COMPARING BUS TRAVEL TIMES AND TRAVEL SPEEDS MODELS AT VARIOUS ANALYSIS LEVELS FOR TRANSIT PLANNING

Yuxuan Wang, Corresponding Author
Department of Civil, Geology and, Mining Engineering
Polytechnique Montréal, Montréal, Québec, Canada, H3T 1J4
Email: yuxuan.wang@polymtl.ca
ORCID: 0000-0002-1512-8846

Catherine Morency, Ph.D.
Department of Civil, Geology and, Mining Engineering
Canada Research Chair on Mobility
Canada Research Chair on Personal Mobility
Polytechnique Montréal, Montréal, Québec, Canada, H3T 1J4
Email: cmorency@polymtl.ca

Martin Trépanier, Ph.D.
Department of Civil, Geology and, Mining Engineering
Interuniversity Research Center on Enterprise Networks, Logistics and Transportation
Polytechnique Montréal, Montréal, Québec, Canada, H3T 1J4
Email: mtrepanier@polymtl.ca

Word Count: 8791 words + 3 table(s) × 250 = 9541 words

Submission Date: August 11, 2024

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 these two measures and models at various analysis levels. In this paper, we try to compare travel times and travel speeds using different 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, new routes and new service hours.

The results show travel time and speed models perform similarly for new service hour scenarios, with the time models performing slightly better. However, speed models tend to perform better for new route scenarios. Both models perform better at more aggregated levels, such as the timepoint-to-timepoint level. For lower levels, both models perform better at inter-stop level than stop-to-stop level, emphasizing the need to include more detailed data to improve the models. Errors calculated from travel speed models are slightly more biased than travel time ones. However, the relative errors from travel time models are larger than the speed 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.

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

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 (*1*) 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 (*2*). 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 (*3*). 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 (*4*). Up until now, transit agencies and academics have mostly focused on travel times adherence and prediction (*5*). 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 we expect 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. Speed 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 the speed is more directly comparable, 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 (6) 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. Thus, we pose the question of whether the travel time and speed models yield similar results for public transit.

With the goal of improving transit travel time models and passenger satisfaction, we aim to help planners better understand transit travel conditions. We develop a framework to allow us to compute and compare two commonly used measures in transit planning, travel times and travel speeds. Then, we model travel times and travel speeds at different analysis levels using two years of archived service delivery data from Montréal, Québec, Canada. Finally, we compare the advantages and disadvantages of these two approaches as well as the different analysis levels, so that we can make recommendations to transit agencies for their 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.

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. (4) 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 termini as well as between various timepoints using measures such as the mean, median, or a given percentile of observed travel times (*7*).

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 (*8*). Most works found 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 (1) calculated the schedule padding using best-case transit travel times recorded. They found 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.

To help provide passenger information, many works have tried to predict transit travel times. Scholars have proposed methods predicting transit travel times. Some of the works used only AVL data (*9, 10*). There are also attempts to incorporate additional datasets to improve the prediction models, such as real-time traffic data (*11*). Using AVL data, it is also possible to evaluate the service delivered to passengers. Wessel et al. (12) proposes a method to retroactively improve the accuracy of transit agencies' GTFS feed by using archived AVL data. Agencies have also started developing passenger-centric performance measures using their AVL and origin-destination data (*13*).

However, travel speed and travel time are related variables, where the travel time equals the travel distance divided by the travel speed. By evaluating speed, 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 could potentially allow us to create schedules or target issues at the systemwide scale. Therefore, operating speed is another commonly used indicator for transit performance evaluations. It is defined as the average speed over a section travelled by the passengers which includes all stops.

Cortés et al. (14) created a classification for average bus operating speeds and identified slow roadway segments for agencies to improve speed. Aemmer et al. (15) aggregated the travel time by roadway segments to calculate the pace (inverse of the speed). The results show buses can more often travel faster than the schedule on a few selected segments. Zhang et al. (16) tested a few factors that could affect bus operating speeds, such as bus lanes, road classifications, geographical area, peak direction, and service types. They found buses on main roads, in outskirt neighbourhoods, during off-peak hours, or in bus lanes tend to travel faster than on other segments.

