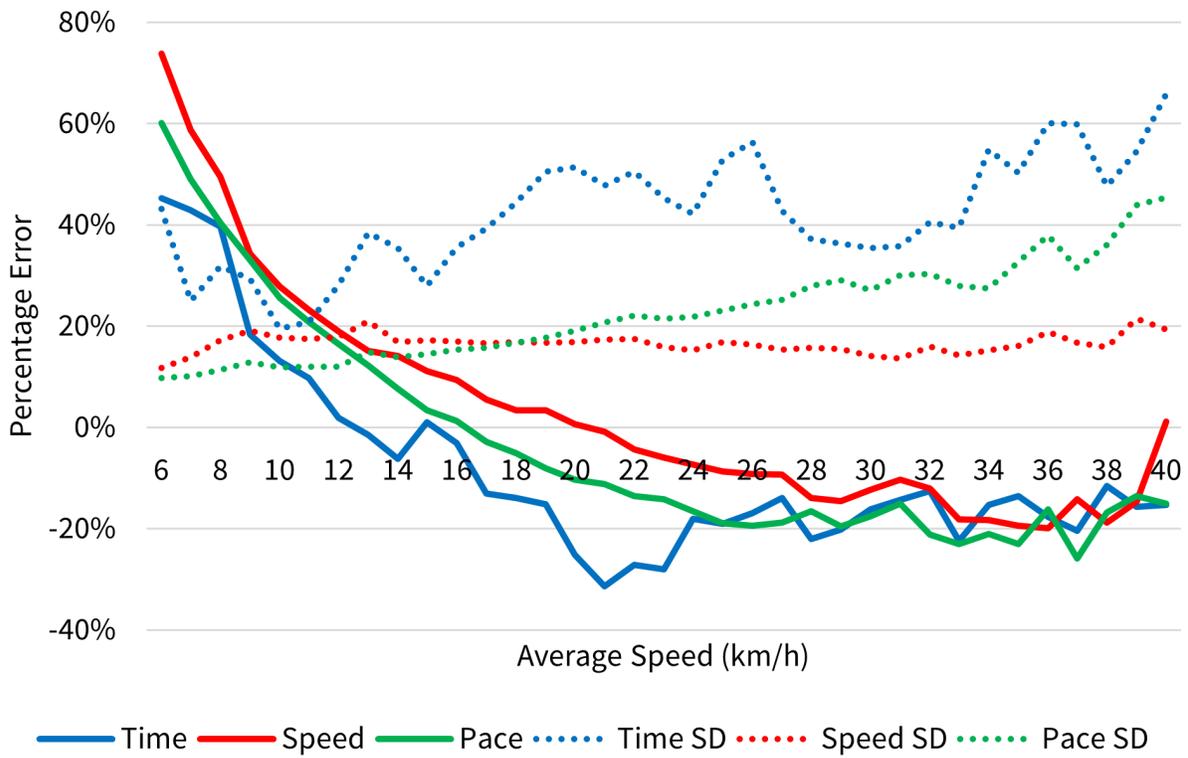
(c) Errors from Normalized Pace Model

Figure 5.2 Error Distributions for Time, Speed, and Pace Models

Figures 5.3a and 5.3b show the modelling errors aggregated by the observed average segment speed. Figure 5.3a shows that the average errors follow a similar conditional average and conditional standard deviation for both the direct and the normalized model results. This makes sense, since the models try to minimize the errors directly.



(a) Error



(b) Percentage Error

Figure 5.3 Errors Aggregated by Observed Average Segment Speed

The percentage errors for the normalized results from speed and pace models follow a more stable conditional average, whereas the direct time model has more varied conditional averages given the segment speed. This illustrates that the direct time models are slightly less reliable in relative measures, which might be again due to the fixed distance coefficients.

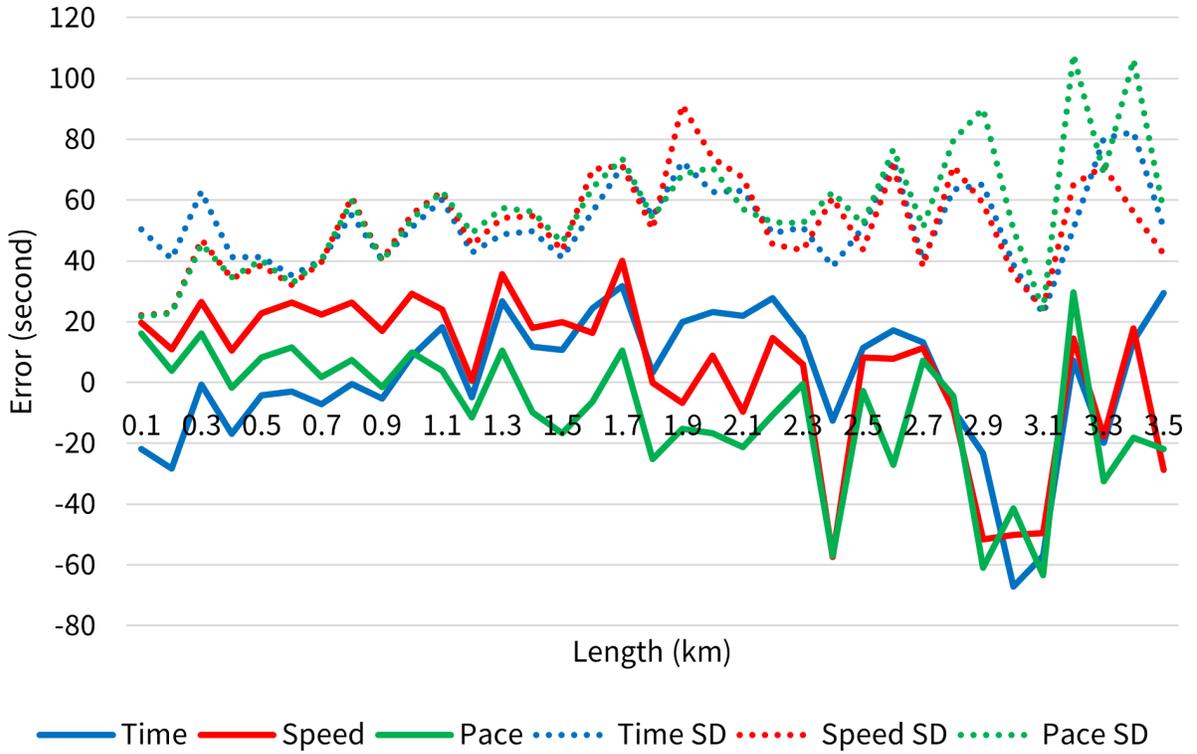
As for the standard deviations of errors, the time and speed models perform mostly with the same trend for medium speeds. However, variation from the time model is higher for slower and faster segments, which once again highlights that time models may be less reliable for certain segments. The standard deviations from the pace model also become significantly larger for faster segments. As segments get faster, the changes in pace become smaller. Since faster segments tend to be long highway sections, small variations in pace could contribute to larger travel time variations.

Despite the similar conditional average percentage errors for the normalized models, the standard deviations become higher for faster segments in the pace model, which again could be smaller pace differences for faster segments. The standard deviations for time models are also consistently higher than the normalized models. We once again believe the length coefficients from the time models may be too restrictive for the models to adapt to different segment lengths or speeds.

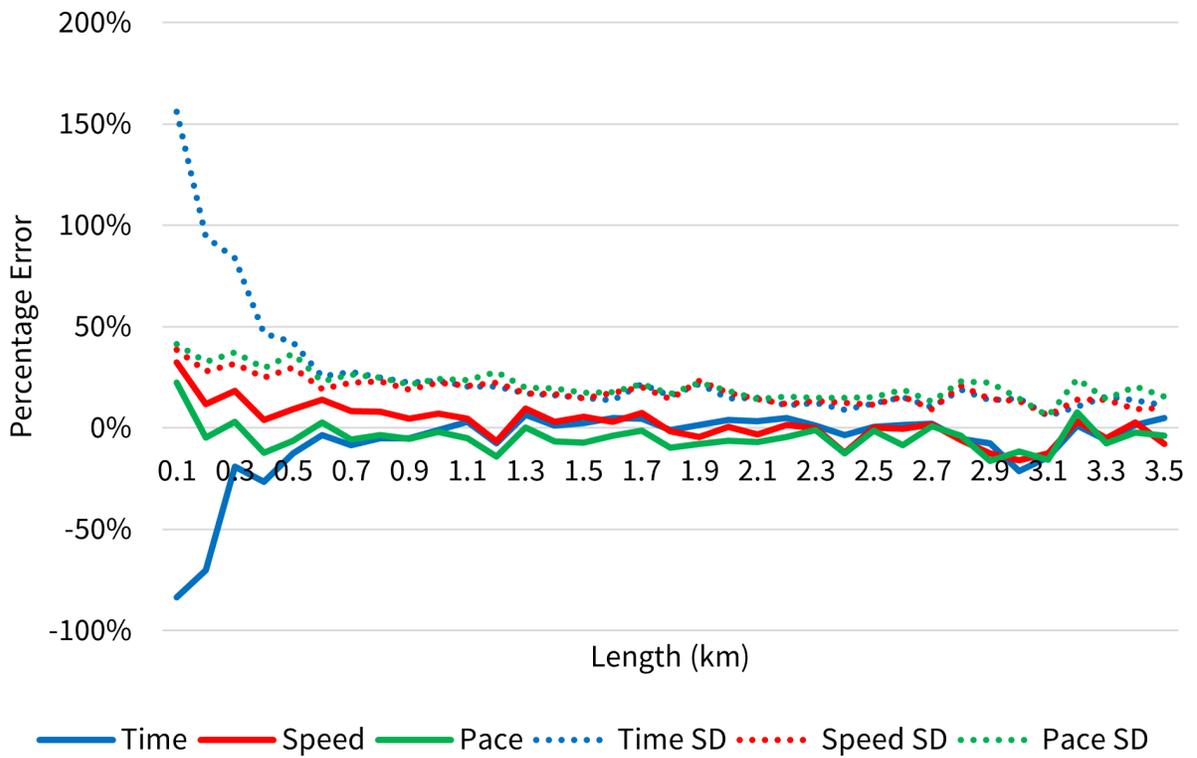
Both figures show that the models underestimate the travel time on slower segments but overestimate on faster segments. This is also expected since the models were given little information on pedestrian counts, congestion, or traffic light timings, which would be more relevant for slower segments, whereas the faster segments typically include long sections on highways with few traffic lights or sections in areas without congestion. We once again highlight the need to include traffic light timings and traffic levels in the modelling process in future works.

Finally, Figures 5.4a and 5.4b show the modelling errors aggregated by the segment length. From Figure 5.4a, we can again see that the errors from all three models follow similar trends for conditional averages. Once again, this makes sense since the models try to minimize the errors. The speed errors are generally stable for segments less than 1.8 km. However, the errors gradually increase for the direct time model and decrease for the pace model. This shows that the time and pace models may have put too much emphasis on segments that are roughly 1.1 km long and became less flexible for other segments.

The standard deviations for all three models follow the same general trend, varying around 40% to 60%. However, the variation becomes larger for longer segments and the pace model has the largest variations for longer segments.



(a) Error



(b) Percentage Error

Figure 5.4 Errors Aggregated by Segment Length

Figure 5.4b, which shows the aggregated percentage errors by segment length, highlights the large differences between the two models for shorter segments typically found on regular local routes. The direct time conditional percentage errors vary to as much as -80%, whereas, the normalized results from speed and pace models have percentage errors around 0% to 30% for short segments, much more stable compared to the direct time model. This indicates that despite the similar aggregated errors between the direct and normalized models, small changes in the model result can lead to relatively larger differences relative to the actual observed values on these shorter segments. For longer segments, the direct and normalized models become much more similar, which again shows the potential biases that the direct time model may place too much emphasis on long segments given their long travel times.

Similar observations can be found for the conditional standard deviations. The conditional standard deviation varies up to 150% for short segments in the time model. The conditional standard deviations for the two normalized models are much more stable across various segment lengths at around 25% to 40%. This observation once again highlights the fact that small changes in the model result can lead to relatively larger differences.

These larger relative errors for shorter segments also make sense. The traffic light waiting times or dwell times at stops become a more significant portion of the travel time for the shorter segments. Thus, we again emphasize the need to include more detailed traffic light timing and ridership data in the models or to create a hybrid model to include them in the analysis. In addition, shorter segment lengths correspond to the local services, where vehicles make every single stop. Typically, local services represent the majority of services provided by transit agencies. This suggests that normalized speed or pace models perform relatively better for shorter segments and local services. Thus, transit planners need to consider their specific planning context, whether they would prefer certain biases in the model or if they are willing to accept the larger relative errors on short segments. Again, since different measures evaluate the results "from different angles" (Kolassa, 2020), transit planners and future researchers need to think more about which measure is more suitable given a specific context.

5.5 Conclusion

To summarize, good travel time estimations are important for both transit agencies and passengers, who rely on good travel time estimations for their decision-making processes. However, travel times are the results of varying speeds, paces, and distances. Given the same speed, longer segments will have longer travel times, whereas shorter segments will have shorter travel times. Similarly, given the same distance, a faster speed will result in shorter

travel times, and a slower speed will result in longer travel times. Hence, it is possible to use time, speed, or pace as inputs for planning purposes.

Most of the previous literature focuses on travel times, and normalized measures, like travel speeds or paces, are typically used to evaluate delivered services or to plan infrastructures. Thus, we raise the question, how do we compare the effectiveness of these common transit measures when we model a transit system?

We hypothesized that normalized measures may be better suited when evaluating transit performances or planning for transit schedules since it does not depend on travel distances like the travel time measures. In addition, there are many analysis levels when analyzing transit services. The current scheduling or service planning practices typically focus on timepoint-to-timepoint or service pattern travel times. For passengers, they typically focus on the travel or arrival times at specific stops, since they do not necessarily travel between timepoints. We also consider that inter-stop level may be more suitable for understanding travel conditions since it does not heavily depend on signal timings and dwell times.

Thus, in this paper, we proposed a framework to compute and compare the travel time, speed, and pace measures commonly used by transit agencies at various analysis levels. To test these measures, we came up with two scenarios. One is to test the expansion of service areas using new routes, and another is to test new service hours for existing services.

Our simple models show that the modelling results are in line with our expectations. Our evaluations show that the non-linear models perform slightly better. Transit travel times, speeds, and paces are greatly impacted by temporal variables, like time of the day, spatial variables, like street classifications and the number of traffic lights, as well as operational variables, such as service frequency and ridership. However, most other spatial variables like land use are not significant for travel time models. The distance coefficients in travel time models are fixed paces, i.e. the time to travel 1 kilometer, which may be too restrictive to deal with the changing segment lengths and speeds in reality.

Spatial, temporal, and operational variables can explain higher analysis levels much better, such as timepoint-to-timepoint and service pattern levels. For lower levels, the inter-stop level performs better than the stop-to-stop level. These results show that the analyses at higher levels may hide more detailed variations at lower levels. Improving the results at these lower levels requires more detailed data and further studies. Therefore, we emphasize the need to further include detailed dwell times and traffic light timings in the modelling processes to improve the models. In addition, since dwell times and traffic light timings are related to the times, whereas travel conditions are related to speed or pace, future researchers could also test a hybrid model using detailed inter-stop speed, traffic light timing, dwell times, etc.

Segments with existing observations can help greatly when trying to estimate travel times for expanding service hours, since they have a segment-specific intercept to take account of the differences between segments. However, the normalized results from speed models typically win for the new route scenarios, given the lack of segment-specific intercepts. We believe planners could more intuitively estimate how fast buses can travel along the segments based on local knowledge and environmental variables when planning for a new route.

We also highlight the shortcomings of using specific aggregated measures in previous literature, since different measures evaluate the results "from different angles" (Kolassa, 2020). A more disaggregated error analysis shows that the speed models tend to underestimate the average travel times since the speeds are less affected by extreme values, such as extreme weather events. Pace models, on the other hand, tend to overestimate average travel times. Direct travel time models tend to have less skewed errors. All three approaches perform similarly in terms of the average errors. However, the speed and pace models perform relatively better and more consistently relative to the actual values. Time models tend to struggle more with faster average speeds and short segments, which can be attributed to the fixed paces in the coefficients. Thus, transit planners and future researchers might want to spend more time experimenting with which measures to choose given a specific planning context.

We acknowledge that this paper is by no means an exhaustive evaluation of all possible measures and models used in transit planning. Our goal here is to introduce additional nuances when planning for a transit network or analyzing the modelling results. Future researchers could easily adopt and expand upon this framework to test new methods with additional variables, such as weather, signal timings, ridership variations, and congestion level to better help agencies plan and react to changes in the network for their operations. In addition, future researchers could also compare and experiment with many other modelling methods, such as time-series and artificial intelligence methods.

Finally, we want to mention that our research is limited in terms of passenger experiences, since we do not have detailed passenger data. It is important to consider the impact of service delivery on passenger experiences, since passengers may shift to other modes if their experiences are bad. We have mostly talked about how transit vehicles may get stuck in traffic. However, for passengers, another potential way to get stuck is missing a transfer. Thus, operational measures like the ones we compared may not necessarily reflect passenger experiences. Therefore, future researchers could also introduce additional analysis levels by incorporating passenger experiences.

Declarations

Data Availability

The authors would like to thank the Société de Transport de Montréal, the city of Montréal, and OpenStreetMap contributors for providing the data used in this research.

The research data are freely available online. However, given that they are updated continuously, we do not guarantee they remain unchanged for future studies. The archived real-time data and the code can be provided upon reasonable request to the authors.

Funding Acknowledgements

This research is funded by the Natural Science and Engineering Research Council of Canada and the Canada Research Chair in Transportation Transformation.

Author Contributions

The authors confirm their contribution to the paper as follows: study conception and design: Yuxuan Wang, Catherine Morency, Martin Trépanier; data collection, analysis and interpretation of results: Yuxuan Wang, Catherine Morency, Martin Trépanier; draft manuscript preparation: Yuxuan Wang, Catherine Morency, Martin Trépanier. All authors reviewed the results and approved the final version of the manuscript.

Compliance with Ethical Standards

The authors declare no conflict of interest.

This research does not involve human participants and / or animals.

This research does not involve informed consent.

CHAPTER 6 ARTICLE 3: DECOMPOSITION AND SENSITIVITY ANALYSIS OF BUS TRAVEL TIMES

Yuxuan Wang^{acde}, Catherine Morency^{acde}, Martin Trépanier^{bce}

^a Department of Civil, Geological and Mining Engineering, Polytechnique Montréal, Montréal, Québec, Canada

^b Department of Mathematics and Industrial Engineering, Polytechnique Montréal, Montréal, Québec, Canada

^c Interuniversity Research Centre on Enterprise Networks, Logistics, and Transportation, Montréal, Québec, Canada

^d Canada Research Chair in Mobility, Montréal, Québec, Canada

^e Canada Research Chair in Transportation Transformation, Montréal, Québec, Canada

Paper accepted for publication on 25 September 2025 for Transportation Research Record

DOI: <https://doi.org/10.1177/03611981251391715>

Abstract

Transit service reliability is important for transit planning and operations as well as passenger experience. Large travel time variations increase operating costs and negatively affect passenger satisfaction. Existing literature focuses on specific aspects of transit travel times but less on how these aspects interact with each other. This paper proposes to combine previous research efforts by further decomposing observed trip travel time into 4 elements using 3 months of archived vehicle location and fare transaction data. Departure times and inter-stop travel times are obtained from vehicle locations. Dwell times at stops are estimated from fare transaction data using a dwell time model. Red light waiting times are calculated using the vehicle locations and estimated signal timing plans. Then, using these as inputs, we identify important trip elements affecting the overall travel time variation, as well as how much variation can be attributed to each trip element using variance-based and one-at-a-time sensitivity analyses.

The overall travel times and red light waiting times are more affected by interaction effects between trip elements, whereas the overall inter-stop times and dwell times are mainly affected by large individual variations. The results suggest that planners must consider potential

chain reactions where small variations in one trip element can lead to significant changes in the overall trip times due to interaction effects with varying cycle lengths in fixed signal timing plans. These findings will help planners better integrate available datasets, carry out comprehensive analyses, and pinpoint the determinants affecting travel time variation on each route.

Keywords: Transit Planning, Transit Travel Times, Travel Time Variations, Sensitivity Analysis

6.1 Introduction

Reliable travel time is important for both transit agency operations and passenger satisfaction. Transit agencies and planners have been striving to achieve better reliability. From the agencies' perspective, travel time reliability affects both vehicle and operator scheduling. Transit planners typically add schedule padding or layover times to account for potential travel time variations, which will, unfortunately, increase the operating costs (Danaher et al., 2020). Missing a layover will also propagate delays to downstream trips and cause driver dissatisfaction issues (Danaher et al., 2020). Unreliable travel times will also force passengers to budget additional time to arrive at their destination on time, which in turn affects their satisfaction and mode choice (Carrel et al., 2013a). Some passengers value reliable service more than service frequency and faster travel times (Balcombe et al., 2004; Daskalakis and Stathopoulos, 2008).

Many studies have looked at the reliability of transit travel times from specific perspectives, such as travel times on various analysis levels (e.g. timepoint to timepoint) (Muller and Furth, 2001), dwell times at stops (Dueker et al., 2004), as well as signal priority measures at intersections (Furth et al., 2025). However, little attention has been paid to how these different elements interact with each other. For example, due to signal synchronization, if the vehicle always arrives at the stop during the red light, maybe dwell time variation becomes less critical. On the other hand, if the vehicle always arrives during the green light, the previous inter-stop traffic and dwell time can become more critical when the driver tries to cross the intersection before the light turns red. Thus, it is still important to consider the interactions between different trip elements as well as to quantify the sensitivity of each trip element.

Hence, we hope to better integrate various data sources already available to most agencies, such as vehicle position and fare collection data. Then, we aim to carry out a more detailed comprehensive analysis that would include all these various aspects to better describe these factors affecting transit travel time variations. Finally, we also attempt to help planners better identify the important determinants affecting travel time variation among these factors on each route, which will help create specific strategies to improve travel time reliability.

This paper aims to combine the previous research efforts by further decomposing the travel times by splitting the overall travel times into various trip elements. Then, we will conduct a sensitivity analysis using variance-based sensitivity analysis and one-at-a-time analysis. This would allow us to answer the two questions: Where do travel time variations come from? How much travel time variation can be attributed to each trip element, i.e. departure time,

inter-stop time, ridership change, and traffic signal timing change?

We propose to answer these two questions by further decomposing observed trip travel times using 3 months of archived transit data from various sources in Montréal, Canada. The existing literature focused on three aspects of the total travel time, inter-stop time (congestion), dwell time (ridership), and red light waiting times (traffic signals). The inter-stop times, i.e. how long the vehicle travelled from one stop to another, are easy to obtain from the vehicle location system. Unfortunately, due to data availability issues, we will create two models to estimate dwell times and red light waiting times. The dwell time, i.e. how long the vehicle is stopped for passenger activities at a given stop, is estimated based on the ridership observations, where we used 25,000 on-board observations to estimate a dwell time model, and the model is applied to the ridership data obtained from automated fare collection where there are no on-board observations. Red light waiting times, on the other hand, are the interaction between the vehicle arrival times at the signal and the signal timing plan, which we will have to calculate separately. The arrival time can be calculated using the departure time plus the travel time up to the given intersection. The traffic signal timing plans, including the timing plan changes, offset, red length, and cycle length, are estimated based on the vehicle location observations using the methodology proposed by Fayazi et al. (2015). Since traffic signal timings are not under the agency's control, we treat them as a fixed input and analyze the variations in red light waiting times as a result.

Finally, we conduct two sensitivity analyses on these decomposed times to demonstrate the importance of each trip element. First, we propose to use the variance-based analysis, followed by the one-at-a-time analysis. The variance-based analysis is a global method that can handle interaction effects and non-linearity among different variables (Saltelli et al., 2008). The results provide two indices regarding the proportion of variation that can be attributed to each trip element, one with the interaction effects and one without. We also perform some example one-at-a-time analyses to demonstrate the non-linearity observed between the variation of a trip element and the overall variation in trip travel times.

The results will enhance our understanding of travel time variations and help planners identify certain locations or trip elements affecting travel time variations. In turn, these insights could help agencies target specific issues, choose the appropriate strategy to improve the reliability of a given route, develop more robust transit schedules, and thus improve passenger experiences.

This paper is structured as follows. We will summarize existing research contexts in the literature review section. Then, we will explain the research framework, data sources, and detailed methodology in the methodology and data section. Next, we will show the result of

our case study using data from Montréal, Canada in the case study section. Finally, in the conclusion and future research section, we summarize our work and results in this paper, as well as showing potential future expansions to this work.

6.2 Literature Review

Transit reliability measures are commonly used by transit agencies in their planning and operations. Academics have also studied specific elements of transit reliability and proposed numerous additional measures. The focus of current literature is typically on different levels of travel time variations, ridership variations which relate to dwell times, and transit priority measures.

Research has focused extensively on transit travel times since the implementation of Automated Vehicle Location systems. Travel time variations are important for transit agencies because they serve as a key scheduling input to determine the schedule padding and layover times needed, which affect vehicle requirements (National Academies of Sciences, Engineering, and Medicine and others, 2013). The variations are typically analyzed at four different levels, line, trip, timepoint to timepoint, and stop to stop levels according to an agency survey (Danaher et al., 2020). The analyses for travel time variations have focused on typical variation measures such as the standard deviation, coefficient of variation, or a predefined percentile (National Academies of Sciences, Engineering, and Medicine and others, 2013).

Common factors affecting transit travel times are typically route length, passenger activity, and number of signalized intersections, dating back to 1984 from Abkowitz and Engelstein (1984). Researchers also show that the number of stops, direction, time-of-the-day, dwell time, and weather variations also have significant effects on route run time (Berkow et al., 2009).

Some researchers have also observed mixture travel time distributions in recent years (Ma et al., 2013; Kieu et al., 2015). A mixture distribution means travel time observations are generated first from a collection of underlying distributions or operating conditions (e.g. free-flow vs congested traffic flows) with various associated probabilities (Ma et al., 2013). Then the value is selected from the selected probability distribution. Travel times following the mixture distribution means that if we put the observed travel times in a histogram, we can observe more than one peak, unlike the Normal distribution or a skewed distribution with one single peak.

Figure 6.1 shows an example of a mixture distribution with multiple peaks in the distribution shape observed on the westbound route 27 departed at 6:30 am in Montréal. In the figure,

we can observe three main peaks at roughly 6:46:30, 6:48:00, 6:49:30, and many smaller peaks earlier and after the three main peaks. Each peak, separated by roughly 1 minute and 30 seconds, would represent an underlying distribution or operating condition. Given that the traffic signal cycle lengths for the last stretch of this route are 90 seconds, we would hypothesize that it relates to whether the bus misses a traffic signal cycle or not.

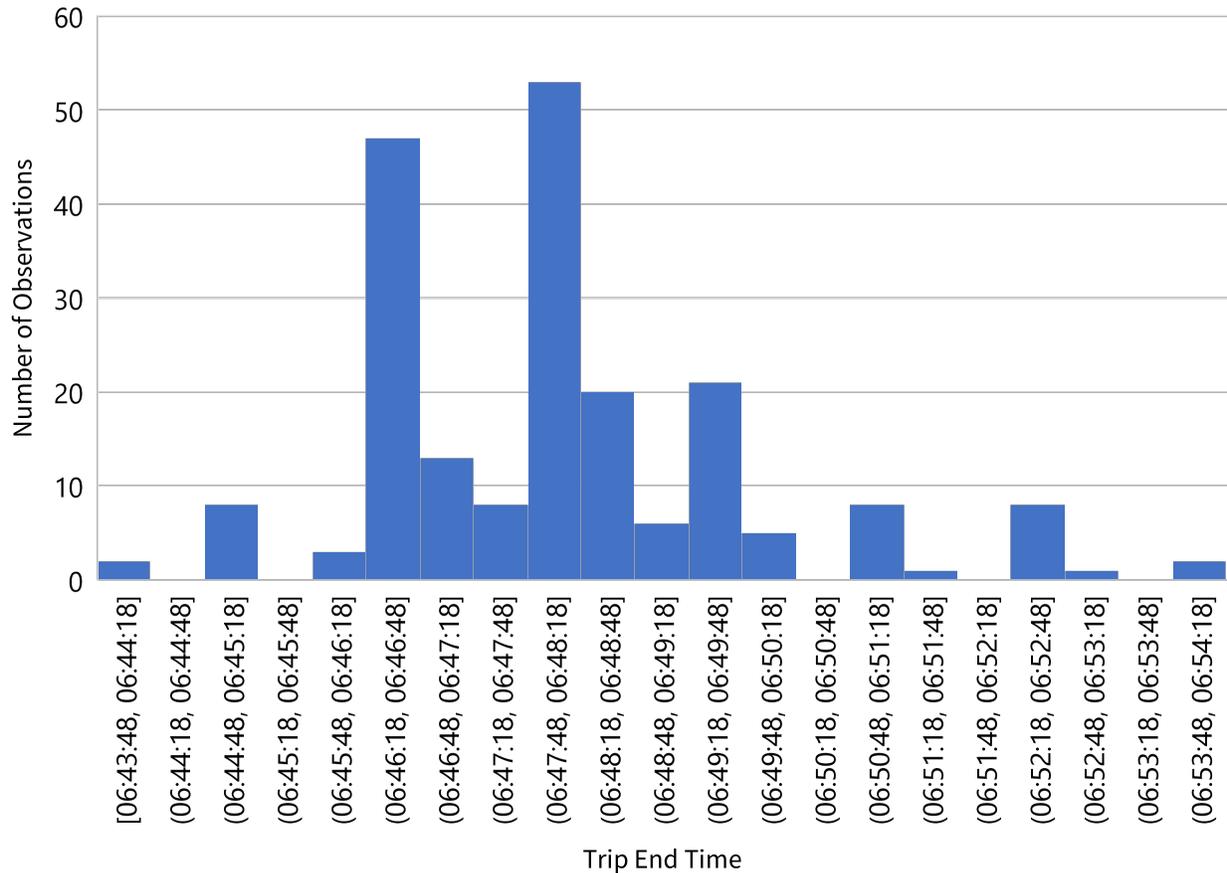


Figure 6.1 Example of mixture travel time distribution observed for trips departed at 6:30 on westbound route 27 in Montréal.

In addition, even the smallest scale so far, namely the stop-to-stop scale, consists of the travel time between the two stops, dwell time at one stop, and potentially multiple traffic signal waiting times. Buses routinely wait for traffic signals, and higher ridership at a stop could also contribute to buses missing a green light. A much finer analysis scale with the ability to better isolate various trip elements and handle interaction effects among these elements is needed to attribute the travel time variation to a specific issue or a combination of issues.

Dwell times have also received high coverage in previous studies. It is defined as the time a vehicle spends at stops for passenger boarding and alighting, typically the time between the

door opening and closing (National Academies of Sciences, Engineering, and Medicine and others, 2013). Similar to travel time studies, dwell time studies have also focused on typical variation measures such as the standard deviation, coefficient of variation, or a predefined percentile (National Academies of Sciences, Engineering, and Medicine and others, 2013). Some studies analyzed stop-level dwell times, which can be compared with vehicle load to understand the source of dwell time variability. The variability could be due to passenger boarding and alighting activities, existing crowding in vehicles which makes boarding and alighting difficult, fare payment methods, and ramp usages (Dueker et al., 2004; Rashidi et al., 2023; Currie et al., 2013). However, there is less attention in the existing literature on how dwell times might impact the overall travel times.

The Transit Capacity and Quality of Service Manual (National Academies of Sciences, Engineering, and Medicine and others, 2013) does not consider the time a vehicle remains stationary at the stops after passenger boardings and alightings as dwell time, such as red light waiting time. Most traffic signal-related studies in the transit context are related to transit signal priorities for transit vehicles. Many studies show positive impacts of transit priority signals on reducing transit travel times (Lin et al., 2015). However, some other studies failed to show any significant travel time gains from signal priority, and they speculated that the no right turn on red policy combined with near side stops hinders bus departures (Diab and El-Geneidy, 2013). Many previous studies also only considered the number of priority signals in the model and didn't include any detailed signal timing information that could have contributed to the situation.

For some cities, especially in downtown areas, the majority of the signalized intersections still use a fixed timing plan. Researchers have suggested that good arrival time predictions are important in these cities for transit priority signals to be effective at reducing travel times (Li et al., 2012). Better inter-stop and dwell time estimations are needed to improve the arrival time predictions. Scheduling strategies can also be adapted to take advantage of transit priority signals (Altun and Furth, 2009). However, these previous studies tend to focus on one specific intersection or a few consecutive intersections on a given corridor, which would have similar base timings. Similarly, some studies only included the number of priority signals to model their effects, which essentially assumes these signals behave similarly. Since buses can make turns and travel through multiple corridors, the signal synchronization and cycle lengths may all be different in reality. Smaller intersections typically have shorter cycle lengths, while large intersections have longer cycle lengths. There is still a need to consider how the travel time and red light waiting times are affected by varying signal cycle lengths and synchronization patterns.

Overall, the literature mostly focused on the travel time impacts from one specific element of the transit system, namely the variation in travel times, dwell times, and signal priorities. There is still less attention on combining these various elements to examine how these travel time elements affect and interact with each other. For example, given a signal synchronization, if the vehicle always arrives at the stop during the red light, dwell time variation can become less critical since the bus is stopped by the signal anyway. On the other hand, if the vehicle always arrives during the green light, the previous inter-stop traffic and dwell time can become more critical when the bus tries to rush through the intersection before the light turns red.

In addition, since transit agencies have limited resources, there is still a need to help planners prioritize their resources. Thus, it is important to determine which trip element is more important on a given route to reduce the overall travel time variations. By evaluating the importance of each trip element, planners can select a good and effective strategy to improve the reliability of a specific route.

Thus, in this paper, we propose a framework to further decompose transit travel times so that the variation of each trip element along the route can be isolated. Given the limitations of earlier studies, we also conduct a sensitivity analysis to rank the importance of each trip element based on its potential impact on the overall travel times. We aim to provide better tools to help planners diagnose and improve transit reliability.

6.3 Methodology and Data

Figure 6.2 shows our research framework. The basic idea is to decompose the trip travel times into the sum of a sequence of times. For this paper, we decompose the trip travel times into three categories, inter-stop times, dwell times, and red light waiting times. To reconstruct the arrival times at each stop, we also need to include departure times in our analyses.

To get these detailed times, we use the General Transit Feed Specification (GTFS), GTFS Real Time (GTFS-RT), OpenStreetMap, Automated fare collection, and some ride-check observation data. The departure times and inter-stop times are directly obtained or calculated from the GTFS-RT feed. Due to the lack of door-closing times, we estimate a simple dwell time model at the stop level using ride-check observation data and apply the model to the fare transaction data. The traffic light settings are also estimated following the methodology proposed by Fayazi et al. (2015) due to data availability issues.

Then the decomposed travel times are grouped by route direction and three different time-of-the-day and used for sensitivity analyses. Here, we focus on both the variance-based analysis

(Saltelli et al., 2008), which is a global method that quantifies the proportion of variance that can be attributed to each variable. We also conduct a one-at-a-time sensitivity analysis to better illustrate the non-linearity between the input variables and the overall travel times. Since sensitivity analyses require varying one or more variables, we can adjust the inter-stop and dwell time variables given the observations. However, since signal timings are mostly fixed and not within the agency's control, we calculate the resulting red light waiting times given the arrival time at the signal and the signal timing plan as input.

We use the data provided by Société de Transport de Montréal on the island of Montréal in Canada as a case study. The system currently has 222 bus lines in operation, about 2000 buses in the fleet, and more than 17,000 published bus trips on average weekdays. More detailed information regarding the methodologies and data used in each step can be found in the subsections below.

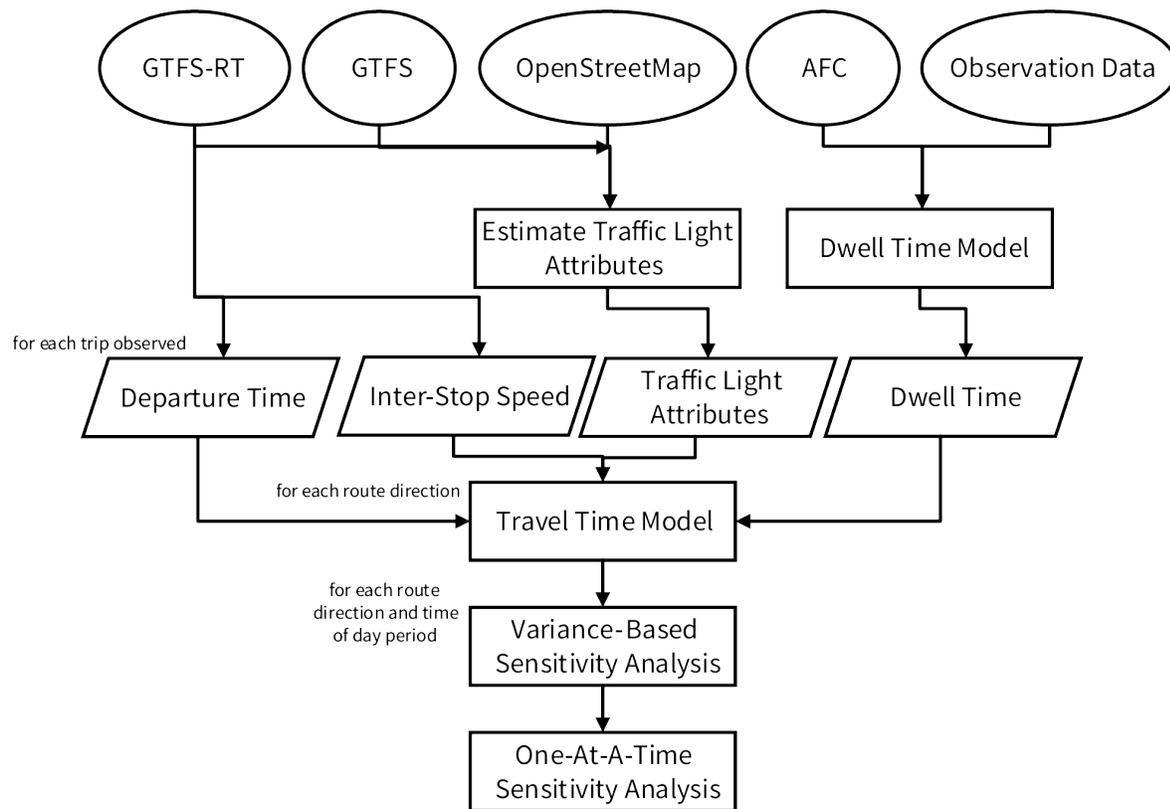


Figure 6.2 Research Framework

6.3.1 Defining the Trip Elements

The GTFS file provides detailed information on the planned services, such as schedules and geographical information for the routes and stops. Since it only publishes transit-related information based on stop arrival and departure times, we need to add more detailed information to better isolate each trip element for our analysis.

To split the routes into these elements, we used the route shape from GTFS and matched traffic light positions from OpenStreetMap, so that we can estimate the red light waiting times in later steps using the vehicle location information. Since there may be additional traffic signals between two scheduled stops and inter-stop time variations will cause the red light waiting times to change for these signals, we will split the stop-to-stop segment at these traffic lights so that we can attribute red light waiting times to each signal along the route.

To reconstruct the travel and arrival times, we need to calculate four categories of information, the departure time from the first stop, a series of inter-stop times, a series of dwell times, and a series of red light waiting times. Since most of our bus stops are on the near-side, i.e. right before the bus enters the intersection, The sequence of these times is typically ordered as follows, first, we depart from stop 1 at a given departure time, followed by the inter-stop time to stop 2, then the dwell time at stop 2 if applicable, and finally the red light waiting time at stop 2 if applicable. Since red light waiting times are the result of vehicle arrival time at the intersection and the timing plan, we will have to calculate them using the arrival time at the signal, based on vehicle position, and the estimated signal timing plan. The pattern then repeats itself for the following stop-to-stop pairs until the service destination. Thus, we define our input variables following the same logic for every trip. After the introduction of these times, we will provide an example to demonstrate the categorization in a separate subsection.

6.3.2 Estimating Traffic Signal Settings

Traffic signal timing plan changes often since traffic levels vary constantly. In addition, due to roadworks, programming errors, or malfunctions, the planned signal timings are not necessarily what is operated. Unfortunately, due to data availability issues, we could not obtain the signal timings in real time. Thus, we will estimate the signal timings also using the aforementioned archived bus trajectory data.

Given the high density of signalized intersections, most of the traffic signals in Montréal use the coordinated timing plan to optimize traffic flow on major streets, as recommended by the National Cooperative Highway Research Program's Signal Timing Manual (Urbanik et al.,

2015). There are three principal parameters for a coordinated timing plan, cycle length, offset, and split. Cycle length is the time for a complete sequence of signal phases at an intersection (Urbanik et al., 2015). In Montréal, it is generally between 60 and 140 seconds, depending on the traffic volumes at the intersection. The offset is defined as the time offset between coordinated phases to a predefined synchronization point (Urbanik et al., 2015). They are used to offset the green or red phases to help vehicles progress through multiple intersections without stopping. The splits are the portion of green light plus clearance time allocated to each phase at an intersection (Urbanik et al., 2015). In this paper, we will focus on the green light time only, since the red times and clearance times have the same effect on buses, i.e. buses are not allowed to cross the intersection.

To estimate these three traffic signal settings, we mainly followed the methodology proposed by Fayazi et al. (2015) with some minimal modifications to handle a few particularities of our local vehicle locations feed. The basic idea is to match stop times at a given traffic light observed over a given time of the day by testing out various cycle lengths and signal offsets to see which combination fits the observations the best. Then, using a moving window, we can detect changes in traffic signal schedules, such as peak schedule vs off-peak schedule. Readers interested in the details can refer to the original paper cited above.

To verify the estimated signal settings, we conducted ride checks on board buses and point checks at intersections. The estimation errors are typically around 3 to 5 seconds. We consider this acceptable, as it is similar to the length of yellow lights and the errors observed in the original paper (Fayazi et al., 2015). As mentioned in the original paper, drivers have different risk tolerances towards yellow lights, some may treat it as a green light and some may treat it as a red light, thus causing some slight discrepancy in the estimated timing (Fayazi et al., 2015). For Montréal, we mainly use fixed signal timing plans with very little flexibility, which allows this estimation method to function well. Unfortunately, the sensitivity for more flexible timing plans and more aggressive transit signal priorities has to be left for future research.

6.3.3 Getting Departure Times and Inter-Stop Times

This step mainly uses the GTFS-RT data, which provides the actual bus arrival and departure times at stops, as well as detailed bus location and speed information around every 5 to 20 seconds. Using this information, we can directly obtain or calculate the departure times and inter-stop times needed for our analyses. Here, we used 3 months of archived data from January 8, 2024 to March 24, 2024.

Since this project focuses on travel times, outlier observations, such as major detours, might

greatly affect the sensitivity of travel time, and agencies are typically aware of these types of travel time variations. Thus, we will remove these outliers from the analysis using Density-Based Spatial Clustering of Applications with Noise (DBSCAN) (Ester et al., 1996), which is a density-based algorithm to identify clusters and outliers in the data. To reiterate, due to various statistical distributions observed in previous research (Kieu et al., 2015), we don't want to make any assumptions about the data distribution. In addition, due to the mixture travel time distribution observed (e.g. Figure 6.1), there may be many small clusters of travel times that might be more extreme but meaningful. Hence, we choose this method for cluster detections and outlier removals. For each trip, we compare the similarities of the trip departure, travel time and delay observation for all trips. The outliers that are not similar to other observations, such as significant delays and unusually long or short travel times, are identified and removed from the analysis.

6.3.4 Estimating Dwell Times

The dwell time is defined as the time for passengers boarding and alighting the bus (National Academies of Sciences, Engineering, and Medicine and others, 2013). However, due to the lack of door-closing times and some drivers leaving the door open for ventilation while waiting for the red lights, we will estimate a dwell time model using ride check observation data.

The ride check observations include detailed door opening times, number of boardings, number of alightings, and any kneeling or ramp usages. The end of dwell time is defined as either the door closing times collected on-board, or 5 seconds after the last passenger boarding or alighting in case the driver leaves the door open. Then, using roughly 25,000 dwell time observations over two years, we estimated a simple dwell time model using linear regression using these observations, following the example from Dueker et al. (2004). For trips without on-board observations, we used the origin-destination from fare transaction data to obtain the passenger counts at each stop, and applied the estimated dwell time model for decomposition.

We acknowledge that the dwell time model is in no way perfect, as there remain some issues with the agency's ridership matching algorithm, our ridership observations can still be improved, and dwell time variations deserve more research on their own. As the agency improves the origin and destination matching algorithm, future researchers could add more variables with regard to the overall passenger flow, ridership variations given bus bunching, trip purposes, and so on. However, due to the time limit, we will settle with these imperfections for the moment and only focus on the sensitivity analyses at a given stop. We will leave improving dwell time models, passenger flow models, understanding ridership variations, and

passenger arrival patterns as a future research task.