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 the two measures in the transportation field, especially from Bauer and Tulic (6), 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, the General Transit Feed Specification (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 (*12*), 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 (*12, 13*).

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. In addition, different measures evaluate the results "from different angles" (*17*), 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 for comparing the travel time and travel speed modelling approaches. 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.

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 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 and speeds. 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.

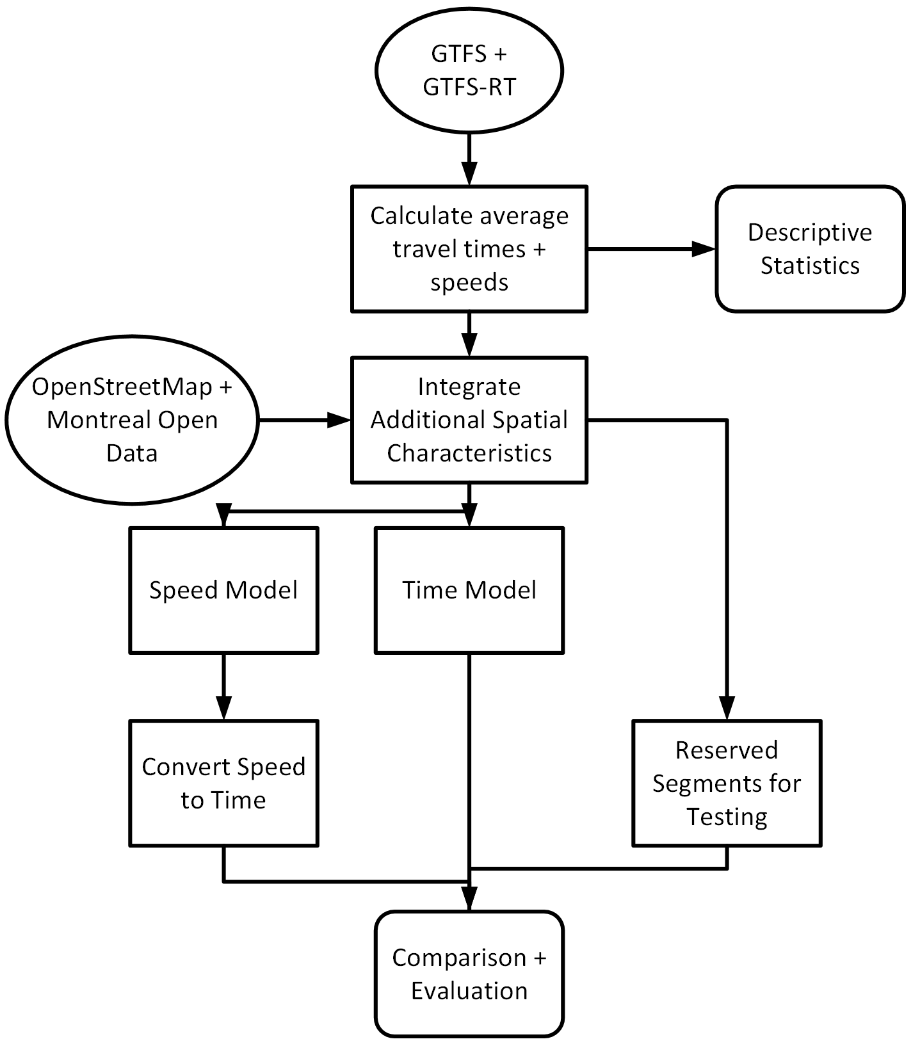


FIGURE 1: Research Framework

Typically, agencies have their own policies on how to analyze travel times (*7*), 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 (*7*). 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 and the open data provided by the city of Montréal. These spatial attributes along with the calculated travel times and speeds are then used as inputs for modelling average travel times and speeds. 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 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 given time of day on a given segment. Thus, we reserve another 10% of the remaining data with various time-of-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.