6.3.5 Example of decomposing a stop-to-stop travel time observation

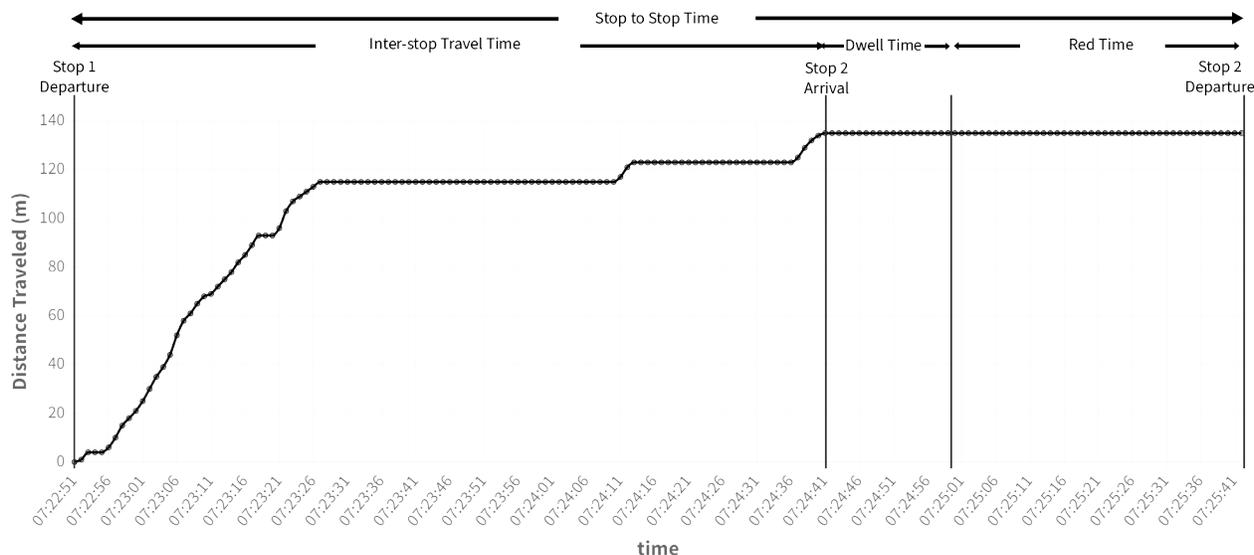


Figure 6.3 Example: Splitting a Stop to Stop Travel Time Observation

Finally, to sum up these detailed time breakdowns, Figure 6.3 shows an example time distance diagram for one stop-to-stop travel time observation. Here, we have a stop-to-stop travel time observation where the bus starts from the terminal (distance 0 m) and travels to stop 2 on the route (140 meters after the terminal). From the time-distance graph, we can see that the stop-to-stop travel time is 2 minutes and 50 seconds from 7:22:51 to 7:25:41. The stop-to-stop travel time is split first into the inter-stop time which is the time it took to travel to the next stop from 7:22:51 to 7:24:41. For our example, despite the bus being 20 meters away from the stop for over a minute, the door did not open due to congestion, so the inter-stop travel time is around 1 minute and 50 seconds. Then, as the bus stops at the second stop, the stopped times are split into dwell times according to the dwell time model, which is estimated to be around 20 seconds, given the estimated passenger boarding and alighting numbers. Then, the bus stays at the near-side stop for the remainder of the red light doing nothing, until the green light activates at 7:25:41 according to the estimated timing plan. Thus, the estimated red light waiting time is around 40 seconds. The decomposition is repeated and performed for every stop-to-stop observation that follows until the end of the route and for every recorded trip on a given route-direction.

6.3.6 Sensitivity Analyses

Sensitivity analyses are typically used to attribute the variation of model outputs to the input space. In this paper, we analyze the sensitivities by route-directions and by three time-of-the-day periods, since the signal synchronization, ridership pattern, and traffic conditions may be different. There are typically two main categories of sensitivity analyses, one-at-a-time and global analyses (Saltelli et al., 2008). We will provide a quick introduction to both methods in this subsection.

Variance-Based Sensitivity Analysis

Due to the complexity and non-linearity typically observed in previous travel time studies, we first consider the variance-based sensitivity analysis (also referred to as the Sobol method). Variance-based sensitivity analysis is a commonly used method that can handle non-linearities between input and output results. It is typically used in analyzing large environmental or biological models (Saltelli et al., 2008), albeit rarely used in transportation fields. It is also included in many statistical or optimization software packages.

A quick summary of the method is that it varies all input variables at the same time. Then, it sends the inputs to the model, e.g. a travel time model, which is considered a black box. Next, it decomposes the total variance of the model output into partial variances, i.e. the percentage of total variance, and attributes the partial variances to individual input variables. Readers who are interested in more detailed derivations and explanations can refer to Saltelli et al. (2008).

To generate the samples needed to calculate these indices, researchers typically conduct a Monte Carlo simulation (Saltelli et al., 2008) to uniformly sample all input space. However, given the various travel time distributions observed from previous studies (Kieu et al., 2015), we decided to use the observed trips as inputs to avoid making assumptions about the data distribution.

The method would then randomly split our trip observations into two matrices \mathbf{A} and \mathbf{B} , with each row representing a trip observation and each column representing a trip element defined above.

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & \dots & a_{1,k} \\ \dots & \dots & \dots \\ a_{N,1} & \dots & a_{N,k} \end{bmatrix}; \mathbf{B} = \begin{bmatrix} b_{1,1} & \dots & b_{1,k} \\ \dots & \dots & \dots \\ b_{N,1} & \dots & b_{N,k} \end{bmatrix} \quad (6.1)$$

Sensitivity analysis attributes the output variance to the variation in some input variables while holding other variables constant. For the variance-based analysis, k matrices $\mathbf{C}_{1\dots k}$ are then generated using matrices \mathbf{A} and \mathbf{B} . Matrix \mathbf{C}_i uses all elements of \mathbf{A} except the column corresponding to variable (or column) i , which is taken from \mathbf{B} , thus allowing the method to calculate the impact on overall results by varying one variable (Saltelli et al., 2008). As an example, Matrix \mathbf{C}_2 is:

$$\mathbf{C}_2 = \begin{bmatrix} a_{1,1} & b_{1,2} & \dots & a_{1,k-1} & a_{1,k} \\ a_{2,1} & b_{2,2} & \dots & a_{2,k-1} & a_{2,k} \\ \dots & \dots & \dots & \dots & \dots \\ a_{N,1} & b_{N,2} & \dots & a_{N,k-1} & a_{N,k} \end{bmatrix} \quad (6.2)$$

In simpler terms, the method uses one observed trip as a base case, then replaces different elements of the trip with the observations from another trip, i.e. what would have happened if the bus were to spend a different amount of time at a given section. If the analysis includes a change to make an extra stop, the inter-stop times cannot be directly applied due to the extra slowdown and acceleration time needed. In this case, we will estimate and include the time changes due to additional or skipped stops. Similarly, new red light waiting times are recalculated using the updated arrival time or signal timing plan.

Finally, the sensitivity indexes are estimated using original travel times observed for trips in \mathbf{A} and \mathbf{B} , and the potential impact calculated from matrices $\mathbf{C}_{1\dots k}$ (Saltelli et al., 2008).

The model produces two main indices, the first-order index and the total-order index. The first-order index is calculated as the percentage of total variance reduced if one variable were fixed (Saltelli et al., 2008). In other words, the first-order index is the percentage of the total variance caused by the variation of a given input variable, without interaction effects. It is formulated as

$$S_i = \frac{Var[E(Y|X_i)]}{Var(Y)} \quad (6.3)$$

Similarly, the total-order index is calculated as the complement of the variance produced by varying all but one variable, normalized by the total variance (Saltelli et al., 2008). In other words, it shows the percentage of the total variance caused by the variation of a given input variable, with interaction effects. It is formulated as

$$S_{Ti} = 1 - \frac{Var[E(Y|X_{j \neq i})]}{Var(Y)} \quad (6.4)$$

These indices lie between 0 and 1, since they represent a percentage of total variance. These indices can also be used for ranking purposes according to the share of total variance attributed to each input. The first-order indexes typically add up to less than 1 due to the exclusion of interaction effects. Similarly, the sum of total-order indices will typically exceed 1 due to counting interaction effects multiple times.

Although it is theoretically possible to calculate any n -th order sensitivity indices, they are not practical to calculate or apply. Assuming there are k total variables, the number of n -th order indices is the number of ways to choose n variables from all k variables, i.e. the combination $C(n, k) = \frac{n!}{k!(n-k)!}$. The computational complexity is generally too expensive to calculate the sensitivity indices among various combinations of variables. Hence, most studies will only focus on the first and total order indices.

One-At-A-Time Sensitivity Analysis

One-at-a-time sensitivity analysis is a typical direct approach to see the effect of changing one input on the output (Czitrom, 1999). The steps are incrementing one input variable while keeping the others constant, then returning the changed input to its original value, and repeatedly changing the other variables in the same way. Sensitivities are typically measured by the partial derivatives, i.e. the amount of change observed in the output by changing the input by 1.

Due to its simplicity, the method does not examine the entire input space like the variance-based method from above, and thus the results do not show the interaction effects and non-linearity (Czitrom, 1999) between variables. Due to its easy-to-understand nature, we decided to include this method to better illustrate the sensitivity results and non-linearity between trip travel times and various trip elements. However, we will not aggregate the results into one number in this paper due to the limitations mentioned above.

6.4 Case Study: Montréal

In this section, we present the sensitivity results using the data from Montréal. First, we will provide a quick overview of the variance-based global analysis results for 20 selected routes operating in the central part of Montréal, so that readers can have an understanding of the most important factor affecting each travel time component. Then, we will use the westbound Route 27, the northbound Route 30, and the eastbound Route 97 as examples to demonstrate more route-specific analyses. The data and conclusions may be specific to Montréal and the routes analyzed, but the readers can nevertheless adapt the methodology

for their cities and draw their own conclusions given their specific local contexts.

6.4.1 Overall Variance-Based Sensitivity Results for 20 Selected Routes

Here, we summarize the most important factors affecting 20 selected routes in the central areas of Montréal. These routes typically travel through areas with a higher population density, mixed-use developments, and higher traffic light density. First, we will show the most important factors for each travel time component, i.e. inter-stop travel times, dwell times, and red light waiting times. Then, we will present the sensitivity results for the overall trip travel times.

Most Important Factors Affecting Inter-stop Travel Times and Dwell Times

The differences between first-order and total-order sensitivities are very small for the overall inter-stop and dwell times. Therefore, the results suggest little interaction effects contributing to the variation of inter-stop times and dwell times.

The important factors affecting the overall inter-stop times are all inter-stop variables. As we categorized the trips by time-of-day periods, the inter-stop travel times are relatively similar. Therefore, they are not as sensitive to the departure time changes as the overall trip times and red light waiting times as we will demonstrate later. The overall inter-stop time is also not very sensitive to the signal timing plan changes, since the vehicles would still need to travel to the next stop in similar traffic conditions.

Similarly, the most important factors affecting the overall dwell times are all ridership variables for the stops. This makes sense since dwell times only include the times for passenger activities as defined by Transit Capacity and Quality of Service Manual (National Academies of Sciences, Engineering, and Medicine and others, 2013) (e.g. Figure 6.3), and we do not yet consider the ridership variations caused by vehicle interactions or schedule adherence issues in this paper. Therefore, we emphasize the future research need to better understand ridership variations caused by vehicle interactions, as well as incorporating passenger arrival patterns into the variance decomposition models.

Most Important Factors Affecting Red Light Waiting Times

Figure 6.4 shows the most important factor contributing to the overall red light waiting time variations. Again, according to the Transit Capacity and Quality of Service Manual (National Academies of Sciences, Engineering, and Medicine and others, 2013), red light waiting times

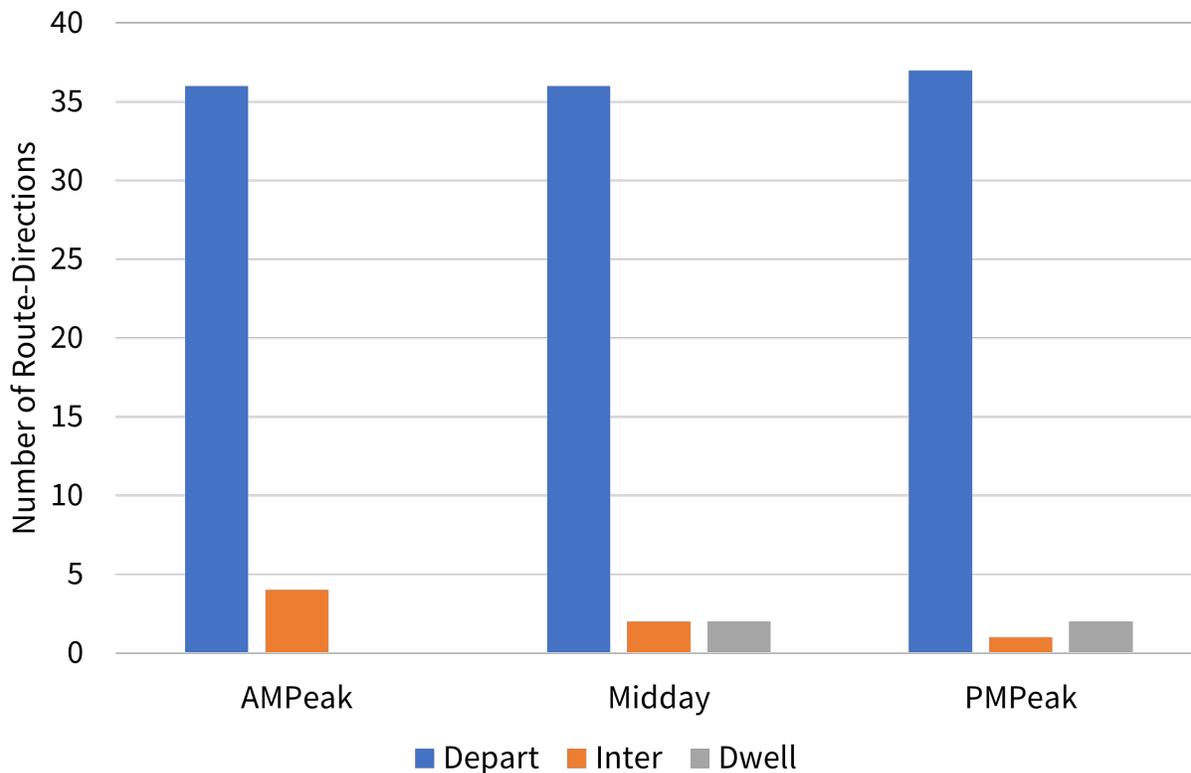


Figure 6.4 Most important factors contributing to overall red light waiting time variations for 20 routes at various time-of-day periods

are the times stopped after passenger boardings and alightings are complete (e.g. Figure 6.3).

The most important factor for the red light waiting times on the 20 routes (40 route-directions) analyzed is almost all departure time changes at the first stop. The number of route-directions with the departure times as the most important factors remains stable throughout the day, roughly 36 to 37. The number of route-directions with other factors as the most important is generally around 3 to 4. As the day goes on, the number of route-directions with inter-stop times as the most important factor slightly decreases, but the number of route-directions with dwell times as the most important factor increases slightly as the ridership variation becomes higher later during the day. Still, it is very important to consider the departure time when analyzing red light waiting time variations.

The 20 routes analyzed in this paper are all located near the center of the city, thus having a higher density of traffic lights. The red light waiting time counts towards 20% to 25% of trip travel times for these 20 routes, and it is the category with relatively high variance. This observation shows that it is really important for planners to choose the right departure

time. If the planners want to reduce the red light waiting time variations, there is a need to consider how departure time changes affect red light waiting time variations when adjusting the schedules.

Currently, the signals follow a fixed timing plan with varying cycle lengths given the importance of an intersection. Larger intersections have longer cycle lengths (e.g. 140 seconds), and quieter intersections have shorter cycle lengths (e.g. 60 seconds). Thus, green waves do not necessarily line up perfectly for the entire bus route at all times. The importance of departure time arises from how synchronized or de-synchronized the signals are. To reiterate, since the travel times can be defined as the departure time from the first stop, plus a series of dwell times, inter-stop times, and red light waiting times, the red light waiting times down the route depend on the departure time at the first stop, and the previous inter-stop, dwell, and red light waiting times. Thus, the departure time at the first stop roughly sets the synchronization pattern further down the route.

Readers can consider a simple example. Buses leave intersection 1 at a perfect 5-minute interval at an intersection with cycle lengths of 100 seconds i.e. every 3 cycles. For example, departure times can be at 0, 5, 10 minutes, etc, past the hour. Then buses travel to intersection 2 with the same 1-minute travel time, thus arriving at the intersection at 1, 6, and 11 minutes past the hour. If the signal cycles at intersection 2 is 120 seconds and the green light starts at 0, 2, 4, 6, 8, 10, and 12 minutes past the hour with 53 seconds green time and 7 seconds for yellow and clearance times, the first and the third bus will have to wait for red lights at intersection 2 while the second doesn't. Hence, every other bus will have to spend longer times trying to pass intersection 2.

Given the importance of departure time on red light waiting times, planners could add adequate schedule padding to ensure the buses have similar travel times or to optimize for specific departure times to include the red light waiting time variations in planned travel times. Another potential strategy to reduce the red light waiting time variations could be a more aggressive signal priority that is applied more in advance to reduce the significance of departure times. Moreover, not all trips are able to depart exactly on time, potentially due to residual delays from earlier trips or waiting for a late passenger, etc. Thus, it is also important for future researchers to find a way to better provide signal priority to buses while considering the varying cycle lengths and potential departure, inter-stop, and dwell time variations.

In addition, a common assumption used when creating schedules and optimizing vehicle assignments is to assume travel times remain the same within a few minutes. Intuitively, traffic congestion levels may stay similar and ridership may remain similar if we change the

departure time by a few minutes. However, as shown here, the red light waiting times are significantly affected by departure time changes. Thus, planners need to be careful when optimizing bus schedules since the constant travel time assumption may not be applicable everywhere.

Most Important Factors Affecting Overall Trip Travel Times

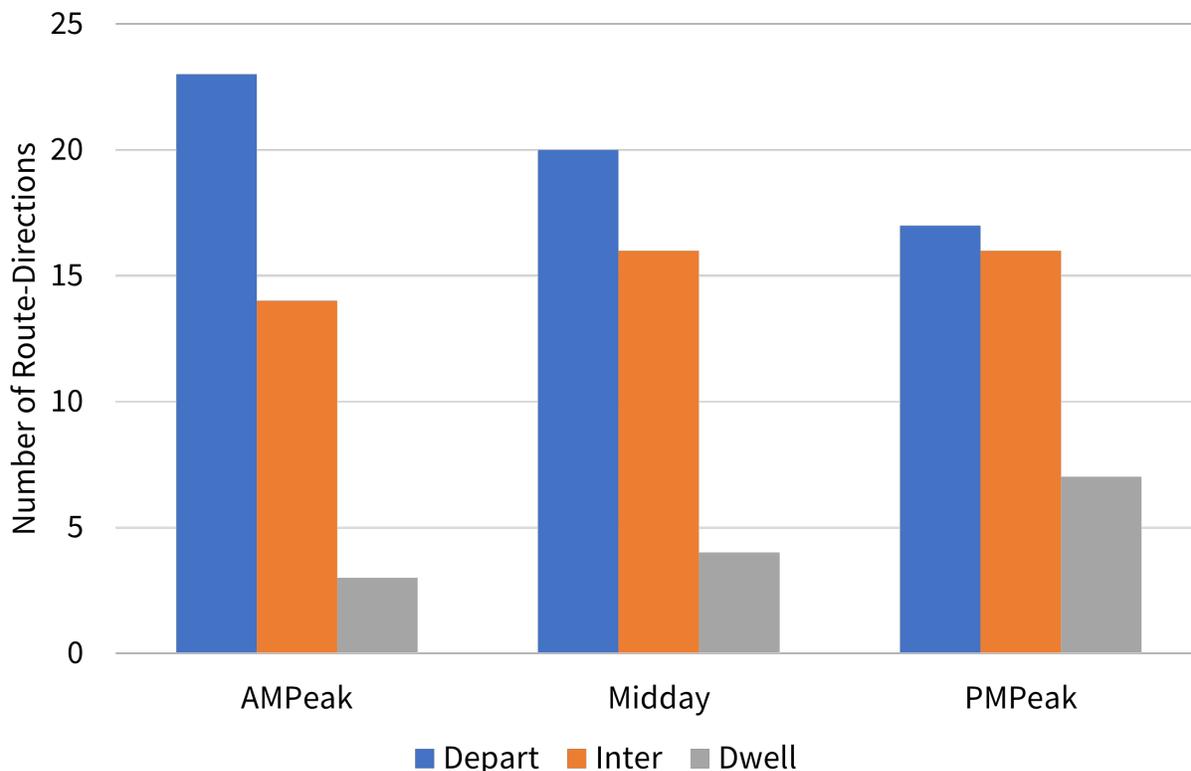


Figure 6.5 Most important factors contributing to overall trip time variations for 20 routes at various time-of-day periods

Figure 6.5 shows the most important factors contributing to overall trip time variations for 20 routes at various time-of-the-day periods. As we can observe from the figure, the most important factor can change throughout the day. Thus, planners may need to consider different strategies to improve transit reliability at various time-of-the-day periods.

The departure time contributes the most to the trip travel time variations for most of the routes analyzed. As shown in earlier subsections, departure times are very important for red light waiting times, but not as important for dwell and inter-stop times. Since red light waiting times are also a big part of the overall travel times, the results suggest that variations

in red light waiting times contribute greatly to the overall travel time variation. Moreover, there are more possibilities for buses to get delayed or run early while interacting with traffic signals or congestion compared to the last stop on the route.

However, the number of route directions with departure time as the most important factor decreases as the day goes on, decreasing from 23 route-directions during the AM peak to 20 route-directions during midday to 17 route-directions during the PM peak. The number of routes having inter-stops as their most important factor grows from 14 to 16 from AM peak to midday. The number of route directions with dwell time as their most important factor grows from 3 to 7 throughout the day. As we will demonstrate in the later sections, inter-stop time and dwell time variations become higher later in the day.

We believe this could be correlated to the travel pattern changes or signal timing plan changes. For the AM peak, the traffic patterns and ridership patterns are relatively stable, since most people are going to work or school, as shops and other destinations are not yet open. As the day goes on, more shops and restaurants begin to open, traffic congestion and ridership can become more fluctuated and significant on some routes or certain sections. Thus, planners may consider other ways to improve reliability, such as implementing bus lanes.

In addition, the changes may correlate to the three signal timing plan changes during the day as well. During the off-peak hours, such as midday, midnight, and weekends, the signals typically use the same timing plan. During the two peaks, the timing plan is changed to facilitate travel towards downtown in the morning or away from downtown in the evening. Therefore, planners also need to consider how travel directions might potentially affect the overall travel times.

6.4.2 Detailed Variance-Based Sensitivity Results for Three Example Routes

In this section, we will use the results from three different routes analyzed to illustrate the different issues and factors affecting transit travel time reliability. These three chosen examples are based on the most important issues discussed in the earlier section and are somewhat representative of other routes in our study. However, planners want to be mindful of different local contexts and configurations for different routes, although the analysis process is similar. First, we will use the westbound route 27 as an example to illustrate the impacts of departure time variations, which interact with signal timing plans and in turn affect the red light waiting times. Then, we will discuss the northbound route 30, which faces congestion variations in a popular shopping area. Finally, we will examine the eastbound route 97, which has high ridership variations at the stop in front of the metro station.

Westbound Route 27: Departure Time

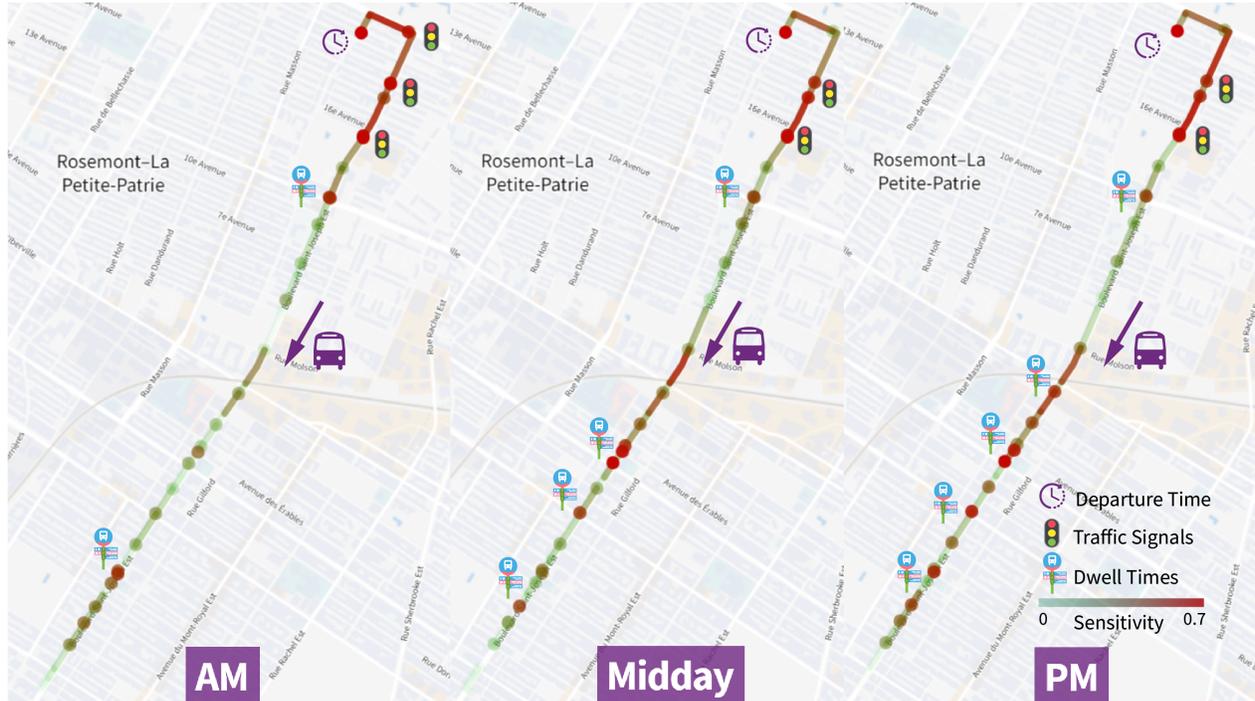


Figure 6.6 Example: Sensitivity Indices for Westbound Route 27

Route 27 is relatively short and straightforward. The 4.1 km route operates on a secondary corridor Saint-Joseph Boulevard through a densely populated neighbourhood on the eastern side of the city. Its daily ridership is an average of 2400 passengers, mainly feeding passengers into the metro service which is its western terminus. Its peak direction headway is 9 to 12 minutes with roughly 40 to 50 passengers on board. For off-peak directions or hours, the headway is typically between 23 to 27 minutes with roughly 25 to 35 passengers on board. The planned one-way travel times are between 19 to 24 minutes. During rush hour, there is also a limited stop service that supplements this local service in the peak direction only with similar headways and ridership numbers. Another factor that makes this route interesting is that it has the highest traffic signal density within the network. On average, there is a signal every 170 meters. However, the distance between signals is only 50 to 75 meters on some sections.

Since we decomposed each stop-to-stop segment into 4 variables, departure time at the first stop, inter-stop times, dwell times, and signal timing plans, the 22-stop service has 83 input variables. As mentioned in the methodology section, we cannot calculate detailed results showing every possible combination of variables and their interactions. To demonstrate the computational complexity, the number of 41st-order sensitivity indexes for this route-

direction would be $C(41, 83)$, which roughly equals to 8.39×10^{23} . Hence, most studies will only report the first and total order sensitivities for each variable (i.e. the sensitivity attributed to 1 variable without or with interaction effects). We will include a map highlighting the most important variables using the total-order sensitivity (Figure 6.6), and we will include examples of more detailed first and total-order results in the explanations.

The sensitivity results for the overall inter-stop and dwell times are similar to the observations in previous sections. From the figure, we can observe that the sensitivity indices are higher near the north-eastern side of the route, at almost 0.7, meaning that with interaction effects, 70% of the total travel time variance can be attributed to the variables and their interactions in this area. Given the small overall travel time variations, at around 2 to 3 minutes, this area alone could be attributed 1.4 to 2.1 minutes of variation. The most important factor is the departure times, the first-order sensitivities are around 8%, and the total-order indices remain around 60% to 70%. This result again suggests that departure time alone doesn't contribute much towards the overall red light waiting time variation. However, the interaction effects of departure time contribute significantly toward the overall travel time variation. Similar to previous sections, the most important factor affecting the overall dwell time variation is the stop with the largest ridership variance, located in the lower part of the map for the AM peak, and the middle of the map for the midday and PM peak, contributing around 50% of the total variance with interaction effects.

For the overall travel time variations, the most important factor for both the first-order and total-order indices is the departure times at the first stop for all three time-of-the-day periods, shown as a dark red circle on top of the figures. Table 6.1 shows the departure time sensitivity results with their equivalence in seconds calculated using the standard deviation. From the table, we can see 8% of the total variance during the AM peak can be attributed to the departure time without interaction effects (first-order). In terms of standard deviation for context, it is roughly 10 seconds. If we include the interaction effects (total-order index), the departure time attributes 67% of the total variance during the AM peak (roughly 83 seconds in terms of standard deviation). However, the sensitivity for departure time changes slightly later in the day. The first-order indices for the midday and pm peak periods are 6% and 8%. The total-order indices for the midday and pm peak periods are 47% and 42%.

Since the overall travel time includes the sum of all red light waiting times, where the importance of departure time dominates, and all inter-stop and dwell times, where departure time is not as significant. This result highlights the importance of red light waiting time variation on the overall travel time variance.

More specifically, the signal cycle lengths along this route are 80, 90, 100, and 120 seconds,

with a least common multiple of 3600 seconds. Thus, for any given departure time within one hour, the green wave patterns and the red light waiting times will be different. We therefore emphasize the need for transit planners to pay more attention to the base timing of traffic signals and their interactions with transit vehicles when analyzing or planning for travel times.

Table 6.1 Departure Time Sensitivity Results for Westbound Route 27

	AM Peak	Midday	PM Peak
First Order (%)	8%	6%	5%
Total Order (%)	67%	47%	42%
Interaction Effects (%)	59%	41%	37%
First Order (sec)	9.92	8.28	7.70
Total Order (sec)	83.08	64.86	64.68
Interaction Effects (sec)	73.16	56.58	56.98

The second most important factor for the overall trip travel times is the traffic light timing changes, shown in the sections and points near the top of the maps. During the study period, there was a traffic light timing plan change due to a newly implemented bike lane as well as improved pedestrian crossings. The change in signal timing contributed on average 2% of the total variance without interaction effects (around 2 seconds in terms of standard deviation), but 20% of the total variance with interaction effects (24 to 30 seconds in terms of standard deviation) for all three time-of-day periods. Thus, a change in traffic light timing may not cause a significant variation in travel times locally, but planners must consider the potential chain reaction it has with downstream sections of the route when evaluating the impacts of signal timing plan changes.

To summarize, the above observations show strong interaction effects between input variables for the overall trip time and the overall red light waiting times, given the larger differences between first-order and total-order sensitivity indices. Thus, the result emphasizes the need for planners to choose the departure times at the first stop carefully and better understand the interaction effects in transit travel times for planning. The results also show little interaction effects for the inter-stop times and dwell times given our assumptions, which makes it easier for planners to improve reliability for times in these two categories.

We will try to demonstrate the potential variations caused by different variables and the non-linear relationship by applying the one-at-a-time analysis on this route as an example in the next section.

Northbound Route 30: Inter-stop Time

Route 30 is a 11.3 km route, running north-south on corridors Saint-Denis and Saint-Hubert. The population density around these two corridors is high. Given the mixed-use zoning commonly found in the central areas of Montréal, these two streets are also busy shopping areas. Hence, this route passes popular destinations for the locals and tourists. However, since this route duplicates part of the metro orange line service and passengers generally prefer the faster congestion-free metro service, the ridership on this route is not high (around 2000 to 2500 during average weekdays) and the service frequency is quite low at around 30 to 35 minutes. As for travel times, the planned times are between 44 to 64 minutes depending on time of the day, and the standard deviation of the overall travel times is around 4 minutes in the morning peak, 5.5 minutes around noon, and 8 minutes at the afternoon peak.

Figure 6.7 shows the general alignment of the route and the sensitivity indices on the map, with the two main problem areas, Plaza Saint-Hubert and Saint-Denis, marked on the map in the central and south side. The departure times (marked as the right-most dot in the figure) are the most important during the morning peak, contributing around 20% of the variation with interaction effects. However, the importance remains relatively stable during the day. The inter-stop times at Plaza Saint-Hubert start to contribute more variance to the overall travel times. The sensitivity results for the inter-stop times at Plaza Saint-Hubert gradually rise during the day, around 30% of the variation could be attributed here during midday and the afternoon peaks.

The main issues identified for this route are the inter-stop times, unsurprisingly due to traffic congestion, since not everyone uses transit to access these two popular areas, despite having good transit connections. In this section, we will use the Plaza Saint-Hubert area as an example, since it is the top issue for the midday and the afternoon peak, although the variation is a lot smaller in the morning peak (total-order index less than 1%). Table 6.2 shows some simple statistics to highlight the day-to-day variations as well as the within-day variations of inter-stop time. From the table, we can observe that the average inter-stop time grows higher as the day progresses, where the morning peak has the lowest averages and standard deviation regardless of day of the week, and the afternoon peak has the highest averages and variation regardless of day of the week. This matches our expectation, since the shops and restaurants in this popular area are not yet open during the morning peak. Therefore, the area attracts fewer visitors and traffic at this time. Around noon, shops and restaurants start to open their doors until late at night, attracting more visitors during the later time of the day, thus increasing the traffic and travel time. There are also obvious within-week variations for the travel times. From the table, we can see that Friday has the

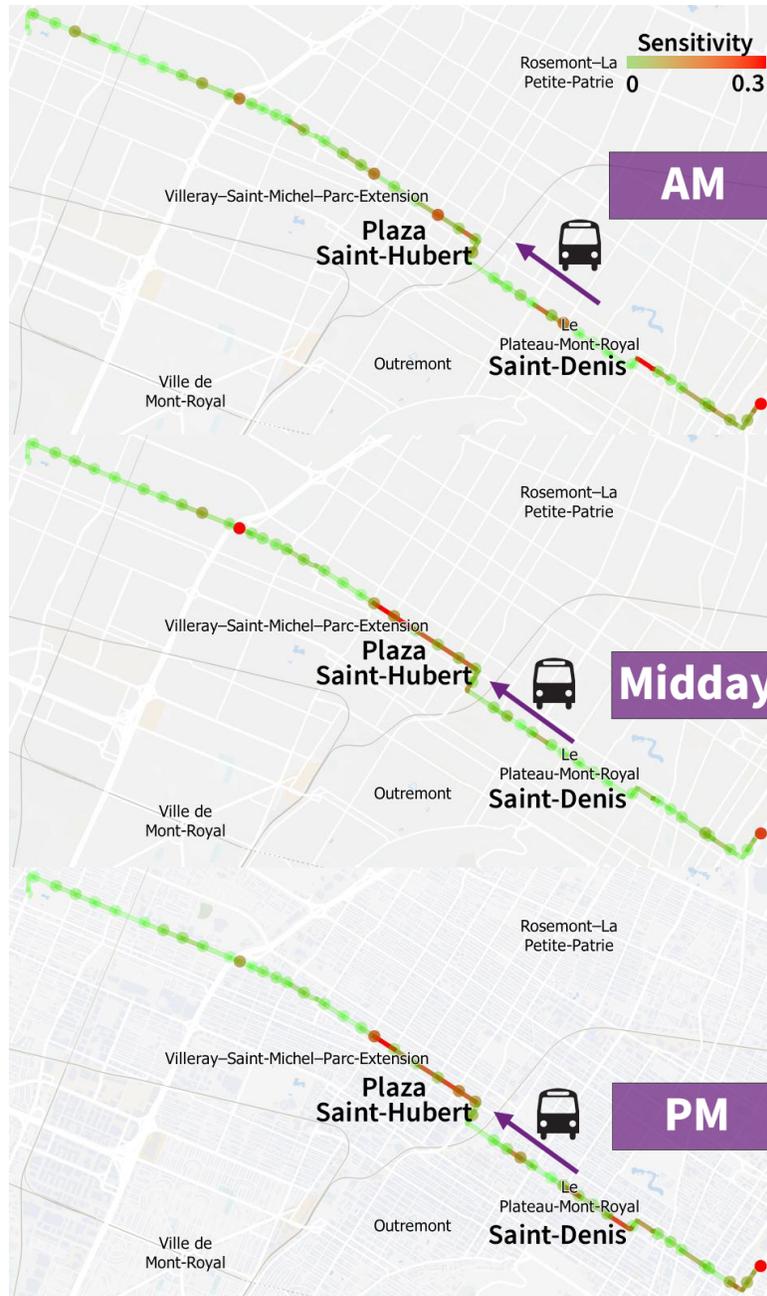


Figure 6.7 Example: Sensitivity Indices for Northbound Route 30

highest travel time averages and variations. We hypothesize that some people are able to work from home on Fridays, therefore having more freedom in their preferred times to visit the area. Similarly, Monday generally has lower averages and standard deviations, which also makes sense since some shops close on Mondays, so there are slightly fewer trip attractions in the area. However, since the agency uses the same schedule for Monday and Friday and Friday has significantly higher (three times the average and standard deviations!) compared to Monday, planners should be ready to deal with such variations when scheduling travel time and recovery times.

Table 6.2 Inter-stop Time Averages and Variations at Different Time of the Day and Day of the Week for Northbound Route 30

Day of the Week	AM Peak		Midday		PM Peak	
	Average	Std. Dev.	Average	Std. Dev.	Average	Std. Dev.
Monday	41.12	10.03	60.21	31.31	75.54	45.33
Tuesday	42.12	13.62	77.99	52.06	78.89	55.77
Wednesday	41.99	15.31	68.44	42.01	76.48	56.24
Thursday	39.51	9.25	70.69	47.42	81.42	57.66
Friday	42.85	20.03	87.28	55.93	125.32	71.85

Given the significant inter-stop travel time changes mentioned before, we want to discuss the overall travel time variations that can be attributed to this short route segment. Table 6.3 shows the sensitivity analyses results for the inter-stop times to the intersection at Saint-Hubert and Jean-Talon (near the top end of the Plaza Saint-Hubert area in Figure 6.7). Again, we can see that the morning peak has the lowest sensitivity results due to the consistent short travel times, at around 2% of the total travel time variation and around 4.5 seconds in terms of time. Then, as the day goes on, the sensitivity for this specific route segment grows as the traffic becomes worse and more varied, and the sensitivity of this segment increases as the total-order index reaches 15% and 73 seconds in terms of time. The total-order index increases especially fast as expected, since the traffic variations in this popular shopping area are low in the morning peak, buses can generally clear the segment without much queuing. In the afternoon, on the other hand, traffic is worse, and buses may need to wait for an extra cycle or two to clear this segment. Hence buses may face a slightly different green wave pattern for the upcoming intersections as the traffic signals are not all synchronized nor have the same cycle lengths. As a result, the departure times matter slightly less for this route in the midday and afternoon peak periods, as the congestion levels may also affect the upcoming signal patterns.

Table 6.3 Inter-stop Time Sensitivity to Saint-Denis and Jean-Talon for Northbound Route 30

	AM Peak	Midday	PM Peak
First Order (%)	0%	2%	3%
Total Order (%)	2%	11%	15%
Interaction Effects (%)	2%	9%	12%
First Order (sec)	0.72	5.95	13.00
Total Order (sec)	4.51	36.53	73.15
Interaction Effects (sec)	3.79	30.58	60.15

Eastbound Route 97: Dwell Time

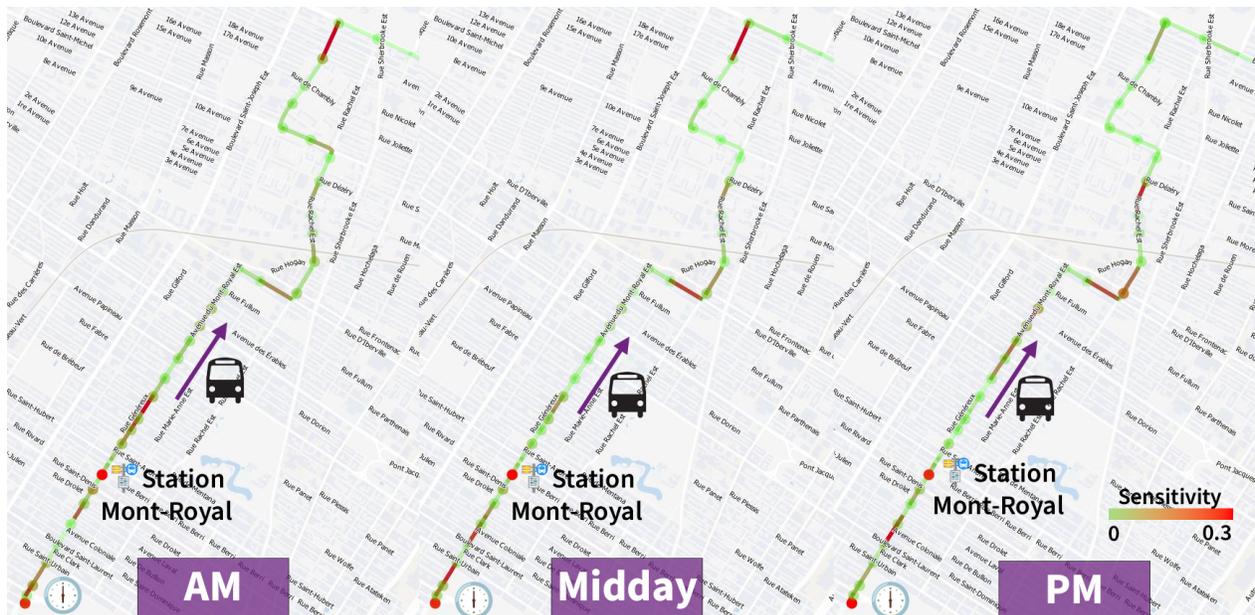


Figure 6.8 Example: Sensitivity Indices for Eastbound Route 97

Route 97 is a 6.7 km route, part of the "frequent" service network, that runs along Avenue du Mont-Royal in Montréal. Avenue du Mont-Royal is an important corridor for the city of Montréal. The population density is relatively high. The western end connects to the Mont-Royal park, allowing easy access for outdoor leisure activities for local residents and tourists. With mixed-use zoning, it is also an important destination for shopping, cultural, and dining opportunities. Hence, it is a popular destination for locals and tourists. With a connection to the metro station at Station Mont-Royal, the route also allows riders from other parts of the city to visit the area. Its daily ridership is around 6000 passengers, and the ridership generally increases during peak shopping season, reaching as high as 9000 passengers. The

headway for the afternoon peak is 7 to 10 minutes, the planned one-way travel times are between 30 to 44 minutes, and some trips consistently have roughly 30 passengers waiting to board here.

Figure 6.8 shows the sensitivity indices on the map, with varied sensitivity results. However, the top 2 factors in terms of the variance attributed are the departure time (the dot on the bottom left) and the dwell time at Station Mont-Royal. The percentage of variance attributed to these two variables remains relatively stable during the day. For the departure time, it is generally between 15% to 20%. For the dwell time at Station Mont-Royal, it is slightly higher at between 17% to 20%. There are also a few sections with higher inter-stop sensitivity. However, the sensitivities are generally related to the congestion and remain less than the dwell time at Station Mont-Royal. Hence, this section will mainly focus on the dwell-time at Station Mont-Royal area as an example.

Here, we use the stop in front of the metro station as an example to illustrate the variations. Table 6.4 shows some simple statistics to highlight the ridership variations for this route. From the table, we can observe that the average ridership grows higher as the day progresses. This matches our expectation, since the shops and restaurants are not yet open during the morning peak, and the routes are used mainly by regular commuters going to work. Around noon, shops and restaurants start to open their doors until late at night, attracting more visitors during the later time of the day periods. The variation of ridership also grows higher, especially during the afternoon peak. The reason could be that people may choose to enjoy leisure activities after work, so that the ridership variation comes from not only the variations of commuters going home but also variations in leisure trips. Again, other cities may have different travel patterns and ridership variations compared to Montréal. Hence, the results shown here are only used to illustrate the method, and other researchers could adopt similar analyses for their specific contexts.

Table 6.4 Ridership Variation on Route 97 Direction East at Station Mont-Royal

Time of the Day	Average	Standard Deviation	Max
AM Peak	11.16	6.21	32
Midday	14.66	7.65	36
PM Peak	24.33	11.29	54

As mentioned before, the most important variable associated with this route, this direction more specifically, is the dwell time variation for the stop directly in front of the metro station. Table 6.5 shows the sensitivity analyses results for the dwell time at Station Mont-Royal, given the standard deviation of the overall travel time is generally around 3 minutes for the morning

peak and midday, then 4 minutes during the evening peak, we also estimated the variance contribution in seconds. From the table, we can see that the total variance attributed to the dwell time at this stop remains relatively stable, around 7% to 11% for first-order and around 17% to 20% for total-order, despite the increase in average ridership and ridership variation later in the day. However, as we convert the percentage of variance attributed to times, we can observe the relatively stable first-order effects but an increase in total-order effects. Similar to the inter-stop times mentioned above, as the ridership increases, especially for the bus with 54 people waiting to board, the longer dwell time may cause the bus to miss one or two traffic light cycles and have to depart at the following traffic light cycle, which would cause a slight deviation in green wave patterns. Hence, the departure time is a close second in terms of the importance of the sensitivity indices, despite more traffic lights being synchronized in longer sections compared to route 27.

Table 6.5 Dwell Time Sensitivity at Station Mont-Royal for Eastbound Route 97

	AM Peak	Midday	PM Peak
First Order (%)	11%	7%	7%
Total Order (%)	17%	20%	18%
Interaction Effects (%)	6%	13%	11%
First Order (sec)	19.11	12.49	17.90
Total Order (sec)	30.73	38.87	43.30
Interaction Effects (sec)	11.62	26.38	25.40

As a limitation, the algorithm to infer boarding and alighting locations for each passenger is still relatively new for the agencies in the region, so the estimations here are not very precise for more detailed analyses on passenger flows. As the agencies improve the algorithm to estimate boarding and alighting locations, future researchers could do more analysis into the associated ridership variations, such as inferring trip purpose, passenger arrival patterns, ridership variation given vehicle bunching, and so on. Then, future studies could improve on this study to be more precise on dwell time variations, which could potentially help agencies better plan for these ridership and dwell time variations.

Discussion

The three case studies, routes 27, 30, and 97, show three different sources of travel time variation. They are somewhat representative of the 20 routes chosen, travelling around the centre of the city with high population density and mixed land use. The variability could be linked to the various specific operational and geographic contexts of the route, e.g. signal

synchronization, ridership, trip purpose, congestion, and land use. The sensitivity analyses across these routes demonstrate the varying influence of departure time, inter-stop time, and dwell time, and emphasize the importance of understanding interaction effects in transit planning. Planners should also be prepared to plan the temporal variations observed, since the trip elements may have different importance during the day.

In all three cases, the total-order sensitivity indices were significantly higher than first-order indices, emphasizing that many transit delays are not caused by single variables alone but by their interactions. For example, departure times determine roughly how the signal timings are likely to be for the trip. Gaining or missing a traffic signal cycle due to traffic or ridership variation would be relative to the green wave pattern determined at the departure time. Hence, these interaction effects highlight the need to tailor transit priority strategies according to the operational and geographical contexts of the route, although the analysis method could be generalized combined with the local contexts and planners' local knowledge.

6.4.3 One-At-A-Time Sensitivity Results for Westbound Route 27

In this subsection, we will attempt to demonstrate the non-linear relationship between each trip step and the trip travel time. As mentioned before, due to the limitation of one-at-a-time analysis, we will not aggregate the results into one number. Like the previous section, we will again use a westbound route 27 trip during the AM peak as an example. Similar analyses can be carried out for the other routes.

Changing Dwell Time or Inter-stop Time

Similar to previous sections, we can also examine the potential impact of ridership variation or inter-stop time variation on trip travel times while keeping the same departure time and traffic signal settings from the same median observation.