Data

We use the bus system of Société de Transport de Montreal 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 no more than every 20 seconds. In this paper, we used the archived data from May 1st, 2021 to March 24th, 2024.

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) (*18*), 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.

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) (*19*) 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 (*19*). 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 the travel times are comparable. The service pattern travel time is defined as the time between the departure from the terminus to the arrival at the ending terminus.

Finally, using these times calculated above and the segment lengths extracted from GTFS, we calculate the corresponding inter-stop speed, stop-to-stop speed, timepoint-to-timepoint speed, and the overall operating speed of the service pattern.

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 (*20*).

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, thus 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 (*21*). 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 forests method originally proposed by Ho (22). 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.

(23) 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.

Input Variables

There are many factors affecting transit travel time and speeds. The Transit Capacity and Quality of Service Manual (*19*) 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 June period as the base case.
- Time of Day. Due to the non-linear nature of traffic and the time periods, we simplified time as a categorical variable. According to the descriptive statistics, the time of day variations also differ given the day of the week. Thus, we include both time of day and the day of the week in our categories. We identified six time of the day categories for weekdays, namely early morning (4 - 6), morning peak (7 - 9), midday (10 - 14), evening peak (15 - 17), evening (18 - 22), and late-night (22 - 4) periods. For weekends, due to the lack of morning peaks, we combined the morning peak and the midday into a morning category. Here we use the Weekday PM peak as the base case.

The spatial variables included are related to street characteristics, land use characteristics, and the population density near a given segment. 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 load, a ranked variable related to the number of people on board. As the number of people on board increases, the time for boarding and alighting generally increases as well. Thus, we include this variable to account for this friction. The data is ranked into five categories, namely empty, many seats available, few seats available, standing room only, and full.
- 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.

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 mean squared error 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.

RESULTS

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

Model Coefficients

Since we used 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 linearity assumption.

Random Effects for Linear Mixed Model

Table 1, 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 (*20*).

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. This means that the individual differences for each segment would contribute to around 5 km/h travel speed differences. 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 (*24*). Here, we can observe that, all models have an ICC above 0.6, 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, with the exception of the speed model at the service pattern level. This means that the individual differences between segments become more important at smaller scales. This highlights the fact that higher analysis levels may hide variations in smaller levels, and more research is needed for stop-stop level scheduling, also pointed out by other researchers (*12, 13*).

TABLE 1: Random Effects for Linear Speed and Time Models

Analysis Level	Speed	Speed Adj. ICC	Time	Time Adj. ICC
Inter-stop	5.26	0.74	19.81	0.87
Stop to Stop	5.70	0.69	21.71	0.79
Timepoint to Timepoint	3.99	0.63	56.82	0.62
Service Pattern	4.34	0.92	315.90	0.62

Fixed Effects for Linear Mixed Model

Table 2 shows all of 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, and the units for times are in seconds. For the description of each variable, please refer to the earlier sections. In addition, we marked variables with p-value less than 0.05 with italic fonts, since most of the values are statistically significant.

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 is lower.

More specifically, for the service period variables, all of the speed coefficients are negative and all of the time coefficients are positive. This is expected since we chose the June or summer schedule as the base case. In other periods, the ridership is typically higher and traffic congestion is generally worse. During the winter, the speeds or times are also affected by adverse weather events like snow storms, resulting in worse travel conditions.

As for the time of days, we can observe that all the speed coefficients are positive and all the time 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

TABLE 2: Fixed Effects for Linear Speed and Time Models

	Speed Models				Time Models			
	Inter	Stop	Timepoint	Route	Inter	Stop	Timepoint	Route
Pop. Intercept	31.38	27.38	14.07	0.63	-13.37	-10.77	58.33	1213.00
Num. Stops	N/A	N/A	-0.17	-0.12	N/A	N/A	2.13	9.93
Period Sep	-0.13	-0.28	0.34	-0.03	0.39	0.78	-2.03	29.67
Period Nov	-0.34	-0.18	0.19	0.02	0.37	0.26	-2.61	25.81
Period Jan	-0.89	-0.26	0.26	0.04	1.11	0.39	-2.12	18.70
Period Mar	-0.82	-0.11	0.23	0.07	0.73	0.45	-1.08	23.71
Week Early AM	2.80	4.72	4.40	4.83	-5.62	-13.21	-71.58	-536.20
Week AM Peak	0.56	0.72	0.93	0.69	-1.64	-3.12	-19.25	-98.43
Week Midday	0.87	1.45	1.47	1.51	-2.58	-5.01	-27.61	-216.90
Week Night	1.81	3.46	3.53	3.29	-4.28	-10.34	-60.45	-385.90
Week Late Night	5.61	8.54	6.01	2.62	-7.27	-17.79	-86.79	51.82
Sat. Early AM	3.94	6.46	6.05	5.89	-6.06	-14.85	-89.62	-615.40
Sat. AM	2.69	3.87	3.79	4.12	-5.06	-10.72	-64.67	-507.70
Sat. PM	1.25	2.20	1.95	1.88	-2.55	-5.84	-34.33	-267.50
Sat. Night	2.04	3.88	3.75	3.61	-4.12	-10.30	-62.49	-448.40
Sat. Late Night	4.78	7.69	5.99	5.39	-6.31	-15.94	-82.72	-588.60
Sun. Early AM	4.34	7.00	6.36	6.25	-6.45	-15.90	-92.27	-635.80
Sun. AM	3.25	4.66	4.50	4.74	-5.68	-12.07	-72.36	-553.90
Sun. PM	1.69	2.64	2.43	2.45	-3.32	-7.11	-42.86	-337.80
Sun. Night	2.41	4.30	4.24	4.06	-4.68	-11.32	-69.74	-499.10
Sun. Late Night	5.89	8.25	6.27	4.29	-6.30	-16.24	-92.67	-318.00
Bus Lane On	-0.57	-0.55	-0.82	-0.63	4.88	6.20	21.83	74.30
Bus Lane Off	-0.43	-0.78	-1.04	-0.84	4.25	5.16	17.12	126.10
Average Delay	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05
Average Load	-0.10	-0.23	-0.07	-0.11	0.08	0.50	1.01	17.95
Average Freq.	-0.10	-0.10	0.00	0.03	-0.03	0.09	0.42	7.78
Num. Turns	-3.39	-2.68	-0.41	-0.07	12.61	13.62	9.67	33.09
Num. Lanes	0.81	0.35	-0.50	-2.57	-2.83	-2.32	-7.03	64.96
Num. Stop Signs	-1.70	-0.98	-0.43	-0.13	5.42	4.71	8.09	11.89
Num. Signals	-1.78	-2.50	-0.57	-0.07	9.15	12.84	14.61	19.44
Speed Limit	0.06	0.05	0.13	0.52	-0.03	-0.03	0.29	-32.98
Local Length	4.52	6.91	3.59	1.52	90.79	81.15	84.06	20.76
Collector Length	6.41	8.83	2.82	0.82	95.35	91.14	87.65	32.97
Secondary Length	7.95	10.58	3.65	0.75	94.27	88.21	79.68	64.86
Primary Length	5.58	8.55	3.68	0.59	106.50	99.48	73.00	58.32
Motorway Length	4.22	4.29	2.59	1.02	47.29	46.43	49.23	78.14
Green Space	0.93	1.40	0.29	-0.13	1.13	0.32	1.66	-0.04
Downtown	-1.28	-0.20	0.08	-1.51	0.60	0.13	4.72	24.54
Retail	-1.33	-1.92	-1.53	1.39	3.20	6.81	23.11	38.63
Industry	2.80	3.44	0.52	-0.74	1.75	-0.17	-3.25	39.22
Pop. Density	-0.11	-0.23	-0.28	-0.25	-0.18	0.15	2.55	2.02
Dist. Downtown	0.22	0.26	0.27	0.10	-0.04	-0.15	-1.95	0.42

Saturdays and Sundays do not have a morning peak, the weekend mornings behave similarly to the early evenings.