As shown in the previous section, the importance of dwell time and inter-stop times is high for a stop at the western half of route 27. Figure 6.9 shows the relationship between a time increase in either dwell or inter-stop times and the resulting red light waiting time and trip travel times.

Given the median observation, the bus completes the stop at 6:41:20 while the traffic light is green, and the traffic light would switch to red at 6:41:30. Hence, the bus has 10 seconds to clear the intersection. If for some reason, the bus spends extra time travelling to the stop or picking up passengers, there are different impacts on the red light waiting times and the travel times.

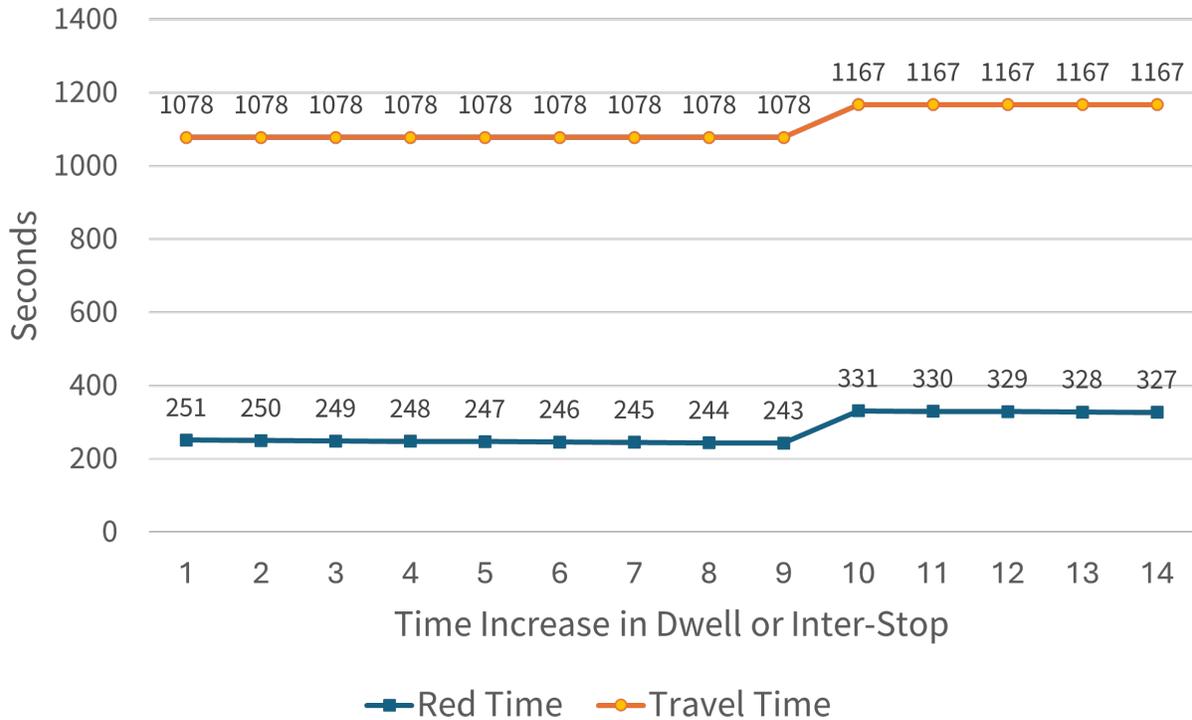


Figure 6.9 One-at-a-time analysis for dwell time or inter-stop time changes

For red light waiting times, if we add less than 10 seconds, the overall red light waiting times decrease. This makes sense because we will wait less at the next red light given the longer time spent on this specific section. For the next traffic signal, the arrival time of the median observation is at around 6:42:00, given the green light comes on at 6:42:18, the red light waiting time is around 18 seconds. If we can pass the current intersection with a small delay, say one second, the bus would then arrive at 6:42:01, which means one second less for the red light waiting time at the next intersection. The overall travel time stays the same if we add less than 10 seconds, the overall trip time stays the same. This makes sense, since we would depart the next intersection at the same green light. Thus, the red light at the next intersection absorbed the extra time needed at this given stop, and there would be no impact on downstream travel times.

If we add 10 seconds, the bus would arrive at 6:41:30, exactly when the red light comes on at this intersection. We would not be able to pass the intersection, and the bus would then depart when the next green light starts at 6:42:33, adding 63 seconds to the total red light waiting time. Then, we would update the potential arrival times for the next stop, which would be 6:43:13. Checking the signal timing plan, the next green phase activates at 6:43:58, resulting in 45 seconds of red light waiting time. The bus would then depart the next inter-

section at 6:54:58. Then, given the 20-second cycle length difference to the third intersection, the bus would pass during the green phase, rather than the red phase in the median observation, reducing 20 seconds of additional red light waiting time. The calculation would repeat itself at every stop and intersection until the bus reaches the terminal. Unfortunately, the bus would not be able to catch up for these additional times, which would give a total of around 88 extra seconds waiting for red lights or extra travel times, similar to a traffic light cycle.

To summarize, the differences between the overall red light waiting time and the overall travel time are similar but not the same. Red time would potentially absorb additional "useful" time spent travelling to the stop or picking up the passenger. Unless the bus misses a green phase, which would then cause an increase in the red time waiting times, then start decreasing again to absorb additional times needed. However, for the overall travel times, it stays relatively stable since it can be considered as the sum of inter-stop, dwell, and red light waiting times. As red light waiting times could potentially absorb some of the extra inter-stop or dwell times, the sums wouldn't change, unless the bus misses a green light and departs one signal cycle late. Planners can potentially look at the risks of not going through an intersection when evaluating the sensitive elements on a given route.

However, a limitation of our study is that the vehicle delays also cause ridership variations. If the vehicle is delayed, it may pick up additional passengers that are supposed to catch the next trip. Similarly, delays on the previous trips might affect the ridership on this trip. Therefore, we need further research into the interaction effects between ridership and delays, the interaction between transit vehicles, as well as the potential passenger arrival patterns at stops to improve our model and analyses.

Changing Traffic Light Settings

As seen from previous sections, the traffic signal timing changes were the second most important factor for this route. During the analysis period, there was one single change in the traffic light timing plan for a newly implemented bike path, as well as an improved pedestrian crossing. Three consecutive signals were affected by this change (the 3rd, 4th, and 5th signals, shown as red horizontal lines, from the start of the route in Figure 6.10). In the new timing plan, there is a dedicated bike and pedestrian signal phase, a shortened green light duration for car traffic, and a shorter overall cycle length. More notably, the cycle lengths were modified from all 100 seconds to 80, 100, and 90 seconds respectively for the three signals. As a result, the green wave patterns for these three signals repeat every 3600 seconds after the change, instead of being synchronized every cycle before the change.

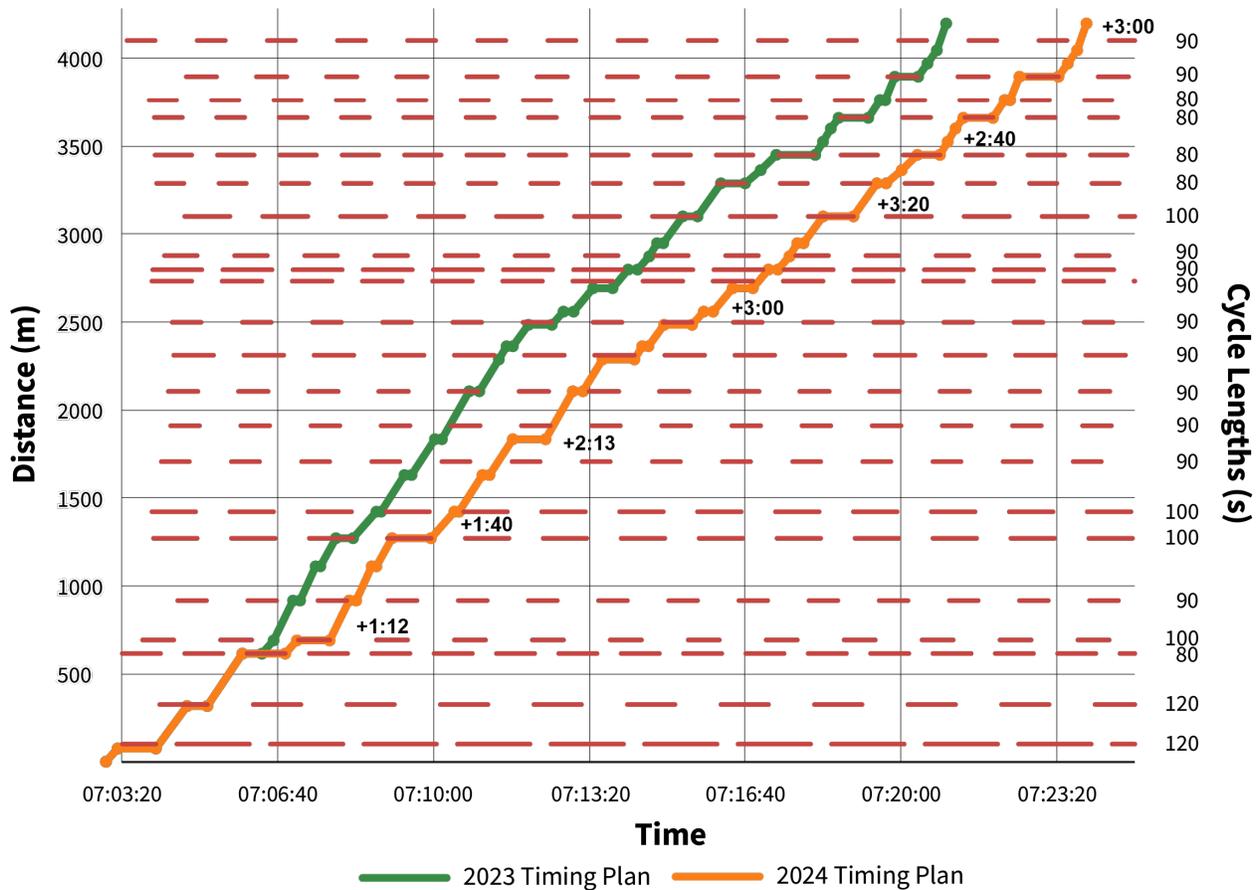


Figure 6.10 One-at-a-time analysis for traffic signal changes

Figure 6.10 shows an observed trip before the timing plan change (green line). The orange line in the figure shows the potential impact on the trip travel time if we impose the same departure time, the same inter-stop time, and the same dwell time on the newly implemented signal timing plan with the red phases shown in red. The time deviations generated and the signal cycle lengths after the timing plan change are also marked on the graph.

Given the median observation before the timing plan change, the bus would complete the stop at the intersection around 614 meters at 7:06:20. Checking the traffic signal timing plan, the light is green, and the bus can pass the intersection without any issue. Since the traffic lights are synchronized before the timing plan change, the bus would then proceed to pass the following intersections without any issue until reaching the stop and intersection at 1270 meters. However, due to the change in timing plan, now the signal is red at this time, hence the bus would wait until the green light comes on at 7:06:50, an additional 30 seconds. In addition, the cycle lengths are now different, the signal would turn red as the bus arrives at the intersection. Hence, the bus cannot immediately pass the intersection 75 meters after,

having to wait another 42 seconds for the red light. Given the extra 1 minute and 42 seconds delay, the bus would miss the green light on the original cycle. When the next green light comes on, the bus will be 1 minute 40 seconds or exactly one signal cycle late. However, since the intersection at 1700 meters received another cycle length change from 100 to 90 seconds, the bus is running one cycle and 10 seconds late. The light would remain green, but it is 10 seconds closer to the red light. However, as the bus approaches the signal at 1900 meters, the extra 10 seconds make it too late, and the bus must wait for another traffic light, further increasing the delay to 2 minutes and 13 seconds. Eventually, the bus could potentially arrive at the terminal two cycles later compared to the original timing plan given the chain reaction effects.

From the figure, we can observe a potential 3-minute increase in overall trip travel times due to the timing plan change. If we analyze the impacts locally at the spots with changed traffic signals, the difference between the two lines is only 1 minute. However, this 1-minute delay would cause the vehicle to potentially delay 2 additional minutes, due to missing more green lights later in the route. Also, from the figure, the bus very nearly misses a few signals after the changed signals. Thus, the inter-stop times or dwell times would also become more important if we consider the interaction effects. In reality, some drivers may drive fast and to avoid hitting the red lights and some may slow down, which may or may not be the desired behaviour. Again, planners want to look at the risks of hitting a red light at intersections and determine what they are going to plan for.

Once again, the cycle lengths are not necessarily the same on the entire route and the signal timing plans were changed. Thus, the green waves are not aligned everywhere along the route. Planners must consider the potential chain reaction effect after the signal changes to the end of the route rather than analyzing the impacts of traffic signal changes locally. Thus, we highlight once again the importance of including more detailed traffic signal configurations and interaction effects in the transit travel time studies.

For the three consecutive intersections used as an example in this section, a simple way to reduce the travel time variance could be to revert back to using the same cycle length of 100 seconds. However, it is not quite possible to synchronize every traffic signal on every traffic corridor in the city, given the number of signalized intersections (around 2600 in Montreal), various intersection configurations, and different travel demands. Synchronizing green waves for transit on a certain direction may also affect travel times for buses on the opposite direction. Hence, transit planners must work more closely together with traffic engineers to determine a better way to synchronize signals for buses and plan for travel times accordingly given these impacts. As mentioned by Furth et al. (2025), the best way to improve transit

travel time is to use both passive signal priority (i.e. setting traffic signals based on bus travel times) and active signal priority (i.e. based on real-time vehicle locations or sensor information).

Changing Departure Time

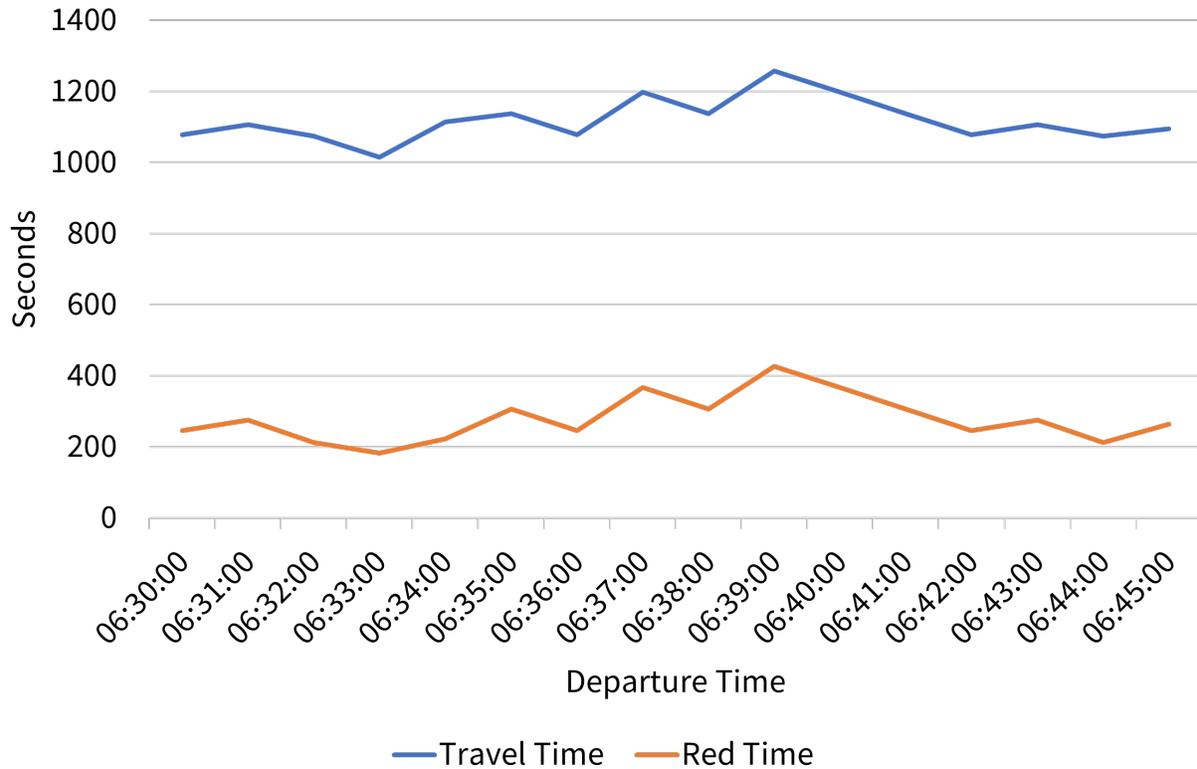


Figure 6.11 One-at-a-time analysis for departure time changes

Sometimes buses did not leave at the intended time due to late arrivals from previous trips or waiting for a running passenger. Some agencies would also use algorithms to modify the departure times by a few minutes to facilitate interlining. However, one of the important variables discussed previously is the departure time, and it was shown to contribute a lot of variation to trip travel times and the red light waiting times. Here, we use a median observation on the westbound route 27 trip departing at 6:30 as an example to demonstrate the impacts of departure time by imposing the same inter-stop time, same dwell time at stops, and keeping the same traffic signal settings. The 6:30 departure has been moved around a few times to facilitate interlining by a few minutes during the past few years. Hence, given the service frequency of 15 to 20 minutes, we will test a 15-minute window for the departure time change as an example.

Figure 6.11 shows the trip travel time and red light waiting time variations given various departure times. The times are calculated similarly to the previous two subsections, i.e. updating the corresponding arrival and departure times as well as checking the signal timing plan to determine the signal state for all stops and signals. Hence, we would not show the detailed calculations for length purposes. From the figure, we can observe that the trip travel time and red light waiting time follow the same trend, which is not surprising as we kept the same inter-stop time and dwell times. This result shows a high correlation between the overall travel times and the overall red light waiting times.

A 4-minute difference can be observed on this 20-minute route between 6:33 and 6:39 departures if we keep all other variables the same. As mentioned previously, the signal cycle lengths along this route are 80, 90, 100, and 120 seconds, with a least common multiple of 3600 seconds. Thus, given any departure time within one hour, the green wave patterns and the red light waiting times will be different, resulting in large travel time and red light waiting time variations shown above. Given the high correlation observed between the overall travel times and the red light waiting times, we once again highlight the impact of departure time changes on the overall trip and red light waiting times and the need for planners to plan carefully given the base signal timings.

Moreover, given that our on-time window is 4 minutes, small departure time changes on this route could also have a huge impact on the on-time performance. In addition, in reality, drivers may speed up to rush through the intersection before the light turns red (the interaction effects), which may or may not be a desired behaviour. Therefore, planners need to carefully consider the consequences and add adequate schedule padding or layover times in case buses cannot depart on time for some reason. When optimizing for interlinings, the assumption that travel times remain constant if departure times are adjusted only by a few minutes is not necessarily a good assumption, as shown in this section. Hence, planners might want to carefully select their departure times when adjusting schedules for interlining or for adjusting service frequency to meet changing passenger demands. Alternatively, planners could also choose to settle on a fixed departure time, so that they will remain unchanged at a given anchor point (the peak load point for example). This way, planners could directly analyze the historical travel time observations without having to deal with the varying underlying green wave patterns.

Discussion

This section highlights the complex, non-linear relationships between individual trip elements, i.e. dwell time, inter-stop time, traffic signal timing, and departure time. We also want to

highlight once again the potential chain reaction and cumulative effects due to individual element variations on the overall transit travel time.

Through one-at-a-time sensitivity analyses using a median trip observation from the west-bound Route 27 during the morning peak, we demonstrate that small variations in any single factor can trigger significant downstream impacts, especially when interacting with the various states of traffic signals. Dwell and inter-stop time variations could be partially absorbed by red light waiting time, unless a green phase is missed, leading to a significant increase in signal delays. Similarly, changes to traffic signal timing plans may break synchronization and create delays at intersections way past the changed sections. Even minor adjustments to departure time can potentially cause up to 4-minute differences in travel time due to the variation in signal cycles. These findings underscore the need for transit planners to account for interaction effects and carefully select departure times when optimizing for interlining. Given the difficulty of synchronizing every traffic signal in the city, we want to emphasize the need for collaborations between transit planners and traffic engineers to optimize signal timing strategies and scheduling practices to improve travel time reliability safely for all road users.

6.5 Conclusion and Future Research

To summarize, transit reliability is important for both transit agencies and passengers. Larger travel time variations affect the scheduling process for transit agencies and trip planning decisions for passengers.

Previous studies mostly focused on either the variation of trip travel times or the variation of specific elements in trip travel times. The main focus on travel time variation has been on the contributing factors and the overall distribution of travel times. Dwell time variation has also been studied, though not directly related to the overall travel times. Traffic signal studies are mostly focused on giving priority to buses with mixed results. We identified that there is still a need to better isolate the variation in each trip element, further decompose the travel time, examine the interactions between these trip elements, and rank the impact of each element on the overall travel time variation.

Aiming to create more comprehensive evaluations of transit travel time variations and to help prioritize the resources for transit planners, we conducted both a variance-based sensitivity analysis and a one-at-a-time analysis for observed transit travel times. More specifically, we further decomposed 3 months of recorded transit travel times from 20 routes in Montréal into several trip elements, i.e. departure time, inter-stop times, dwell times, and red light

waiting times. Departure time and inter-stop times can be obtained directly from the vehicle location data. However, due to the data limitation, the traffic signal settings and dwell times are estimated using various combinations of vehicle location data, fare transaction data, and ride-check observation data.

The results show that we can better decompose the travel times by adding more details on traffic signal settings and ridership variations, and that the variations in specific trip elements can be better isolated. The detailed data and sensitivity analyses allowed us to answer the question of where the travel time variances come from. The results for the 20 analyzed routes show that the overall travel time variations can come from all four elements identified, i.e. departure time variations, inter-stop time (congestion) variations, dwell time (ridership) variations, and signal timing plan changes. However, as to how much variation can be attributed to each step, the results would vary at different time of the day periods, from route to route, from segment to segment, and the interaction effects among different elements are very significant.

The most important element for half of the 20 routes analyzed in our case study is the departure time at the first stop, which determines how synchronized the traffic signals are further down the route, given the fixed coordinated timing plan with varying cycle lengths. In addition, most of the variation in the overall travel time and red light waiting time came from the interaction effects, since red light waiting times are the interaction effects of when the bus arrives at the signal and the signal timing plan. The inter-stop travel times and dwell times, however, are affected less by the interaction effects and can mostly be ranked by the individual variations. Hence, areas with high ridership and congestion variations tend to have high impacts on the overall travel time as expected. The one-at-a-time analysis also demonstrated the non-linear relationship between the variation of each trip element and the trip travel time variation.

Planners need to choose the departure time carefully and better understand the base signal timings before designing or adjusting service schedules. Slightly modifying the departure time to facilitate interlining is not necessarily a good idea without a better understanding of the potential impact on the overall travel times. As we demonstrated, the historical vehicle location and travel time observations are conditional given the fixed signal plan and the varying cycle lengths. A slight change in departure time can potentially cause a large travel time variation due to the changes in signal synchronization patterns. Planners could either optimize for specific departure times rather than applying the same travel time everywhere, or add adequate schedule padding to account for the travel time variations which would increase operating costs. Similarly, small changes in signal timing plan changes can have a

chain-reaction effect on the overall travel times. The travel time variations in the affected area can be small, but the impact on the whole trip can be large due to the interaction effects. Thus, planners need to include more detailed signal timing plans and red light waiting time information when scheduling or planning for transit travel times.

To conclude, we want to again highlight our contribution in the data integration, which includes common data sources widely available to the agencies, as well as the analysis scale, which is more detailed compared to the existing studies. Since the current literature tends to focus on one specific aspect of travel time, our research also adds to the literature on how to combine these different elements in a more comprehensive study and how these different elements interact. The basic framework and the results help identify important determinants for travel time variations for practitioners and future research directions for researchers.

However, there are many limitations and potential future research that are needed. The first is related to our analysis scale. We have only analyzed three months of recent data, yet we would need more observations to better reduce the confidence intervals for less frequent routes during off-peak times. There is also a need for a longitudinal study to discover potential trends and changes in the sensitivity results to better inform transit planners on the changes in the system. With more data on more routes, it will also be possible to examine potential temporal and spatial patterns affecting transit travel time reliability, which may in turn help the planning process for new schedules or routes.

Our methodology does not currently consider the interaction between buses, since delays can affect the number of boardings and alightings, where the vehicle may pick up additional passengers that are supposed to take the following vehicle. Our methodology also only considers the overall dwell time variations within a time of the day period without considering detailed overall passenger flow variations and the impacts on travel times. Unfortunately, due to the data limitations, we cannot carry out a more detailed ridership analysis or estimate a more comprehensive model here to include the overall passenger flow, various day-to-day ridership variations, and seasonal ridership patterns. Similarly, there are various studies on the interaction of bus bunching, irregular headways, and ridership variations (Arriagada et al., 2019), yet there is still a need to incorporate these vehicle and ridership interactions into the various variance analysis models to determine the interaction effects and reliability impacts of these factors. In addition, despite there is some research on the passenger arrival rates (Luethi et al., 2007), it is not easy to determine the passenger arrival rates at bus stops using the fare collection data available to us, nor to add this information into variance decomposition models. As the agency in the region improves the ridership matching algorithm, some of these limitations could be addressed by future researchers. By incorporating this information,

future researchers could improve the accuracy of the travel time models, better pinpoint issues on the routes, and better inform planning and operation decisions.

In addition, since our traffic signals mainly use fixed timing plans with very little flexibility for transit signal priorities, there is a need to better understand the interaction between buses and non-fixed timing signals using empirical data, as the majority of the current literature is based on simulations. This way, we could potentially compare various transit systems to identify pros and cons in various configurations to help improve transit reliability. Similarly, future research could find a way to better give signal priority to buses in travelling down a coordinated corridor with densely configured signals.

Acknowledgements

The authors would like to thank the Société de Transport de Montréal and Autorité Régionale de Transport Métropolitain for providing data access.

Author Contributions

The authors confirm their contribution to the paper as follows: study conception and design: Yuxuan Wang, Catherine Morency, Martin Trépanier; data collection, analysis and interpretation of results: Yuxuan Wang, Catherine Morency, Martin Trépanier; draft manuscript preparation: Yuxuan Wang, Catherine Morency, Martin Trépanier. All authors reviewed the results and approved the final version of the manuscript.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research is funded by the Natural Science and Engineering Research Council of Canada and the Canada Research Chair in Transportation Transformation.

Data Accessibility Statement

Most of our research data are freely available online, such as the GTFS, GTFS-RT, and OpenStreetMap files. Researchers can access the most up-to-date versions directly from the related websites. The archived GTFS data, GTFS-RT data, and the code used in the study can be provided upon reasonable request to the authors. As we continue to improve the precision of our estimates and algorithms, we do not guarantee the exact values remain unchanged. However, we cannot provide the fare transaction data due to potential privacy issues in accordance with our data agreement with the transit agencies.

CHAPTER 7 ARTICLE 4: IMPROVING BUS TRAVEL TIME ESTIMATIONS BY INCORPORATING DETAILED TRAVEL TIME COMPONENTS AND INTERACTION EFFECTS

Yuxuan Wang^{acde}, Catherine Morency^{acde}, Martin Trépanier^{bce}

^a Department of Civil, Geological and Mining Engineering, Polytechnique Montréal, Montréal, Québec, Canada

^b Department of Mathematics and Industrial Engineering, Polytechnique Montréal, Montréal, Québec, Canada

^c Interuniversity Research Centre on Enterprise Networks, Logistics, and Transportation, Montréal, Québec, Canada

^d Canada Research Chair in Mobility, Montréal, Québec, Canada

^e Canada Research Chair in Transportation Transformation, Montréal, Québec, Canada

Paper submitted on 13 November 2025 for the Journal of Public Transportation

Abstract

Accurate travel time estimations are critical for public transit performance and passenger satisfaction. Typical travel time analyses often rely on higher analysis levels, hiding the detailed variations in different travel time components, namely, inter-stop travel times, dwell times, and traffic signal delays. Recently, more detailed analysis levels and methods have been proposed, but with limited insights on how these components can be integrated into travel time estimations.

This paper builds on our previous work on travel time decomposition and sensitivity analysis by extending the framework to a travel time estimation context. We discovered significant interaction effects between different trip components in our previous work, and interactions with traffic signals could create significant overall travel time variations Wang et al. (2025). In this paper, we take the opportunity to provide more high-level system-wide summaries on traffic signal impacts. In addition, we propose a travel time model that incorporates these detailed travel time components, and we evaluate the accuracy of the proposed model using data from future service periods.

Results demonstrate that estimating travel times using detailed travel time components can

potentially improve accuracy and responsiveness to potential changes in the network. The proposed model could reasonably reproduce observed travel times with good precision and could potentially be used to expand and improve interline optimization algorithms. These findings further highlight the importance for planners to analyze travel times in more detail and to incorporate detailed travel time components into planning contexts to deliver more adaptive and responsive services.

Keywords: Public Transit, Bus Travel Times, Travel Time Reliability, Travel Time Estimation

7.1 Introduction

Reliable and precise public transit travel time estimations are essential for public transit systems. For agencies, travel time estimates determine vehicle allocation, operator assignments, and overall service reliability, which are related to the operating costs. Inaccurate travel times can lead to delays, increased operating costs, and driver retention issues due to missing their breaks (Danaher et al., 2020). For passengers, good travel time estimates are helpful for trip planning and affect passenger mode choices. Reliable travel time estimates also improve passenger satisfaction (Carrel et al., 2013a).

In practice, travel times are generally analyzed at an aggregated level, often at the timepoint level in North America (Danaher et al., 2020), without fully accounting for the variability introduced by individual trip components such as dwell times, inter-stop travel times, and delays at traffic signals. These simplifications, while practical to analyze, may hide more detailed variations due to the complexity of real-world operation conditions, potentially leading to large overall variations which may affect passenger experiences.

Recent advancements in literature have created opportunities for more granular analyses of bus operation data, such as more detailed dwell time impacts (Gallo and Sacco, 2022) and signal delays (Furth et al., 2025). While prior research has examined travel time reliability and its components individually, limited attention has been given to how these components interact and how such detailed insights can be integrated into planning processes, especially when estimating travel times. For example, the interaction with traffic signals can significantly influence overall trip duration if the bus just misses out on a green light. Yet current travel time estimation frameworks rarely incorporate these dynamics. This issue becomes more important in dense urban environments, where traffic signals are frequent, congestion varies, and ridership patterns fluctuate throughout the day.

Building on our previous work that decomposed bus travel times into distinct components and analyzed their contributions towards the overall travel time variation (Wang et al., 2025), this paper advances the discussion by exploring how decomposed travel times can potentially be used in various planning and travel time estimation scenarios. Specifically, we propose a simple yet flexible travel time model that uses these detailed components and their interactions in travel time estimations. We then compare the travel times estimated with the actual observations to evaluate the accuracy. We also test the performance of estimated times in future service periods to see the potential of using the model for planning. This approach allows planners to test various operational scenarios, evaluate the sensitivity of travel time to changes in departure times, congestion, ridership, and traffic signals. By doing so, we aim

to bridge the gap between detailed operational data and practical travel time analyses and help planners better react to changes in the network.

In addition to extending the methodological framework, we also wish to take this opportunity to provide additional system-wide empirical results to complement our earlier study, which were based on a few selected routes. While our previous work focused on decomposing travel times and ranking the sensitivity of individual trip components, the current work expands the scope by presenting system-wide descriptive statistics on signal delays across our bus network, quantifying their contribution to both total travel time and variability, given the relatively few empirical evaluations in the literature.

Our contributions in this paper are first, the integration of detailed components into a travel time model. We demonstrate how departure times, inter-stop travel times, dwell times, red light waiting times, and their interactions can be integrated in a simple travel time model. Next, we contribute to the development of a scenario-based travel time model. By using these detailed travel time components, we introduce a model that enables planners to simulate different changes in the network individually, such as departure time changes, and assess their impacts on travel times more proactively.

The paper is organized in the following ways. In section 7.2, we summarize existing literature on bus travel time studies and current scheduling practices. Next, in section 7.3, we describe our research framework, methodology, and data used. We also introduce the proposed travel time model incorporating these detailed travel time components in section 7.3. Then, in section 7.4, we present the results from Montréal, Canada as a case study to evaluate the potential use cases in travel time analyses and estimation scenarios. Finally, in section 7.5, we provide a short summary to conclude this paper.

7.2 Literature Review

Travel time is an important measure for transit agencies and passengers. Scholars have mainly focused on evaluating the overall travel times. More recently, there has been more detailed research on specific components of travel times, i.e. inter-stop times, dwell times, and red light waiting times. However, there was less literature focusing on the interactions between these components and how to improve travel time estimations while incorporating these additional details.

The intense focus on travel time studies started after the early adoption of Automatic Vehicle Locating systems, which allow transit agencies to better monitor the performance of their services. Early proposals to incorporate these data into operation planning processes can

be traced to the early 2000s (e.g. Furth (2000)). Travel times and their variations are important inputs for transit agencies since they determine the vehicle requirements, the schedule padding, and layover times needed, which in turn affect the operating costs National Academies of Sciences, Engineering, and Medicine and others (2013). In practice, travel times and travel time variations are typically analyzed at four different levels, line, trip, timepoint to timepoint, and stop to stop levels according to an agency survey Danaher et al. (2020).

Good travel time and variation estimates are important inputs in service planning, especially in scheduling. As for the general scheduling practice, Coleman et al. (2018) provided a summary of a typical scheduling process at New York City Transit. In general, the route performances are reviewed at various fixed intervals for different types of routes and schedules to facilitate driver sign-ups for their work. If passenger levels exceed a predefined agency standard, service frequency is adjusted. Travel times are also analyzed using an agency standard, and then adjusted both between the two terminals as well as between various timepoints using measures such as the mean, median, or a given percentile of observed travel times (Furth, 2000). Once the headway and travel times are set based on ridership and the average round-trip times, travel time variations are used as additional inputs for setting terminal recovery times. Good schedules should allow enough running time to operate the route, account for congestion and signal delays, and provide enough recovery time to absorb potential delays at the terminal so that most buses can start the next trip on time (Furth and Muller, 2007).

Using the archived vehicle location data, agencies could also retroactively improve the accuracy of transit agencies' schedules. Using the observed vehicle location data, Wessel et al. (2017) created a bus timetable representing service delivered, which can then be used in assessing transit performance, reliability, and accessibility issues. Since the practice generally focuses on the more aggregated level for travel time analyses, typically at the timepoint level. Passengers, on the other hand, require arrival and departure times for their specific stop served by a certain trip. Thus, there is a discrepancy between general planning practice and what's shown to the passengers (Wessel et al., 2017).

There are also attempts to focus on more detailed estimations for individual trip time components, i.e. inter-stop, dwell, and red times. For red light waiting times, most literature relates to traffic signal priority. Much of the literature showed positive impacts on travel times and variances from implementing signal priority Lin et al. (2015). However, some research did not show a significant impact due to various reasons, such as near-side stops combined with no turn on red policy causing cars to block bus departures (Diab and El-Geneidy, 2013). Researchers have suggested that good arrival time predictions are important for transit priority

signals to be effective at reducing travel times Li et al. (2012); Furth et al. (2025). Better inter-stop and dwell time estimations are needed to improve the arrival time predictions. Planners should also adapt their scheduling practice to take advantage of the signal priorities based on the conditions to receive signal priorities. Altun and Furth (2009) developed simulation models to test the impact of schedule aggressiveness on bus running times, headway regularity, and crowding. They found that conditional signal priority, which only gives priority to late buses, combined with more aggressive schedules and strategic holding, can lead to a reduction in travel times.

For dwell time at stops, the literature generally focused on modelling the determinants of dwell times and the variability of dwell times. The time variations are typically attributed to passenger boardings and alightings, in vehicle crowding causing friction for passenger movements, fare collection methods, and ramp usages Dueker et al. (2004); Rashidi et al. (2023); Currie et al. (2013). Less attention is paid towards the dwell time impact on the overall travel times. More recently, Gallo and Sacco (2022) proposed an optimization of bus dwell times to reduce the red times at the following intersection. However, their method only accounts for one intersection, which showed a reduction in red times. Yet, it is still necessary to consider the possible impacts for the future stops and intersections as well as the impact of various signal timing parameters. In addition, it is still important to consider how to incorporate dwell times and ridership levels into the planning and travel time analysis process.

Given there is a lack of literature focusing on how various trip components interact with each other causing overall travel time variations, we proposed to use more detailed travel time observations in variance decomposition sensitivity analyses (Wang et al., 2025). In our previous paper, we decomposed 3 months of bus travel times into several trip components, i.e. departure time, inter-stop times, dwell times, and red light waiting times. Departure time and inter-stop times can be obtained directly from the vehicle location data, whereas the traffic signal settings and dwell times are estimated using various combinations of vehicle location data, fare transaction data, and ride-check observation data. We showed that variations in specific trip components can be better isolated by adding more details on traffic signal settings and ridership estimations. The analyses also show that the overall travel times and red light waiting times are more affected by interaction effects between trip components, mainly due to the dense traffic signals and varying cycle lengths in our fixed signal timing plans where the green wave pattern rarely lines up. The overall inter-stop times and dwell times are mainly affected by large individual variations of inter-stop travel times and ridership respectively. Thus, planners must consider potential chain reactions where small variations in one trip component can lead to significant changes in the overall trip times.

To summarize, the existing literature has mostly focused on evaluating the overall travel time or one specific component of the overall travel time. Our previous study examined how various travel time components affect the overall travel time and interact with each other as well as ranking the sources of variations for each route. However, since transit agencies have limited resources, there is still a need to help planners prioritize their resources, e.g. adjusting their travel time estimates to take advantage of the more detailed information in various planning scenarios.

Given that the practice generally does not focus on incorporating signal timings in travel time estimations and analyses, we aim to provide more information with respect to the system-wide red light waiting times and sensitivity results as additional evidence to support the decision-making process. Aiming to provide better tools to help planners and react to changes in the transit network, we will also attempt to propose and evaluate a simple travel time model that takes advantage of the extra information we could obtain from more detailed information.

7.3 Methodology and Data

The methodology generally follows and expands on the methods we started to develop in our previous paper (Wang et al., 2025). Here, we include a short summary. In addition, we further explore the possibility of using these detailed times in a planning context by proposing a simple scenario-based travel time model while taking into account the decisions operators make while operating the trips.

7.3.1 Decomposed Travel Times

Overall, we proposed to analyze three categories of detailed travel time components, inter-stop times, which relate to congestion, dwell times, which relate to passenger activities, and red light waiting times, which relate to traffic signal impacts. We would also need to include the departure times as an input to reconstruct the arrival times at traffic signals to estimate traffic signal states (Wang et al., 2025). Hence, there are four types of inputs in our analysis.

We use the data provided by Société de Transport de Montréal (STM) on the island of Montréal in Canada as a case study. The system currently has 222 bus lines in operation, about 1850 buses in the fleet, and more than 17,000 published bus trips on average weekdays.

We used the General Transit Feed Specification (GTFS) feed to obtain the route shape as well as stop locations. However, since GTFS does not provide traffic signal locations, we matched the route shape with OpenStreetMap to obtain detailed traffic signal locations on

the route. Since there may be zero or many traffic signals between two passenger stops, we then consider the traffic signals as operational stops with no passenger activities to better isolate the time wasted waiting for traffic lights to turn green.

The travel time of a given trip could be decomposed as the departure time, plus a series of inter-stop travel times, plus a series of dwell times, and plus a series of red light waiting times. The departure time and the series of inter-stop travel times associated with a trip can be obtained directly from the vehicle location logs. However, the dwell times and the red light waiting times are less obvious given that the majority of the stops in Montréal are near-side. Therefore, the stopped time would include both dwell and red light waiting times. We proposed to separate them using an estimated dwell time model and estimated traffic signal parameters.

To estimate the dwell time model, we mainly followed the previous work from Dueker et al. (2004). The observations used as inputs to estimate the linear regression model come from our long-term ride check observations, which include roughly 25,000 detailed door opening times, number of boardings, number of alightings, and any kneeling or ramp usages. Since the agency uses the passenger fare transaction data to estimate the trip origin and destinations, we then aggregated these estimations to calculate stop-level boarding and alightings to estimate dwell times.

In Montréal, most of the traffic signals follow fixed timing coordinated plans due to the high density of traffic signals. To estimate these three traffic signal settings, i.e. cycle lengths, offsets, and green splits, we mainly followed the methodology proposed by Fayazi et al. (2015) with some minimal modifications to handle a few particularities of our local vehicle locations feed. The basic idea is to calculate arrival and departure times at a given traffic light observed over a period, then test out various cycle lengths and offsets to see which combination matches the observations the best. Readers interested in the details can refer to the original paper cited above. The estimated times are also validated during our ride checks.

Here, we reproduce our example in our previous paper (Wang et al., 2025) to demonstrate how the times are calculated. In Figure 7.1, we have a stop-to-stop travel time observation where the bus starts from the terminal and travels to stop 2 on the route. The overall stop-to-stop travel time is 2 minutes and 50 seconds. Then, we decomposed the travel times into the inter-stop time, i.e. the time between stop 1 departure and stop 2 arrival, which is 1 minute 50 seconds. In this example, despite the bus being a few meters away from the stop for over a minute, the door did not open due to congestion. Then, as the bus stops at the second stop, the stopped times are split into dwell times and red light waiting times. According to

the dwell time model and the estimated passenger boarding and alighting numbers, the dwell time at this stop is around 20 seconds. Then, the bus stays at the near-side stop for the remainder of the red light, until the green light activates 40 seconds later according to the estimated timing plan. The decomposition is repeated and performed for every stop-to-stop observation for a given trip.

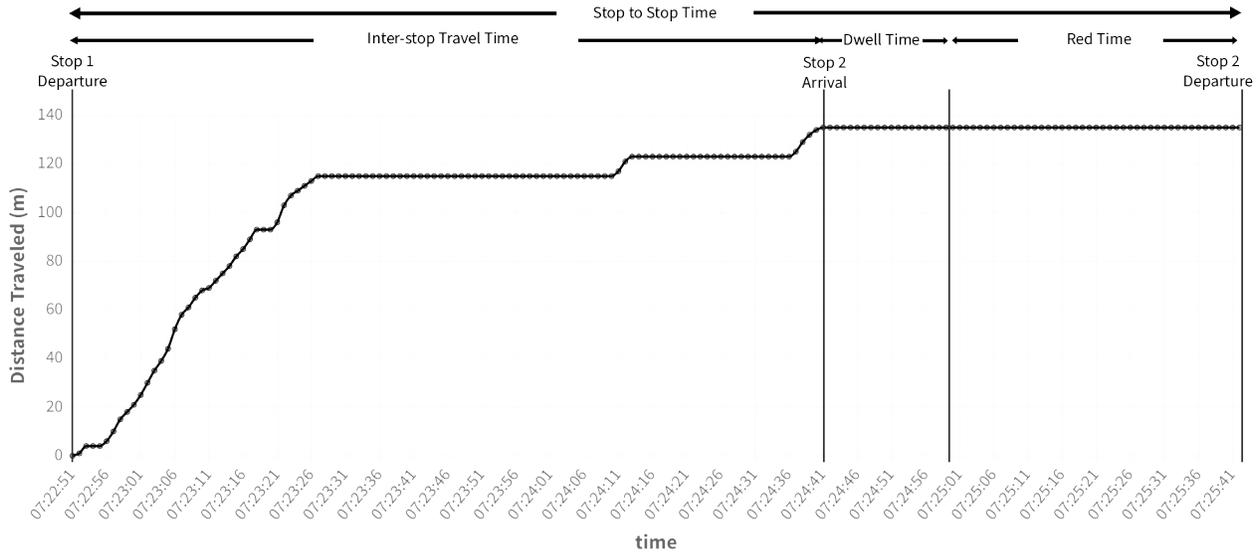


Figure 7.1 Example: Splitting a Stop to Stop Travel Time Observation Reproduced from Wang et al. (2025)

7.3.2 Short Summary on Variance-based Global Sensitivity Analysis Method

In our previous work, global sensitivity analysis was used to identify which individual component of bus travel time, such as inter-stop times, dwell times, and red light waiting times, contributes most to overall travel time variability. The global sensitivity analysis is based on a variance decomposition framework, where total travel time is expressed as a function of multiple interacting travel time component variables. Unlike one-at-a-time sensitivity analysis which cannot handle non-linearity and interaction effects, variance-based global analysis accounts for both individual contributions and interaction effects among variables (Saltelli et al., 2008).

The results from the analysis typically include first-order and total-order indices. The first-order index measures the percentage of overall variance attributed to a single input variable without interaction effects. The total-order index is the percentage of the overall variance attributed to a single variable, including both its direct effect and its interactions with other

variables. The differences with and without interaction effects allowed researchers to determine not only which factors are influential on their own but also the ones that become critical when interacting with the others (Saltelli et al., 2008). The model inputs included decomposed travel time components derived from vehicle location data, fare transactions, and estimated signal timing plans, while the output was the overall trip travel time (Wang et al., 2025).

7.3.3 Proposed Travel Time Model

To evaluate the effectiveness and potential of using these detailed times in planning contexts, travel time analyses more specifically, we propose a simple travel time model in this paper following the sequences of events outlined above.

First, a vehicle would depart at the given departure time from the first stop, since the passenger activities at the starting terminal are typically considered as part of the recovery time (National Academies of Sciences, Engineering, and Medicine and others, 2013). Then, we would add the inter-stop travel time to the next location that potentially requires the operators to stop, i.e. a passenger stop or a traffic signal. We check if passengers are waiting or if the light is red respectively. If passengers are waiting in the given planning scenarios, we add the estimated dwell times. If the traffic signal is red, we add the remaining red time. If there are no passengers or if the light is green, the vehicle would pass right away with no time added. Then, repeatedly, we would add the inter-stop times to the next location to check for passengers or traffic signals.

In short, this simple model would cumulatively sum up the individual travel time components to obtain the overall travel time, which is simple enough to be implemented as a small computer program or in Excel. Readers can find a more detailed example calculation in our case study section. This way, planners could test out various planning scenarios, such as evaluating different congestion levels by adjusting inter-stop times, evaluating ridership change impacts by adjusting passenger patterns, and testing new traffic signal configurations by adjusting signal timing plans.

7.4 Case Study: Montréal, Canada

In this section, we will provide more system-wide results to complement our previous research findings as well as to demonstrate the impacts of traffic signals on the bus routes. In addition, we will provide a few travel time scenarios to evaluate the potential usefulness of applying the detailed times in our simple travel time model.

7.4.1 System-Wide Descriptive Statistics for Red Times

First, as mentioned in the previous paper, there is a significant variation in travel times that can be attributed to red time variations (Wang et al., 2025). Since the current literature has relatively fewer empirical evaluations on the signal delays using real-world observations, we will summarize the impacts of red times for all the daytime routes in Montréal as a demonstration.