An increase in the number of stops on a segment can also result in longer travel times 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 there is no passenger 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 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 streets with bus lanes are more congested than those without bus lanes. In addition, if the bus lanes are in service, the negative impact on bus speeds is generally smaller, bringing these segments more in line with less congested segments without bus lanes.

As for the average load and frequency, they are all negative for speed and mostly positive for time. Again, this is expected, since they correlate to the ridership and traffic congestion. More ridership and congestion means buses will spend more time not moving, thus reducing the speed and increasing the time.

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 lanes is positive for bus speeds. More lanes mean wider streets, which typically correlate to more traffic and higher speeds. Similarly, the speed limit is also positive for bus speeds. However, it is 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. 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 in case 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, higher ridership, and higher pedestrian counts in these areas. Thus, it is important for agencies to improve service in these areas to improve passenger experiences. Interestingly, most land use variables are not significant in the time models, except retail and industry land uses.

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. Interestingly, these two variables are not significant for the time models.

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.

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 scenarios to test the models by holding back some data from the overall dataset. One scenario is service expansion onto a new route where there is no existing data. Another scenario is expanding the service hours on an existing segment. We use two modelling methods, linear mixed model and mixed effect random forest. For each method, we tested the models using both scenarios. Since the time and speeds are not directly comparable, for the speed model, we then converted the speed results to travel times and named the indirect modelling result for short in this section.

Aggregated Measures

Table 3 shows the results of the commonly used aggregated error measures of each 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.

Overall, we can observe that 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 two to three 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.

The direct time models perform better for existing segments in new service hours, whereas the indirect speed models perform better for reserved segments on new routes on smaller scales. In our models, we included both speed-related, such as streetscapes and land use, as well as time-related variables, such as traffic signals which are related to a fixed time 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.

Despite missing some detailed time information, the inter-stop models, which doesn't include the dwell times and signal waiting times, perform better than the stop-to-stop models. To improve these models, future work can consider adding this missing information by creating a hybrid model and combining both time and speed models.

For higher levels, namely timepoint-to-timepoint and the service pattern levels, the differences between time and speed models become smaller, especially in the new routes scenarios. 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