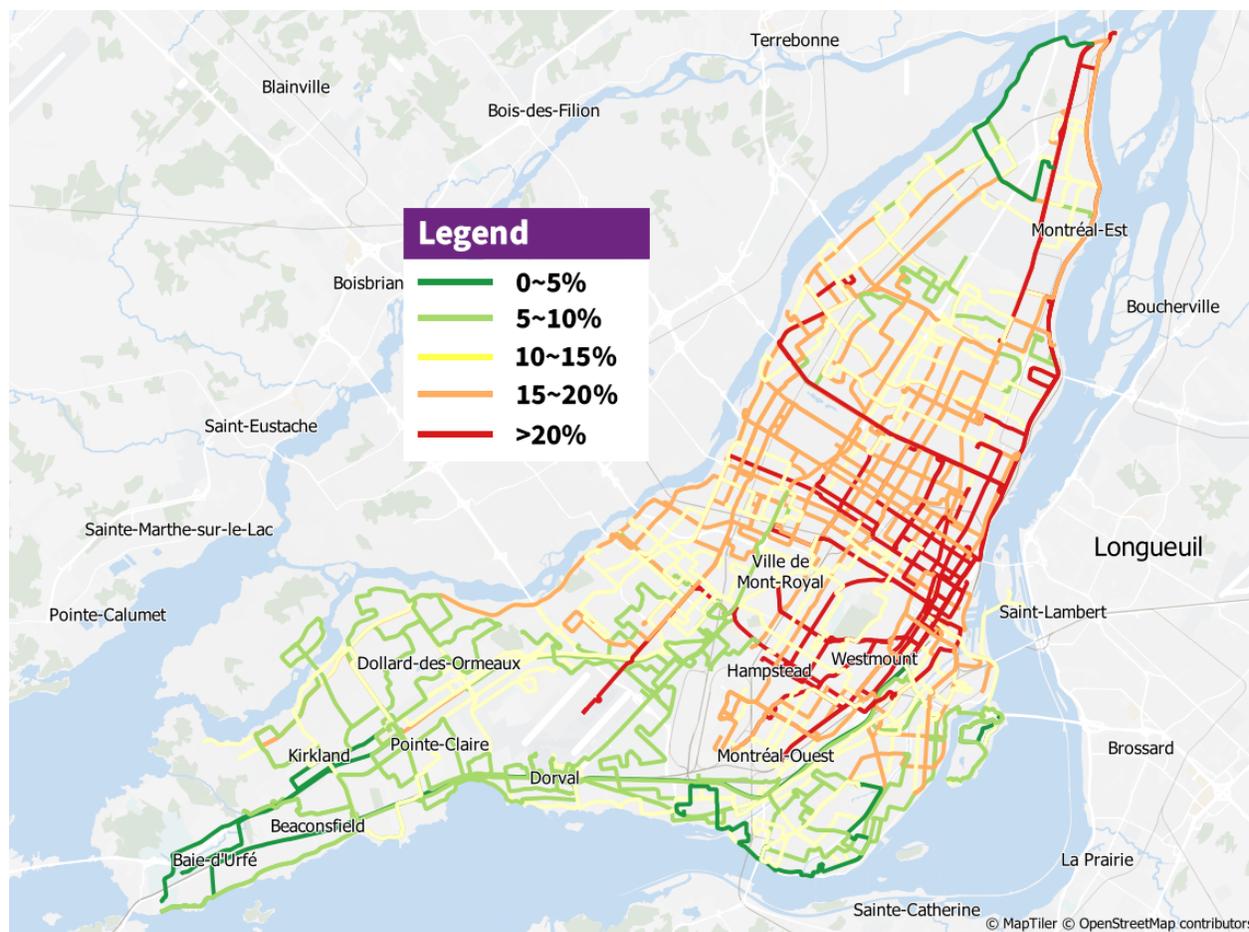


Figure 7.2 Percent of Total Travel Time Attributed to Red Times for Weekday services from Jan. 2024 to Mar. 2024

Figure 7.2 shows the overall red times as a percentage of total travel times. We can observe that many routes in the central area have a higher amount of red times compared to the western part of the island. The red times for many routes in the central area account for more than 20% of total travel times, due to the higher traffic signal density, sometimes as dense as 70 meters between signals, resulting in very slow operating speeds for the bus routes in this area. For the express routes that operate on highways or routes in the suburbs,

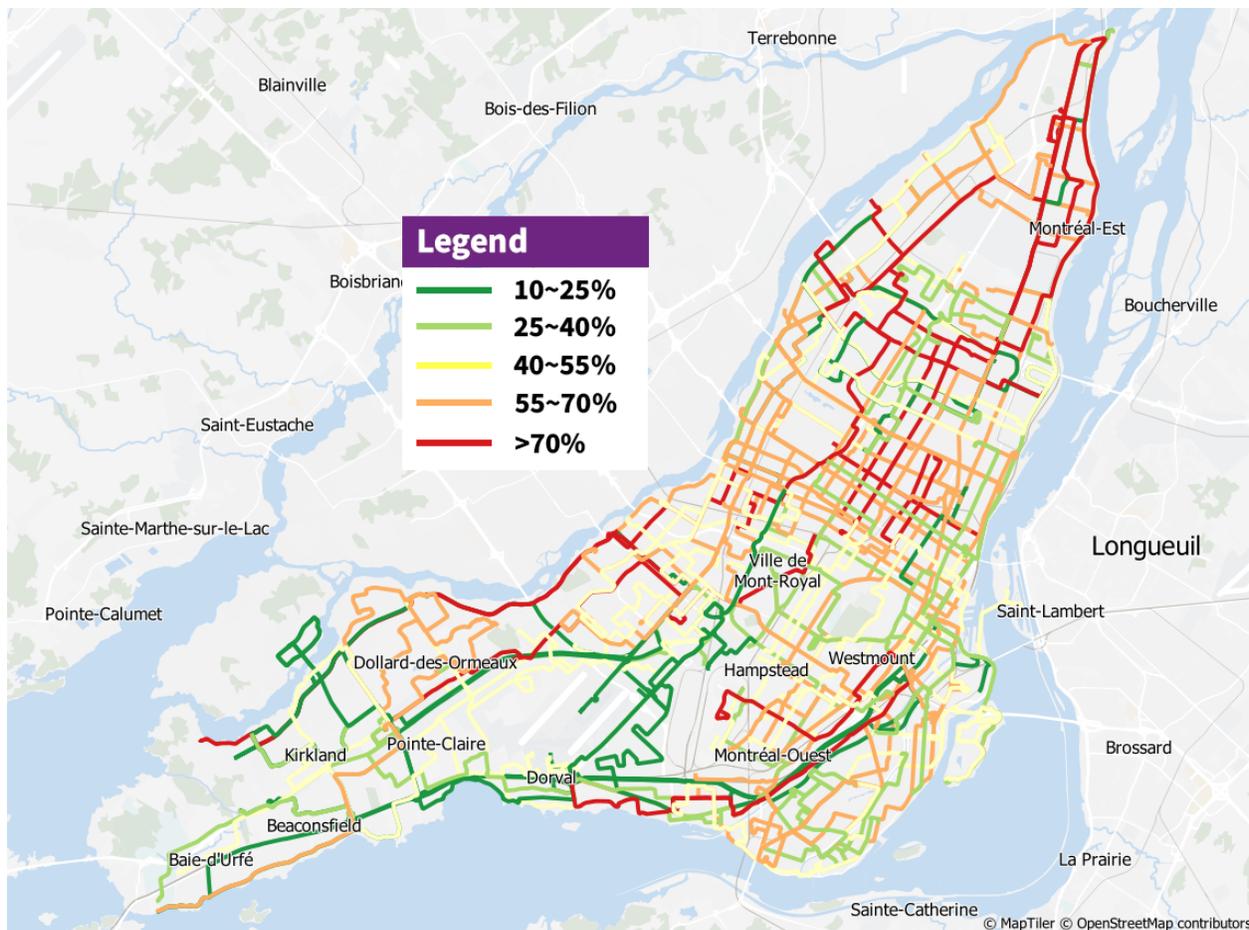


Figure 7.3 Red Time Variations as a Percentage of Total Travel Time Variation for Weekday Services from Jan. 2024 to Mar. 2024

especially the west island, the red times only account for less than 10% of total travel times, since there are fewer traffic signals in the suburbs and on highways. Unless the service runs on major suburban corridors, then red times could contribute to around 15% of total travel time, due to longer distances between traffic signals and relatively free travel conditions.

For all daytime routes, red times account for 15% of the travel times on average at the network level for the early 2024 observations, which equals 1624 service hours per weekday. Hence, a significant portion of STM's total operating budget for the bus network, around 900 million dollars in 2025, can be attributed to wasted time waiting for traffic lights to turn green. Similarly, given the estimated origin and destinations for passenger trips, we can estimate that red times account for around 19,572 passenger-hours per day, which has a very significant impact on the passengers. Given the higher ridership counts in the central area, combined with the dense traffic signals and slow operating speeds, the impacts from

red light waiting times are very significant in terms of potential passenger experiences on crowded buses.

As pointed out in Furth et al. (2025), near-zero signal delays and faster travel times for transit vehicles are achievable goals which require both passive and active signal priority controls. Therefore, at least in our case study in Montréal, there is a significant time and cost saving potential from reducing red times for buses. Better travel times and more reliable services would increase both passenger satisfaction and operational efficiency. The time saved could either be used to run extra services or to give more breaks to the operators to improve their work satisfaction. Hence, we emphasize the importance of transit planners cooperating with the city's traffic engineers to continuously evaluate the impact of red times on travel times and find solutions to reduce the red times for buses. Similarly, given that cities generally have their own specific signal timing guidelines and the relatively fewer practice and research literature on evaluating empirical red times, we want to call for more empirical evaluation of traffic signal design on bus operations, which could help improve our understanding of the impacts related to various signal timing strategies.

Similarly, we calculated the ratio between the total travel time variations and the red time variations. Note that, due to interaction effects and correlation between each individual time component, the variance cannot be directly added and the sum of detailed variances will not equal 1. Figure 7.3 shows the estimated amount of travel time variations attributed to red time variations. From the figure, we can observe that most routes are above 55%, i.e. the red time variations account for 55% of the total travel time variations. Hence, a large portion of the travel time variations can be considered as gaining or missing traffic light cycles. The express routes and routes in the west island still have a lower percentage here, due to the fewer traffic signals again.

However, compared to the contribution to the overall travel times, the contribution towards the overall travel time variation is more evenly distributed. Routes that operate in the central areas have relatively less significant travel time variation attributed to the red time variations compared to routes that operate slightly outside of the central area or on major suburban corridors. This is mainly due to inter-stop travel time variations, i.e. congestion, becoming more significant in central areas, especially shopping areas, despite the densely configured traffic signals. Given the observations from our previous paper that traffic signals do not have the same cycle lengths, the car queue lengths may also vary due to varying green wave patterns, sometimes causing extra vehicles to block the bus access to near-side stops, forcing the bus to wait another traffic signal cycle. Congestion, more commonly found in the central area, could also cause buses to get stuck for one or two traffic signal cycles

at a time, hence slightly lowering the significance of red time variations. Routes that are slightly outside of the downtown core or on major suburban corridors have more car-centric design with many lanes and relatively less congestion, but still have many traffic signals, thus the contribution from red time variation remains high. This result does not mean red time variations are not significant for routes in the central core. It simply means that extra congestion would lower the proportion of red time variations contributing towards the overall travel time variations. Once again, we want to emphasize the need for transit planners and traffic engineers to collaborate and lower the variation in red light waiting times for transit vehicles, as these variations would cause the overall travel time reliability issues, which would incur extra operating costs.

7.4.2 Potential Travel Time Estimation Scenarios for Westbound Route 27

In this section, we will try to test the possibility of using the decomposed travel times and add detailed signal settings, departure times, and ridership variations into the travel time modelling process. We will include some example calculations using detailed data from Route 27, which is a short and straight route connecting a dense residential area with the metro system.

Simulating different travel conditions using the original data

First, we want to test if our proposed travel time model would generate estimates close to the observation, given the most relevant and comprehensive data. More specifically, we try to reconstruct the travel times from early 2024 and generate a few travel time estimates using different travel conditions given various thresholds. Then, we will compare the estimates to the observed value to examine if our travel time model generated results similar to the observed.

Here, we use the median travel condition but with ridership variation as an example to demonstrate the calculation process for one specific trip, i.e. we test out the model using three ridership scenarios with median travel conditions, same departure times, and same traffic signal settings. Figure 7.4 shows the estimated vehicle trajectory in a time-distance from the travel time model using the decomposed observations for the trip departing at 6:30 on route 27 heading west. Since the traffic light states depend on the arrival times at the signals, we will show the actual time in the graph. The travel times (marked at the top right corner) can be calculated by subtracting the trip departure time at 6:30 from the terminal arrival times. The orange line shows the estimated trip trajectory with median ridership, where the green line has the lower ridership at the 20th percentile (roughly 5 to 7

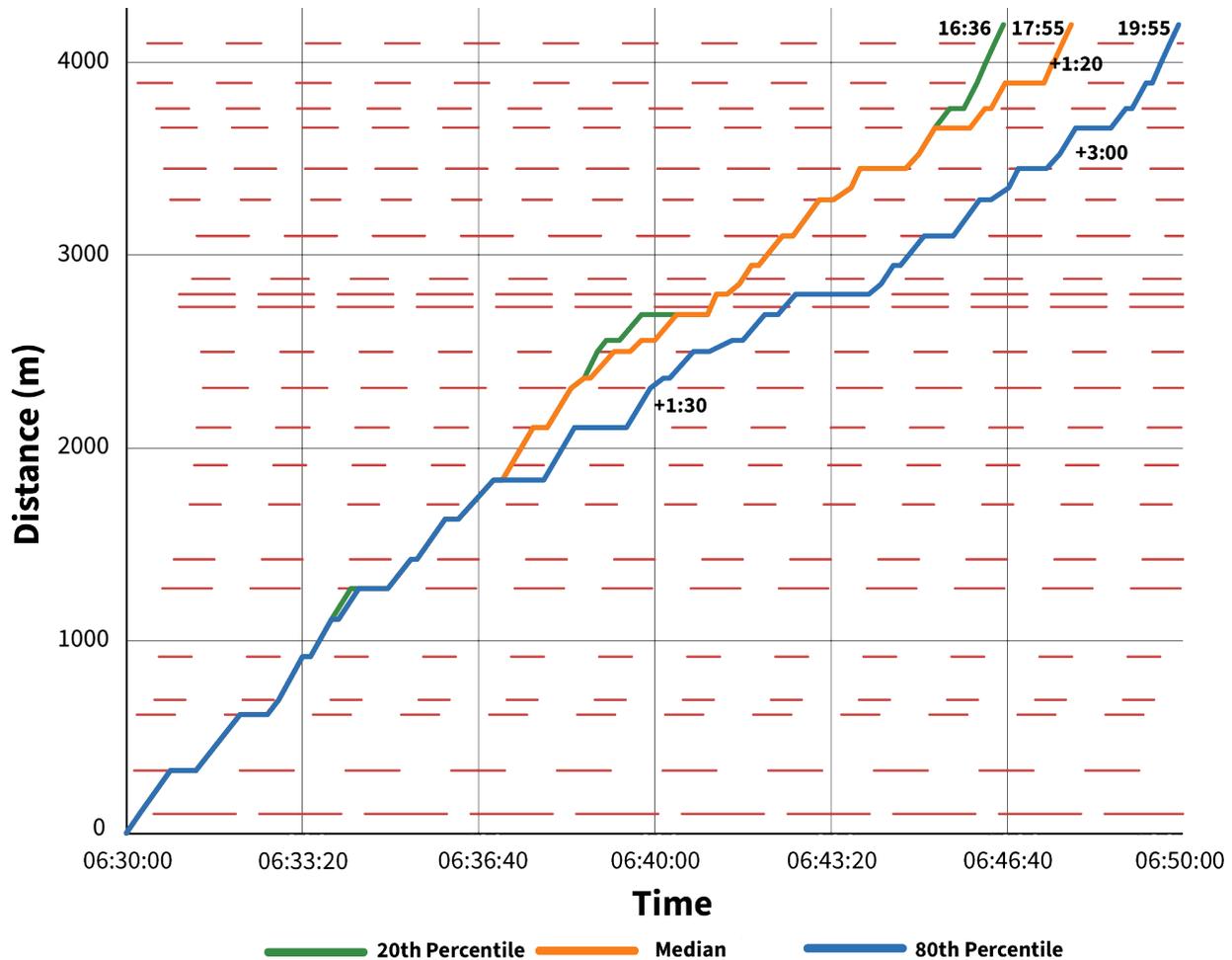


Figure 7.4 Estimated Bus Trajectory for Three Ridership Scenarios on Westbound 27 Departing at 6:30 in Time-Distance Graph

passengers per trip) and the blue line has higher ridership at the 80th percentile (roughly 5 to 7 passengers per trip).

Starting from the terminal at 0 meters, the vehicle departs at 6:30, 20 seconds later at around 100 meters, we check the traffic signal state, which is green, allowing us to pass with no additional time added. Then another 35 seconds later, the bus would arrive at around 300 meters, where the bus would pick up 0 to 2 passengers depending on the ridership scenario, translating to 0 to 12 seconds of dwell time. However, after the passengers have boarded at this near-side stop, the vehicle could not depart since the traffic light remains red until 6:31:10. Therefore, the three scenarios would still depart at the same time despite the additional passengers boarding. Hence, the travel time up until stop 2 is 1 minute 10 seconds. Then, the pattern repeats itself, where we add around 40 seconds of travel time to

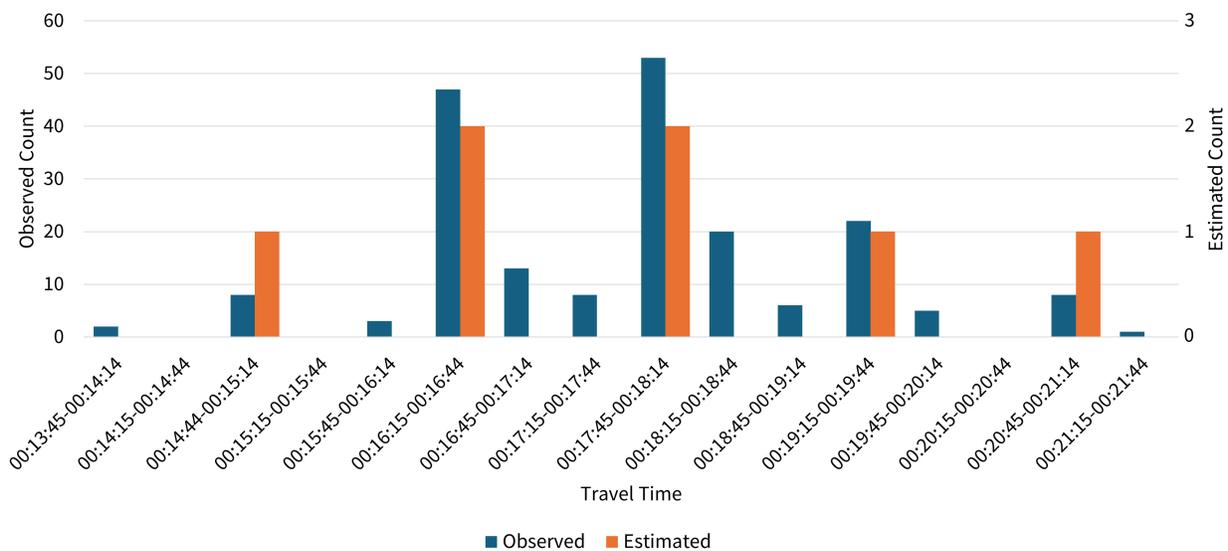


Figure 7.5 Travel Times Observed and Simulated for Trips Departing at 6:30 on Westbound Route 27 in Montréal

stop 3 at around 600 meters, the dwell time needed to pick up passengers at stop 3, checking the signal state, and departing when the signal turns green.

Repeating the pattern to account for all inter-stop, dwell, and red times, the terminal arrival times are at around 6:46:36, 6:47:55, and 6:49:55, translating to travel times of 16 minute 36 seconds, 17 minutes 55 seconds, and 19 minutes and 55 seconds, corresponding to the three main "peaks" in the actual observations shown in Figure 7.5. The potential to save time with fewer ridership is about 1 minute 20 seconds. However, the extra time spent on more ridership is about 2 minutes. Therefore, we show once again the non-linearity observed in the travel time changes due to various ways and possibilities to miss a traffic signal cycle. Thus, ridership variations may not impact the overall travel time equally.

More specifically, as we can observe, the scenario with fewer passengers can skip a stop at around 1100 meters from the start. However, due to the upcoming red lights, the less than 10-second gain was quickly erased by the red light. Before the vehicle reaches the stop at around 1900 meters from the start, the three scenarios generally behave similarly. However, at this stop, the first two scenarios could quickly depart from the stop, barely passing the following green signal. The scenario with slightly more ridership was unfortunately stopped by the traffic signal, potentially causing the vehicle to be one cycle out of sync with the other two scenarios. Similarly, as the vehicle progresses, gaining a signal cycle from a shorter stop at 2300 meters was erased 300 meters later due to long red times. Since the ridership near the beginning of the route stays quite stable and the second half of the ridership varies more,

we can see more obvious splits among the three scenarios in the latter half of the route.

Table 7.1 shows a few additional scenarios that we used to test the overall travel time estimations with varying ridership and congestion levels. The inputs to the travel time models were inter-stop travel times and the ridership patterns based on the 20th, 50th, and 80th percentiles. The 20th percentile is chosen to reflect free-flow travel conditions, and the 80th percentile is chosen to reflect more congested conditions. Similarly, the 20th percentile ridership is chosen to reflect a quiet day, potentially with more people working from home, and the 80th percentile is chosen to reflect a busy day, potentially with many people going to work. We also included the median observations to test if the estimated travel times correspond well to the recorded median times. Since our goal is to test if the travel time model corresponds well to the original data, we kept the same traffic signal settings and departure times from early 2024 to evaluate if these inter-stop and ridership variations would be similar to the recorded terminal arrival times. We have added the distribution of estimations in Figure 7.5 to help indicate the relation with actual observations.

Table 7.1 Estimated Travel Times Given Congestion and Ridership Scenarios

Travel Conditions	Ridership	Travel Time
20th percentile	20th percentile	15 min 09 sec
20th percentile	50th percentile	16 min 29 sec
50th percentile	20th percentile	16 min 36 sec
50th percentile	50th percentile	17 min 55 sec
80th percentile	50th percentile	18 min 05 sec
50th percentile	80th percentile	19 min 55 sec
80th percentile	80th percentile	21 min 02 sec

By comparing the table with the histogram of the actual distribution of terminal arrival times for the trip departure at 6:30, we can see that the travel times correspond well to the actual observations. From the graph, we can observe that the three most likely travel times, around 16 minutes 30 seconds, 18 minutes, and 20 minutes, are very close to the times estimated. For the first major "peak" of travel times at around 16 minutes and 30 seconds, it roughly corresponded with the faster travel combined with median ridership or with the median travel condition but with less ridership. Similarly, the second "peak" at around 18 minutes is similar to the scenarios with median travel conditions and ridership or with slower travel conditions but median ridership. Please note that even though we are slowing down and taking on the same passengers, the impacts are relatively small, where the trip would terminate with a 10-second difference in both scenarios. Again, given our previous discussions, traffic signals could sometimes absorb the travel time or dwell time variations

until the vehicle is on the verge of missing or passing a green signal, which would cause a large increase in travel time variation due to red times. For the third peak at around 20 minutes, it is similar to median travel conditions combined with more passengers. For the two extreme cases, faster travel with low ridership and slow travel with high ridership, they generally align with the observations right before and right after the three main "peaks". Given the similarity between the estimated times and the observed times, we think planners could potentially test various ridership and travel time scenarios based on previous ridership observations during the planning process to evaluate if the scheduled travel times and recovery times are adequate. They could also examine the potential impacts from service changes, e.g. evaluating if headway-related ridership pattern differences would impact the overall travel times. As we mentioned, the differences from adjusting the thresholds are not necessarily the same.

Hence, by generating various vehicle trajectories and travel time estimates, planners could better communicate what they have planned. They could also better evaluate the terminal recovery times and better tie the recovery times needed to different travel conditions and ridership levels. Finally, since the base traffic signal synchronization patterns were determined at departure, planners could use these trajectory estimations to better adjust traffic signal synchronizations. Signal priority could also be more proactive and apply changes in advance to ensure vehicles do not get left behind from a longer stop (e.g. the blue higher ridership line in Figure 7.4). Traffic signals could equally slow down a vehicle when it is operating ahead of schedule to provide a more on-time service to all passengers.

Estimating travel time for one future westbound trip

In this section, we test the possibility of using the decomposed travel times to improve travel time estimations for one departure on route 27 based on the January 2024 observations for future service changes. After the previous study, there are a few service changes. First, since the agency performs interline optimizations which automatically adjust the departure times by a few minutes, there are a few departure time changes. In addition, a few traffic signal timing plans were adjusted either temporarily or permanently. We would apply the same travel time and ridership observations to the new departure times and traffic signal timings.

Table 7.2, the observed median travel time for all the studied sign-up periods, including the scheduled departure times, and observed travel times for each period, the estimated travel times for periods after January 2024, as well as the differences between the estimated and observed travel times in seconds both in terms of the trip level differences and in terms of the average differences at stop level which passengers care the most.

Table 7.2 Estimated and Observed Median Travel Times for Westbound Route 27 Departing Around 6:30

Period	Departure	Time Obs.	Time Est.	Route Diff.	Average Stop Diff.
2024 Jan	06:30:00	16 min 55 sec	-	-	-
2024 Mar	06:30:00	17 min 15 sec	16 min 49 sec	26 sec	18 sec
2024 Jun	06:28:00	16 min 46 sec	17 min 04 sec	-18 sec	13 sec
2024 Sep	06:27:00	18 min 19 sec	18 min 04 sec	15 sec	24 sec
2024 Nov	06:27:00	18 min 21 sec	18 min 04 sec	17 sec	32 sec
2025 Jan	06:32:00	18 min 16 sec	17 min 56 sec	20 sec	24 sec
2025 Mar	06:36:00	19 min 13 sec	19 min 27 sec	-14 sec	15 sec
2025 Jun	06:28:00	16 min 56 sec	17 min 02 sec	-8 sec	15 sec

Given the potential impact of departure time on travel times shown in the previous studies and sections, due to variation in traffic signal synchronizations. It is important to note that the absolute majority of the observed trips departed within a few seconds of the scheduled departure time, which would allow us to make direct comparisons. Shortly before the scheduled departure, the on-board driver console would play a message to allow the passengers and drivers to get ready for the departure. Then, the driver console would also sound a beep to indicate the exact departure time, allowing precise departures.

From the table, we can see that the average travel time is almost 17 minutes from the observations in early 2024. If we are assuming the same travel time of 17 minutes for planning purposes, the travel times could be fine, since the differences are at around 20 seconds. Given that the interval of vehicle position updates typically ranges from 5 to 20 seconds, this difference is not very significant. However, there are several departure times and signal timing changes that could contribute to the increase in travel times, and applying the same observed travel times could make it harder for buses to remain on time on this route. There may be potential problems for the September 2024, November 2024, and January 2025 periods, as the observed median travel time becomes around 1 minute and 25 seconds longer than the early 2024 observations. The result would be worse for the March 2025 period, where the observed travel times are 2 minutes 20 seconds longer than the early 2024 observations. For trips that operate on days with relatively higher ridership, the difference could be even more significant due to the possibility of longer dwell times at stops, causing the bus to miss additional traffic signals, as demonstrated in the previous section. Given the on-time window used at the agency is from 1 minute early to 3 minutes late, directly using early 2024 observations for future planning could potentially be problematic. In addition, the layover or recovery times at the end of the route are typically set at 10% of the travel times per labour agreement. For

this route, it is set at 2 to 3 minutes, increasing the possibility of having potential cascading delays for downstream trips in the vehicle trip block or vehicle run.

Using the same median inter-stop times and passenger counts at each stop from the January to March 2024 period, we estimated the travel times for the following periods with the corresponding signal timings and departure times. In the table, we can see that the estimated travel times are around 20 seconds off from the observed values. For the periods with longer observed travel times, the estimations also captured the extra times needed, where 1 minute 10 seconds were added to periods between late 2024 and early 2025, and 2 minutes 30 seconds were added to periods starting in March 2025. Although the observed times in January 2024 are closer to the observed times in the March and June periods in 2024 as well as the June 2025 period, the estimated times can be a more flexible method to estimate travel times while also keeping a reasonable error margin at around 20 seconds. Even if we look at the stop-level travel time differences for passengers making their trip plans, the average stop-level differences are generally within 30 seconds, with individual differences varying between a few seconds and a minute. Again, at the trip level, the difference of 20 seconds is generally similar to the update frequency for vehicle locations. At the stop-level, the planners could create a more defined and accurate schedule for passengers, given the fact that most North American bus schedules are not planned at the stop level, yet passengers need these times to plan for their trips (Wessel et al., 2017). More granular vehicle movement data might potentially help improve the precision estimation, although 20 seconds may not be very significant for scheduling, as most of the departure times are generally communicated in minutes. Hence, by incorporating the new departure times and traffic signal timings, the estimated travel times can better account for the travel times needed to complete the trip. This would also allow planners to quickly evaluate if the terminal recovery or layover times are enough to absorb potential delays and ensure the next trips start on time.

In practice, service schedules needed to be developed in advance to facilitate operator signing up for their shifts. When developing schedules in advance, we understand that planners may not have access to the latest or the most accurate traffic signal timing plans, travel time changes potentially due to roadwork-related lane closures, and ridership patterns potentially due to modifying service frequency for example. In this paper, our goal is to argue that planners should do these more detailed analyses to better understand the travel conditions and continuously update their travel time estimates based on the latest information. More collaboration between various agencies and departments is needed to make these upcoming changes available for bus scheduling purposes. This way, service planners could better react to potential upcoming issues, such as traffic signal timing changes and departure time changes, causing extra travel times and increasing the possibility of delays. The up-to-date information

could help planners find solutions to improve reliability more rapidly and help better inform the drivers to make their decisions on the road to provide good service to passengers.

Estimating for all future westbound 27 trips

As demonstrated in the previous section, if we could improve the travel time estimations for upcoming service changes for one trip, given the updated departure times and traffic signal timings, the logical next step is to test the model on all departures for the same route travelling in the same direction.

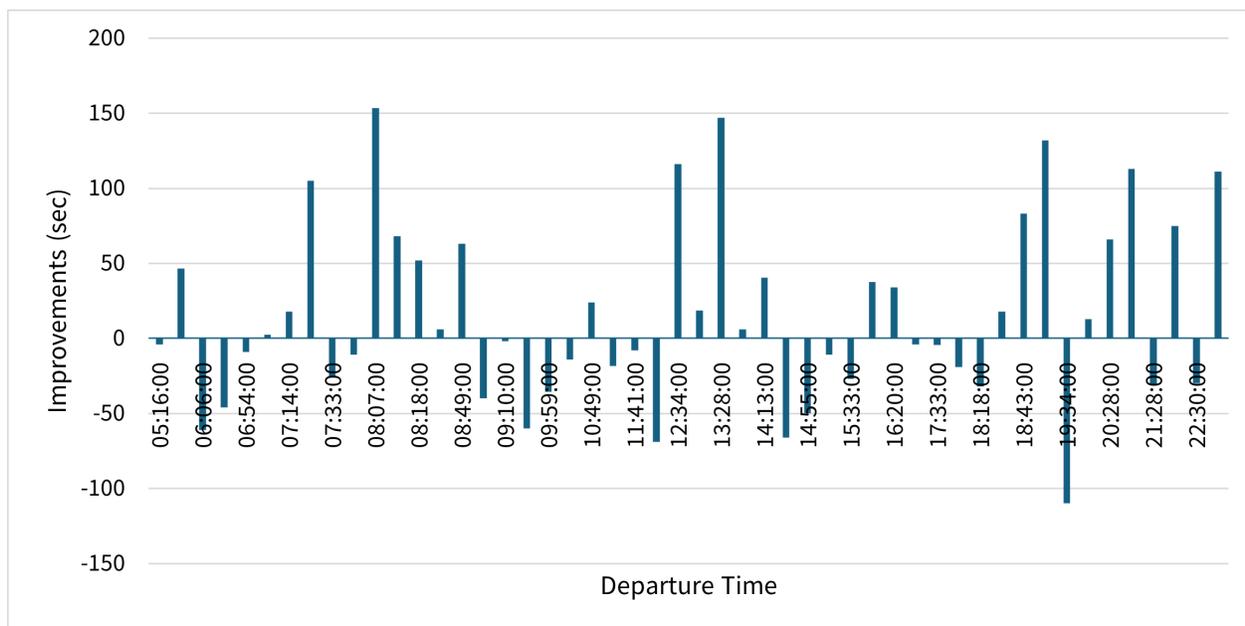


Figure 7.6 Travel Time Estimation Improvements Compared to Averages for Westbound 27 Trips Throughout the Day

Figure 7.6 shows the improvements obtained from estimations based on updated departure times and traffic signal timings compared to simply applying the average trip travel times across all westbound route 27 departures during the day for the September service change in 2024. We chose this period because many departure times and traffic signals were adjusted, which would allow us to test our estimations.

To summarize the figures, among the 70% of the trips that have improvements, the estimates are mostly around one minute closer to the real observed travel time for morning and evening services compared to simply using the averages. Several departures have more than 2 minutes of improvements, highlighting the need to improve our travel time estimates. Once again, given the labour agreement, the minimum scheduled layover time for this route is around 2

to 3 minutes. Hence, 2 minutes closer to the real observation is significant for planning in this context.

On the other hand, there are fewer improvements during the late morning and pm peak trips, and that translates to around 30% percent trips without improvements. Similar to the demonstration in the earlier section, the estimated travel times for these trips without improvements are still reasonably close to the actual value. During these times, the differences between the two models are generally less than a minute, and the difference between the estimated travel times and the observed values is generally around 20 to 40 seconds. For three trips, the estimates are significantly worse by around 1 minute 20 seconds, similar to a traffic signal cycle, compared to using the average. By looking at the stop-level data, the relatively worse observations are typically caused by driver changes in the middle of the route, waiting at extra traffic lights, and ridership variations. As a result, the difference between estimated travel times and the observed travel times is also large, at around 1 minute 10 seconds to 1 minute 20 seconds for these trips, again similar to a traffic signal cycle.

The model we proposed using the decomposed times couldn't account for mid-route driver changes, since run-cutting and driver sign-ups happen after the schedule and the travel times have been created. Therefore, there are additional times stopped at the shift change location that the model using decomposed travel times couldn't plan for. However, the shift changes tend to happen on similar trips for the following service periods. Thus, the average travel times could better account for the additional time needed, despite the fact that most scheduling processes do not include driver changes mid-route. For passengers planning their trip or passengers on board, these additional mid-route driver changes could contribute to travel time variations and could result in missed transfers and so on. Therefore, we want to highlight the need to account for driver changes when planning new schedules and when communicating travel times to passengers.

Another difficulty observed when modelling travel times at this level of detail can be attributed to the various operator behaviours. Once again, we come back to the sensitivity analysis in our previous paper that overall travel times are really sensitive to the traffic light synchronization (Wang et al., 2025). There are various driver preferences when encountering a yellow light or when there are only a few seconds of green light left. Some may be more inclined to depart immediately as long as they are allowed to, and some may choose to wait and depart at the next cycle, potentially resulting in roughly 1 minute 30 seconds of travel time variation. Using the decomposed travel times, planners could then explain what they have planned and what they assumed more clearly to the drivers, so that when drivers could adapt their driving to account for various changes encountered on the road, such as extra con-

gestion and passengers. Again, more granular, higher-resolution vehicle location logs could help improve the estimates. However, to accommodate variations in travel conditions and operator behaviours, planners could potentially test out various thresholds using the model and plan for adequate travel times and recovery times between trips to ensure reliable operations. Planners could also consider implementing more aggressive signal priority strategies to ensure buses depart during the intended traffic signal cycle, e.g. more red time for early buses and more green time for late buses.

It is also difficult to account for ridership changes without a detailed ridership prediction due to the data limitations for our project. There is also a rush-hour only limited-stop service that complements this route, and there are also as many departure time changes related to headway variations. The interactions between vehicles could mean passengers have two potential options, the local service scheduled to depart right in front of the limited-stop service could mean some limited-stop passengers would catch the local service instead, affecting the dwell time and the overall travel time. Similarly, departure time changes may require passengers to shift to a later departure or an earlier departure, again, affecting dwell times and the overall travel times. Hence, we want to highlight the future research needs to include detailed ridership variations in the travel time model.

Adjusting departure times to find better estimated travel times

Since planners often optimize for interlining by adjusting departure times by a few minutes, as shown in the earlier sections where the departure times were modified frequently. We want to test if planners could use similar approaches to incorporate the departure time-related travel time changes into the optimization process. Planners could potentially optimize and move the departure times by a few minutes so that buses can have a better traffic signal synchronization to depart to reduce travel times, or to improve travel time reliability given our reasonable success in estimating travel times from earlier sections. Here, we will limit the adjustments to plus or minus 2 minutes, which equals the maximum traffic signal cycle length on this route. We also limit the adjustments to whole minutes, since planned terminal departure times are all on top of the minute at 0 seconds. In other words, we would check if better estimated travel times could be obtained if we leave one traffic cycle earlier or later.

Figure 7.7 shows the number of trips for each potential adjustment to obtain a better traffic light synchronization. As we can see, there could be better traffic signal synchronizations and reduced red times for 84% of the future trips by adjusting departure times by a minute or two to align with potentially better traffic signal synchronization patterns.

Figure 7.8 shows the potential travel times gained by adjusting departure times. As we can

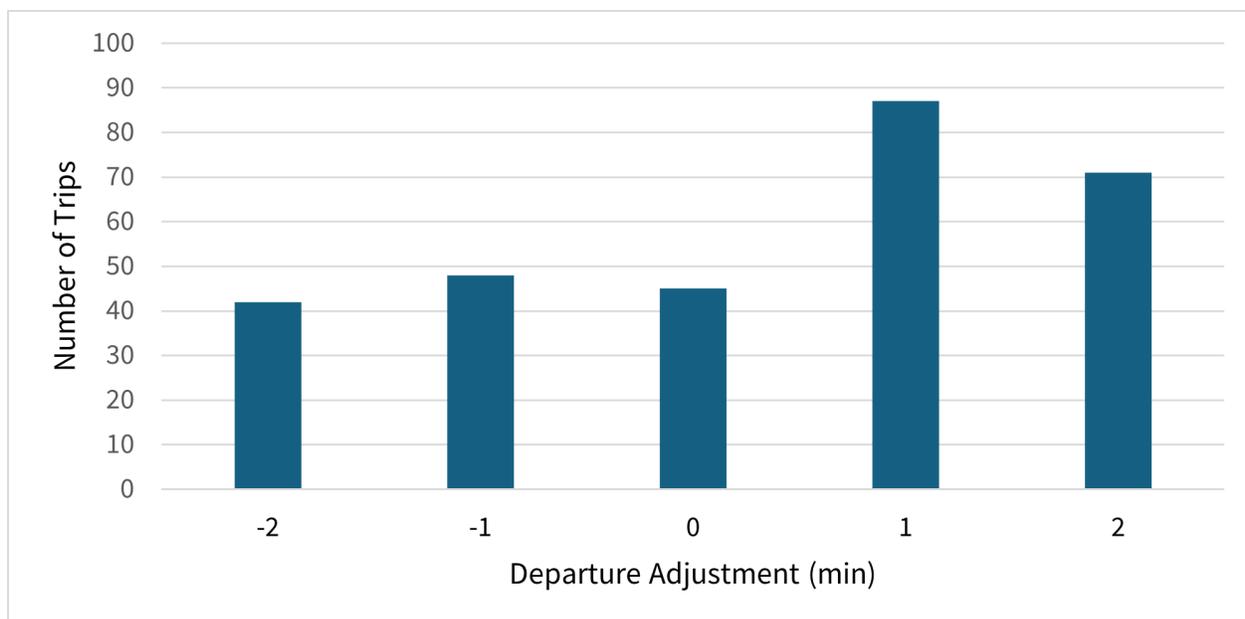


Figure 7.7 Departure Time Adjustment to Improve Estimated Travel Times

see in the graph, around one-third of trips could save between 1 minute and 1 minute 30 seconds alone, which roughly corresponds to one traffic light cycle. It is also common for trips to save between 2 minutes and 2 minutes 30 seconds, which is roughly two traffic light cycles. Otherwise, the savings are generally less than 1 minute. If the original intended departure times are extremely unlucky given the traffic signal synchronizations, there could be a 4-minute gain for this 20-minute route, which is very significant for saving costs and ensuring drivers have enough breaks between trips.

Hence, we believe we could potentially perform these types of optimizations to improve travel times or the reliability of travel times. Optimization algorithms could potentially incorporate these estimates while adjusting departure times for interline operations. However, as mentioned in previous sections, most schedules are developed months before the actual service dates to allow drivers to sign up for their work. It is less likely that the planners have access to the most up-to-date information related to traffic signal timings. Therefore, planners should collaborate with the city's traffic engineers to find the best solutions to ensure fast and reliable bus trips. Optimization algorithms should also adapt, given the uncertainty in travel time variations, to ensure good and reliable travel times.

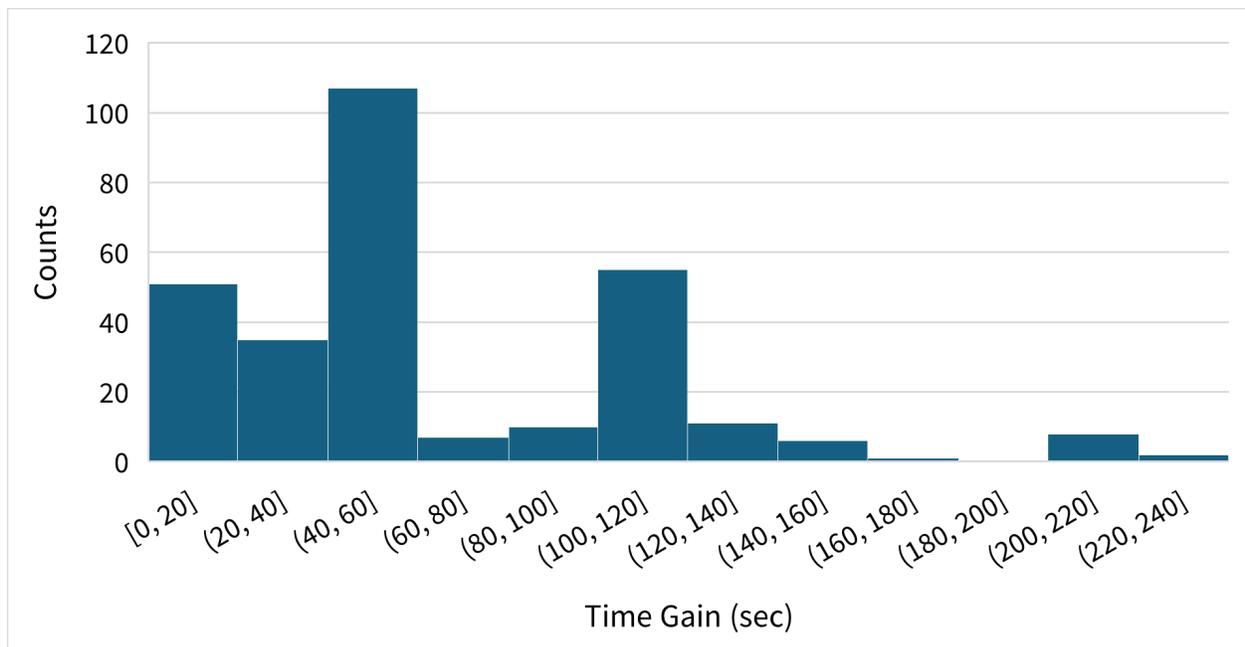


Figure 7.8 Potential Travel Time Gained from Departure Time Adjustments

7.5 Conclusion and Next Steps

This paper explored how detailed travel time components, i.e. departure time, inter-stop times, ridership estimations, and traffic signal timing plans can be integrated into transit planning to improve the accuracy and reliability of bus travel time estimates. Building on our previous work, which focused on variance decomposition and sensitivity analysis, we extended the framework to a travel time estimation context by proposing a scenario-based travel time model. This model enables planners to simulate different levels of congestion and ridership conditions, evaluate the impacts of departure time adjustments, and incorporate detailed operational factors into different travel time estimates.

Our case study in Montréal demonstrated that red light waiting times represent a significant proportion of total travel time and variability, particularly in dense urban areas with frequent signals. The expanded system-wide sensitivity analysis for all daytime routes revealed that departure time remains the most influential factor for overall red time variations on most of the routes. Given the red times' significance in overall travel time and overall travel time variations, departure times are still the primary determinant for overall travel time variations. To address red time impacts on bus services, planners can adjust departure times to align with green waves, modify traffic signal parameters, or implement more aggressive signal priority strategies, ideally combining these approaches for optimal results. Future research

can explore scenario-based travel time estimations we proposed in this paper to evaluate how such adjustments in scheduling and signal timing can improve bus performance under various constraints.

Then, we proposed a scenario-based travel time model that uses the detailed travel time components, and we tested the potential use cases for the model. We showed that incorporating detailed inter-stop, dwell, and signal delay components can produce travel time estimates that closely align with observed values, even under changing operating conditions. The model performs generally better compared to simply using the observed average travel times in the planning context. However, more information is needed to improve the accuracy and the usefulness of the model, such as mid-route driver change and the most up-to-date signal timings for example. Finally, our experiments highlighted the potential benefits of incorporating departure time-related travel time changes into inter-line optimization processes to improve travel times and reliability, with possible savings of one to two minutes per trip.

Despite these results, several limitations remain. Our model does not fully capture mid-route driver changes, operator behaviour differences, or ridership interactions between overlapping services on the same corridor. The high possibility of not having the most up-to-date traffic signal data, congestion levels, and detailed ridership forecasts could constrain the precision of our estimates. Hence, we emphasize the need for traffic engineers to collaborate with transit planners to improve communications on and plan for upcoming changes. Future research should also focus on integrating predictive ridership models, higher-resolution vehicle location data, and real-time signal control strategies into the travel time models to improve the accuracy and the responsiveness to changes in the network. Further work is also needed to improve interline optimization algorithms to incorporate these detailed travel time components into large-scale scheduling and analysis processes.