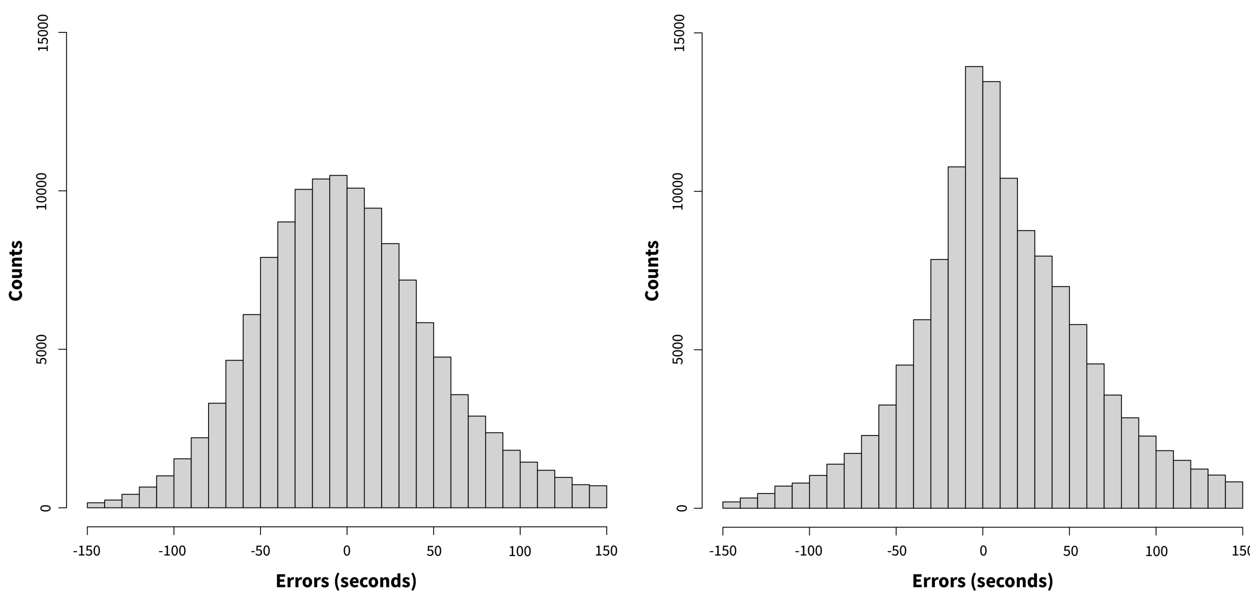
TABLE 3: Model Error Measures

		New Hours		New Routes		New Hours		New Routes	
		Linear	Forest	Linear	Forest	Linear	Forest	Linear	Forest
		Inter-stop				Timepoint			
Speed	R^2	0.83	0.86	0.27	0.30	0.82	0.86	0.52	0.59
	RMSE	3.05	2.75	5.59	5.49	2.99	2.68	5.01	0.46
	MAE	2.16	1.93	4.33	4.31	1.92	1.66	3.67	3.32
	MAPE	0.08	0.07	0.16	0.16	0.11	0.09	0.21	0.19
Time Indirect	R^2	0.96	0.97	**0.72**	**0.75**	**0.92**	0.93	0.81	**0.85**
	RMSE	8.51	7.18	**12.13**	**11.35**	**45.13**	41.99	65.54	**59.14**
	MAE	3.89	3.59	**5.97**	**5.82**	**24.78**	21.92	**44.04**	**40.49**
	MAPE	0.10	0.09	**0.16**	**0.16**	**0.11**	0.10	**0.18**	**0.17**
Time Direct	R^2	**0.96**	**0.97**	0.54	0.67	0.92	**0.94**	**0.83**	0.82
	RMSE	**8.42**	**7.78**	15.64	13.21	45.20	**39.40**	**61.63**	63.17
	MAE	**3.43**	**3.11**	9.03	8.80	26.39	**20.74**	44.27	44.15
	MAPE	**0.09**	**0.08**	0.30	0.28	0.14	**0.09**	0.27	0.24
		Stop to Stop				Service Pattern			
Speed	R^2	0.82	0.86	0.21	0.31	0.93	0.97	0.79	0.80
	RMSE	3.85	3.45	7.41	6.96	1.37	0.97	2.33	2.25
	MAE	2.82	2.49	5.94	5.54	0.83	0.61	1.83	1.68
	MAPE	0.13	0.12	0.37	0.33	0.05	0.04	0.10	0.09
Time Indirect	R^2	0.85	0.94	**0.40**	**0.45**	0.89	0.95	0.83	**0.90**
	RMSE	13.66	11.48	**23.62**	**22.79**	287.87	194.55	338.13	**254.51**
	MAE	7.98	6.99	**15.47**	**14.56**	**102.38**	**72.94**	**232.54**	**185.34**
	MAPE	0.15	0.13	**0.27**	**0.26**	**0.05**	**0.03**	**0.11**	**0.09**
Time Direct	R^2	**0.95**	**0.95**	0.30	0.41	**0.89**	**0.96**	**0.84**	0.86
	RMSE	**11.23**	**10.60**	27.51	23.47	**284.69**	**179.73**	**327.80**	305.56
	MAE	**6.77**	**5.76**	19.00	16.29	124.34	80.70	259.09	244.18
	MAPE	**0.15**	**0.13**	0.45	0.41	0.07	0.04	0.14	0.15

likely to indicate different "winners" in the same category. 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 the indirect result from 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 (*25*), 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 don't 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" (*17*). 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 that a few cases where the error measures indicated different "winners". In this subsection, we will use the direct and indirect 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.



(a) Direct Time Model

(b) Indirect Results from Speed Model

FIGURE 2: Error Histograms

Figure 2a and 2b show the error distributions for the direct time results and indirect time results from the speed model. To help readers see the differences, we used the same x and y scale for these two figures.

The errors for the direct time model 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 for indirect time results from speed models are more centered around 0, but skewed towards the right. In other words, the indirect models tend to underestimate the average travel times. In addition, speeds are bounded values between zero and the top speed of the vehicles, and times values do not have an upper bound. Thus, the congested observations may have more impact on the average time measures.

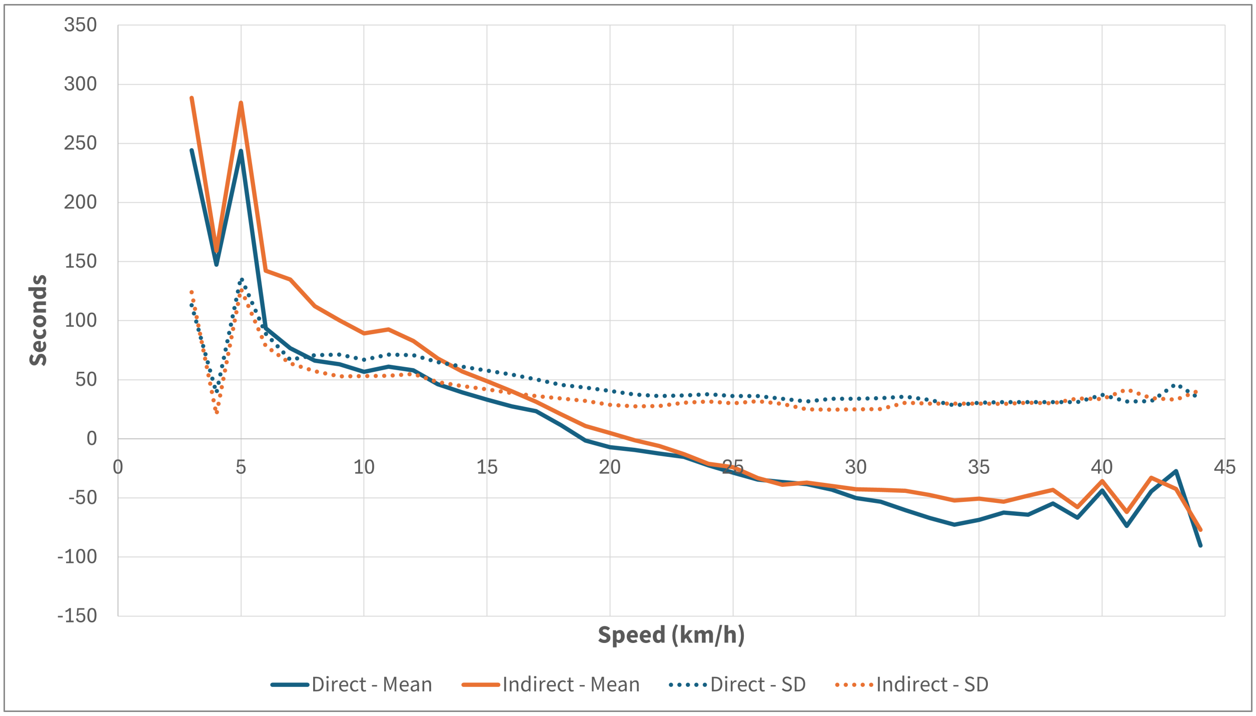
Figures 3a and 3b show the modelling errors aggregated by the observed average segment speed. Figure 3a shows that the errors follow a similar conditional average and conditional standard deviation for both the direct time model and the indirect results from the speed model. This makes sense, since the models try to minimize the errors. However, the percentage errors for the indirect results from speed models follow a more stable conditional average and conditional standard deviation, whereas the direct time model has more varied conditional averages and conditional standard deviations, especially for faster segments. We once again believe the length coefficients (the pace, inverse of the speed) from the time models may be too restrictive for the models to adapt to different segment lengths or speeds.