Acknowledgements

The authors would like to thank the Société de Transport de Montréal for providing data access. This research is funded by the Natural Science and Engineering Research Council of Canada and the Canada Research Chair in Transportation Transformation.

CHAPTER 8 ARTICLE 5: SPATIAL-TEMPORAL EQUITY ASSESSMENT OF PUBLIC TRANSPORT TRAVEL TIME RELIABILITY

Yuxuan Wang^{acde}, Catherine Morency^{acde}, Martin Trépanier^{bce}

^a Department of Civil, Geological and Mining Engineering, Polytechnique Montréal,
Montréal, Québec, Canada

^b Department of Mathematics and Industrial Engineering, Polytechnique Montréal,
Montréal, Québec, Canada

^c Interuniversity Research Centre on Enterprise Networks, Logistics, and Transportation,
Montréal, Québec, Canada

^d Canada Research Chair in Mobility, Montréal, Québec, Canada

^e Canada Research Chair in Transportation Transformation, Montréal, Québec, Canada

Conference paper submitted on 31 October 2025 for the 2026 World Conference on
Transport Research in Toulouse, France

Abstract

Public transport reliability is important for the agencies and passengers. Unreliable services affect passengers' satisfaction and mode choice. Most of the literature focuses on evaluating reliability for actual passengers or the accessibility measures to assess equity. Some passengers may still be forced to travel between less popular origin-destination pairs with long travel times. It is relevant to assess the reliability between all origin-destination pairs within the service area and evaluate the temporal and spatial distribution of reliability indices.

We propose to simulate potential passenger trips across the entire service area and compare the planned travel times against delivered travel times based on archived vehicle position to determine the reliability across the region. Using discrete metrics like on-time performance and continuous metrics like buffer times, we discover spatial and temporal variations of these reliability indices for a given origin and across the entire region.

We found that various reliability thresholds affect the reliability outcomes significantly. Reliability for some routes remain constant all day, either good or bad, but some change throughout the day. The number of reliable routes for each origin zone varies slightly across the region, where the numbers remain more consistent for central areas than for the suburbs.

Keywords: Public Transport, Passenger Travel Times, Travel Time Reliability, Equity

8.1 Introduction

Public transport agencies continuously monitor and try to improve their service reliability, which is important for both service operations and passenger experiences. Unreliable services increase operating costs for agencies due to the additional schedule padding or layover times (Wessel and Widener, 2017). Similarly, lengthy wait times, in-vehicle delays, missed transfers, and being left behind at stops due to crowding negatively affect passengers' satisfaction and their future intentions to ride public transport (Carrel and Walker, 2017).

Many studies have examined the reliability of public transport travel times from both operational and passenger perspectives. For operations, the researchers have focused on various elements of travel times at various analysis levels, such as timepoint to timepoint travel times typically used for scheduling (Muller and Furth, 2001), dwell times at stops (Dueker et al., 2004), as well as signal delays at intersections (Altun and Furth, 2009).

For passengers, having customer-centric travel time metrics, such as excess buffer times and excess waiting times (Uniman et al., 2010), can help improve public communication to riders, help measure how passengers are affected by service disruptions, improve transparency, and give agencies metrics that better reflect their priorities towards passengers (Halvorsen et al., 2019). However, these studies tend to focus on the recorded passenger trips and existing passenger demands.

Public transport planners have also looked at opportunity-based metrics, such as connectivity and accessibility (Handley et al., 2019), to examine the equity aspects of a given public transport network. Researchers are also adapting these opportunity-based metrics to include the service reliability variations (Liu et al., 2023). Yet, these studies tend to focus on the accessibility for a specific time budget.

Some riders may get stuck using public transport, doing extra-long commutes with low service frequency every day. Reliability issues have been found to reduce accessibility measures, contributing to the underestimation of potential equity issues (Liu et al., 2023; Lee and Kim, 2023). In addition, for riders with different mode choices, a lower public transport demand for a certain origin-destination combination could be due to unreliable travel times forcing passengers to use other modes (Carrel and Walker, 2017). With most passenger travel time studies focusing on services experienced by actual passengers, it is still important to assess the reliability between all origin-destination pairs as if passengers were to travel on public transport across the entire service area, the potential spatial and temporal distribution of travel time reliability in the region, and the potential impacts on different population segments.

With the goal of improving our understanding of the spatial and temporal patterns of public transport reliability, as well as attempting to recognize potential equity issues related to service reliability, this paper analyzes the reliability from a given location to all possible destination zones as if passengers were to make these trips regularly. This provides more detailed information to planners on potentially unreliable origin-destination pairs as well as when and where, in the network, the reliability issues occur.

In this paper, we propose a framework to explore the reliability of the overall passenger travel times. More specifically, we simulate the planned routes with the fastest travel times given different departure times between all origin and destination pairs using 3 months of archived Automated Vehicle Location data from Montréal, Canada. Next, we compare the planned travel times and the operated travel times that would have been experienced by passengers for all weekday services within our analysis period. These reliability indicators are then stored in an origin-destination matrix, which then allows us to discover the spatial and temporal patterns for public transport service reliability.

The paper is organized in the following ways. In section 8.2, we summarize existing literature on passenger-centric travel time reliability metrics, their planning applications, and other opportunity or potential-based studies. Next, in section 8.3, we describe our research framework, methodology, and data used. Then, in section 8.4, we present the results from Montréal, Canada as a case study. Finally, in section 8.5, we provide a quick summary to conclude this paper.

8.2 Literature Review

Numerous indicators and diagnostic processes have been proposed to improve public transport operations. However, since vehicle-based travel time measures do not show how many people were affected and for how long, researchers have argued that the focus should shift more to passengers (Hendren et al., 2015). Passenger-based measures can improve transparency and give agencies metrics to better reflect their priorities towards passengers. Agencies could also use these metrics to understand how incidents and service adjustments affect passengers (Halvorsen et al., 2019). However, these passenger-based metrics are harder to translate to improve operations, as they only illustrate the impacts but not the cause of these impacts for corrective actions (Danaher et al., 2020).

A passenger trip involves access time from the origin to the boarding stop, waiting time, in-vehicle travel time, potential transfer time, and access time from the alighting stop to the destination. Agencies typically analyze the overall passenger travel times from the initial

access to the system until the passenger exits the system (Danaher et al., 2020).

Similar to on-time performance in operational metrics, there are schedule-based metrics for passengers which can be traced all the way back to Buneman (1984). They calculated the percentage of passenger trips being delivered on time based on the overall passenger delays in minutes. Similarly, Zhao et al. (2013) and Halvorsen et al. (2019) calculated the excessive overall travel times, excess wait times, and excess in-vehicle times compared to planned travel times. In addition, Hendren et al. (2015) created a dashboard to provide personalized travel time reliability summaries, which are more closely aligned with actual passenger experiences.

Another popular approach is to use variation-based measures, since passengers typically plan their trips based on worst-case scenarios (Danaher et al., 2020). These measures are calculated based on historical distributions of travel times, such as the average travel time, median travel time, and various pre-defined travel time percentiles (Hendren et al., 2015). Reliability buffer time is a common measure to calculate the extra time a passenger has to plan in addition to the usual travel time, to arrive at the destination on time with a certain level of tolerance. It typically compares the 90th to 95th percentile travel time to the mean and median travel times, which would represent the extra time passengers have to plan to have at most 2 disruptions in a month (Uniman et al., 2010).

To generalize these passenger-based measures for multimodal trips, Dixit et al. (2019) integrated multiple modes into the buffer time measurements and compared the variability of different components of buffer times, wait times, in-vehicle times, and transfer times. Bagherian et al. (2016) also analyzed the temporal variability of passenger travel times at the stop level, line level, and network level.

For agencies that require passengers to tap-in and tap-out, passenger entry and exit information is easier to obtain, which is why they are heavily present in the literature. Agencies without tap-out data typically generate a representative ridership model for an average week-day using the fare transaction data, which is then matched to scheduled and delivered trips (e.g. (Halvorsen et al., 2019)).

Some researchers have also tried to ignore the observed demand by incorporating public transport travel time reliability into accessibility measures, given an origin and a time budget. They showed a reduction in accessibility measures using the delivered service compared to the planned service due to delays and missed transfers (Liu et al., 2023; Bills and Carrel, 2021). Some have also found studies using planned travel times underestimate the inequity in public transport accessibility, for example to healthcare, compared to using the actual arrival and departure times (Lee and Kim, 2023). Hence, travel time reliability is also an important factor when considering social equity impacts of public transport.

Public transport travel time reliability is also a significant factor for passenger mode choices. People tend to orient their lives around a certain mode, e.g. by public transport, and the decision is relatively stable (Carrel and Walker, 2017). In case of unreliability, people can budget extra travel times. However, in the long term, passengers affected by unreliability might switch to other modes, and people who stopped using public transport services are less likely to return unless major improvements are made (Carrel and Walker, 2017). In addition, some captive public transport users may have fewer choices and end up not travelling if the services are not reliable. Since the aforementioned studies are all based on observed demands and the long-term reliability impact on passenger mode choice, it is still important to consider the potential public transport travel time reliability between origins and destinations with less observed demand, since unreliability could be a contributing factor for lower demand.

Hence, in this paper, we consider the travel time reliability to all possible destinations in the region equally, regardless of time budgets and regardless of actual demand. This can help provide more detailed information to planners on unreliable origin-destination pairs, as well as when and where the reliability issues occur. We also attempt to discover potential spatial and temporal patterns related to any reliability variations that passengers would experience in different areas of the city.

8.3 Methodology and Data

As noted in the previous sections, we examine travel time reliability across all possible origin-destination pairs, independent of time budgets or travel demand. These indices can reflect the possibility that people rely on public transport for less common trips, potentially due to less fortunate economic situations. We also explore the spatiotemporal variations in overall public transport travel time reliability, as these patterns can provide insight into the potential impacts on different population segments.

Figure 8.1 shows a summary of our research framework. In general, we use a hexagonal grid system to evenly sample 691 origin-destination zones across the service area. Each hexagon's centroid is linked to the nearest publicly accessible street block for routing, and public transport trips between all zones are generated using planned GTFS (General Transit Feed Specification) data. Actual travel times are assessed using GTFS Real Time data from January to March 2024. Then, using the actual travel times, we calculate the reliability measures, on-time performance and buffer time. Finally, we evaluate the spatial and temporal distribution of these measures across the service area.

More specifically, in this paper, we use the data from Société de transport de Montréal (STM)

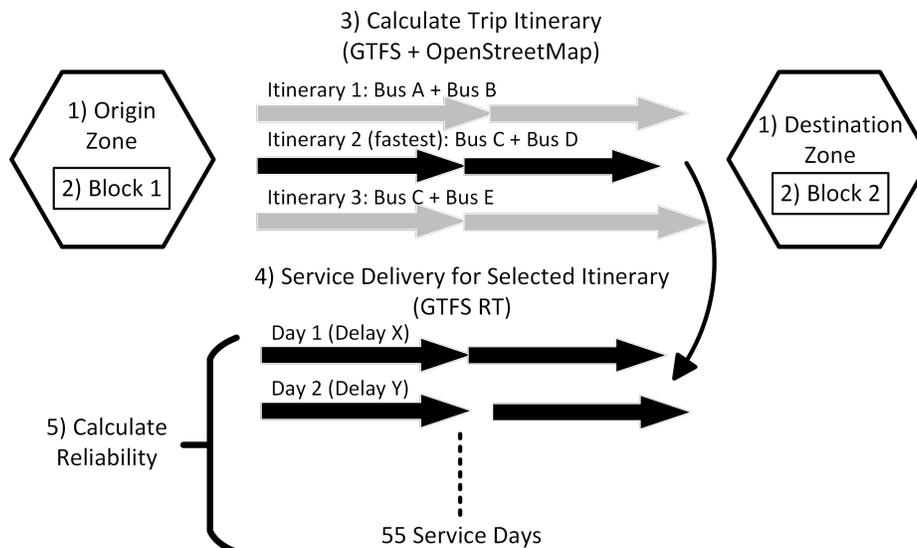


Figure 8.1 Research Framework

as an example to demonstrate our ideas. The public transport services include a bus system involving 225 bus lines, 1849 buses, 8673 bus stops, operating more than 17,000 planned bus trips on average weekdays, and a metro system involving 4 lines, 69 kilometres of tracks, 68 stations, and 999 train cars. There are other operators that run commuter rails and suburban buses mostly to destinations outside the island of Montréal. Therefore, we will not include them as part of our study. Future research could include these services and focus more on the metropolitan area as a whole.

Since we are considering travel time reliability across the region that is not necessarily covered by observed demand, we decide to sample origin and destination points across the region due to the size of the network and assuming the reliability would be similar for two consecutive bus stops on the same line. In this paper, we choose to sample the entire service area using hexagon grids (step 1 in Figure 8.1), since the distances between neighbouring centres of each cell are the same. As a demonstration for this paper, we used hexagon cells with 500 metres edges, or roughly 800 metres between neighbouring centres. The distances correspond roughly to the distance between streets that bus routes operate, and roughly a 10-minute walking distance within each zone. We chose these distances so that there is equal spatial representation of the entire service area, and there is enough detail to understand the variation of travel time reliability without spending too much computing power. Future research could nonetheless test more detailed sampling resolutions and make weighted comparisons based on populations or the number of points of interest.

Since the centroids of the hexagons are not necessarily located in publicly accessible areas, we sample a street block closest to the centroid as the origin and destination point for the routing engine to calculate the routes based on planned services (step 2 in Figure 8.1). In total, there are 691 hexagon cells in the service area, thus 691 x 691 origin-destination pairs. However, around 30 hexagon cells lie entirely within major industrial complexes or large infrastructure zones (e.g. airports or railway yards), which are not accessible to the public. Hence, these zones are removed from the analyses.

Then, using these origins and destinations, we use the GTFS feed to generate all possible trip plans between a given origin-destination pair for weekdays (step 3 in Figure 8.1). Next, we assume a potential passenger selects the route with the fastest planned travel time as a reference trip to evaluate the reliability for the given origin-destination pair at a given hour. Since passengers generally dislike long walks and waiting times, we also include a multiplier for walk times (1.7) and wait times (1.8) to discourage excessively long walks or transfers based on the values estimated by Wardman (2004). We do not consider the cost as a factor here since the agency operates on a flat-fare basis.

For each selected trip plan, we calculate the actual travel times for the reference trips during the analysis period (from January to March 2024 as an example) using archived GTFS Real Time feeds (step 4 in Figure 8.1). As indicated in previous studies, large portions of passengers typically arrive 1 to 2 minutes before the scheduled departure time, even for frequent routes with headways as low as 5 minutes (Luethi et al., 2007). Since we rarely have frequent routes with less than 5-minute headways, for simplicity, we assume the passengers arrive 3 minutes in advance at the departure stop and strictly follow the planned route for the trip. Future researchers could test out various trip planning parameters and calculate the sensitivity of the results.

To compare the delivered service and the planned service, we consider the reliability metrics similar to on-time performance and the buffer time metrics (step 5 in Figure 8.1). The on-time performance is calculated as the percentage of potential trips with arrival time within a given threshold, such as no more than 3 or 5 minutes later than the planned time, as per a survey of various North American public transport agencies (National Academies of Sciences, Engineering, and Medicine and others, 2013). We will use 3 minutes as our base case for analysis, which is similar to the agency's standard, and test out a few additional thresholds as a discussion.

Since passengers plan according to the worst-case scenarios to ensure arriving at their destinations on time, we also include the continuous buffer-time metric used by previous researchers, e.g. Dixit et al. (2019). It is defined as the difference between a given high percentile (typi-

cally 90th to 95th percentile) and the median, which would show the difference between the worst-case scenarios and the typical experience for the journey.

Given the various passenger tolerances for delays, the perceived reliability may be different. Hence, we will compare and discuss the impacts of a few thresholds on the reliability results in our discussion to demonstrate the impacts, since the results depend on the chosen threshold. Since the agency's standard for late departure is more than 3 minutes after the scheduled time, and the agency aims to have 80% of the services delivered on time, we will use 3 minutes for our on-time performance and 80th percentile for buffer time metrics as a base case for analyses to better align with the agency's standard.

Next, following the thresholds similar to the agency's standards, we compare the likelihood of the route being on-time for one origin, Polytechnique Montréal as an example, to all other destination zones at various reliability metric thresholds and departure times. Similarly, we compare the overall number of destinations with reliable routes across the day.

Finally, we compare the number of destinations with reliable routes for all zones with various departure times by aggregating the results for individual zones to uncover potential spatial-temporal variations. Using the number of destinations for each departing zone, we use the Gini Index to measure the distribution inequality within the entire service area. The Gini index is a measure of statistical dispersion commonly used to quantify income or wealth inequality within a population, where 0 represents perfect equality and 1 indicates maximum inequality. It is widely applied in economics and social sciences to assess disparities across different groups or countries (Ceriani and Verme, 2012).

We also intersect the hexagon cells with the census tracts from 2021 to obtain basic demographics information, such as the population density, assuming the population within each census tract is evenly distributed. Since the reliability results for neighbouring zones may have spatial correlations, e.g. similar route choices resulting in similar reliability indices and neighbouring zones have similar demographics, we then fit a simple spatial lag model to discover the impacts on various population segments while accounting for any potential spatial correlations. The spatial lag model assumes that the outcome in a region depends not only on local explanatory variables (X), but also on the outcomes in neighbouring areas (Wy). It accounts for spatial dependence by including a spatially lagged dependent variable as a predictor, capturing the influence of neighbouring units on a given observation (Anselin, 1988). The model is formulated as:

$$y = \rho Wy + X\beta + \epsilon$$

where y is the dependent variable, ρ is the spatial autoregressive coefficient measuring the

strength and direction of spatial dependence, W is the spatial weights matrix identifying the neighbours, X is a matrix of independent or explanatory variables, β is the vector of regression coefficients for the independent variables, and ϵ is a vector of the error terms.

8.4 Case Study: Montréal, Canada

In this section, we show the results from our case study using the data from Montréal as a demonstration. First, we present the descriptive summaries and potential variations using various reliability metric thresholds. Then, using the thresholds close to the ones used by the agency, we discuss the reliability variations from one origin, Polytechnique Montréal, to all other destinations. Then, we examine the number of destinations with reliable routes for each origin zone. Finally, we examine whether the number of destinations with reliable routes is evenly distributed for various original zones and population segments.

8.4.1 Impact of Various Thresholds on Reliability Metrics

In this section, we quickly explore the impact of various thresholds on our chosen reliability metrics since passengers may have different tolerances or expectations. For example, passengers typically have a stricter reliability requirement for their trips to work or school than trips to shopping or leisure activities, given the fixed start time at work or school (Carrel and Walker, 2017). Hence, to provide an example here, we calculate the reliability metrics given various thresholds to examine the impacts using Polytechnique Montréal as the origin.

For on-time performance, we chose 1 minute, 3 minutes, and 5 minutes as the thresholds. In other words, a trip is considered on-time if the passenger arrives no more than 1 minute, 3 minutes, or 5 minutes later than the planned arrival time. We chose the 3-minute threshold since it corresponds to the agency's policy for considering a trip being on-time, the 5-minute threshold since it relates to the threshold of other agencies in North America (Danaher et al., 2020), and a 1-minute threshold for a potential passenger who is really strict about being on-time. As a reminder, the agency's target is to have better than 80% of the trips departing no more than 3 minutes late at timepoints, and the actual on-time performance at timepoints during the study period is 79.6%.

The top half of figure 8.2 shows the resulting on-time performance using these three thresholds for routes departing from Polytechnique to other destination zones departing at 17:00. In general, as expected, we can observe that, more and more destinations are considered to have reliable routes when we relax the thresholds.

More specifically, when the threshold is set to 1 minute, the on-time performance tends to

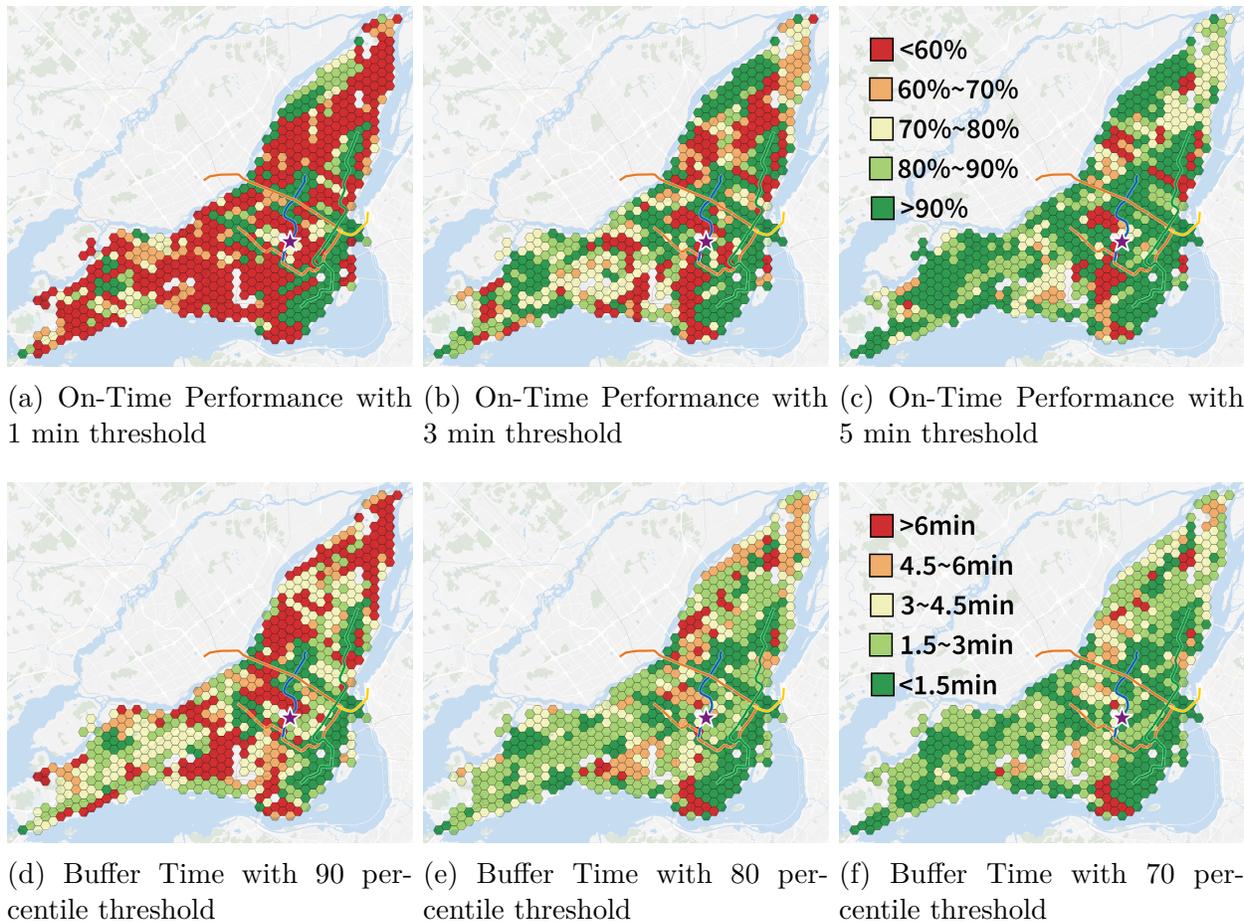


Figure 8.2 Reliability for routes from Polytechnique Montréal (purple star) to other destinations departing at 17:00 with various thresholds, by indicator (Jan-Mar 2024 data)

become a bit extreme. Most of the destinations have routes with either very good or very bad on-time performance. Destinations near the metro (marked by green, orange, yellow, and blue lines) tend to be more reliable than the other destinations, which is to be expected since the metro has no congestion issues causing significant delays. Despite having a few service interruptions during the study period, the percentage of the trips delivered on time is still very high. A few neighbourhoods on the northern edge or the western side of the island also have routes with good arrival on-time performance, which could be attributed to the higher service frequency during peak hours. Passengers may be able to catch an earlier bus than planned, or use a long highway express route from the metro, which could arrive early when leaving the highways.

On the other hand, when the threshold is set to 5 minutes, most of the destinations would have routes with acceptable on-time performance, except for a few neighbourhoods with

consistent congestion issues or long-term construction projects. Notably, the areas directly southwest and northwest of the university have significant congestion issues due to the street being a narrow commercial street with higher trip attractions, while also being a bottleneck due to limited street connectivity caused by highways and railways. Despite the presence of peak-hour bus lanes, compliance is relatively low, further contributing to the potential delays. Areas close to the northeastern terminus of the metro blue line also have significant delays, which could be attributed to long-term construction detours and stop cancellations due to the metro extension projects.

As for the buffer time metrics, we chose 90%, 80%, and 70% as the thresholds. Once again, since the agency's target is to have 80% of the buses departing no later than 3 minutes after the scheduled time at timepoints, we chose the 80% threshold as a reference, which would also correspond to having 2 late trips per week, assuming one round-trip per weekday, i.e. 10 trips per week. Similarly, we chose the 70% and 90% thresholds to represent potential passengers who are more forgiving of unreliability and passengers who are more strict for travel time reliability, which is more closely aligned with the existing literature. These two thresholds correspond to one or three late arrivals per week on average, assuming one round-trip during weekdays.

Similar to the on-time performance metric, when the potential passenger becomes more forgiving, i.e. a lower percentile for the buffer time, less buffer time is needed to reach a destination. When the potential passenger is really strict on the buffer time threshold, more buffer times are needed. The routes to destinations along the metro still have low buffer time, once again showing the reliability of the metro system. However, unlike the on-time performance, there is not a significant reduction of buffer times after lowering the threshold from 80% to 70%, suggesting the spread of the empirical travel time observations is low between the two thresholds.

Another observation is that the two metrics may not show the same reliability issues. Using the northern edge of the island as an example, the on-time performance metric shows that most of the potential passengers would arrive less than 3 minutes later than the scheduled time. However, the buffer time metric shows a large buffer time (around 4 to 6 minutes). Again, the services are frequent during peak hours. A small travel time variation in the first leg of the trip may cause passengers to catch an earlier bus, thus saving a few minutes of transfer time and the overall travel time. Thus, passengers may change their expectations based on their past journey experiences. If they are sometimes able to catch an earlier bus than expected to their destination, would they change their expectations and behaviours to consider the earlier bus as the one they are expected to catch? What would they consider as

“reliable”? If the delivered services can consistently get the passengers to their destination early or late with a low travel time variation, regular passengers may have a different expectation towards their travel time reliability compared to a passenger without route knowledge (Carrel and Walker, 2017). Future research could potentially consider incorporating these considerations while analyzing the results.

8.4.2 Spatial-Temporal Changes of Reliability

In this section, we look at the potential travel time reliabilities departing from Polytechnique Montréal to other destination zones in the region. As mentioned in the previous section, the passengers’ tolerances towards unreliability (different metric thresholds) may affect their experience. Pending more information regarding the distribution of passenger preferences, in this section as a demonstration of the analysis, our thresholds mainly adhere to thresholds similar to the agency’s standards, i.e. having 80% of the trips depart no more than 3 minutes later than scheduled.

Figure 8.3 shows the potential on-time performance and buffer times measures from Polytechnique Montréal to other destinations departing at 7:00, 12:00, and 17:00. Each hexagon represents a potential destination zone, and the colour of each zone reflects the route reliability to the given zone at the given departure time.

From the figures, we can observe that most of the destinations have routes that can be considered reliable, given the agency’s standards in general. The buffer times measures meet the agency’s standards more often compared to the on-time performance, i.e. more late arrivals compared to the schedule, but the variation of travel times is generally small. Again, we can observe some consistent delays for routes to certain destinations.

However, the spatial-temporal distribution of reliability to different destinations is not evenly distributed. Similar to the previous section, destinations near the metro tend to have consistently on-time service with low travel time variations. There are a few destinations with consistently late services and larger travel time variations, which is not surprising given the local knowledge that these areas have major construction and detours. The reliability patterns to different destinations also change throughout the day. For example, the on-time performance to destinations directly southwest of Polytechnique is generally above 80% at 07:00, but it decreases to less than 60% later in the day. Similarly, the buffer times are less than 3 minutes, but they increase to 4 or 5 minutes later in the day. As mentioned in the previous subsection, this area has limited road connectivity due to highway and railway crossings, and the route goes through a dense shopping area, which further increases the congestion impacts in this area. Hence, the agency might want to focus more on the given

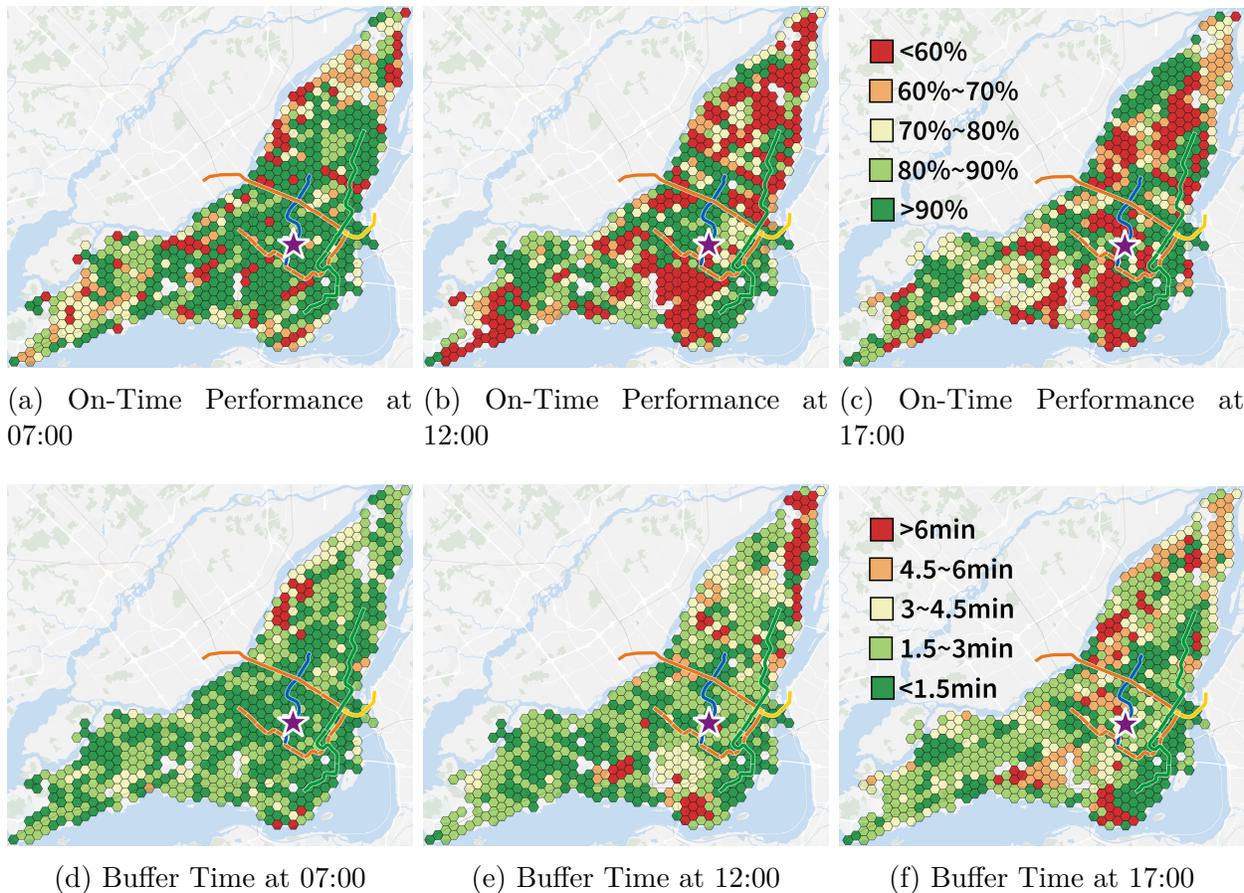


Figure 8.3 Reliability for the overall route travel times from Polytechnique (purple star) to other destinations for three departure times, by indicator (Jan-Mar 2024 data)

local context and specific time periods when revising the schedule. For example, given that most of the shops don't open in the early morning, the agency might want to modify the schedule to better reflect the travel time needed or adjust the bus lane hours to help reduce bus delays.

To analyze these reliability metrics in more detail, we could also look at the reliability of different legs of a passenger journey, which includes a series of wait times and in-vehicle times. Here, we include a high-level summary of the overall wait times and in-vehicle times, since some destinations may have a direct connection and some routes may involve multiple transfers.

Figure 8.4 shows the reliability of different legs or components of passenger journeys in terms of the overall wait times and in-vehicle times, using the same reliability metrics and thresholds. Again, more destinations have in-vehicle times or wait times that are considered

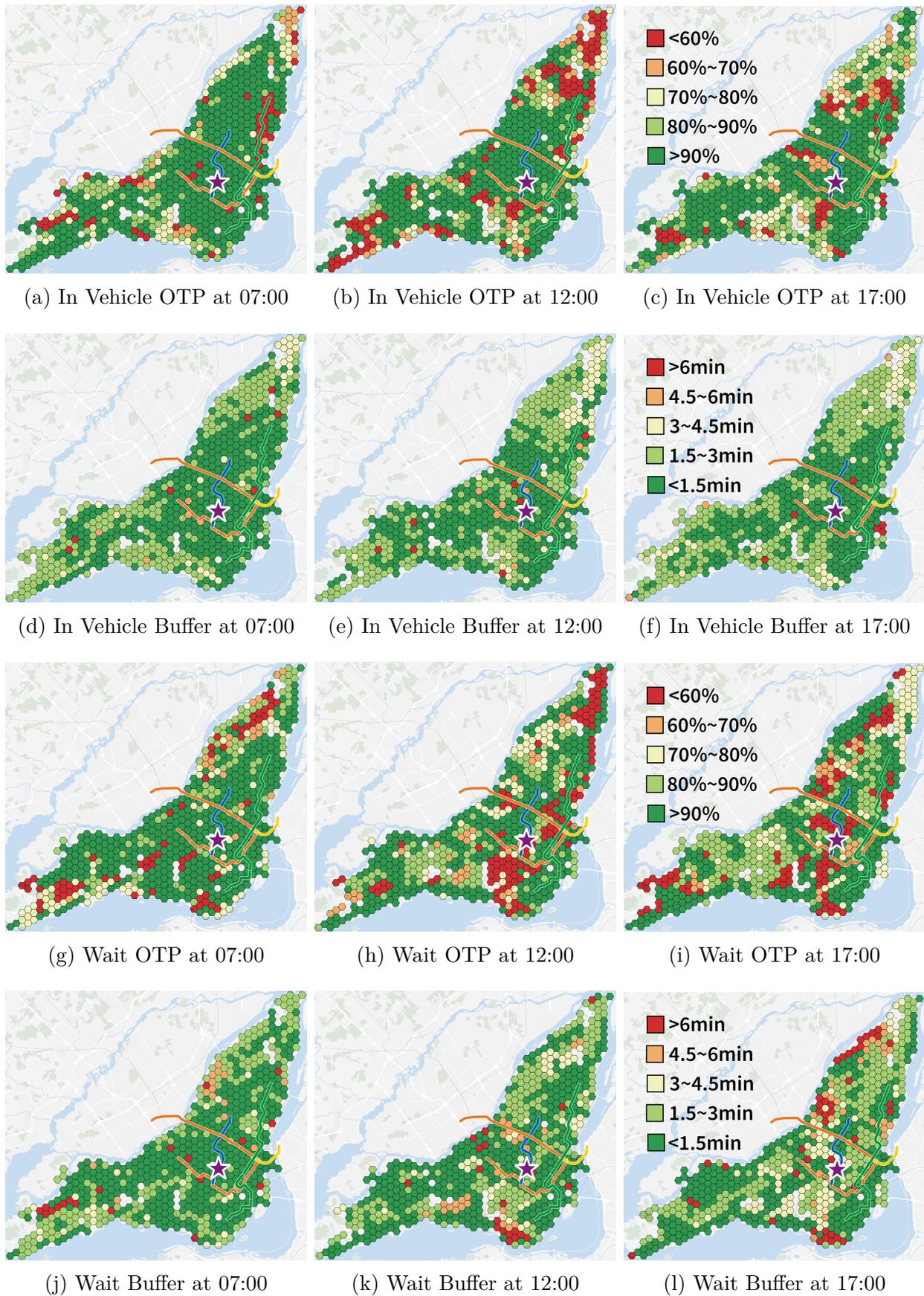


Figure 8.4 Route reliability breakdown from Polytechnique Montréal to other destinations at different times of the day, by indicator (Jan-Mar 2024 data)

reliable if we use the buffer time metrics, demonstrating consistent delays.

For the in-vehicle times, the on-time performances are more extreme, where the performance is either really good or really bad. Again, the spatial patterns change during the day. For example, there are consistent delays at all three departure times towards the northern edge of the island. The destinations south of the university have longer in-vehicle times around midday and the afternoon peak.

The buffer time metrics for in-vehicle times, on the other hand, are generally within the chosen threshold of 3 minutes, as most destinations are marked as green, i.e. meet the threshold similar to the agency's standard. However, the central areas of the island are more reliable. Since the university is located on a metro line, journeys to destinations in the central region typically start at the metro station, followed by a short bus ride. As expected, the buffer times for these destinations are low, with many routes having buffer times less than 1.5 minutes. For longer journeys toward destinations on the edge of the island, buffer times tend to be longer, since the journeys are more reliant on the bus network, which may be affected by more varied congestion levels. Hence, regular passengers may adopt different expectations for short or longer trips. The perceived reliability can be left for future research.

For the on-time performance of overall wait times, the spatial pattern becomes more varied compared to the in-vehicle times. There are more areas with lower performing wait times, and there are more areas with performances slightly less than the target of 80%. Similar to the in-vehicle time observations, there are consistent delays at all three departure times towards the northern edge of the island. The destinations south of the university have longer wait times around midday and the afternoon peak.

The buffer time metrics for wait times generally align better with the chosen threshold, similar to the in-vehicle times. The routes with longer wait times typically align with the problematic spots in the wait time on-time performance. However, unlike the buffer times for the in-vehicle times, the wait times for many areas of the island become less reliable throughout the day, since the bus reliability tends to get more varied in the afternoons. Hence, the agency might want to better adjust the schedule to reduce passenger wait times. Given that the agency manages 225 bus routes, synchronizing transfers for all of the routes may not be possible. The agency could possibly suggest a minimum transfer time at transfer points to better account for potential transfer needs, given the travel time and reliability variations observed.

Figure 8.5 shows the number of destinations with routes that are considered reliable by hour, i.e. the metrics are within the agency's standards in terms of the overall travel times for the entire journey, in vehicle times, and wait times. Again, for on-time performance, the target

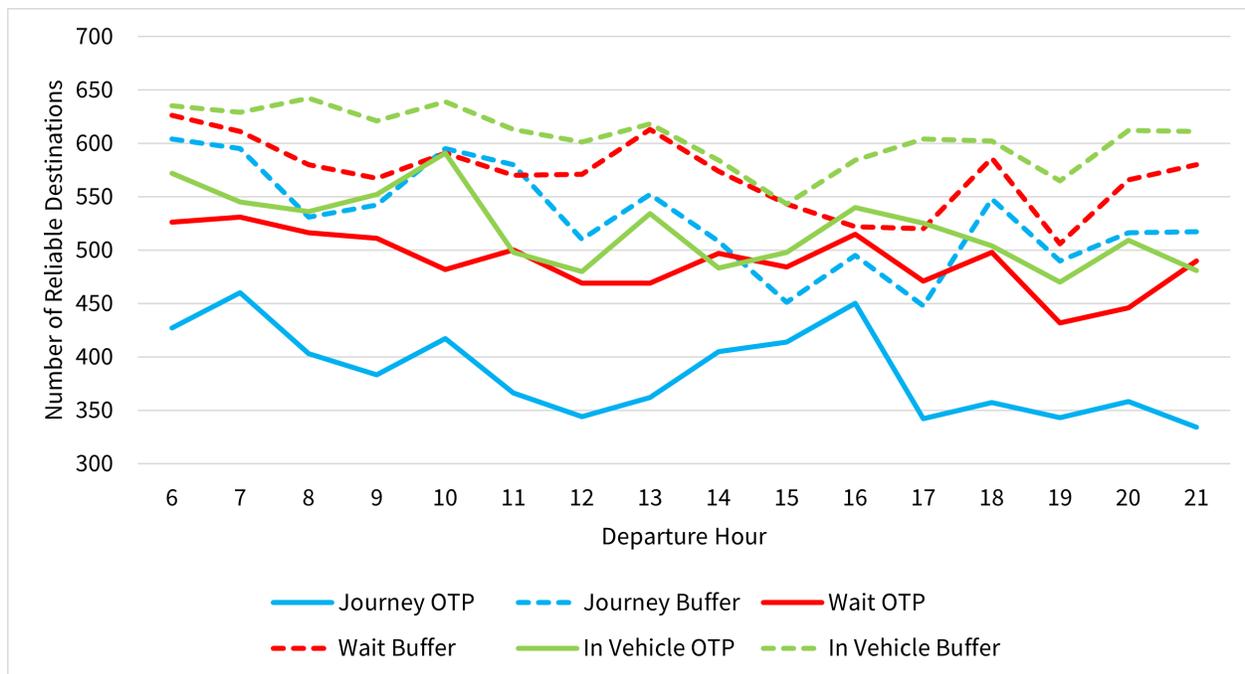


Figure 8.5 Number of Destinations with Routes Considered Reliable Departing from Polytechnique Montréal, by indicator (Jan-Mar 2024 data)

is 80% given the 3-minute threshold, and for buffer time, the target is 3 minutes given the 80th percentile threshold.

In general, the overall travel times do not perform as well as the wait times nor the in-vehicle times. This makes sense, since we used the same standards for all three categories and the overall travel times can be affected by both wait and in-vehicle times. For the temporal patterns, we can observe less reliable routes during the two peak hours and at noon for the lunch break. The afternoon peak has the worst performance, which is related to more varied congestion and passenger demand patterns.

For the in-vehicle times, it is considered the most reliable category compared to the wait times and the overall travel times. The on-time performance metrics also show small decreases in two peak hours and at noon, similar to the overall travel time observations, but there is an additional decrease around 2 or 3 in the afternoon. The buffer times generally remain stable from the morning peak until noon, then also decrease at around 2 or 3 pm. Given the local context that the university is surrounded by many other elementary, middle, and high schools, it is not surprising that the in-vehicle times, especially bus travel times, degrade around the end of the school day, where there is a lot more pedestrian traffic as well as vehicular traffic around the university and schools.

The wait times perform slightly worse than the in-vehicle times and remain generally more stable all day than the overall travel times and in-vehicle times, suggesting a generally consistent waiting experience all day. The wait times are slightly more likely to be longer around noon and after the pm peak, but there are slightly more variations (i.e. buffer time) in the morning and the afternoon peaks.

In general, given our chosen thresholds, i.e. the times are less than 3 minutes longer than planned 80% of the time, the buffer times for all three categories behave better than the on-time performance at nearly all of the sampled departure times. Again, this suggests consistent delays, which regular passengers with route knowledge may have been used to compared to passengers making a spontaneous trip to a new area. Hence, we once again want to highlight the future need to examine the travel preferences and delay tolerances for these two types of passengers.

Since most of the passenger-based reliability research focuses on the existing demand with highly aggregated temporal and spatial scales, we want to demonstrate the extra details that planners could observe given our relatively smaller analysis scale. However, as previous research also pointed out, these extra details may come with a much longer computing time requirement. In addition, as we have shown that wait times perform generally worse than the in-vehicle times, we want to emphasize the need to incorporate wait time reliability in public transport service planning. Since larger agencies typically do not operate on a hub-and-spoke model due to the variation in coverage, headways, and travel times, sometimes it is impossible to synchronize all transfers. Hence, more strategies are needed for the agency to also improve the wait time and transfer time reliabilities.

8.4.3 Number of Potential Destinations with Reliable Routes for All Origins

Figure 8.6 shows the number of destinations with routes considered reliable given our on-time performance and buffer time metrics, for each origin zone. Again, we use the thresholds similar to the agency's standard. For buffer times at the 80th percentile, if the variation is less than 3 minutes, the route is considered reliable. For on-time performance, if 80% of the trips arrived less than 3 minutes late, the route is considered reliable.

From the figure, we can observe that most origins have reliable routes to 300 to 500 destinations (50% to 70% of all possible zones) during the observation period. In general, we can observe some spatial and temporal changes of the potential destinations with reliable routes. In addition, given our thresholds, the buffer time metric would consider more routes as reliable. Thus, the potential issues for public transport riders are more likely to be consistent delays, rather than large travel time variations. The agency might want to adjust the

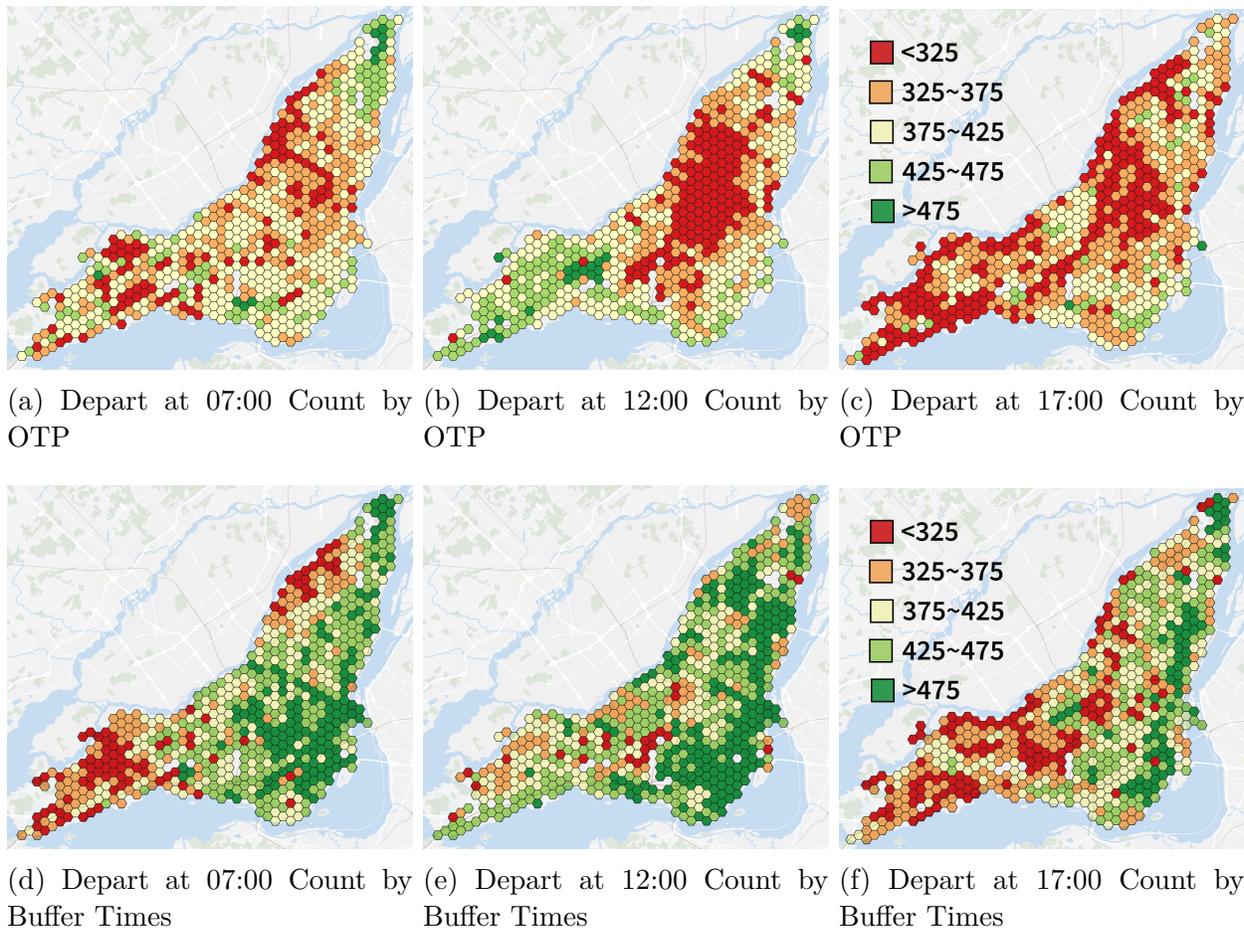


Figure 8.6 Number of Reliable Destinations for each Zone at Different Times, by indicator (Jan-Mar 2024 data)

schedule or implement various public transport priority measures to improve performance, such as dedicated lanes and signal priority. We have included an example for the bus service delivery performance observed at noon in Figure 8.7 to help readers understand our operational contexts.

There are areas with consistently good performance in both metrics, for example the downtown area (central eastern edge of the island). This is to be expected as the bus network is better configured to feed passengers into the metro network, which in turn focuses on getting people to and from downtown. Again, the metro services are generally very reliable, with three out of four metro lines that focus on the downtown passenger demands. Although the bus routes that run through the downtown area are slower and less reliable midday (Figure 8.7), there are more rush-hour bus lanes with longer service hours to help with the long express buses to various suburbs. Hence, it is not a surprise that this area performs generally

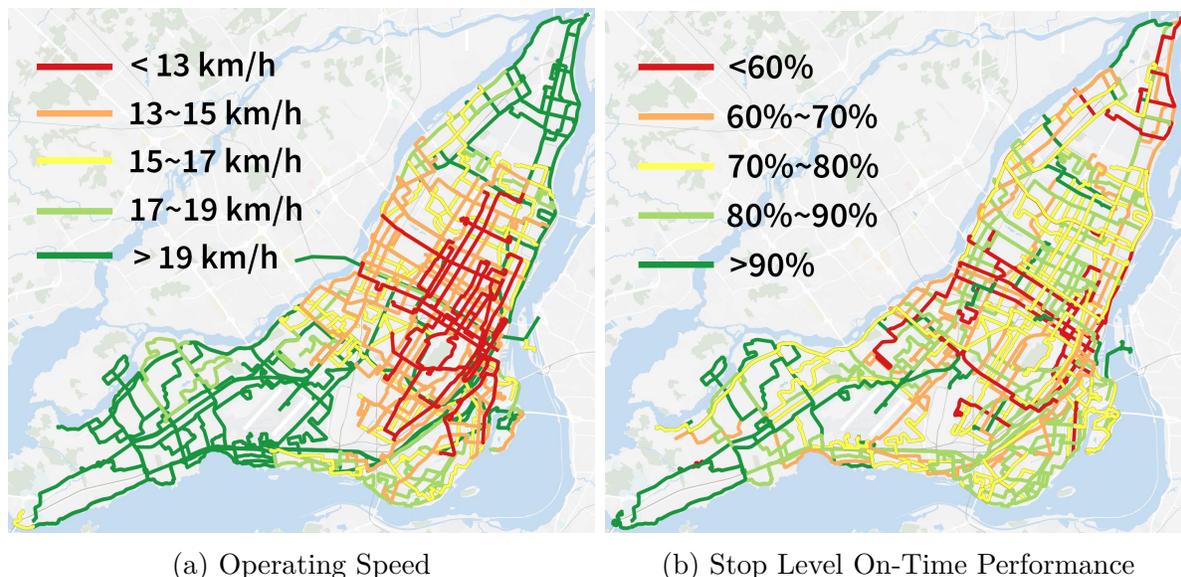


Figure 8.7 Bus Service Performance by Route for 12:00 Departures (Jan-Mar 2024 data)

well in terms of having more destinations with reliable routes.

There are some origin zones where the potential destinations with reliable routes change throughout the day for both metrics, such as the western edge of the island. In the morning and the afternoon peak hours, the reliability of the routes tends to decrease for both metrics. However, for departures around noon, the reliability of the routes is relatively better. Given the local knowledge that the west island has more suburban developments, i.e. higher single-family housing and car-oriented street designs. Some local and express bus services are forced to take detours and use the highway system due to the lack of local street connections. Hence, it is not surprising that the public transport service there is more affected by congestion during peak hours and services off-peak generally perform well (e.g. Figure 8.7). In addition, since the network is more oriented for people going to and from the metro, which sends people to downtown, it may not necessarily be easy for people travelling in the reverse commute direction or from these suburbs to other suburban areas during rush hours. Hence, these suburban areas tend to have lower public transport modal share compared to the rest of the island, and the modal share is even lower for suburb-to-suburb demands (Autorité régionale de transport métropolitain, 2023). The difficulty of travelling from suburb to suburb, which typically requires two or more transfers, has also been demonstrated by other researchers in other metropolitan areas such as Palm and Farber (2024). Future researchers could attempt to determine to what extent the relatively low travel time reliability contributed to the low modal share, since the two could form a vicious cycle where low demand reduces the urgency

to improve services in the area, causing further degradation in reliability, which forces people to shift to other modes, further decreasing the demand.

Similarly, there are origins where the potential destinations with reliable routes are consistently high or consistently low. For example, the near northeastern part of the island has fewer than 375 potential destinations with routes that are consistently reliable in terms of on-time performance. These areas have higher population density and ridership, which would mean late arrivals would affect more passengers in reality. Access in these areas typically requires more transfers as the northeast-southwest connections are not straightforward due to the presence of large infrastructures, industries, and parks. Most bus routes in the area are short and also tend to terminate at the nearby metro station with no through running. Thus, many transfers are required for the passengers in this area. Furthermore, passengers originating from the west or the northeastern side of the island have access to long highway express buses which feed into the metro system, allowing them to bypass all the slow and less reliable routes in this area (Figure 8.7) when travelling to other destinations unless this area is their destination. However, for passengers originating in this area, the trips to all destinations must travel through this area on local streets with slow bus routes, low mid-day service frequency, and dense traffic signals, further degrading the travel time and reliability (Figure 8.7). Better bus route configurations, aggressive bus priority measures, and potentially new metro expansions in the area might be able to help improve the service reliability in the area.

As for the buffer times in the same northeastern area, the number of potential destinations with reliable routes is generally high (>425) except during the afternoon peak, suggesting consistent delays with less travel time variations. Again, there are many construction-related detours during the study period. These detours could consistently add travel times on one or both route directions, causing significant delays but not necessarily a large travel time variation off-peak. Again, potential passengers with local knowledge would know about the detours and may plan accordingly. For passengers who are making a spontaneous trip to this area without local knowledge, their experiences would be very different. Since consistent delays negatively impact everyone, we want to emphasize the need for agencies to quickly update the expected travel times for passengers, as well as for the bus operators to be more flexible with service adjustments during a long-term detour. It is also important for future researchers to consider the possible impacts on passenger mode choices and equity impacts given various passenger expectations and behaviours. How many passengers have good route knowledge? How would passengers' route knowledge help them make decisions or affect their expectations for service delivery?

8.4.4 Potential Equity Implications for the Number of Destinations with Reliable Routes

To examine the potential equity impact on different population segments, we intersected each hexagon with the 2021 census tracts and calculated the demographic attributes assuming the population within each tract is distributed evenly. Then, we fitted a simple spatial lag model as well as calculated the Gini index to examine the evenness of distributions.

The Gini index for the number of destinations a zone has reliable access to is roughly 0.09, 0.07, and 0.10 for the buffer time metric (Table 8.2) for the three chosen departure times at 7:00, 12:00, and 17:00. As the Gini Index lies within 0 and 1, where 0 represents perfect equality and 1 indicates maximum inequality, these values suggest that the overall number of destinations with reliable routes for each zone is more evenly distributed during the midday period, with the peak hours being slightly less evenly distributed. This is also shown by the earlier results, where more zones would have destinations with less varied travel times, i.e. buffer times less than 3 minutes.

On the other hand, the on-time performance metric shows a different evenness pattern (Table 8.2). The Gini index for the number of destinations with reliable routes for each zone is roughly 0.7, 0.9, and 0.9 for our three chosen departure times at 7:00, 12:00, and 17:00. Here, the morning peak is the more evenly distributed time period. As seen from earlier sections, most buses operate on time, probably because most passengers are regulars going to work or school with less varied congestion conditions. Around noon and the afternoon peaks, the delays become slightly more unevenly distributed overall.

The two measures do not yield the same evenness of distribution. The more evenly distributed time periods are generally the times when more zones have reliable access to various destinations. Hence, agencies should continue to strive to understand the specific reasons that cause delays or large travel time variations, which would help improve their performance and even out the reliability distribution. In addition, passengers who are used to travelling at a specific time may have a different expectation given their past experiences, also pointed out by other researchers (Carrel and Walker, 2017). As shown here, each departure time is associated with its own pros and cons. People who have gotten used to travelling mostly on time during morning peaks may have completely different expectations compared to people who have gotten used to travelling during the congested afternoon peak. Thus, future research could potentially look into these various expectations and improve our understanding of perceived reliability by passengers.

To see if various population segments have a similar number of destinations with reliable

routes, we used the simple spatial regression model, as well as accounting for potential spatial correlations. The inputs include some basic demographic and socioeconomic information as independent variables (x), the population (by 1000 people per square kilometre), percent population with bachelor's degree, percent minorities, median age, and median annual income (by 1000 dollars). The dependent variables (y) are the number of potential destinations with reliable routes in terms of buffer time or on-time performance. A summary of the dependent and independent variables is in Table 8.1.

Table 8.1 Summary of independent and dependent variables

	Min	Max	Average	Standard Deviation
Number of Reliable Routes - OTP	272	476	386.20	65.36
Number of Reliable Routes - Buffer	319	567	453.71	80.25
Population Density (1,000 people)	0.16	19.95	4.32	3.58
Percent Bachelor Degree	0.09	0.70	0.35	0.14
Percent Minority	0.06	0.85	0.36	0.15
Median Age	28.11	66.00	41.87	4.74
Median Income (\$1,000)	26.83	84.00	41.16	7.34

The model results (Table 8.2) show that the significance and the sign of each variable vary across the day, once again showing that the potential reliability has different impacts at different times.

Table 8.2 Spatial regression model results for the number of destinations with reliable routes (Jan-Mar 2024 data)

Departure	Buffer Time			On Time Performance		
	07:00	12:00	17:00	07:00	12:00	17:00
Intercept	183.50 ^{***}	206.69 ^{***}	167.64 ^{***}	215.66 ^{***}	43.78 [*]	171.20 ^{***}
Population Density	1.53 ^{***}	2.63 ^{***}	1.87 ^{***}	-1.02 [*]	-0.61	-0.36
Percent Bachelor Degree	0.30 [*]	0.05	0.08	-0.16	0.00	0.17
Percent Minority	-0.21	-0.50 ^{***}	-0.40 ^{***}	-0.19	0.01	-0.12
Median Age	-1.27 ^{***}	-0.07	-0.18	-0.68 [*]	0.64 [*]	-0.32
Median Income	-0.44	-0.28	-0.02 ^{**}	0.07	0.37	-0.79 [*]
Gini Index	0.0891	0.0700	0.0969	0.0696	0.0920	0.0931

^{***} $p < 0.01$, ^{**} $p < 0.05$, ^{*} $p < 0.1$

For departures at 7:00, the intercept and three other variables, population density, percent bachelor's degree, and median age, are significant ($p < 0.1$) in the buffer time model. The

coefficients for both population density and percent bachelor's degree are positive, meaning zones with higher population density and a more educated population tend to have slightly more reliable routes to other destinations. The older population are related to fewer destinations with reliable routes, which makes sense given that there is a small correlation where the older population tend to live in the suburbs with lower population density. Around noon, population density still remains positive, but only one other variable is significant, the percentage of minority populations. Areas with higher proportions of minority population relate to a slight decrease in the number of destinations with reliable routes. In the evening peak, the population density and percentage of minority populations are still significant. Similar to earlier departures, higher population density correlates to a small increase, and a higher percentage of minority population correlates to a small reduction in the number of destinations with reliable routes. In addition, the median income is slightly significant with a negative coefficient, suggesting that the higher-income population are negatively affected during the afternoon peak. People with higher incomes tend to live in areas with limited or less straightforward road access, which could restrict direct public transport routing and increase potential unreliability.

As for the on-time performance models, for 7:00 departures, only population density and median age are slightly significant, which is not surprising, as the number of destinations with reliable routes is the most evenly distributed according to the Gini Index. Again, most routes can adhere to the schedule during the morning peak, since in reality, the passengers at this time are regulars going to work or school, and there is less variation in congestion levels. As the day goes on, there are even fewer significant variables. Notably, at noon, one variable is significant other than the intercept. Zones with older populations tend to have more destinations with reliably on-time routes compared to other zones. Again, since the older population is slightly correlated to the suburbs, which generally have fewer traffic congestion issues compared to more central locations. In the evening, only the median income is significant, as it relates to fewer destinations having reliable routes. Similar to the buffer time observations, they tend to live in areas with less direct access, which could restrict the public transport route options.

To summarize, different times of the day correspond to different impacts on different population segments. In reality, more people would want to go downtown to work during the morning peak and return home from downtown during the afternoon peak. Hence, the route network is more optimized to focus on commuters to and from downtown, where the peak directions have more frequent services than the opposing directions. Trips outside peak hour and peak directions tend to be affected more negatively (e.g. departing from the suburbs in the afternoon peak). Hence, planners should also consider improving the reliability for

passengers not travelling in peak directions or travelling from suburb to suburb. Similarly, in reality, people returning home from central locations tend to transfer from frequent metro services to less frequent bus services. Gaining or missing a transfer could cause larger travel time variations for passengers. Future research could also potentially reverse our research and consider each zone as an arrival zone and count potential origins with reliable routes. Buffer time models tend to have more significant variables than the on-time performance variables. Given the fact that most agencies optimize for on-time operations and the difficulty of minimizing passenger travel time variations, our research again highlights the need for planners to consider potential passenger impacts when adjusting services, such as the impact on waiting and transfer times.

8.5 Conclusion and Next Steps

To summarize, public transport reliability is important for both agencies and passengers. Larger travel time variations affect the trip planning decisions and mode choices for passengers. If the services are consistently unreliable, people might change to other modes, and it is difficult to get people back to using public transport (Carrel and Walker, 2017).

Previous studies mostly focused on the reliability of observed trips, either using tap-in and tap-out data or using an inferred ridership model. Thus, they did not show the reliability of less observed destinations, which might still be important for passenger mode choice and satisfaction.

Aiming to create a more comprehensive view of public transport reliability, we simulated potential passenger trips and calculated travel time reliability assuming all origin-destination pairs are equally important. We routed passengers to all destination zones from a given origin based on the schedule and selected the fastest route in a given hour as a reference route. Then, we checked the reliability for the reference route using schedule-based (on-time performance) and variability-based (buffer time) measures. Finally, we summarized these reliability indicators for a given origin and for all origins.

Our first analysis evaluates on-time performance and buffer time reliability using varying thresholds to reflect different potential passenger experiences. Strict reliability thresholds yield results that are more extreme, i.e. either really reliable or really unreliable. Results show that reliability improves as thresholds become more forgiving as expected. Areas around metro stations have consistently performed better since there is no congestion in the metro system. Notably, the two metrics do not always align, as some areas show good on-time performance but high buffer time, likely due to frequent service sometimes allowing early

connections.

Using a similar threshold as the agency, we analyzed in detail the spatial and temporal patterns of the overall travel time reliability departing from Polytechnique Montréal. While most routes meet agency standards, there are still some spatial and temporal variations in reliability to different destinations. Buffer time metrics generally show better performance than on-time metrics, suggesting small travel time variations but with consistent delays given the agency's standards. Routes to destinations near the metro are typically more reliable, while routes to destinations near construction sites or further away from the origin experience more variability. Reliability shifts throughout the day, with morning routes generally more dependable than those in the afternoon.

A breakdown of the overall travel time, wait times, and in-vehicle times shows that in-vehicle times are more reliable compared to the wait times. In-vehicle times are more affected during the peak hours by congestion as expected. However, wait time variations are generally more stable but degrade slightly during midday and the afternoon peak. Overall travel times perform worse than individual components, reflecting compounded effects of delays. The findings highlight the importance of considering both travel time variations (buffer times) and delays (on-time performance). The study emphasizes the need for planners to factor in wait time reliability and transfer synchronization when developing their service plans to help improve passenger experience, particularly in complex public transport networks, due to the computational intensity and more varied route configurations for these detailed analyses.

For all of the potential origins within the service area, the analysis reveals that most origin zones in the network have reliable access to 300–500 destination zones (50–70% of all possible zones), with buffer time metrics generally identifying more routes as reliable than on-time performance metrics given the agency's standards. This suggests that riders are more often affected by consistent delays than unpredictable travel times, highlighting the need for schedule adjustments and priority measures to reduce travel time. The spatial-temporal reliability patterns also vary by time of day, with afternoon peak hours showing uneven delays for more origins given the more varied traffic conditions and passenger demands. Downtown consistently performs well due to robust infrastructure and service design, while areas in the west and near northeastern zones show more variability, often due to complex suburban street layouts, congestion, or transfer needs. Since the network is focused on getting people to and from the central area, it is not necessarily easy for people to travel in the direction opposite to the peak hour traffic or to travel from suburbs to other suburbs. Given the low observed demand modal share for suburb-to-suburb trips, we emphasize the need to consider the potential impacts on passengers who need to make these trips.

Equity analysis using the Gini index shows that buffer time reliability is relatively evenly distributed, especially at noon, while on-time performance shows a different pattern, where it is more evenly distributed for the morning peak. Regression modelling incorporating demographic data indicates that reliable access varies across population segments at different times of the day. Thus, the potential service reliability impacts different population segments differently throughout the day. These findings highlight the importance of considering not only operational reliability but also passenger experience, expectations, and social equity in public transport planning.

However, there are many limitations and potential future research that are needed, since this paper is a demonstration of our idea. Future researchers could nevertheless test out various trip planning parameters, route choices, passenger arrival patterns at origin, and various reliability measure thresholds. This additional information could help discover issues faced by potential passengers during their trip, as well as the sensitivity of various reliability metrics. Similarly, future studies could test out other types of sampling methods, potentially with accurate trip production and attraction information to better understand the reliability impacts on different population segments and trip purposes. Future research could look more into the various expectations and perceptions of passengers, as we demonstrated that different departure times and departure areas may be associated with different reliability, and regular passengers may have gotten used to these particularities compared to a passenger making a spontaneous trip on a new route. Finally, future studies could also concentrate on suburb-to-suburb travels to determine how low observed demands, low modal shares, and low reliability interact with each other given the difficulty of making these trips.

Acknowledgements

The authors would like to thank the Société de Transport de Montréal for providing data access. This research is funded by the Natural Science and Engineering Research Council of Canada and the Canada Research Chair in Transportation Transformation.

CHAPTER 9 DISCUSSION

This chapter discusses the findings from the five research papers, analyzing the progression from foundational statistical analyses to practical planning tools and equity evaluations. It demonstrates how each paper builds upon the others and contributes to a deeper understanding of transit travel times, highlighting the importance of detailed analysis levels and the role of traffic signals and ridership variations in shaping reliability. The chapter also highlights limitations and future research directions, emphasizing the need for continuous data improvement, cross-agency collaboration, and a stronger focus on passenger experiences in transit planning.

9.1 Overview of the Projects

The overarching goal of this study is to provide transit agencies with tools and knowledge to improve service delivery, optimize operations, and better serve diverse populations. Hence, this document presents five interconnected research projects that collectively aim to improve the understanding, modelling, and planning of transit travel times. Figure 9.1 present a high-level relationship between different parts of the study.

Each paper builds upon the insights of the previous paper and adds more details to the analysis levels to provide a more comprehensive understanding of transit travel time characteristics and their broader impacts. The research progresses from statistical characteristics to practical planning applications and passenger impact assessments.

At a high-level, each paper contributed to the general research question on system-wide multi-level analysis and spatial-temporal analysis to varying degrees. Each paper also corresponds to their specific questions on statistical distributions, means, and variances of transit travel times, modelling travel times using detailed trip components, and the potential passenger and equity impacts caused by the detailed travel time variations.

Paper 1 identifies future research needs by examining the overall statistical distribution characteristics of transit travel times at various analysis levels, since previous literature often assumes or oversimplifies travel time distributions, possibly due to low data resolution. It identifies that mixture distributions are prevalent across three analysis levels, stop-to-stop, timepoint, and route levels, influenced by factors such as traffic signals and ridership variations. There are two fundamental properties used to describe the statistical distributions from Paper 1, mean and variance, which will be explored in Papers 2 and 3.

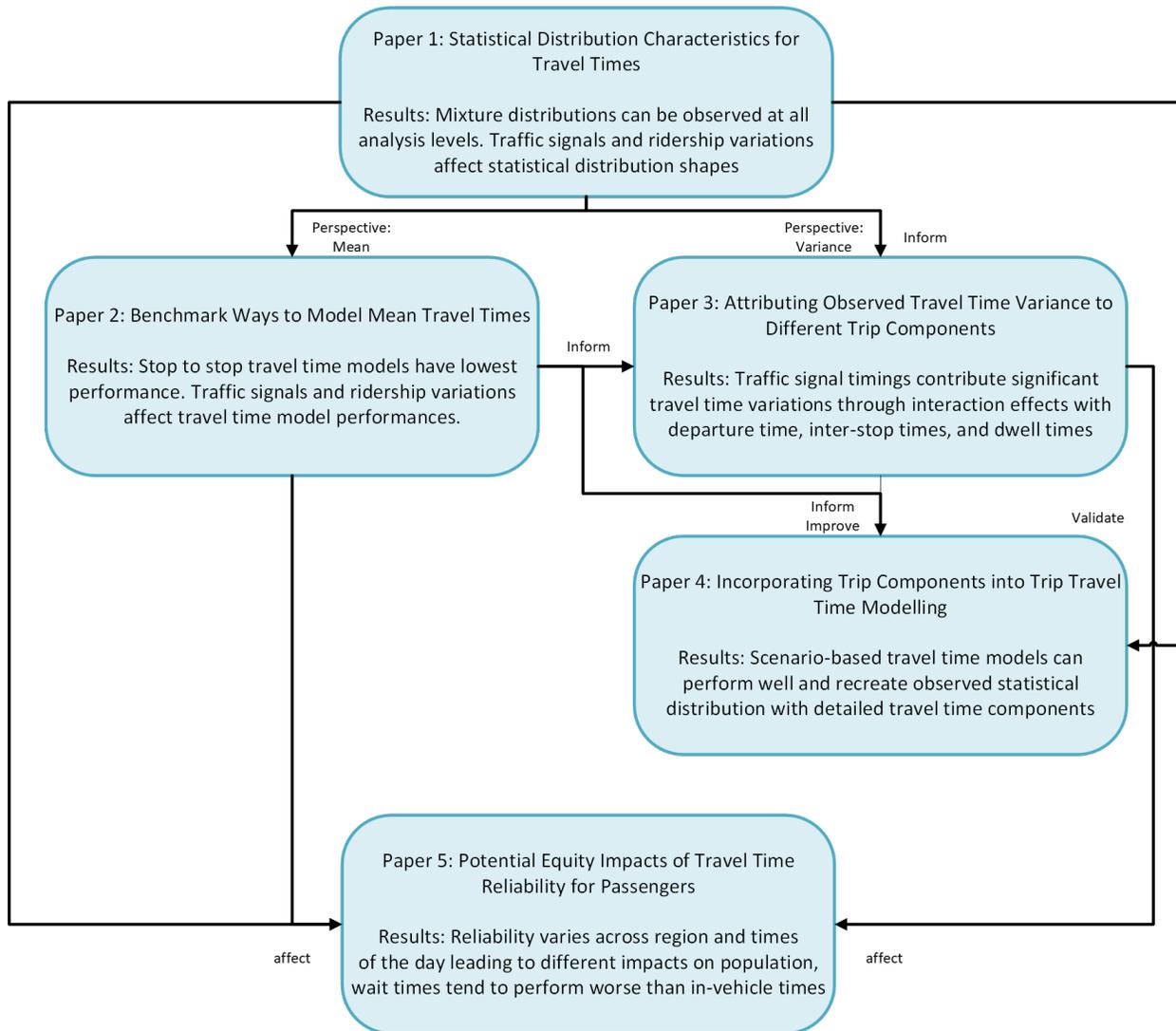


Figure 9.1 High Level Summaries and Relationships Between the Papers

Paper 2 focuses on modelling the mean travel times using streetscape, operational, and temporal data. The analysis levels get more detailed, where the inter-stop level is added to the analysis. It reveals that the stop-to-stop travel time model performs poorly, yet the inter-stop level performs well, mainly due to the importance of traffic signals and large ridership fluctuations creating difficulty for modelling. However, the inter-stop analysis level may still include any traffic signals in between two stops.

Informed by the results from Papers 1 and 2 on the potential impacts of traffic signals and ridership variations, Paper 3 aims to mitigate these issues by including estimations for these two variables. Paper 3 also focuses on the sources of travel time variance, which is the other fundamental property of a travel time distribution. It further decomposed travel times by considering traffic signals as an operational stop without passengers, and traffic signal, ridership, and congestion impacts can now be fully isolated and analyzed. Global sensitivity analysis also attributed variance contributions to specific travel time components. The result shows significant impacts from traffic signal timings and their interactions with other factors like departure time, inter-stop times, and dwell times.

This granular understanding of variance informs the next stage of modelling. Paper 4 takes the insights from Papers 2 and 3, modelling detailed means and variances, and incorporates detailed trip components into travel time models. Again, this paper considers the traffic signals as an operational stop, which allows the travel time model to react to individual component changes or overall systemic changes. By comparing to the results and knowledge gained in Paper 1 on the statistical distribution, Paper 4 demonstrates that scenario-based models can effectively replicate observed statistical distributions, offering a more flexible approach to modelling overall travel times.

The smallest analysis levels gradually got more and more detailed as the research progressed. The first paper was at the stop-to-stop level, which included the dwell times at the stop and all red light waiting times. The second paper was at the inter-stop level, which does not include any dwell time, but it includes all red light waiting times between the two stops. The third paper was at a more detailed "inter-stop" level, where the traffic signals are considered as a virtual operational stop with no ridership and with their own departure and arrival times, which allows for better isolation of congestion effects, signal delays, and dwell times. Paper 4 kept the same level of detail when trying to predict travel times for future periods.

Finally, given all the factors that impact the travel times at various analysis levels, Paper 5 focuses on the potential passenger impacts and equity implications of travel time reliability. Passengers do not necessarily board and alight at timepoints where planners define the bus schedule. Hence, the analysis level shifted again to different combinations of origin and

destinations along a route. It shows how reliability varies across regions and times of day, affecting different population groups. The study also finds that wait times are generally less reliable than in-vehicle times, raising concerns about equitable access to reliable transit.

Together, these papers form a series of projects that move from foundational statistical analyses to practical operation planning and equity-focused applications. Each subsequent paper not only incorporates insights from previous ones but also introduces additional methodological enhancements that improve the applicability of the models. A central theme across all papers is the gradual refinement and deepening of analysis levels, each adding more granularity and precision to the study of transit travel times. The analytical framework evolved from relatively more aggregated levels to more disaggregated component level models in Papers 3 and 4, where traffic signals and ridership impacts are independently estimated and isolated. The various analysis levels used in these projects show different views on travel times, and more detailed analysis levels help indicate specific factors affecting travel times. Hence, these projects enable multi-level understanding of transit travel times. As a result, planners and researchers gain knowledge on both broad patterns and fine-grained issues of transit reliability. This also allows planners to better focus their resources and target specific problems on a given route, supporting more effective interventions to improve travel times and reliability. The figure also visually shows the progression of the path taken, where each study informs the next and contributes to a more comprehensive understanding of transit travel times.

9.2 Detailed Discussion of Research Works

To summarize, good and reliable transit travel times are important for both transit agencies and passengers, who rely on travel times for their decision-making processes. Based on the existing literature, this research attempted to address two general and five specific research questions to improve the understanding and planning of transit travel times:

- General Question 1: How do different analysis levels compare in assessing systemwide transit performance, and what are their implications for transit planning and scheduling?
- General Question 2: How do spatial and temporal variations affect transit travel times and reliability across a transit system?
- Specific Question 1: Which segments have travel times that follow mixture statistical distributions, and which operational or environmental factors could explain mixture statistical distributions?

- Specific Question 2: What are the pros and cons of different related dependent variables when modelling average travel times?
- Specific Question 3: How much travel time variance can be attributed to specific trip components?
- Specific Question 4: How can detailed trip components and their interactions be incorporated into travel time models to improve the accuracy of performance prediction under different planning scenarios?
- Specific Question 5: What is the potential travel time reliability if passengers were to travel on public transport across the entire service area?

Aiming to provide more comprehensive evaluations of transit travel times and to provide more insights to the transit agencies for planning purposes, we conducted a systemwide multi-level study to further our understanding in the statistical distribution, the mean, and the variance of transit travel times. We then proposed a travel time model to integrate the knowledge gained in these studies into a simple scenario-based travel time model for planning purposes. Aiming to better understand potential passenger impacts due to unreliability, we simulated travel times for passengers, identified potential spatial and temporal changes related to the observed travel times, and evaluated potential equity impacts on different population segments. The following subsections will provide a more detailed discussion on how each paper contributes to each question.

9.2.1 General Question 1: How do different analysis levels compare in assessing systemwide transit performance, and what are their implications for transit planning and scheduling?

Across the five papers, multiple analysis levels were explored to evaluate transit performance, ranging from inter-stop segments to full routes and systemwide passenger experience simulations. Paper 1 demonstrated that different levels, i.e. stop, timepoint, and route levels, reveal mixture distribution patterns in travel time statistical distributions. Paper 2 compared modelling outcomes across analysis levels using different dependent variables, travel time, speed, and pace, revealing that higher-level analyses may be more suitable for high-level infrastructure and cost analyses, but they may mask detailed variability that may affect operations and passenger experiences. Finer levels require more detailed data but offer more accurate predictions. Papers 3 and 4 extended this by incorporating detailed trip components into performance assessments, showing that modelling at the component level, such as adding

signal delay and dwell time, offers a more accurate understanding of trip travel time variability and operational impacts. Finally, Paper 5 introduced a passenger-centric perspective by evaluating travel time reliability across the entire service area, offering an additional analysis level based on potential user experience.

Together, these system-wide analysis with various analysis levels help contribute to the literature a comprehensive framework for understanding and improving transit travel time reliability by illustrating specific issues facing specific routes. The main findings highlights that choosing the appropriate analysis level is critical for transit planning, more detailed levels provide deeper insight into variability and component-level interactions, while higher levels hides detailed variations but support broader high-level planning decisions.

9.2.2 General Question 2: How do spatial and temporal variations affect transit travel times and reliability across a transit system?

All five papers show spatial and temporal variations in transit travel times and their reliability to various degrees. Paper 1 showed that mixture distributions, associated with various underlying traffic conditions, are often concentrated in specific spatial zones, such as corridors with traffic signals and during temporal peaks like weekday rush hours. Paper 2 confirmed that time of day, street classification, and operational features, such as frequency and ridership, affect average travel times and the performance of different modelling approaches. Paper 3 found that variation in signal timing patterns makes travel times even more temporally dependent. Small changes in departure times can lead to significant impacts downstream due to interaction effects. By expanding the results, Paper 4 further demonstrated that red light waiting times are significantly higher in the central areas with dense traffic signals, while less travel times are attributed to red lights in the suburbs. The performance of our simple travel time model also emphasized that travel time modelling depends on accurately capturing spatial and temporal contexts of specific trip components, such as detailed ridership and traffic signal timing. Paper 5 brought these themes together by showing how travel time reliability varies across neighbourhoods and times of day, with suburban areas having worse reliability during peak hours, dense central neighbourhoods with no metro service having consistently low reliability all day, and downtown areas having better reliability given better metro coverage.

By analyzing these spatial and temporal variations, this research contributes to when and where unreliable or slow travel conditions occur, helping planners anticipate and address issues that disproportionately affect certain areas or populations. Together, these results collectively highlight the need for planners to incorporate detailed spatial and temporal di-

mensions when evaluating or predicting transit performance. Overlooking these factors risks hiding detailed variations affecting planning and operations or underestimating potential inequities and passenger experiences.

9.2.3 Specific Question 1: Which segments have travel times that follow mixture statistical distributions, and which operational or environmental factors could explain mixture statistical distributions?

The questions on the statistical distributions of transit travel times mainly relate to the first paper. Previous literature pointed out the need to develop more precise models related to the statistical distribution of transit travel times, since some literature has observed mixture transit travel time distributions using more granular observations.

Aiming to comprehensively identify the mixture travel time distributions and to provide factors relating to these mixture distributions to the transit agencies, this study included a systemwide study at three analysis levels, stop, timepoint, and route levels, to illustrate the presence as well as the spatial-temporal patterns of these mixture distributions. Then, classification models classified whether a given segment would follow mixture distribution, given streetscape, temporal, and operational characteristics, which allowed us to examine the importance of each predictor.

The results show the presence of mixture distributions in all analysis levels, stop, timepoint, and route level. The percentage of timepoint pairs and stop pairs that follow mixture distributions stays relatively stable throughout the day. However, the variation at the service pattern level is higher, potentially due to the smaller sample size. The weekend services also tend to have fewer segments identified as mixture distribution, which could also be due to the smaller sample size or due to the lower traffic congestion and demand patterns. The geographical distribution shows that segments near traffic lights or on major transportation corridors tend to have more mixture distributions. Neighbourhoods with more passenger demands, higher population density, and mixed land use could have more segments identified as mixture distribution.

Using various classification models, we showed that non-linear methods and similarity-based models work the best in classifying the shape of transit travel time distributions. The results highlight the need to consider the non-linearity and suggest that similar segments tend to behave similarly. Planners could potentially try to address travel time issues using similar strategies for similar segments at the same time, given that similarity-based models worked well. The regression models highlighted the potential effects of passenger demand variations, weekday services, operator preferences, vehicle-to-vehicle interactions, traffic lights,

and segment length variables.

Hence, this research contributes the literature on the mixture travel times by identifying and analyzing the presence of mixture distributions across multiple levels of the transit system, including stop-to-stop, timepoint-to-timepoint, and route levels. It fills the gap in the literature on how different environmental, spatial, and temporal variables, such as varying traffic signals and ridership variations, combine to create multimodal distributions that complicate service reliability. Through spatial and temporal analysis, the study shows where and when these mixture distributions occur, offering insights for future research to improve service reliability.

9.2.4 Specific Question 2: What are the pros and cons of different related dependent variables when modelling average travel times?

Improve the understanding of average travel time model mainly relates to Paper 2. It looks at the mean, which is a fundamental property of the statistical distributions identified from Paper 1. Most of the previous literature focuses on travel times at timepoint levels, and normalized measures, like travel speeds or paces, are typically used to evaluate delivered services or to plan infrastructures at route levels. Built upon Paper 1, where we learned traffic signals and ridership variations contribute to mixture travel time distributions, this paper further improves the analysis resolution by adding the inter-stop analysis level, which does not include dwell times at stops and signal delays at near-side stops, which is the majority given the local context. In addition, travel times are the result of varying speeds or paces given the fixed distances in fixed route operations.

Thus, aiming to help planners better choose the models or measures for their specific planning applications, in the second contribution, we proposed a framework to compute and compare the travel time, speed, and pace measures commonly used by transit agencies at four analysis levels, inter-stop, stop, timepoint, and route levels. To test these measures, we came up with two planning scenarios, one is to test the expansion of service areas using new routes, and another is to test new service hours for existing services. Finally, we discuss the aggregated results to show the overall performances for each model and the disaggregated results to demonstrate various biases related to each model.

The models show that the non-linear travel time models perform slightly better. Transit travel times, speeds, and paces are greatly impacted by temporal variables, like time of the day, spatial variables, like street classifications and the number of traffic lights, as well as operational variables, such as service frequency and ridership. However, most other spatial variables, like land use, are not significant for travel time models at a higher level, but are

more significant in finer analysis levels. The distance coefficients in travel time models may be too restrictive to deal with the changing segment lengths and speeds in reality.

Spatial, temporal, and operational variables can explain higher analysis levels much better, such as timepoint-to-timepoint and service pattern levels. For lower levels, the inter-stop level performs better than the stop-to-stop level. These results show that the analyses at higher levels may hide more detailed variations at lower levels. Improving the results at these lower levels requires more detailed data and further studies on dwell times and signal waiting times.

Segments with existing observations can help greatly when estimating travel times for expanding service hours, since they have a segment-specific intercept to take account of the differences between segments due to the lack of detailed dwell time and signal waiting time data. However, the normalized results from speed models perform better for the new route scenarios, given the lack of segment-specific intercepts. We believe planners could more intuitively estimate how fast buses can travel along the segments based on local knowledge and environmental variables when planning for a new route, which is in line with the current practice.

We also highlight the shortcomings of using specific aggregated measures in previous literature, since different measures evaluate the results "from different angles" (Kolassa, 2020) and that various error measures indicated different "winners" when evaluating model performances. A more disaggregated error analysis shows that the speed models tend to underestimate the average travel times since the speeds are less affected by extreme values, such as extreme weather events. Pace models, on the other hand, tend to overestimate average travel times. Direct travel time models tend to have less skewed errors. Although all three approaches perform similarly in terms of the average errors, the speed and pace models perform relatively better and more consistently relative to the actual values. Time models tend to struggle more with faster average speeds and short segments, which can be attributed to the fixed paces in the coefficients.

The methods and results contributes to the literature that we calculate and compare aggregated model performances using the observed travel times, speeds, and paces integrated with environmental and operational characteristics from a wide range of segments in the network, offering a more comprehensive view of the travel time models. The second contribution is that we summarized the disaggregated model performances from the aforementioned model results to discover potential biases related to each dependent variable modelled. These aggregated and disaggregated results allow us to discuss the advantages and disadvantages of these various modelling approaches, analysis levels, and evaluation criteria, so that we can

make recommendations to other researchers and planners when deciding future modelling approaches.

9.2.5 Specific Question 3: How much travel time variance can be attributed to specific trip components?

Improving our understanding of travel time variances mainly relates to Paper 3, which is another fundamental property of statistical distributions identified in Paper 1. Previous studies mostly focused on either the variation of overall trip travel times or the variation of specific elements in trip travel times, such as dwell times or signal waiting times. Given the importance of traffic signals and ridership variations identified in Paper 1 and 2, there is a need to better isolate the signal delays and dwell times to once again improve the data resolution on these two issues in our travel time analyses. There is still a need to better isolate the variation in each trip element, discover how these various elements interact with each other, and rank the impact of each element on the overall travel time variation.

Aiming to create more comprehensive evaluations of transit travel time variations and to help prioritize the resources for transit planners, we further decomposed recorded transit travel times from 20 routes in Montréal into several trip elements, i.e. departure time, inter-stop times, dwell times, and red light waiting times. Then, using these times, we conducted both a variance-based sensitivity analysis and a one-at-a-time analysis for observed transit travel times to demonstrate the importance of various trip elements.

The results from the third paper show that we can better decompose the travel times by adding more details on traffic signal settings and ridership variations, and that the variations in specific trip elements can be better isolated. From the variance-based analyses, we quantified and ranked the contribution of each trip element to the overall trip travel time variations. The red light waiting times contributed the most variation to the trip travel time variations for most of the routes analyzed in our case study. The most important element observed in our case study is the departure time at the first stop, which determines how synchronized the traffic signals are further downstream on the route given the fixed timing plan with varying cycle lengths used in Montreal. In addition, most of the variation in the overall travel time and red light waiting time came from the interaction effects. The inter-stop travel times and dwell times, however, are affected less by the interaction effects and can mostly be ranked by the individual variations. The one-at-a-time analysis also demonstrated the non-linear relationship between the variation of each trip element and the trip travel time variation.

Planners need to choose the departure time carefully and better understand the base signal timings before designing or adjusting service schedules. Slightly modifying the departure

time to facilitate interlining is not necessarily a good idea without a better understanding of the potential impact on the overall travel times. A slight change in departure time can potentially cause a large travel time variation due to the changes in signal synchronization patterns. Thus, the historical vehicle location and travel time observations are conditional given the fixed signal plan and the varying cycle lengths. Planners could either optimize for specific departure times rather than applying the same travel time everywhere, or add adequate schedule padding to account for the travel time variations which would increase operating costs. Similarly, small changes in signal timing plan changes can cause a chain-reaction effect on the overall travel times, where the resulting travel time changes can be small locally, but the impact on the whole trip can be large due to the interaction effects for the sections downstream. Thus, planners need to include more detailed signal timing plans and red light waiting time information when scheduling or planning for transit travel times.

The results adds to the literature by introducing more detailed analysis levels, where traffic signals as an operational stop. The decomposition of travel times help better isolate and pinpoint specific issues related to congestion, ridership, and signal delays, which allows analysts to clearly quantify the impacts of congestion, ridership variation, and red lights. The importance of including traffic signal timings in travel time models also offers future research directions to improve travel time models and travel time reliability.

9.2.6 Specific Question 4: How can detailed trip components and their interactions be incorporated into travel time models to improve the accuracy of performance prediction under different planning scenarios?

The fourth paper extends the third paper to travel time planning and estimation contexts. Given the previous results highlighting the importance of traffic signal impacts and ridership variations and the lack of literature using detailed travel time components in travel time modelling, we proposed a simple scenario-based travel time model that calculates trip components given a planning scenario, i.e. departure time, inter-stop times, dwell times, and red light waiting times. We also tested the model performances given various planning scenarios in the observation period and future service periods. This paper also provides more empirical system-wide evidence on traffic signal impacts on the overall travel times as well as updates the sensitivity results in paper 3.

The fourth paper shows that using the detailed trip components in travel time modelling is feasible with acceptable model performance given various planning scenarios. We showed that incorporating detailed inter-stop, dwell, and signal delay components can produce travel time estimates that closely align with observed values, even under changing operating condi-

tions. The model performs generally better compared to simply using the observed average travel times in the planning context. This model enables planners to better evaluate different operational conditions and their impacts on the overall travel times. However, more information is needed to improve the accuracy and the usefulness of the model, such as mid-route driver change and the most up-to-date signal timings for example. Planners could also use the model to evaluate various strategies to reduce the red time impacts on bus travel times, such as using a combination of strategies like departure time changes, signal parameter changes, and more aggressive signal priorities.

The contributions are first, the integration of detailed components into a simple travel time model. We demonstrate how departure times, inter-stop travel times, dwell times, red light waiting times, and their interactions can be integrated in a simple travel time model while accounting for the decisions drivers make on the road. Next, we contribute to the development of a scenario-based travel time model. By using these detailed travel time components, we introduce a model that enables planners to simulate different changes in the network collectively or individually, such as departure time, congestion, or ridership changes, and assess their impacts on travel times to address potential issues more proactively.

9.2.7 Specific Question 5: What is the potential travel time reliability if passengers were to travel on public transport across the entire service area?

Given all the factors that affect transit travel times and reliability, the fifth paper intended to better understand the potential equity impacts of transit reliability on potential passengers by answering the specific question 5. Previous studies mainly focused on the existing demands and observed passenger travel times. However, there is limited information on whether passengers who are stuck with low-demand routes and passengers on popular routes share similar experiences.

We simulated the potential travel options from evenly sampled hexagon grids across the entire service area and calculated the travel time reliability if passengers were to travel across the whole region evenly and regularly. We routed passengers to all destination zones from a given origin based on the schedule and selected the fastest route in a given hour as a reference route. Then, we checked the reliability for the reference route using schedule-based (on-time performance) and variability-based (buffer time) measures. Finally, we summarized these reliability indicators for a given origin and for all origins.

The fifth paper on the potential reliability for passengers shows that strict reliability thresholds would yield results that are more extreme, i.e. either really reliable or really unreliable, more forgiving thresholds would result in better reliability metrics as expected. On-time per-

formance and buffer time metric results do not necessarily align, and planners want to look at both metrics to determine the exact nature of reliability issues given an origin-destination pair. Some suburban areas show good on-time performance but high buffer time, likely due to catching an earlier transfer trip.

Given the Montréal local context, routes originating near the metro are typically more reliable, while origins near construction sites or destinations further away from the origin experience more variability. Reliability shifts throughout the day, with morning trips generally more dependable than those in the afternoon. A breakdown of journey time into wait times and in-vehicle times shows that in-vehicle times are more reliable compared to the wait times. In-vehicle times are more affected during the peak hours by congestion as expected. Wait time variations perform worse than in-vehicle times and are generally more stable with small degradations during midday and the afternoon peak. Since wait times and transfer times are generally not planned or analyzed, we highlight the need for agencies to incorporate these analyses to improve passenger wait time performance. Overall journey times perform worse than individual components, reflecting compounded effects of delays.

A simple spatial regression model incorporating demographic data and Gini indices indicates that reliable access varies across population segments at different times of the day. Central areas, e.g. downtown, have better access to reliable routes given that the transit network is designed to get people to and from central areas. Suburb-to-suburb travels tend to be more difficult. Thus, the potential service reliability impacts different population segments differently throughout the day.

The contribution is to remove the passenger demands and sample the service area evenly. This would help discover potential equity issues, especially for passengers who are stuck using transit on long or low-demand journeys, since the existing literature focuses on the observed demands. The second contribution is the spatial and temporal analysis, which help discover the reliability issues for a specific origin and for all origins.

9.3 Summary of Limitations and Directions for Future Research

In summary, while this research provides a multi-level, systemwide understanding of transit travel times and reliability, several limitations may affect its applicability. A recurring theme across all papers is the mismatch between the granularity required to model passenger experiences accurately and the aggregated levels typically used in transit planning. Limitations in data quality and availability, particularly regarding ridership patterns, real-time signal timings, and driver behaviours, will affect the generalizability of the models. Method-

ologically, the studies face constraints related to certain simplification assumptions despite the more detailed analysis levels, as well as limited temporal coverage for detailed analysis levels due to higher computational requirements. Furthermore, while the research touches on passenger reliability, it is limited by the lack of detailed behavioural data, assuming uniform expectations across the network. As for the results, the applicability elsewhere may be limited, since other cities may have vastly different transit planning or signal timing guidelines for example. Future work should focus on increasing data resolution, incorporating more detailed modelling levels, and bridging the gap between operational metrics and passenger experiences to enhance both planning practices and service delivery.

More specifically, across all five papers, a key limitation is the disconnect between the analysis levels used for planning and the levels at which passengers experience service, which relates to the first general question. Planning is often done at higher levels, i.e. timepoint or route levels, but passenger reliability is perceived at finer-grained levels such as stops or inter-stop segments. This mismatch limits the ability of existing models to fully capture passenger experiences. Future research should aim to integrate finer-scale analyses into system-level planning, particularly to capture the detailed variations caused by ridership variations and signal settings, which become more significant at smaller scales and may cause travel time variations downstream. Higher-level analyses can hide important local variations in travel times, so we emphasize the need to model transit performance at more detailed levels and improve the granularity of our understanding of transit travel times. We also want to highlight the need for agencies and equipment suppliers to continuously improve the data granularity and quality to enable these detailed analyses.

The studies were also limited in their ability to capture and respond to spatial and temporal variations related to the second general question, especially for changes in operational contexts such as the implementation of new bus lanes, signal priority strategies, or evolving ridership patterns. The more detailed analyses did not cover many changes in signal timing, service adjustments, or rider behaviour shifts, which limited the prediction accuracy for planning in future or modified contexts. Future work should use longitudinal data to further evaluate temporal changes, such as the effectiveness of bus priority strategies over time. More detailed spatial data on infrastructure, land use, and traffic signal timings can help explain potential geographic disparities in reliability. Finally, future research should test different sampling approaches for transit travel times while accounting for differing passenger expectations and mobility patterns, which vary by location and time of day.

Finally, we want to mention that our research is limited in terms of passenger experiences, since we do not have much passenger-related data. Passenger activities relate to the dwell

times, which are important to model travel times for service planning. Passenger data are also important to evaluate the impact of service delivery on passenger experiences, since passengers may shift to other modes if their experiences are bad and they won't come back to transit unless major improvements are made (Carrel et al., 2013a). For passengers, a trip may involve two or more vehicles and the perceived travel times may vary in different parts of the trip. Thus, operational measures like the ones evaluated in most of the papers may not necessarily reflect passenger experiences and perceptions completely. Therefore, future researchers could also introduce additional analysis levels to better understand passenger experiences. Transit agencies could also potentially use detailed travel time analyses to inform passengers on potential issues on the route or the likelihood of getting to their destinations on time.

More specifically, for the first specific question on travel time distributions, there are two specific limitations with regard to Paper 1. One is with regard to the methodology, where we aggregate the entire dataset without considering the service changes that occurred during the study period. Additional research could examine potential contributions from seasonal variations and changes in infrastructure to mixture distributions. Another limitation is that our classification models only included a binary outcome variable, due to the smaller sample size with a higher number of statistical components. Future research could analyze the higher components in more detail. Similarly, due to page limits for the paper, we couldn't discuss more on the validation of different models, such as detailed correlations, units of the inputs, relative effects, and the interpretations of the model coefficients. Thus, readers should pay attention to these details when interpreting the results or when trying to replicate the studies. Finally, we only selected four classification models, from a wide range of categories. Future research could test additional classification models and compare their performances.

As mentioned in the general results and limitations, a potential expansion of this research is to use more detailed data on transit ridership, traffic counts, and traffic light settings to help infer the reason causing the slower travel conditions on each trip or segment. Using identified causes, agencies can target specific areas and potentially "nudge" the travel times under slower conditions towards faster conditions. Similarly, there are other occasional factors causing slower travel conditions beyond transit agencies' control, such as major events, traffic incidents, and weather. By incorporating these data, agencies could evaluate potential response strategies to increase the reliability and resilience of a given transit network. These expansions would help reduce overall travel times, improve vehicle travel conditions, and improve service consistency.

For the second general question on modelling the mean travel times, there are still needs

to further include detailed dwell times and traffic light timings in the modelling processes to improve the models, especially at the stop-to-stop level as they become more significant. Future researchers could easily adopt and expand upon this framework to test new methods with additional variables, such as weather, signal timings, ridership variations, and congestion levels to better help agencies plan and react to changes in the network for their operations. In addition, future researchers could also compare and experiment with many other modelling methods, such as using time-series methods to include more detailed seasonal and year-over-year variations or artificial intelligence methods to include more detailed environmental variables.

Given the limitations from the first two papers on traffic signals and ridership variations, we attempted to add estimated traffic signal timing and dwell time to better understand the travel time variations, i.e. specific question 3. The first is related to our analysis scale. We have only analyzed three months of recent data due to data limitations and compute complexity, yet we would need more observations to better reduce the confidence intervals for less frequent routes during off-peak times. There is also a need for a longitudinal study to discover potential trends and changes with regard to which element got better or got worse in the sensitivity results to better inform transit planners on the changes in the system or the effectiveness of bus priority strategies. With more data on more routes, it will also be possible to examine more detailed temporal and spatial patterns affecting transit travel time reliability, which may in turn help the planning process for new schedules or routes.

The second limitation is related to the dwell time models. Our methodology does not consider the variation in ridership patterns given the limitations of our AFC data. As the matching algorithm improves, it would be interesting to better incorporate the detailed passenger pattern into the research. Our methodology does not currently consider the interaction between buses which relates to more detailed ridership and dwell time variations. Since delays can affect the number of boardings and alightings, especially for more frequent services, the vehicle may pick up additional passengers who are supposed to take the following vehicle. The interactions between vehicles, such as vehicles on the same route, common trunk sections served by various routes, as well as local and express services running parallel, should be better understood to improve the accuracy of the dwell time model, as well as to inform planning and operation decisions.

There is also a limitation with regard to the traffic signal models. Since our traffic signals mainly use fixed timing plans with very limited split changes for transit signal priority, there is a need to better understand the interaction between buses and actuated non-fixed timing signals using empirical data as the majority of the current literature is based on

simulations. More detailed signal waiting time data could also help inform traffic engineers when determining what to simulate, as well as to calibrate their existing models. This way, we could potentially compare and identify pros and cons in various signal timing configurations to help improve transit reliability. Similarly, future research could find a way to better give signal priority to buses when travelling down a coordinated corridor with densely configured signals.

When incorporating detailed trip components into travel time modelling (specific question 4), similar limitations exist, since Paper 4 builds on the previous research. If the goal is to predict future travel times, more up-to-date information is required, such as the latest signal timing and the latest departure times. The high possibility of not having the most up-to-date traffic signal data, congestion levels, and detailed ridership forecasts could constrain the precision of our estimates. Hence, we emphasize the need for traffic engineers to collaborate with transit planners to improve communications on and plan for upcoming changes. The travel time model does not fully capture mid-route driver changes, operator behaviour differences, or ridership interactions between overlapping services on the same corridor. More research is needed to incorporate mid-route driver change times into a travel time model, since in practice, run cutting and driver sign-ups happen after the travel times have been set, yet these times will have a general impact on passenger travel times. Future research should also focus on integrating predictive ridership models, higher-resolution vehicle location data, and real-time signal control strategies into the travel time models to improve the accuracy and the responsiveness to changes in the network. Further work is also needed to improve interline optimization algorithms to incorporate these detailed travel time components into large-scale scheduling and analysis processes.

When evaluating potential passenger experiences for our specific question 5, future researchers could look into various passenger expectations and perceptions with regard to reliability issues, since our paper applied the same expectation for every origin-destination pair. Similarly, there are various trip planning parameters that would affect the chosen routes and departure times. These parameters and metric thresholds could potentially affect the results. Therefore, future research could potentially determine the sensitivity of research results given the variations in passenger expectations. Similar to the third paper, researchers could do more research into ridership pattern variations and passenger arrival rates at stops, since regular passengers with route knowledge might adapt their behaviour based on their experiences. In turn, they could test out different trip planning and preference parameters based on this research, which might better reflect reality and the varied passenger expectations. Finally, this paper did not include the numbers of various destinations for each analysis zone, which could better reflect the passenger demands and experiences in reality. However, due to our

relatively small spatial resolution, large amounts of zones only contain residential land uses, which are typically not considered as points of interest or destinations, but transit trips to visit friends or family do happen frequently and the reliability should be considered. Future researchers could factor in more detailed and comprehensive trip attractions and generations, as well as detailed population information, such as modal availabilities, economic factors, and other special needs, into the analyses of reliability impacts.

9.4 Detailed Discussions on the Methodological Implications

To summarize the first four papers, the smallest analysis levels gradually got more and more detailed. The first paper was at the stop-to-stop level, which included the dwell times at the stop and all red light waiting times between the departure times. The second paper was at the inter-stop level, which does not include any dwell time, but it still includes all red light waiting times between the departure at the first stop and the arrival at the second stop. The third paper was at a more detailed "inter-stop" level, where we considered the traffic signals as a virtual operational stop with no ridership and with their own departure and arrival times. Hence, for the third paper, the "inter-stop" only includes the travel times between the stops or operational stops, and it is much clearer to isolate congestion effects from signal delays and dwell times.

Once again, we want to highlight the common theme from all three papers: small local variations may cause a global chain reaction and the higher levels of analysis hide more detailed variations that affect passenger experiences. Hence, for high-level planning purposes such as infrastructure planning, it is okay to stay at a higher level. However, for schedule revisions, operation optimizations, and transit preferential treatments, it is important to pay closer attention to the more detailed results as they align better with passenger experiences and can better indicate specific problems to the planners.

There have been many calls to move forward towards stop-level scheduling, which passengers rely on since most of them don't travel from timepoint to timepoint (Wessel et al., 2017). However, as we have shown, the most difficult analysis level for modelling and analysis is the stop-to-stop level. Red times may be evened out for higher-level analysis. Missing one green light may not be a big impact at the route level, but missing one green light becomes more significant at the stop level. The importance of dwell and red time modelling has been a common theme in all three papers. Yet, our proposed travel time model from paper 4 can combine departure, inter-stop, dwell, and red times together and can provide estimates that are closely aligned with the observed values under different operational contexts. Hence, for scheduling and improving reliability purposes, the higher-level analyses might not be a good

idea since they hide detailed variations in red times, dwell times, and congestion levels. We once again emphasize the need for agencies to continuously improve the data granularity, which would allow future researchers or analysts to focus on these detailed variations.

For future researchers, I want to highlight the importance of understanding the process which generated the observed data, since it directly influences the selection of models, the validity of assumptions, and the applicability of the results. Observing the data generation process, such as travel times, is essential for developing accurate and reliable models. By directly observing how higher-level travel time data are generated, such as by riding a bus or observing buses get stuck at intersections, we can then look at the data for clues with a better understanding of assumptions, contexts, limitations, and potential sources of error that are present in the data. When this understanding is strong, even simple methods used in my projects, from the early 1960s to 1980s, can perform surprisingly well because they align closely with the data generation process. Without the thousands of hours spent on buses, the inspirations for my research projects would not have happened.

As shown in the four papers, there are assumptions in transit planning that are not necessarily valid, such as assuming travel times remain constant when moving departure times by a few minutes to facilitate interlining. Similarly, the results in my papers are based on data generated in Montreal and may not be directly applied elsewhere, since other cities may have vastly different operational contexts, such as the traffic signal timing guidelines. Insight into how travel time observations were generated, what the underlying patterns are, and what potential biases might be present, such as red and dwell times, we can improve our data collection process to include more diverse contexts and improve our models to be more applicable elsewhere.

Although collecting and adding these ridership and traffic signal details is not a simple task, which includes merging many different datasets from different sources, the methodologies proposed in the papers could be easily integrated into existing data standards and research frameworks. There are various attempts to generalize the GTFS standard to include more operational information (e.g. the Transit Operational Data Standard). For example, the traffic signals could be directly added to the stop times table in GTFS as well as the trip updates in GTFS-RT as an extension, as shown in the overview of the methodology, which would allow more detailed calculations mentioned in the papers.

Similarly, the railway industry has long been using the concept of operational stops for scheduling and optimization purposes. A single-track railway means there is only one track between two stations or operational stations (passing loops). The passing loops on single-track railways is a place where there is more than one track to allow trains running in opposite

directions to pass each other. Trains going in the same direction can also overtake at these operational stations. A given train cannot depart unless the passing train has cleared the track block (Lamorgese and Mannino, 2015). Railway timetables are developed to include various line and train speed restrictions and these waiting times in the passing loops (i.e. scheduling for red lights).

Unfortunately, bus systems, especially when operating in mixed traffic, tend to be more affected by traffic signals and other traffic, resulting in slower travel times (averages) and higher travel time variations (variances). Researchers could consider expanding the dataset to include operational stops and adopt some ideas from railway operations to examine the bus operations with more detailed traffic signal and ridership information similar to our efforts in papers 3 and 4. Then, researchers could better analyze the issues facing bus planning and operations, such as optimizing bus schedules and better predicting bus travel times.

In railway operations, delays also occur quite often. Various models exist to find rescheduling and rerouting decisions to avoid route conflicts and to minimize overall delays on single-track railways (Mannino and Nakkerud, 2023; Lamorgese and Mannino, 2015). Once again as mentioned in the results, traffic signals cause a significant variation in travel times, which is a headache for modelling. Similar concepts from railway operations can be adopted for real-time signal priority for buses while accounting for delays for other road users. Signal timings can be evened out slightly a few cycles before the vehicle arrives, so that transit vehicles can go through without a significant waiting time as well as minimizing impacts for other traffic. These fine-tuning and optimization methods require good predictions of bus arrival times and trajectory (Furth et al., 2025; Li et al., 2012), which we have helped contribute to in paper 4. Regardless of railway operation or bus operation, it is important to consider the small local impacts when rescheduling as well as the global chain-reaction effects on the entire line operation, as shown in our papers as well as the railway studies cited above.

Finally, we want to highlight the importance of evaluating potential passenger experiences, since they are the reason why we provide transit services in the first place. Future research should consider adding more information with regard to the passenger impacts. For example, for traffic signals, planners could evaluate the amount of passenger delays versus the amount of vehicle delays. If a bus with 50 people on board during rush hour has to wait 1 minute for the red lights at an intersection, the signal delays for the passengers are 50 minutes, which is significant. For some cities, signal priority is only given to late buses or signal priority is not applied to consecutive buses (Altun and Furth, 2009), it is important to consider whether planners want to use signal priority and risk running late for passengers or have more on-time arrivals but at the expense of not benefiting from signal priority.

Similarly, when optimizing for interlining, departure times are adjusted by a few minutes with the assumption that the travel times do not change much. As we have shown, this assumption is not necessarily a good one. The significant signal delays, frequent changes in departure times, and the resulting travel time changes will affect how passengers use and perceive the services.

For passengers, especially those commuting to work, school, or time-sensitive appointments, unreliable travel times can be more disruptive than a slightly longer but dependable journey. If buses or trains frequently run late or have large travel time variations, riders may be forced to leave earlier than necessary to ensure on-time arrivals, effectively extending their perceived travel time and reducing the appeal of transit altogether (Carrel and Walker, 2017). It is necessary to incorporate potential passenger travel time impacts and reliability into these operational optimization efforts, such as interlining and transit priority treatments. The optimizations could be great on paper for operational efficiency, but they could have potential negative impacts on passenger experiences and perceptions, which would make these efficiency gains less useful in the real world. Thus, it is important to constantly evaluate the passenger experiences and assess the changes in behaviours due to various changes and optimization efforts in the transit network.

9.5 General Discussions on the Policy Implications

Based on the research projects conducted here, we can see that, at least in Montréal, the common issues related to the travel times and their variations are traffic signal settings and inter-stop travel time variations. Thus, improving transit travel times and reducing travel time variations will be an all-hands-on-deck situation spreading across various agencies and departments and will be more difficult if there is less political will to do so (Arias et al., 2021).

Starting simple, the transit agencies, suppliers, and future researchers should continuously improve the quality, reliability, and granularity of various existing transit data sources, since, again, higher-level analyses mask more detailed variations affecting operations and passenger experiences. More detailed analyses of various existing datasets can reveal additional factors impacting transit travel times and their variations. The most sophisticated data analysis methods are only as good as the data feeding into them. Good quality data can provide information for planners to better adjust service and for passengers to plan their trips. Well-run services and accurate information to passengers can build public trust and support when advocating for more funding given the difficult financial contexts in public transit around the world.

There should be a specific emphasis on evaluating ridership variations and signal waiting times to better inform transit planning and operations. As highlighted in various sections above, there is less attention in the existing literature that focuses on these two factors. Given the limited financial resources available to the agencies in most places, having these more detailed data could help the agency to prioritize their efforts, focus on certain issues, and find the solutions with the best value for the resources invested.

Obtaining data from different sources to perform the detailed analyses may also be difficult, since it may relate to multiple departments. Some agencies, especially smaller agencies, don't have the ability to analyze data in-house and often rely on the equipment suppliers to manage the data and do the analyses for them. In these cases, data requests may take longer, potentially limiting the agency's ability to better react to changes in the network. If agencies are given poor-quality results or no data at all, they can't respond quickly to demand variations, travel time changes, or potential routing changes due to detours, which might lead to wasted resources, deteriorated services, and loss of public trust. Hence, we want to emphasize the importance of data sharing between departments and data ownership for the public transit agencies.

Unfortunately, there was less information related to the passenger experiences during my study, since the development of more detailed passenger trip estimation is still in progress. There is still a need for planners to include more detailed passenger experiences and impacts in the analysis so that transit agencies can attract new passengers and retain existing passengers, since passengers are the reason why we run transit services. Again, given the literature review, passengers value their time and delays differently during various stages of their trip (Wardman, 2004). By including more detailed information, agencies could also find solutions that improve passenger perceptions.

In addition, once planners understand the congestion, traffic signal settings, and ridership variations in more detail, planners can better react to potential changes in the network and surrounding environments. Planners can also better communicate to operators what the expectations of the schedule are that was planned. This is important since schedules are planned in advance to facilitate driver sign-ups, and operational contexts may change after the driver bookings are finalized. Also, not every operator has the same driving habits or expectations as highlighted in our fourth paper. Communicating these expectations helps drivers make various decisions while delivering the service, such as how to take advantage of various transit priority measures since they vary from cities to cities, corridors to corridors.

The needs of a transit system are constantly changing due to everyday changes like detours and long-term factors like urban growth, seasonal fluctuations, economic changes, and tech-

nology advancements. Continuous performance monitoring allows agencies to better adapt to changes in ridership or traffic conditions. Once again, as demonstrated in the results, routes have very specific issues given the local operating contexts. Transit priority initiatives, such as signal priority, require constant performance monitoring and calibrations as needed. Since the traffic changes constantly, the implemented priority strategies need to adapt too.

Given the lack of empirical evaluation of transit priority measures, the current studies are generally for a given route or a portion of a corridor. We emphasize the needs for planners and researchers to systematically and empirically evaluate the transit performance with or without these measures. Then, sharing the results and lessons learned from different implementations to better support future research and planning.

For example, the bus rapid transit route in Cleveland, Ohio has almost all the things you could ask for in a bus rapid transit route when it began operating, such as a dedicated center-running bus lane and off-board fare payment. However, the actual travel times were 10 minutes (33%) longer than what was promised, due to the lack of signal priority. The new signal control systems were paid for by the transit agency and the federal government, but after car drivers complained, it was turned off by the city's traffic division and reverted to the conventional timing plans (Schmitt, 2014). The transit agency did not complain because the ridership was up anyway. Similarly, the light rail in Houston, Texas, lost its signal priority around September 2025, which contributes to around 10-minute delays per trips, 5000 hours total for passengers at a few intersections downtown (Beaubien, 2025). Unlike Cleveland, Houston also organized a race between pedestrians and light rail vehicles to raise awareness of the impacts, with the light rail barely beating the pedestrians.

Similar observations can be made for Montréal, where most of the signals follow the conventional fixed-timing coordinated plan, which is in accordance with various design guidelines built to facilitate continuous car travel. Hence, there is very little flexibility left to facilitate the bus travel time variations with limited signal priority. To recover the coordination, the signals may also ignore the priority requests coming right after.

Researchers could develop methods to support transit priority while limiting impacts on other road users. However, any fancy method would not beat the off button. Hence, transit planners should continuously monitor the detailed performances and transit agencies could be more transparent about the impacts from various factors on transit operations, such as signal delays, congestion, and ridership variations.

Being transparent on the delays for light rail passengers from losing traffic signal priorities, Houston successfully raised public attention on this issue, and regular citizens have also helped advocate for turning the signal priority back on (Beaubien, 2025). Similarly,

some agencies (e.g. CTA, MBTA and Southeastern Pennsylvania Transportation Authority (SEPTA)) are very transparent about the funding issues they are facing. They educate people on what's at stake, such as eliminated routes, shortened service hours, and increased fares. They also tell concerned passengers up-front on how and where to speak up for transit, such as at legislator meetings, public meetings, and budget hearings.

Again, the results from this thesis demonstrated that transit travel times are affected by various factors typically managed by different departments and agencies. Unfortunately, the difficulty is that public transit agencies do not directly manage various factors that directly impact transit travel times and their reliability, such as street designs, signal timings, and road works. These detailed analyses and results on transit travel time impacts can help change the narrative, show the challenges faced by the agencies and passengers, and demonstrate how good the service could become. Using all the newly gained knowledge and tools, more people could help advocate for more resources to improve the financial and operational situation faced by most transit agencies.

Planners also need to understand the assumptions related to different optimization methods to inform their own planning process, since, once again, the local contexts may be vastly different. It requires testing and evaluating to determine the impacts of various new methods. Researchers could spend decades improving existing solutions and developing various new solutions, yet if the new methods are not tested and implemented or if they are turned off, there will not be any improvements.

Agencies could test pilot projects and demonstrations to help illustrate how good services could become, such as temporary bus lanes and signal priorities, e.g. Ferenchik (2019). Even if the test results are not so great, planners should still post the detailed results. The lessons learned from both success and failure stories need to be recorded and celebrated, since we know better now. Again, given the lack of empirical evaluation on the detailed factors affecting transit travel times, these projects could help share the lessons learned, which could help researchers improve their methods and help other cities avoid similar mistakes.

Continuous performance monitoring and calibration of these new methods are once again key to ensuring that the new pilot projects or methods function well, remain efficient, responsive, and effective while meeting the varying passenger demands and traffic conditions. Agencies should track various detailed performance indicators, such as red times and ridership, which can help identify issues early, take corrective actions quickly, and avoid long-term service deterioration or inefficiencies. Constant calibration of travel times could help agencies calibrate their optimizations or models more frequently, which allows them to allocate buses and drivers more efficiently, implement bus lanes or signal priorities where it is needed the most,

and improve overall system efficiency, especially given the financial constraints.

Finally, similar to the methodology implications, it is important for transit researchers, planners, policy makers to go out, ride the buses, ride the trains, and observe the specific aspects that could be useful to improve our understanding of the system. My projects could never have happened without the operators allowing me to ask them questions, sit on the bus, and observe for hours. Also, as shown in the paper on potential passenger travel time reliability experiences, the potential reliabilities are not necessarily the same for everyone for any given moment. Researchers and planners could attempt to uncover the potential issues by coming up with travel scenarios that do not necessarily fit their original intention.

We could potentially draw some parallels from software development. Negative testing is often used to test an application to improve the likelihood that an application works as intended and can handle unexpected input by potential users, such as edge cases, invalid data, or out-of-range inputs. Similarly, transit planners could adopt similar processes to test the network performance for edge cases or unexpected use cases. This might include navigating multiple transfers using night buses, using wheelchairs, waiting for a bus during severe snowstorms, travelling on routes with detours, or travelling during major events.

Public transit planners play a crucial role in designing systems that move millions of riders, even a tiny percentage of riders travelling on edge case scenarios can translate to potentially hundreds or thousands of real-world passengers. For planners to truly understand the challenges faced by the public, it's essential that they ride the buses and evaluate the system in various ways. While commuting to work or school, there may be moments when the system demonstrates its robustness or reveals its weak spots. Planners and researchers can spend time thinking about what aspect makes the current ride smooth or frustrating, then apply these experiences when planning for future services. Testing these edge case scenarios not only uncovers hidden flaws but also helps ensure that the system works not just for the average commuter, but for everyone. Experiencing the system in person provides insights regarding whether the data captured the experiences and whether the assumptions made during the planning process are applicable. By doing these, planners and researchers will gain empathy for passengers which helps create more effective and passenger-centred improvements.

To conclude, transit travel times and their reliability are important for everyone, transit agencies, passengers, and the society as a whole. The thesis highlights the need to implement more granular analyses and continuous performance monitoring at the agencies for transit travel times and their variations. To achieve this, it is important to collaborate and test various solutions to improve to transit services, which would in turn help reduce the car dependency in our society. People are more open to transit when they can see tangible

improvements in service quality and experience the benefits (Carrel and Walker, 2017). Before asking people to drive less, give them something better to say yes to.

CHAPTER 10 CONCLUSION

This final chapter summarizes the importance of public transit, existing research, as well as the contributions towards methodological contributions, results, and broader implications through the five research papers presented in this thesis.

Public transit is important when addressing environmental, economic, and social challenges, particularly in the context of climate change and sustainable development. Public transit help reduce greenhouse gas emissions, promote walkable urban forms, lower transportation costs, boost local economies, and enhance equity. However, transit agencies face various challenges such as increasing travel times, funding shortfalls, aging infrastructure, climate-related disruptions, labour shortages, and negative public perceptions, all worsened by the COVID-19 pandemic. The research aims to address these challenges by improving the understanding of transit travel times, integrating automated data sources, and developing tools to enhance service reliability and efficiency to make public transit more attractive for passengers.

The literature review highlights how transit travel time performance and reliability are defined, measured, and perceived, revealing a lack of standardized analysis levels and a gap between agency and passenger perspectives. Reliable travel times are essential for operational efficiency and passenger satisfaction, as unreliability increases costs and reduces ridership. Automated data sources, such as AVL, APC, and AFC systems, can offer more detailed insights into transit operations and passenger behaviours. The literature showed various factors affecting travel times, including congestion, ridership, and signal delays, while signal priority and bus lanes help improve travel times. Gaps remain in understanding more detailed components of travel time, the spatial and temporal variability of reliability, and the equity implications.

The methodology chapter outlines a research framework designed to address the gaps identified in the literature. The study is structured around two general and five specific research questions that compares system-wide spatial and temporal variations at four analysis levels, i.e. inter-stop, stop-to-stop, timepoint-to-timepoint, and route levels. By integrating various automated data sources, GTFS, GTFS-RT, AFC, OpenStreetMap, census data, along with ride-check observations, the methodology matched planned and actual travel time data, analyzed statistical distributions, benchmarked travel time models, attributed variance to specific trip components, and simulated passenger experiences. By leveraging existing data and developing detailed models, the research seeks to deliver actionable insights to enhance service reliability, reduce operational costs, and improve passenger satisfaction in public tran-

sit.

Paper 1 identified mixture travel time distributions across three analysis levels, stop-to-stop, timepoint, and route levels. The presence of mixture distributions are related to temporal, spatial, and operational factors such as time of the day, traffic signals, ridership fluctuations, and land use characteristics. Classification models show that non-linear and similarity-based models perform best. Mixture distributions are more common in dense, high-density areas with more traffic signals and ridership, as well as during peak periods. This paper highlights the research needs for better understanding the complexity of transit travel time distributions and more detailed modelling.

Paper 2 focused on modelling average travel times using both direct and normalized approaches, i.e. travel time, speed, and pace, across four analysis levels, inter-stop, stop-to-stop, timepoint, and route levels. Models were tested using two planning scenarios, expanding service hours and planning new routes, with both aggregated and disaggregated error metrics. Inter-stop level models outperform stop-to-stop models due to reduced influence from dwell times and signal delays. Speed and pace models are more robust for new route planning, while direct travel time models perform better when historical data is available. Models also have different biases given various error measures emphasizing the importance of selecting the right modelling approach based on context and data availability.

Paper 3 decomposes trip travel times into four components, departure time, inter-stop travel time, dwell time, and red light waiting time. Variance-based and one-at-a-time sensitivity analyses quantified the contribution of each component to the overall travel time variations. Departure times interacted with varying signal cycle lengths contributed the most variance for more than half of the routes analyzed. It underscores the importance of signal timing and synchronization, as small changes in departure time or signal settings can lead to large variations in travel time. It also highlights the need to account for detailed data when designing schedules, adjusting traffic signal timing, and identifying sources of unreliability in transit systems.

Paper 4 extends the insights from Paper 3 by developing a scenario-based travel time model incorporating detailed trip components to estimate travel times under various planning contexts. The model reconstructs bus trajectories and are tested against observed data from different service periods. The model can effectively replicate observed travel times and outperform traditional average-based models, especially in scenarios involving changes in operational conditions. Using detailed trip components in travel time modelling provides more flexibility for planners to better adapt to various changes in the system, such as new signal timing plans or schedule adjustments.

Paper 5 shifts the focus to passenger experiences and simulated passenger travel time reliability across the entire service area. Using on-time performance and buffer time as reliability metrics, it shows reliability varies by perspectives, by location, and by time of day. Downtown areas tend to have more reliable service given that it is the focus of the entire transit system, while suburb-to-suburb trips are less dependable especially for off-peak hours and directions. Wait times are generally less reliable than in-vehicle times and overall journey times suffer from compounded delays, highlighting the need to improve wait times in transit planning. Spatial regression analysis showed that access to reliable transit is slightly unevenly distributed across population segments, raising equity concerns.

Overall, the five papers collectively contribute to a multi-level, data-driven understanding of transit travel times, moving from statistical analysis to practical modelling and equity-focused applications. The research demonstrates that higher-level analyses often hide significant local variations, and that detailed, component-level modelling can provide more accurate, targeted, and actionable insights to help improve travel times and reliability. The studies revealed the importance of traffic signals and ridership variations in affecting travel time reliability. The research recommended improving data granularity, cross-agency collaboration, and continuous performance monitoring to better enhance transit services.

REFERENCES

- Abdelhalim, A. and Zhao, J. (2024). Computer vision for transit travel time prediction: an end-to-end framework using roadside urban imagery. *Public Transport*.
- Abdy, Z. R. and Hellinga, B. R. (2011). Analytical method for estimating the impact of transit signal priority on vehicle delay. *Journal of Transportation Engineering*, 137(8):589–600.
- Abkowitz, M. and Engelstein, I. (1984). Methods for maintaining transit service regularity. *Transportation Research Record*, 961:1–8.
- Abkowitz, M., Slavin, H., Waksman, R., Englisher, L. S., Wilson, N. H., et al. (1978). Transit service reliability. Technical report, United States. Urban Mass Transportation Administration.
- Aemmer, Z., Ranjbari, A., and MacKenzie, D. (2022). Measurement and classification of transit delays using GTFS-RT data. *Public Transport*, 14(2):263–285.
- Al-Deek, H., Sandt, A., Alomari, A., and Hussain, O. (2017). A technical note on evaluating the effectiveness of bus rapid transit with transit signal priority. *Journal of Intelligent Transportation Systems*, 21(3):227–238.
- Al-Jumailey, S. I. (2011). Effect of operation policies on dwell time. In *2011 International Conference on Electric Technology and Civil Engineering (ICETCE)*, pages 1194–1196.
- Alam, O., Kush, A., Emami, A., and Pouladzadeh, P. (2021). Predicting irregularities in arrival times for transit buses with recurrent neural networks using GPS coordinates and weather data. *Journal of Ambient Intelligence and Humanized Computing*, 12(7):7813–7826.
- Altun, S. Z. and Furth, P. G. (2009). Scheduling buses to take advantage of transit signal priority. *Transportation Research Record*, 2111(1):50–59.
- Ameijeiras-Alonso, J., Crujeiras, R. M., and Rodríguez-Casal, A. (2019). Mode testing, critical bandwidth and excess mass. *Test*, 28:900–919.
- An, J., Liu, Y., and Yang, X. (2014). Measuring route-level passenger perceived transit service reliability with an agent-based simulation approach. *Transportation Research Record*, 2415(1):48–58.

- Anselin, L. (1988). *Spatial econometrics: methods and models*, volume 4. Springer Science & Business Media.
- Arias, D., Todd, K., Krieger, J., Maddox, S., Haley, P., Watkins, K. E., and Berrebi, S. (2021). Using GTFS to calculate travel time savings potential of bus preferential treatments. *Transportation Research Record*, 2675(9):1643–1654.
- Arriagada, J., Gschwender, A., Munizaga, M. A., and Trépanier, M. (2019). Modeling bus bunching using massive location and fare collection data. *Journal of Intelligent Transportation Systems*, 23(4):332–344.
- Autorité régionale de transport métropolitain (2023). Enquête origine-destination 2023. Technical report.
- Bagherian, M., Cats, O., van Oort, N., and Hickman, M. (2016). Measuring passenger travel time reliability using smart card data. In *TRISTAN IX: Triennial Symposium on Transportation Analysis, Oranjestad, Aruba*.
- Balcombe, R., Mackett, R., Paulley, N., Preston, J., Shires, J., Titheridge, H., Wardman, M., and White, P. (2004). *The demand for public transport: a practical guide*. Transportation Research Laboratory.
- Barabino, B., Lai, C., Casari, C., Demontis, R., and Mozzoni, S. (2017). Rethinking transit time reliability by integrating automated vehicle location data, passenger patterns, and web tools. *IEEE Transactions on Intelligent Transportation Systems*, 18(4):756–766.
- Barker, J., Biehler, A., Brown, L., Clark, W., and Ekern, D. (2008). NCHRP report 616: Multimodal level of service analysis for urban streets (tech. rep.). *Washington, DC: Transportation Research Board*.
- Bates, J., Polak, J., Jones, P., and Cook, A. (2001). The valuation of reliability for personal travel. *Transportation Research Part E: Logistics and Transportation Review*, 37(2):191–229. Advances in the Valuation of Travel Time Savings.
- Bauer, D. and Tulic, M. (2018). Travel time predictions: should one model speeds or travel times? *European Transport Research Review*, 10(2):46.
- Beaubien, G. (2025). Houston light rail commuters experience delays on red line following signal change. *Houston Public Media*.

- Beduhn, T., Khani, A., and Hickman, M. (2015). Using transit ITS data for modeling network reliability and the impact on passenger route choice. In *The 6th International Symposium on Transportation Network Reliability*.
- Berkow, M., El-Geneidy, A. M., Bertini, R. L., and Crout, D. (2009). Beyond generating transit performance measures: Visualizations and statistical analysis with historical data. *Transportation Research Record*, 2111(1):158–168.
- Bertini, R. L. and El-Geneidy, A. (2003). Generating transit performance measures with archived data. *Transportation Research Record*, 1841(1):109–119.
- Bertini, R. L. and El-Geneidy, A. M. (2004). Modeling transit trip time using archived bus dispatch system data. *Journal of Transportation Engineering*, 130(1):56–67.
- Bills, T. S. and Carrel, A. L. (2021). Transit accessibility measurement considering behavioral adaptations to reliability. *Transportation Research Record*, 2675(5):265–278.
- Boudabous, E., Karaa, M., Sboui, L., Montecinos, J., and Alam, O. (2024). Analyzing public transit schedule deviations: A case study on Montreal using real-time data. In *2024 IEEE 27th International Symposium on Real-Time Distributed Computing (ISORC)*, pages 1–6.
- Bowman, L. A. and Turnquist, M. A. (1981). Service frequency, schedule reliability and passenger wait times at transit stops. *Transportation Research Part A: General*, 15(6):465–471.
- Buneman, K. (1984). Automated and passenger-based transit performance measures. *Transportation Research Record*, 992:23–28.
- Carrasco, N. (2012). Quantifying reliability of transit service in Zurich, Switzerland: Case study of bus line 31. *Transportation Research Record*, 2274(1):114–125.
- Carrel, A., Halvorsen, A., and Walker, J. L. (2013a). Passengers’ perception of and behavioral adaptation to unreliability in public transportation. *Transportation Research Record*, 2351(1):153–162.
- Carrel, A., Lau, P. S., Mishalani, R. G., Sengupta, R., and Walker, J. L. (2015). Quantifying transit travel experiences from the users’ perspective with high-resolution smartphone and vehicle location data: Methodologies, validation, and example analyses. *Transportation Research Part C: Emerging Technologies*, 58:224–239. Big Data in Transportation and Traffic Engineering.

- Carrel, A., Mishalani, R. G., Wilson, N. H., and Attanucci, J. P. (2013b). A framework for evaluating operations control on a metro line: integrating multiple perspectives and automatically collected train and passenger movement data. *Public Transport*, 5(3):149–176.
- Carrel, A. and Walker, J. L. (2017). Understanding future mode choice intentions of transit riders as a function of past experiences with travel quality. *European Journal of Transport and Infrastructure Research*, 17(3).
- Casello, J. M., Nour, A., and Hellinga, B. (2009). Quantifying impacts of transit reliability on user costs. *Transportation Research Record*, 2112(1):136–141.
- Cats, O. and Gkioulou, Z. (2017). Modeling the impacts of public transport reliability and travel information on passengers’ waiting-time uncertainty. *EURO Journal on Transportation and Logistics*, 6(3):247–270.
- Ceder, A. (1987). Methods for creating bus timetables. *Transportation Research Part A: General*, 21(1):59–83.
- Ceder, A. A., Hadas, Y., McIvor, M., and Ang, A. (2013). Transfer synchronization of public transport networks. *Transportation Research Record*, 2350(1):9–16.
- Ceriani, L. and Verme, P. (2012). The origins of the gini index: extracts from *variabilità e mutabilità* (1912) by corrado gini. *The Journal of Economic Inequality*, 10(3):421–443.
- Chakrabarti, S. (2017). How can public transit get people out of their cars? an analysis of transit mode choice for commute trips in Los Angeles. *Transport Policy*, 54:80–89.
- Chen, C., Skabardonis, A., and Varaiya, P. (2003). Travel-time reliability as a measure of service. *Transportation Research Record*, 1855(1):74–79.
- Chen, G., Huang, L., Liu, H., and Yang, X. (2013). Effects of bunching at stops on bus dwell time. In *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, pages 1499–1502.
- Chen, X., Cheng, Z., Jin, J. G., Trépanier, M., and Sun, L. (2023). Probabilistic forecasting of bus travel time with a Bayesian Gaussian mixture model. *Transportation Science*, 57(6):1516–1535.
- Coleman, M., Tarte, L., Chau, S., Levine, B., and Reddy, A. (2018). A data-driven approach to prioritizing bus schedule revisions at New York City Transit. *Transportation Research Record*, 2672(8):86–95.

- Cortés, C. E., Gibson, J., Gschwender, A., Munizaga, M., and Zúñiga, M. (2011). Commercial bus speed diagnosis based on GPS-monitored data. *Transportation Research Part C: Emerging Technologies*, 19(4):695–707.
- Cover, T. and Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27.
- Currie, G., Delbosch, A., Harrison, S., and Sarvi, M. (2013). Impact of crowding on streetcar dwell time. *Transportation Research Record*, 2353(1):100–106.
- Currie, G. and Mesbah, M. (2011). Visualization of geographical information system and automatic vehicle location data to explore transit performance. *Transportation Research Record*, 2216(1):59–66.
- Czitrom, V. (1999). One-factor-at-a-time versus designed experiments. *The American Statistician*, 53(2):126–131.
- Danaher, A., Wensley, J., Dunham, A., Orosz, T., Avery, R., Cobb, K., Watkins, K., Queen, C., Berrebi, S., Connor, M., et al. (2020). *Minutes Matter: A Bus Transit Service Reliability Guidebook*. Washington, DC: The National Academies Press.
- Daskalakis, N. G. and Stathopoulos, A. (2008). Users’ perceptive evaluation of bus arrival time deviations in stochastic networks. *Journal of Public Transportation*, 11(4):25–38.
- Diab, E. I. and El-Geneidy, A. M. (2013). Variation in bus transit service: understanding the impacts of various improvement strategies on transit service reliability. *Public Transport*, 4(3):209–231.
- Diab, E. I. and El-Geneidy, A. M. (2015). The farside story: Measuring the benefits of bus stop location on transit performance. *Transportation Research Record*, 2538(1):1–10.
- Ding, J., Yang, M., Wang, W., Xu, C., and Bao, Y. (2015). Strategy for multiobjective transit signal priority with prediction of bus dwell time at stops. *Transportation Research Record*, 2488(1):10–19.
- Dion, F., Rakha, H., and Zhang, Y. (2005). Integration of transit signal priority within adaptive traffic signal control systems. In *84th Annual Meeting of the Transportation Research Board, Washington, DC*.
- Dixit, M., Brands, T., van Oort, N., Cats, O., and Hoogendoorn, S. (2019). Passenger travel time reliability for multimodal public transport journeys. *Transportation Research Record*, 2673(2):149–160.

- D'Souza, C., Gardner, K., Hounsell, N., and Shrestha, B. (2010). New developments for bus priority at traffic signals in London using iBus. In *IET Road Transport Information and Control Conference and the ITS United Kingdom Members' Conference (RTIC 2010). Better transport through technology*, page 72.
- Du, Y., Deng, F., Liao, F., and Ji, Y. (2017). Understanding the distribution characteristics of bus speed based on geocoded data. *Transportation Research Part C: Emerging Technologies*, 82:337–357.
- Duddu, V. R., Pulugurtha, S. S., and Najaf, P. (2019). Using link-level archived automatic vehicle location data to assess transit system LOS at bus-stop level. *Transportation in Developing Economies*, 5(2):1–9.
- Dueker, K. J., Kimpel, T. J., Strathman, J. G., and Callas, S. (2004). Determinants of bus dwell time. *Journal of Public Transportation*, 7(1):21–40.
- El-Geneidy, A., Horning, J., and Krizek, K. (2007). *Using archived ITS data to improve transit performance and management*. Minnesota Department of Transportation.
- El-Geneidy, A. M., Horning, J., and Krizek, K. J. (2011). Analyzing transit service reliability using detailed data from automatic vehicular locator systems. *Journal of Advanced Transportation*, 45(1):66–79.
- El-Geneidy, A. M. and Vijayakumar, N. (2011). The effects of articulated buses on dwell and running times. *Journal of Public Transportation*, 14(3):63–86.
- Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, number 34, pages 226–231. AAAI Press.
- Fayazi, S. A., Vahidi, A., Mahler, G., and Winckler, A. (2015). Traffic signal phase and timing estimation from low-frequency transit bus data. *IEEE Transactions on Intelligent Transportation Systems*, 16(1):19–28.
- Ferrenchik, M. (2019). Test finds bike lane helps cota buses downtown. *The Columbus Dispatch*.
- Furth, P. G. (2000). *Data analysis for bus planning and monitoring*. Transportation Research Board.

- Furth, P. G. and Muller, T. H. J. (2006). Service reliability and hidden waiting time: Insights from automatic vehicle location data. *Transportation Research Record*, 1955(1):79–87.
- Furth, P. G. and Muller, T. H. J. (2007). Service reliability and optimal running time schedules. *Transportation Research Record*, 2034(1):55–61.
- Furth, P. G., Saeidi-Razavi, R., Obeng-Amoako, N. D., and Tahmasebi, M. (2025). Tsp-friendly underlying traffic signal control: An essential complement to transit signal priority. volume 5.
- Gallo, F. and Sacco, N. (2022). On optimizing bus dwell times to reduce the probability of stopping for a red light at intersections. In *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, pages 1637–1644.
- Gittens, A. and Shalaby, A. (2015). Evaluation of bus reliability measures and development of a new composite indicator. *Transportation Research Record*, 2533(1):91–99.
- Godfrey, J. et al. (2024). Transit workforce development challenges and mitigation practices. Technical report, San Jose State University. College of Business. Mineta Transportation Institute.
- Goodwin, P. and Lawton, R. (1999). On the asymmetry of the symmetric MAPE. *International Journal of Forecasting*, 15(4):405–408.
- Gouvernement du Québec (2021). Inventaire québécois des émissions atmosphériques.
- Graves, E., Zheng, S., Tarte, L., Levine, B., and Reddy, A. (2019). Customer journey time metrics for New York City bus service using big data. *Transportation Research Record*, 2673(9):1–10.
- Gurmu, Z. K. and Fan, W. (2014). Artificial neural network travel time prediction model for buses using only GPS data. *Journal of Public Transportation*, 17(2):45–65.
- Hajjem, A., Bellavance, F., and Larocque, D. (2014). Mixed-effects random forest for clustered data. *Journal of Statistical Computation and Simulation*, 84(6):1313–1328.
- Halvorsen, A., Wood, D., Stasko, T., Jefferson, D., and Reddy, A. (2019). Passenger-centric performance metrics for the New York City Subway. *Transportation Research Record*, 2673(1):417–426.
- Han, Q., Liu, K., Zeng, L., He, G., Ye, L., and Li, F. (2020). A bus arrival time prediction method based on position calibration and LSTM. *IEEE Access*, 8:42372–42383.

- Handley, J. C., Fu, L., and Tupper, L. L. (2019). A case study in spatial-temporal accessibility for a transit system. *Journal of Transport Geography*, 75:25–36.
- Hartigan, J. A. and Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics*, 13(1):70–84.
- He, L. and Trépanier, M. (2015). Estimating the destination of unlinked trips in transit smart card fare data. *Transportation Research Record*, 2535(1):97–104.
- Hendren, P., Antos, J., Carney, Y., and Harcum, R. (2015). Transit travel time reliability: Shifting the focus from vehicles to customers. *Transportation Research Record*, 2535(1):35–44.
- Ho, T. K. (1995). Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*, volume 1, pages 278–282 vol.1.
- Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied logistic regression*, volume 398. John Wiley & Sons.
- Hox, J., Moerbeek, M., and Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. Routledge.
- Huo, Y., Zhao, J., Li, W., and Hu, X. (2014). Measuring bus service reliability: An example of bus rapid transit in Changzhou. *Journal of Public Transportation*, 17(2):113–133.
- Isukapati, I. K., Rudová, H., Barlow, G. J., and Smith, S. F. (2017). Analysis of trends in data on transit bus dwell times. *Transportation Research Record*, 2619(1):64–74.
- Jang, W. (2010). Travel time and transfer analysis using transit smart card data. *Transportation Research Record*, 2144(1):142–149.
- Jara-Díaz, S. and Tirachini, A. (2013). Urban bus transport: open all doors for boarding. *Journal of Transport Economics and Policy (JTEP)*, 47(1):91–106.
- Jenelius, E. (2018). Public transport experienced service reliability: Integrating travel time and travel conditions. *Transportation Research Part A: Policy and Practice*, 117:275–291.
- Kamga, C. and Eickemeyer, P. (2021). Slowing the spread of COVID-19: Review of “social distancing” interventions deployed by public transit in the united states and canada. *Transport Policy*, 106:25–36.

- Kathuria, A., Sekhar, C. R., and Parida, M. (2018). Examining failure rate dynamics to estimate capacity of bus rapid transit system corridor. *Proceedings of the Institution of Civil Engineers - Transport*, 173(5):333–344.
- Kieu, L. M. (2015). *Utilising big transit data for transfer coordination*. PhD thesis, Queensland University of Technology.
- Kieu, L.-M., Bhaskar, A., and Chung, E. (2015). Public transport travel-time variability definitions and monitoring. *Journal of Transportation Engineering*, 141(1):04014068.
- Kieu, L. M., Bhaskar, A., Cools, M., and Chung, E. (2017). An investigation of timed transfer coordination using event-based multi agent simulation. *Transportation Research Part C: Emerging Technologies*, 81:363–378.
- Kimpel, T. J., Strathman, J., Bertini, R. L., and Callas, S. (2005). Analysis of transit signal priority using archived TriMet bus dispatch system data. *Transportation Research Record*, 1925(1):156–166.
- Kimpel, T. J., Strathman, J. G., and Dueker, K. (2000). Time point-level analysis of passenger demand and transit service reliability.
- King County Metro (2021). Transit speed and reliability guidelines and strategies.
- Kolassa, S. (2020). Why the “best” point forecast depends on the error or accuracy measure. *International Journal of Forecasting*, 36(1):208–211.
- Kopsacheilis, A., Politis, I., and Georgiadis, G. (2023). Assessment of bus speed influencing factors through the exploitation of machine learning techniques. *Transportation Research Procedia*, 69:751–758.
- Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974.
- Lamorgese, L. and Mannino, C. (2015). An exact decomposition approach for the real-time train dispatching problem. *Operations Research*, 63(1):48–64.
- Le, H. T. and Carrel, A. L. (2021). Happy today, satisfied tomorrow: emotion—satisfaction dynamics in a multi-week transit user smartphone survey. *Transportation*, 48(1):45–66.
- Lee, J. and Kim, J. (2023). Social equity analysis of public transit accessibility to healthcare might be erroneous when travel time uncertainty impacts are overlooked. *Travel Behaviour and Society*, 32:100588.

- Levine, J. C. and Torng, G. (1994). Dwell-time effects of low-floor bus design. *Journal of Transportation Engineering*, 120(6):914–929.
- Levinson, H. (1991). Synthesis of transit practice 15: Supervision strategies for improved reliability of bus routes. *TRB, National Research Council, Washington, DC*.
- Li, J., Wang, W., van Zuylen, H. J., Sze, N. N., Chen, X., and Wang, H. (2012). Predictive strategy for transit signal priority at fixed-time signalized intersections: Case study in Nanjing, China. *Transportation Research Record*, 2311(1):124–131.
- Li, T., Meredith-Karam, P., Kong, H., Stewart, A., Attanucci, J. P., and Zhao, J. (2021). Comparison of door-to-door transit travel time estimation using schedules, real-time vehicle arrivals, and smartcard inference methods. *Transportation Research Record*, 2675(11):1003–1014.
- Lin, Y., Yang, X., Zou, N., and Franz, M. (2015). Transit signal priority control at signalized intersections: a comprehensive review. *Transportation Letters*, 7(3):168–180.
- Lin, Y., Yang, X., Zou, n., and Jia, L. (2013). New passive transit signal priority control strategy for the bus vehicles at urban arteries. *Journal of Northeastern University (Natural Science)*, 34(9):1227.
- Litman, T. (2015). *Evaluating public transit benefits and costs*. Victoria Transport Policy Institute Victoria, BC, Canada.
- Liu, L. and Miller, H. J. (2021). Measuring risk of missing transfers in public transit systems using high-resolution schedule and real-time bus location data. *Urban Studies*, 58(15):3140–3156.
- Liu, L., Porr, A., and Miller, H. J. (2023). Realizable accessibility: evaluating the reliability of public transit accessibility using high-resolution real-time data. *Journal of Geographical Systems*, 25(3):429–451.
- Liu, R. and Sinha, S. (2007). Modelling urban bus service and passenger reliability. Copyright held by the authors and the Institute of Transport Studies, University of Leeds.
- Loukaitou-Sideris, A., Wasserman, J., Ding, H., and Caro, R. (2023). “it is our problem!”: Strategies for responding to homelessness on transit. *Transportation Research Record*, 2677(2):1200–1214.

- Luethi, M., Weidmann, U., and Nash, A. (2007). Passenger arrival rates at public transport stations. In *TRB 86th Annual Meeting Compendium of Papers*, pages 07–0635. Transportation Research Board.
- Ma, J., Chan, J., Ristanoski, G., Rajasegarar, S., and Leckie, C. (2019). Bus travel time prediction with real-time traffic information. *Transportation Research Part C: Emerging Technologies*, 105:536–549.
- Ma, X. and Wang, Y. (2014). Development of a data-driven platform for transit performance measures using smart card and GPS data. *Journal of Transportation Engineering*, 140(12):04014063.
- Ma, Z., Ferreira, L., and Mesbah, M. (2013). A framework for the development of bus service reliability measures. In *Proceedings of the 36th Australasian Transport Research Forum (ATRF), Brisbane, Australia*.
- Ma, Z., Ferreira, L., and Mesbah, M. (2014). Measuring service reliability using automatic vehicle location data. *Mathematical Problems in Engineering*, 2014(1):468563.
- Mai, E., List, G., and Hranac, R. (2012). Simulating the travel time impact of missed transit connections. *Transportation Research Record*, 2274(1):69–76.
- Mandelzys, M. and Hellinga, B. (2010). Identifying causes of performance issues in bus schedule adherence with automatic vehicle location and passenger count data. *Transportation Research Record*, 2143(1):9–15.
- Mandelzys, M., Hellinga, B., and Eng, P. (2010). Automatically identifying the causes of bus transit schedule adherence performance issues using AVL/APC archived data. *Transportation Research Board of the National Academies*.
- Mannino, C. and Nakkerud, A. (2023). Optimal train rescheduling in Oslo central station. *Omega*, 116:102796.
- Martens, K., Golub, A., and Robinson, G. (2012). A justice-theoretic approach to the distribution of transportation benefits: Implications for transportation planning practice in the united states. *Transportation research part A: policy and practice*, 46(4):684–695.
- Massachusetts Bay Transportation Authority (2017). Variance of bus dwell times due to fare media.
- Mazloumi, E., Currie, G., and Rose, G. (2010). Using GPS data to gain insight into public transport travel time variability. *Journal of Transportation Engineering*, 136(7):623–631.

- McLeod, F. and Hounsell, N. (2003). Bus priority at traffic signals—evaluating strategy options. *Journal of Public Transportation*, 6(3):1–14.
- Morpace International, I. (1999). TCRP report 47: A handbook for measuring customer satisfaction and service quality.
- Muller, T. H. and Furth, P. G. (2001). Trip time analyzers: Key to transit service quality. *Transportation Research Record*, 1760(1):10–19.
- National Academies of Sciences, Engineering, and Medicine (2006). *Using archived AVL-APC data to improve transit performance and management*. Washington, DC: The National Academies Press.
- National Academies of Sciences, Engineering, and Medicine and others (2013). *Transit Capacity and Quality of Service Manual, Third Edition*. Washington, DC: The National Academies Press.
- Ngan, V., Sayed, T., and Abdelfatah, A. (2004). Impacts of various parameters on transit signal priority effectiveness. *Journal of Public Transportation*, 7(3):71–93.
- Noland, R. B. and Polak, J. W. (2002). Travel time variability: A review of theoretical and empirical issues. *Transport Reviews*, 22(1):39–54.
- OpenStreetMap contributors (2023). Planet dump retrieved from <https://planet.osm.org> . <https://www.openstreetmap.org>.
- Palm, M. and Farber, S. (2024). 6 prospects for public transit equity in the twenty-first century: Lessons from COVID-19’s impacts on Canada’s largest metropolis. *Urban Mobility*, page 101.
- Parbo, J., Nielsen, O. A., and Prato, C. G. (2014). User perspectives in public transport timetable optimisation. *Transportation Research Part C: Emerging Technologies*, 48:269–284.
- Perk, V. A., Flynn, J. A., Volinski, J., et al. (2008). Transit ridership, reliability, and retention. Technical report, National Center for Transit Research (US).
- Quinlan, J. (1987). Simplifying decision trees. *International Journal of Man-Machine Studies*, 27(3):221–234.
- Rashidi, S., Ataeian, S., and Ranjitkar, P. (2023). Estimating bus dwell time: A review of the literature. *Transport Reviews*, 43(1):32–61.

- Russo, A., Adler, M. W., and van Ommeren, J. N. (2022). Dedicated bus lanes, bus speed and traffic congestion in Rome. *Transportation Research Part A: Policy and Practice*, 160:298–310.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008). *Global sensitivity analysis: the primer*. John Wiley & Sons.
- Schmitt, A. (2014). The ridiculous politics that slow down America’s best BRT route.
- Shalaby, A. and Farhan, A. (2003). Bus travel time prediction model for dynamic operations control and passenger information systems. *Transportation Research Board*, 2.
- Sheffield, M. H., Schultz, G. G., Bassett, D., and Eggett, D. L. (2021). Sensitivity analysis of the transit signal priority requesting threshold and the impact on bus performance and general traffic. *Transportation Research Record*, 2675(5):149–163.
- Siddiq, F., Wasserman, J. L., Taylor, B. D., and Speroni, S. (2023). Transit’s financial prognosis: Findings from a survey of US transit systems during the COVID-19 pandemic. *Public Works Management & Policy*, 28(4):393–415.
- Smith, H. R., Hemily, B., and Ivanovic, M. (2005). Transit signal priority (TSP): A planning and implementation handbook.
- Strathman, J. G., Dueker, K. J., Kimpel, T., Gerhart, R., Turner, K., Taylor, P., Callas, S., Griffin, D., and Hopper, J. (1999). Automated bus dispatching, operations control, and service reliability: Baseline analysis. *Transportation Research Record*, 1666(1):28–36.
- Transport for London (2022a). QSI Performance Results Route Results for London Bus Services Third Quarter 2021-22.
- Transport for London (2022b). Underground services performance.
- Transportation Research Board and National Academies of Sciences, Engineering, and Medicine (2006). *Fixed-Route Transit Ridership Forecasting and Service Planning Methods*. The National Academies Press, Washington, DC.
- Traut, E. J. and Steinfeld, A. (2019). Identifying commonly used and potentially unsafe transit transfers with crowdsourcing. *Transportation Research Part A: Policy and Practice*, 122:99–111.
- Trépanier, M., Morency, C., and Agard, B. (2009). Calculation of transit performance measures using smartcard data. *Journal of Public Transportation*, 12(1):79–96.

Turnquist, M. A., Blume, S. W., et al. (1980). Evaluating potential effectiveness of headway control strategies for transit systems. *Transportation Research Record*, 746(1):25–29.

Uniman, D. L., Attanucci, J., Mishalani, R. G., and Wilson, N. H. M. (2010). Service reliability measurement using automated fare card data: Application to the London Underground. *Transportation Research Record*, 2143(1):92–99.

Urbanik, T., Tanaka, A., Lozner, B., Lindstrom, E., Lee, K., Quayle, S., Beaird, S., Tsoi, S., Ryus, P., Gettman, D., et al. (2015). *Signal timing manual*, volume 1. Transportation Research Board Washington, DC.

U.S. Department of Transportation (2024). Status of the Nation’s Highways, Bridges, and Transit: Conditions & Performance. 25rd Report to Congress, Washington DC.

van Exel, N. J. A. and Rietveld, P. (2010). Perceptions of public transport travel time and their effect on choice-sets among car drivers. *Journal of Transport and Land Use*, 2(3/4):75–86.

Vasudevan, M. (2005). Robust optimization model for bus priority control under arterial progression. *ProQuest Dissertations and Theses*, page 174.

Vlachou, K., Collura, J., and Mermelstein, A. (2010). Planning and deploying transit signal priority in small and medium-sized cities: Burlington, Vermont, case study. *Journal of Public Transportation*, 13(3):101–123.

Vuchic, V. R. (2017). *Urban transit: operations, planning, and economics*. John Wiley & Sons.

Walker, J. (2012). *Human Transit: How clearer thinking about public transit can enrich our communities and our lives*. Island Press.

Wang, Y. (2020). A methodology for identifying inconsistencies between scheduled and observed travel and transfer times using transit avl data: Framework and case study of Columbus, OH.

Wang, Y., Carrel, A. L., and Robert, W. (2021). Disaggregating variability in projected benefits of transit state of good repair due to uncertainty of travel demand forecasts. *Journal of Transportation Engineering, Part A: Systems*, 147(11):04021081.

Wang, Y., Morency, C., and Trépanier, M. (2025). Decomposition and sensitivity analysis of bus travel times. Manuscript accepted for publication.

- Wardman, M. (2004). Public transport values of time. *Transport Policy*, 11(4):363–377.
- Wardman, M. and Whelan, G. (2011). Twenty years of rail crowding valuation studies: Evidence and lessons from british experience. *Transport Reviews*, 31(3):379–398.
- Webb, A., Kumar, P., and Khani, A. (2020). Estimation of passenger waiting time using automatically collected transit data. *Public Transport*, 12(2):299–311.
- Weitz, R. (2008). Who’s afraid of the big bad bus? NIMBYism and popular images of public transit. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability*, 1(2):157–172.
- Wessel, N., Allen, J., and Farber, S. (2017). Constructing a routable retrospective transit timetable from a real-time vehicle location feed and GTFS. *Journal of Transport Geography*, 62:92–97.
- Wessel, N. and Widener, M. J. (2017). Discovering the space-time dimensions of schedule padding and delay from GTFS and real-time transit data. *Journal of Geographical Systems*, 19(1):93–107.
- Westinghouse Air Brake Company and United States Dept. of Housing (1968). *Study of Evolutionary Urban Transportation*, volume 3. US Department of Housing and Urban Development.
- Wu, K., Lu, M., and Guler, S. I. (2020). Modeling and optimizing bus transit priority along an arterial: A moving bottleneck approach. *Transportation Research Part C: Emerging Technologies*, 121:102873.
- Xin, J. and Chen, S. (2016). Bus dwell time prediction based on KNN. *Procedia Engineering*, 137:283–288. Green Intelligent Transportation System and Safety.
- Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M., and Price, A. L. (2014). Advantages and pitfalls in the application of mixed-model association methods. *Nature Genetics*, 46(2):100–106.
- Zhang, L., Weng, J., and Chen, Z. (2014). Characteristic analysis of bus travel speed on commuting corridors based on GPS data. In *CICTP 2014: Safe, Smart, and Sustainable Multimodal Transportation Systems*, pages 1443–1453. ACSE.
- Zhang, S., Fu, H., and Huang, C. (2020). Estimating bus speed distribution of bimodal traffic using vine copula. In *2020 IEEE 5th International Conference on Intelligent Transportation Engineering (ICITE)*, pages 133–140.

Zhao, J., Frumin, M., Wilson, N., and Zhao, Z. (2013). Unified estimator for excess journey time under heterogeneous passenger incidence behavior using smartcard data. *Transportation Research Part C: Emerging Technologies*, 34:70–88.

Zhou, G. and Gan, A. (2005). Performance of transit signal priority with queue jumper lanes. *Transportation Research Record*, 1925(1):265–271.

Ziedan, A., Brakewood, C., and Watkins, K. (2023). Will transit recover? a retrospective study of nationwide ridership in the united states during the COVID-19 pandemic. *Journal of Public Transportation*, 25:100046.

APPENDIX A CORRELATION MATRIX FOR MODEL INPUTS

Due to the length limits for the original papers, there wasn't enough space to demonstrate the correlations for the variables in the modelling steps. Therefore, the correlation matrices for Papers 1, 2, and 5 are shown here in the appendix for reference. Although there are some general correlations between variables, the correlations are still within acceptable ranges for modelling, given the variety of routes, operational, and streetscape characteristics included in the study. A few coefficients higher than 0.5 or lower than -0.5 are underlined in the tables. Higher analysis levels have more higher coefficients that are barely within acceptable ranges, which could have an impact on the model results. Future researchers should check for potential multicollinearity issues in their analyses.

Table A.1 Correlation Matrix for Paper 1 Stop Level

	Sat.	Sun.	E.AM	A.Peak	Eve.	Late	A.Load	SD.Load	Freq.	A.Delay	SD.Delay	%Board	Turns	Signals	Lanes	Spd	BusLane	Oneway	L-Res.	L-Col.	L-Sec.	L-Pri.	L-Hwy	Dense	Mix	Centre	Res.	
Sunday	-0.47																											
Early AM	0.01	-0.01																										
AM Peak	-0.01	-0.01	-0.19																									
Evening	0.00	0.01	-0.16	-0.20																								
Late	0.00	0.01	-0.12	-0.15	-0.13																							
Avg.Load	-0.06	-0.12	-0.24	0.03	-0.09	-0.12																						
SD.Load	-0.03	-0.16	-0.28	-0.01	-0.06	-0.12	<u>0.64</u>																					
Freq.	-0.03	-0.14	-0.22	0.01	-0.03	-0.08	0.38	0.43																				
Avg.Delay	0.00	-0.02	0.12	-0.05	-0.01	0.00	-0.10	-0.06	-0.09																			
SD.Delay	-0.01	-0.06	0.15	-0.05	-0.01	0.02	-0.10	-0.07	-0.08	<u>0.56</u>																		
% Board	-0.03	-0.07	-0.25	0.03	-0.04	-0.09	0.44	0.45	0.23	-0.06	-0.11																	
Turns	-0.01	-0.02	-0.02	0.01	0.00	-0.01	-0.07	-0.06	-0.08	0.00	-0.01	-0.06																
Signals	-0.01	0.00	0.04	-0.01	-0.01	-0.01	0.08	0.09	0.01	0.00	-0.03	0.25	0.30															
Lanes	0.00	0.00	0.04	-0.01	-0.01	0.00	0.06	0.06	0.07	0.01	0.04	0.08	-0.04	0.23														
Speed	-0.01	-0.01	0.05	0.00	-0.01	-0.01	0.08	0.08	0.09	0.00	0.03	0.05	0.06	0.17	0.37													
BusLane	-0.11	-0.11	-0.02	0.04	-0.02	-0.02	0.11	0.11	0.17	0.01	0.05	0.07	-0.01	0.08	0.12	0.10												
Oneway	0.00	0.00	0.04	-0.01	-0.01	-0.01	0.07	0.06	0.07	0.00	0.01	0.10	0.08	0.21	-0.04	0.38	0.12											
L-Residential	0.00	0.00	-0.06	0.01	0.01	0.01	-0.08	-0.09	-0.09	-0.01	-0.02	-0.11	0.09	-0.26	-0.38	-0.45	-0.07	-0.22										
L-Collector	0.00	0.00	-0.04	0.00	0.01	0.00	-0.05	-0.06	0.00	0.01	0.02	-0.09	-0.04	-0.15	-0.14	-0.05	-0.04	-0.19	-0.31									
L-Secondary	0.00	0.01	0.06	-0.01	-0.01	0.00	0.06	0.07	0.04	-0.01	-0.02	0.15	-0.08	0.26	0.29	0.19	0.07	0.23	-0.38	<u>-0.55</u>								
L-Primary	0.00	0.00	0.03	0.00	-0.01	0.00	0.10	0.09	0.06	0.01	0.01	0.09	-0.05	0.15	0.27	0.26	0.04	0.19	-0.14	-0.21	-0.25							
L-Highway	-0.01	-0.01	0.01	0.00	0.00	0.00	0.02	0.03	-0.01	0.00	0.01	-0.01	0.30	0.11	-0.02	0.32	0.00	0.12	-0.05	-0.06	-0.03	-0.03						
Density	0.02	0.02	0.04	-0.01	0.00	0.00	0.06	0.08	0.00	-0.01	-0.08	0.31	-0.10	0.26	0.02	-0.22	-0.01	-0.04	-0.03	-0.10	0.18	0.01	-0.09					
Mixed	-0.01	-0.01	0.00	0.00	0.00	0.00	0.03	0.03	0.01	-0.01	0.00	0.04	0.23	0.28	0.07	0.19	0.05	0.16	-0.06	-0.06	0.06	0.02	0.14	-0.11				
Centre	0.00	0.00	0.02	-0.01	0.00	0.00	-0.05	-0.03	-0.02	0.01	-0.01	0.01	0.21	0.25	0.05	0.06	0.01	0.07	-0.05	-0.03	0.03	0.02	0.20	0.13	0.02			
Residential	-0.01	-0.01	0.00	0.00	0.00	-0.01	0.01	0.02	0.00	-0.01	0.00	0.04	0.40	0.35	-0.01	0.09	0.00	0.02	0.02	-0.01	-0.02	-0.02	0.33	0.01	-0.02	-0.05		
Industrial	-0.01	-0.01	0.00	0.00	0.00	-0.01	0.02	0.03	-0.02	0.00	0.01	-0.05	0.47	0.34	0.01	0.18	0.02	0.06	-0.04	-0.02	-0.03	0.00	0.35	-0.15	0.21	0.07	0.44	

Table A.2 Correlation Matrix for Paper 1 Timepoint Level

	Sat	Sun	E.AM	A.Peak	P.Peak	Eve.	Late	A Load	SD.Load	Freq.	A.Delay	SD.Delay	% Board	Turns	Signals	Lanes	Spd	BusLane	Oneway	L-Res.	L-Col.	L-Sec.	L-Pri	L-Hwy	Mix	Centre	Indu.	Density	
Sunday	-0.44																												
Early AM	0.02	-0.01																											
AM Peak	-0.02	-0.02	-0.17																										
PM Peak	-0.02	-0.01	-0.17	-0.25																									
Evening	0.01	0.01	-0.13	-0.19	-0.20																								
Late	0.01	0.02	-0.10	-0.15	-0.15	-0.12																							
Avg Load	-0.04	-0.17	-0.23	0.02	0.24	-0.09	-0.14																						
SD Load	-0.06	-0.15	-0.20	0.04	0.22	-0.11	-0.15	<u>0.68</u>																					
Frequency	-0.01	-0.12	-0.21	0.02	0.16	-0.04	-0.08	<u>0.51</u>	0.50																				
Avg Delay	0.12	0.02	0.05	-0.13	0.04	0.03	0.04	-0.08	-0.10	-0.11																			
SD Delay	0.02	-0.06	0.05	-0.07	0.00	0.01	0.04	-0.05	-0.06	-0.08	<u>0.62</u>																		
% Board	-0.07	-0.14	-0.29	0.09	0.20	-0.08	-0.19	<u>0.66</u>	<u>0.67</u>	0.47	-0.16	-0.13																	
Turns	-0.02	-0.02	0.06	0.00	-0.01	-0.01	-0.01	-0.13	-0.10	-0.13	0.09	0.09	-0.25																
Signals	0.00	0.01	0.22	-0.04	-0.05	-0.03	-0.02	-0.01	0.00	0.02	0.10	-0.02	0.12	0.17															
Lanes	0.01	0.01	0.02	0.00	-0.01	0.00	0.01	0.09	0.11	0.11	-0.03	0.02	0.15	-0.17	0.18														
Speed	-0.04	-0.04	0.04	0.01	0.01	-0.02	-0.02	0.11	0.05	0.05	-0.11	0.04	-0.01	0.10	0.04	0.23													
BusLane	-0.19	-0.19	-0.06	0.11	0.09	-0.05	-0.04	0.17	0.16	0.19	-0.08	0.06	0.17	-0.07	0.02	0.14	0.14												
Oneway	-0.02	-0.02	0.06	-0.01	0.00	-0.02	-0.02	0.08	0.03	0.03	-0.08	-0.02	0.09	-0.08	0.12	-0.11	<u>0.51</u>	0.13											
L-Residential	0.02	0.01	-0.06	0.00	0.01	0.01	0.01	-0.10	-0.07	-0.09	0.00	0.00	-0.09	0.22	-0.27	-0.20	-0.41	-0.10	-0.24										
L-Collector	0.01	0.00	-0.05	0.00	0.01	0.01	0.00	-0.06	-0.01	-0.08	0.02	0.03	-0.15	0.09	-0.18	-0.28	-0.17	-0.05	-0.21	-0.08									
L-Secondary	0.00	0.01	0.06	-0.01	-0.02	-0.01	0.00	-0.04	-0.03	0.04	0.00	-0.05	0.14	-0.18	0.32	0.12	-0.02	0.04	0.10	-0.34	<u>-0.52</u>								
L-Primary	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.14	0.16	0.12	-0.01	-0.03	0.15	-0.22	0.12	0.36	0.05	0.04	0.10	-0.09	-0.24	-0.37							
L-Highway	-0.02	-0.02	0.03	0.00	0.00	-0.01	-0.01	0.08	-0.03	0.00	-0.05	0.02	-0.02	0.18	-0.09	0.02	<u>0.63</u>	0.04	0.37	-0.22	-0.12	-0.12	-0.12						
Mixed	-0.02	-0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.03	-0.10	-0.03	0.03	-0.03	-0.01	0.06	0.19	0.05	0.22	-0.02	-0.06	0.09	-0.06	-0.01					
Centre	0.00	0.01	0.01	-0.01	0.00	0.00	-0.01	-0.14	-0.16	-0.02	0.13	-0.02	-0.08	-0.07	0.16	0.05	-0.13	-0.01	0.03	-0.12	-0.01	0.07	0.03	0.06	-0.23				
Industrial	-0.05	-0.05	0.00	0.02	0.02	-0.02	-0.02	0.03	-0.01	-0.07	-0.04	0.08	-0.16	0.26	-0.15	0.00	0.47	0.11	0.21	-0.11	-0.02	-0.15	-0.03	0.40	0.09	-0.20			
Pop. Density	0.03	0.03	0.02	-0.02	-0.02	0.00	0.01	0.00	0.00	0.11	0.07	-0.10	0.26	-0.35	0.17	0.03	<u>-0.50</u>	-0.06	-0.18	0.10	-0.11	0.24	0.06	-0.28	-0.19	0.28	<u>-0.58</u>		
Num Stops	0.02	0.03	0.16	-0.03	-0.04	-0.01	0.00	-0.03	0.06	-0.03	0.07	0.03	-0.09	<u>0.52</u>	<u>0.61</u>	0.04	-0.04	-0.08	-0.12	0.00	0.09	0.05	0.00	-0.24	-0.03	-0.15	-0.03	-0.18	

Table A.3 Correlation Matrix for Paper 1 Route Level

	Sat	Sun	E.AM	A.Peak	P.Peak	Eve.	Late	A Load	SD.Load	Freq.	A.Delay	SD.Delay	% Board	Turns	Signals	Lanes	Spd	BusLane	Oneway	L-Res.	L-Col.	L-Sec.	L-Pri	L-Hwy	Mix	Centre	Indu.	Density		
Sunday	-0.44																													
Early AM	0.02	-0.01																												
AM Peak	-0.02	-0.02	-0.17																											
PM Peak	-0.02	-0.01	-0.17	-0.25																										
Evening	0.01	0.01	-0.13	-0.19	-0.20																									
Late	0.01	0.02	-0.10	-0.15	-0.15	-0.12																								
Avg Load	-0.04	-0.17	-0.23	0.02	0.24	-0.09	-0.14																							
SD Load	-0.06	-0.15	-0.20	0.04	0.22	-0.11	-0.15	<u>0.68</u>																						
Frequency	-0.01	-0.12	-0.21	0.02	0.16	-0.04	-0.08	<u>0.51</u>	0.50																					
Avg Delay	0.12	0.02	0.05	-0.13	0.04	0.03	0.04	-0.08	-0.10	-0.11																				
SD Delay	0.02	-0.06	0.05	-0.07	0.00	0.01	0.04	-0.05	-0.06	-0.08	<u>0.62</u>																			
% Board	-0.07	-0.14	-0.29	0.09	0.20	-0.08	-0.19	<u>0.66</u>	<u>0.67</u>	0.47	-0.16	-0.13																		
Turns	-0.02	-0.02	0.06	0.00	-0.01	-0.01	-0.01	-0.13	-0.10	-0.13	0.09	0.09	-0.25																	
Signals	0.00	0.01	0.22	-0.04	-0.05	-0.03	-0.02	-0.01	0.00	0.02	0.10	-0.02	0.12	0.17																
Lanes	0.01	0.01	0.02	0.00	-0.01	0.00	0.01	0.09	0.11	0.11	-0.03	0.02	0.15	-0.17	0.18															
Speed	-0.04	-0.04	0.04	0.01	0.01	-0.02	-0.02	0.11	0.05	0.05	-0.11	0.04	-0.01	0.10	0.04	0.23														
BusLane	-0.19	-0.19	-0.06	0.11	0.09	-0.05	-0.04	0.17	0.16	0.19	-0.08	0.06	0.17	-0.07	0.02	0.14	0.14													
Oneway	-0.02	-0.02	0.06	-0.01	0.00	-0.02	-0.02	0.08	0.03	0.03	-0.08	-0.02	0.09	-0.08	0.12	-0.11	<u>0.51</u>	0.13												
L-Residential	0.02	0.01	-0.06	0.00	0.01	0.01	0.01	-0.10	-0.07	-0.09	0.00	0.00	-0.09	0.22	-0.27	-0.20	-0.41	-0.10	-0.24											
L-Collector	0.01	0.00	-0.05	0.00	0.01	0.01	0.00	-0.06	-0.01	-0.08	0.02	0.03	-0.15	0.09	-0.18	-0.28	-0.17	-0.05	-0.21	-0.08										
L-Secondary	0.00	0.01	0.06	-0.01	-0.02	-0.01	0.00	-0.04	-0.03	0.04	0.00	-0.05	0.14	-0.18	0.32	0.12	-0.02	0.04	0.10	-0.34	<u>-0.52</u>									
L-Primary	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.14	0.16	0.12	-0.01	-0.03	0.15	-0.22	0.12	0.36	0.05	0.04	0.10	-0.09	-0.24	-0.37								
L-Highway	-0.02	-0.02	0.03	0.00	0.00	-0.01	-0.01	0.08	-0.03	0.00	-0.05	0.02	-0.02	0.18	-0.09	0.02	<u>0.63</u>	0.04	0.37	-0.22	-0.12	-0.12	-0.12							
Mixed	-0.02	-0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.03	-0.10	-0.03	0.03	-0.03	-0.01	0.06	0.19	0.05	0.22	-0.02	-0.06	0.09	-0.06	-0.01						
Centre	0.00	0.01	0.01	-0.01	0.00	0.00	-0.01	-0.14	-0.16	-0.02	0.13	-0.02	-0.08	-0.07	0.16	0.05	-0.13	-0.01	0.03	-0.12	-0.01	0.07	0.03	0.06	-0.23					
Industrial	-0.05	-0.05	0.00	0.02	0.02	-0.02	-0.02	0.03	-0.01	-0.07	-0.04	0.08	-0.16	0.26	-0.15	0.00	0.47	0.11	0.21	-0.11	-0.02	-0.15	-0.03	0.40	0.09	-0.20				
Pop. Density	0.03	0.03	0.02	-0.02	-0.02	0.00	0.01	0.00	0.00	0.11	0.07	-0.10	0.26	-0.35	0.17	0.03	<u>-0.50</u>	-0.06	-0.18	0.10	-0.11	0.24	0.06	-0.28	-0.19	0.28	<u>-0.58</u>			
Num Stops	0.02	0.03	0.16	-0.03	-0.04	-0.01	0.00	-0.03	0.06	-0.03	0.07	0.03	-0.09	<u>0.52</u>	<u>0.61</u>	0.04	-0.04	-0.08	-0.12	0.00	0.09	0.05	0.00	-0.24	-0.03	-0.15	-0.03	-0.18		

Table A.4 Correlations Matrix for Paper 2 Stop Level

	Mar	Jun	Sep	Nov	B.Lane	Late	AMPeak	Mid	Eve	AM E	PM E	Eve E	Freq	Turns	Lanes	Spd	StopSigns	Signals	L.Res	L.Col	L.Sec	L.Pri	L.Hwy	Park	CV	Retail	Ind.	Density	
June	-0.28																												
Sep	-0.28	-0.22																											
Nov	-0.29	-0.22	-0.22																										
BusLane	0.00	0.00	0.00	0.00																									
Late	0.00	0.00	0.00	0.00	0.03																								
AMPeak	0.00	0.00	0.00	0.00	-0.04	-0.05																							
Midday	0.00	0.00	0.00	0.00	0.00	-0.04	-0.08																						
Evening	0.00	0.00	0.00	0.00	0.00	-0.04	-0.08	-0.07																					
AM Weekend	0.00	0.00	0.00	0.00	0.00	-0.04	-0.07	-0.07	-0.07																				
PM Weekend	0.00	0.00	0.00	0.00	0.00	-0.04	-0.07	-0.07	-0.07	-0.07																			
Eve Weekend	0.00	0.00	0.00	0.00	0.00	-0.04	-0.07	-0.07	-0.07	-0.07	-0.07																		
Frequency	0.02	-0.04	-0.01	-0.01	0.12	-0.08	0.13	0.03	-0.03	-0.03	0.09	-0.04																	
Turns	0.00	0.00	0.00	0.00	0.01	-0.01	0.02	0.01	0.01	0.00	0.00	0.00	-0.08																
Lanes	0.00	0.00	0.00	0.00	0.43	0.05	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	0.09	0.00															
Speed	0.00	0.00	0.00	0.00	0.17	0.03	0.01	0.00	-0.01	-0.01	-0.01	-0.01	0.05	0.04	0.30														
Stop Signs	0.00	0.00	0.00	0.00	-0.19	-0.03	0.02	0.01	0.01	0.00	0.01	0.01	-0.04	0.17	-0.33	-0.13													
Signals	0.00	0.00	0.00	0.00	0.22	0.03	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	0.04	0.22	0.32	0.06	-0.37												
L.Residential	0.00	0.00	0.00	0.00	-0.14	-0.03	0.02	0.02	0.01	0.00	0.01	0.00	-0.08	0.27	-0.28	-0.12	0.28	-0.14											
L.Collector	0.00	0.00	0.00	0.00	-0.10	-0.02	0.02	0.01	0.00	0.00	0.00	0.00	-0.03	0.17	-0.19	-0.07	0.22	-0.06	-0.17										
L.Secondary	0.00	0.00	0.00	0.00	0.12	0.04	0.01	0.00	-0.01	-0.01	-0.01	-0.01	0.01	0.20	0.32	0.17	-0.18	0.45	-0.15	-0.26									
L.Primary	0.00	0.00	0.00	0.00	0.22	0.01	0.00	0.00	0.00	0.00	-0.01	0.00	0.05	0.02	0.27	0.12	-0.12	0.22	-0.07	-0.13	-0.10								
L.Highway	0.00	0.00	0.00	0.00	0.03	0.01	0.01	0.01	0.00	0.00	0.00	0.00	-0.01	0.34	0.05	0.12	0.01	0.29	0.19	0.05	0.23	-0.01							
Park	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	-0.03	-0.02	0.02	-0.02	-0.02	0.08	0.05	-0.01	0.03						
CentreVille	0.00	0.00	0.00	0.00	0.00	0.02	-0.01	0.00	0.00	0.00	0.00	0.00	0.03	0.04	0.04	0.13	-0.11	0.15	-0.02	-0.05	0.04	0.02	0.02	0.04					
Retail	0.00	-0.01	0.00	0.00	0.07	0.02	-0.01	0.00	0.00	0.00	0.00	0.00	0.03	0.07	0.09	-0.05	-0.12	0.29	-0.08	-0.01	0.11	0.01	0.05	-0.10	0.03				
Industry	0.00	0.00	0.00	0.00	0.00	-0.01	0.05	0.02	0.00	-0.01	-0.01	-0.01	-0.06	0.13	0.02	0.20	0.00	-0.09	0.10	0.09	0.07	0.05	0.09	-0.04	-0.07	-0.13			
Density	0.00	-0.01	0.00	0.00	0.02	0.01	-0.04	-0.02	0.00	0.00	0.00	0.00	0.05	-0.11	-0.06	-0.25	-0.15	0.26	-0.08	-0.15	-0.02	-0.03	-0.04	-0.13	0.18	0.21	-0.42		
Dist CV	0.00	0.01	0.00	0.00	-0.05	-0.01	0.03	0.01	0.00	0.00	0.00	0.00	-0.06	0.01	0.01	0.13	0.22	-0.33	0.07	0.19	-0.05	0.00	-0.02	0.02	-0.30	-0.12	0.12	<u>-0.55</u>	

Table A.5 Correlation Matrix for Paper 2 Timepoint Level

	Mar	Jun	Sep	Nov	Late	AMPeak	Mid	Eve	AM E	PM E	Eve E	Freq	Turns	Lanes	Spd	StopSigns	Signals	L.Res	L.Col	L.Sec	L.Pri	L.Hwy	Park	CV	Retail	Ind.	Density	
June	-0.35																											
Sep	-0.23	-0.22																										
Nov	-0.25	-0.24	-0.16																									
Late	0.02	-0.01	0.01	0.00																								
AMPeak	-0.03	-0.03	0.01	0.01	-0.04																							
Midday	-0.02	0.00	-0.01	-0.03	-0.04	-0.08																						
Evening	-0.02	0.02	0.00	-0.02	-0.03	-0.08	-0.07																					
AM End	-0.01	0.02	0.01	0.01	-0.03	-0.08	-0.07	-0.07																				
PM End	-0.01	-0.01	0.00	-0.01	-0.03	-0.08	-0.07	-0.07	-0.07																			
Eve End	-0.03	0.05	0.01	-0.01	-0.03	-0.07	-0.07	-0.07	-0.06	-0.06																		
Frequency	-0.08	-0.08	-0.02	-0.04	-0.09	0.17	0.03	-0.02	-0.03	0.12	-0.03																	
Turns	-0.01	0.02	0.01	0.02	0.02	0.01	0.00	0.00	0.00	0.00	0.00	-0.11																
Lanes	0.00	0.02	0.00	-0.02	0.03	0.00	-0.01	-0.01	-0.01	-0.01	-0.01	0.05	-0.18															
Speed	0.00	0.00	0.02	-0.02	0.03	0.02	0.00	-0.01	-0.02	-0.02	-0.01	0.02	-0.02	<u>0.58</u>														
Stop Signs	0.00	0.02	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.00	-0.07	0.49	-0.41	-0.25													
Signals	0.02	-0.02	-0.01	0.01	0.09	-0.02	-0.01	-0.02	-0.02	-0.02	-0.02	-0.03	0.09	0.29	0.18	-0.27												
L.Residential	-0.01	0.03	0.01	0.02	-0.01	0.01	0.01	0.01	0.01	0.00	0.01	-0.09	<u>0.53</u>	-0.32	-0.17	<u>0.58</u>	-0.16											
L.Collector	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	-0.04	0.24	-0.27	-0.13	0.44	-0.03	0.03										
L.Secondary	0.03	0.00	0.00	0.01	0.08	0.00	-0.01	-0.01	-0.02	-0.02	-0.02	-0.04	0.05	0.27	0.29	-0.13	<u>0.50</u>	-0.20	-0.24									
L.Primary	0.01	0.00	0.00	-0.02	0.02	0.01	0.01	-0.01	-0.01	-0.01	-0.01	0.01	-0.06	0.30	0.22	-0.14	0.33	-0.05	-0.13	-0.15								
L.Highway	0.00	-0.01	0.01	0.01	0.03	0.01	0.00	-0.01	-0.01	-0.01	-0.01	-0.02	0.28	0.08	0.36	-0.01	0.11	0.02	0.02	0.16	-0.02							
Park	0.02	-0.02	0.01	0.01	0.01	-0.01	0.00	0.00	0.00	0.00	0.00	-0.03	0.09	-0.07	-0.01	0.10	0.01	0.03	0.13	0.09	0.04	0.01						
CV	-0.01	-0.03	0.00	0.00	0.03	-0.01	0.00	-0.01	0.00	0.00	0.00	0.01	0.02	0.03	-0.02	-0.19	0.20	-0.11	-0.09	0.02	0.01	0.07	0.02					
Retail	-0.02	0.01	-0.01	0.00	0.03	0.00	0.00	-0.01	-0.01	-0.01	-0.01	-0.03	0.05	0.10	0.02	-0.06	0.31	-0.06	0.06	0.16	-0.01	0.06	-0.13	0.07				
Industry	0.02	0.01	0.00	0.01	0.02	0.03	0.00	0.00	-0.01	-0.02	-0.01	-0.07	0.25	0.01	0.22	0.07	0.03	0.16	0.11	0.14	0.02	0.15	-0.02	-0.13	-0.07			
Density	-0.01	-0.02	-0.02	-0.01	0.00	-0.03	-0.01	0.00	0.00	0.01	0.01	0.09	-0.21	-0.04	-0.31	-0.22	0.19	-0.16	-0.22	-0.07	-0.02	-0.09	-0.20	0.24	0.16	-0.41		
Dist CV	-0.01	0.04	0.03	-0.01	0.00	0.04	0.02	0.00	0.00	0.00	-0.01	-0.09	0.05	0.02	0.23	0.28	-0.30	0.14	0.26	-0.04	0.02	-0.03	0.09	-0.40	-0.08	0.09	<u>-0.58</u>	

Table A.6 Correlation Matrix for Paper 2 Route Level

	Mar	Jun	Sep	Nov	Late	AMPeak	Mid	Eve	AM E	PM E	Eve E	Freq	Turns	Lanes	Spd	StopSigns	Signals	L.Res	L.Col	L.Sec	L.Pri	L.Hwy	Park	CV	Retail	Ind.	Density	
June	-0.33																											
Sep	-0.23	-0.22																										
Nov	-0.26	-0.24	-0.17																									
Late	-0.02	-0.01	0.04	0.03																								
AMPeak	0.00	-0.01	0.03	-0.05	-0.06																							
Midday	-0.04	-0.02	-0.04	0.00	-0.05	-0.08																						
Evening	-0.01	0.02	-0.02	0.01	-0.05	-0.08	-0.07																					
AM End	0.01	0.04	0.01	0.01	-0.05	-0.08	-0.07	-0.06																				
PM End	-0.02	0.01	0.00	-0.03	-0.05	-0.08	-0.07	-0.07	-0.06																			
Eve End	-0.02	0.06	0.02	-0.03	-0.05	-0.07	-0.06	-0.06	-0.06	-0.06																		
Frequency	-0.04	-0.05	0.03	-0.11	-0.13	0.18	0.05	0.02	-0.03	0.11	0.00																	
Turns	0.01	0.01	0.02	0.04	0.04	-0.03	0.00	0.01	0.01	-0.02	0.01	-0.13																
Lanes	0.00	-0.03	0.03	-0.01	0.02	0.02	0.03	-0.02	-0.03	0.00	-0.02	0.07	-0.21															
Speed	-0.02	0.01	0.02	0.03	0.01	0.03	0.02	-0.01	-0.03	0.01	-0.02	-0.01	-0.11	<u>0.67</u>														
Stop Signs	0.01	0.01	0.01	0.03	0.02	-0.03	-0.01	0.02	0.01	-0.01	0.04	-0.05	<u>0.66</u>	-0.32	-0.18													
Signals	0.03	-0.07	0.04	0.01	0.10	-0.06	-0.01	-0.02	-0.02	-0.02	-0.03	0.01	0.07	0.30	0.12	-0.14												
L.Residential	-0.01	0.01	-0.01	0.05	-0.01	-0.02	0.01	0.02	0.02	0.00	0.02	-0.10	<u>0.67</u>	-0.31	-0.23	<u>0.64</u>	-0.21											
L.Collector	0.03	0.01	0.04	-0.01	0.03	-0.02	-0.01	-0.01	0.00	-0.02	0.03	-0.05	0.34	-0.23	-0.15	<u>0.57</u>	0.04	0.07										
L.Secondary	0.05	-0.01	0.04	-0.01	0.11	-0.04	-0.02	-0.01	-0.03	-0.03	-0.03	-0.07	0.14	0.40	0.38	-0.04	<u>0.60</u>	-0.22	-0.05									
L.Primary	0.01	-0.05	0.02	0.00	0.01	0.03	0.01	-0.01	-0.01	0.00	-0.01	0.06	-0.05	0.27	0.17	0.02	0.36	0.02	-0.09	-0.12								
L.Highway	-0.03	0.02	-0.01	0.04	-0.02	0.01	0.01	0.00	-0.01	0.03	-0.01	-0.02	0.06	0.28	<u>0.67</u>	-0.02	-0.10	-0.10	-0.08	0.18	-0.09							
Park	0.02	-0.02	0.02	-0.02	0.02	-0.01	-0.03	0.00	0.01	-0.01	0.01	0.05	0.09	-0.10	-0.10	0.04	0.21	0.03	0.01	0.10	0.23	-0.12						
CV	0.02	-0.06	0.02	-0.01	-0.01	-0.02	-0.01	-0.01	0.01	0.02	-0.02	0.00	0.04	-0.01	-0.14	-0.34	0.38	-0.19	-0.24	0.11	-0.01	0.01	0.26					
Retail	0.02	-0.05	-0.02	0.02	0.03	-0.03	0.01	0.00	0.00	0.00	0.01	0.02	0.17	0.08	0.09	0.17	0.19	0.08	0.09	0.12	0.10	0.05	-0.14	0.00				
Industry	-0.02	0.03	0.06	0.04	0.00	0.00	0.00	0.02	0.00	-0.01	-0.01	-0.11	0.26	0.16	0.24	0.18	-0.02	0.19	0.04	0.21	0.05	0.22	0.11	-0.14	0.02			
Density	-0.03	-0.03	-0.04	-0.02	-0.01	-0.01	0.01	-0.01	0.01	0.01	-0.01	0.08	-0.27	-0.27	-0.48	-0.35	0.14	-0.23	-0.27	-0.19	-0.09	-0.28	-0.09	0.34	0.05	<u>-0.53</u>		
Dist CV	-0.01	0.04	0.00	0.02	0.00	0.05	0.00	0.01	0.00	-0.02	0.02	-0.04	0.08	0.07	0.19	0.38	-0.35	0.22	0.37	-0.13	0.07	-0.03	-0.16	<u>-0.57</u>	0.00	0.12	<u>-0.57</u>	

Table A.7 Correlation Matrix for Paper 5

	Pop. Density	% Bach.	% Minoriy	Avg age
% Bach.	0.11			
% Minoriy	0.11	-0.25		
Avg age	-0.44	-0.06	-0.34	
Avg income	-0.36	<u>0.63</u>	<u>-0.54</u>	0.34