Both figures show the models would underestimate the travel time on slower segments but overestimate on faster segments. This is 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.

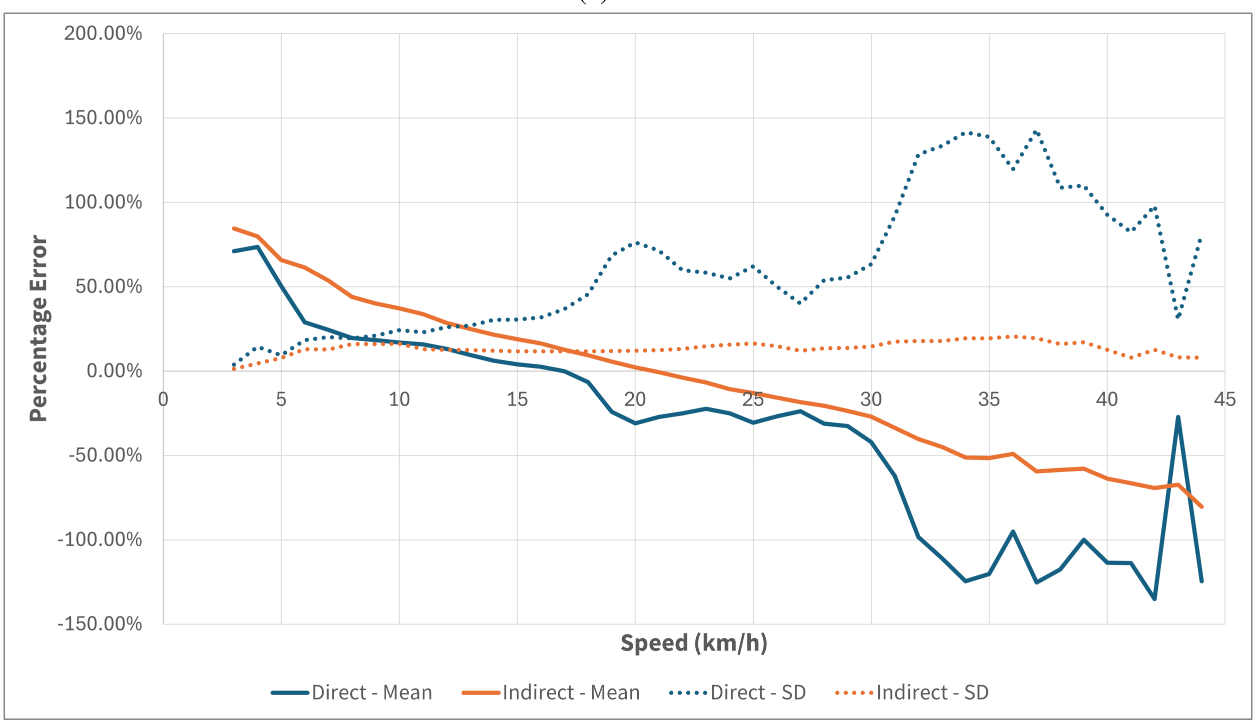
Finally, Figures 4a and 4b show the modelling errors aggregated by the segment length. From Figure 4a, we can again see that the errors from both models follow similar conditional averages and conditional standard deviations for longer segments. Once again, this makes sense since the models try to minimize the errors, which might reduce the accuracy for shorter segments, as the indirect results from speed models are closer to 0 for short segments less than 500 meters.

However, Figure 4b, which shows the percentage errors, highlights the large differences between the two models for shorter segments typically found on local services. The direct time results vary to as much as -100% and the conditional standard deviation varies up to 150% for short segments. Whereas, the indirect results from speed models have stable percentage errors, around 18% for both conditional average and conditional standard deviation, much more stable compared to the direct time model. This indicates that despite the close conditional average errors between the two 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 two 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.

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. 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 speed models perform relatively better for shorter segments and local services. Thus, transit planners need to consider their specific planning context,

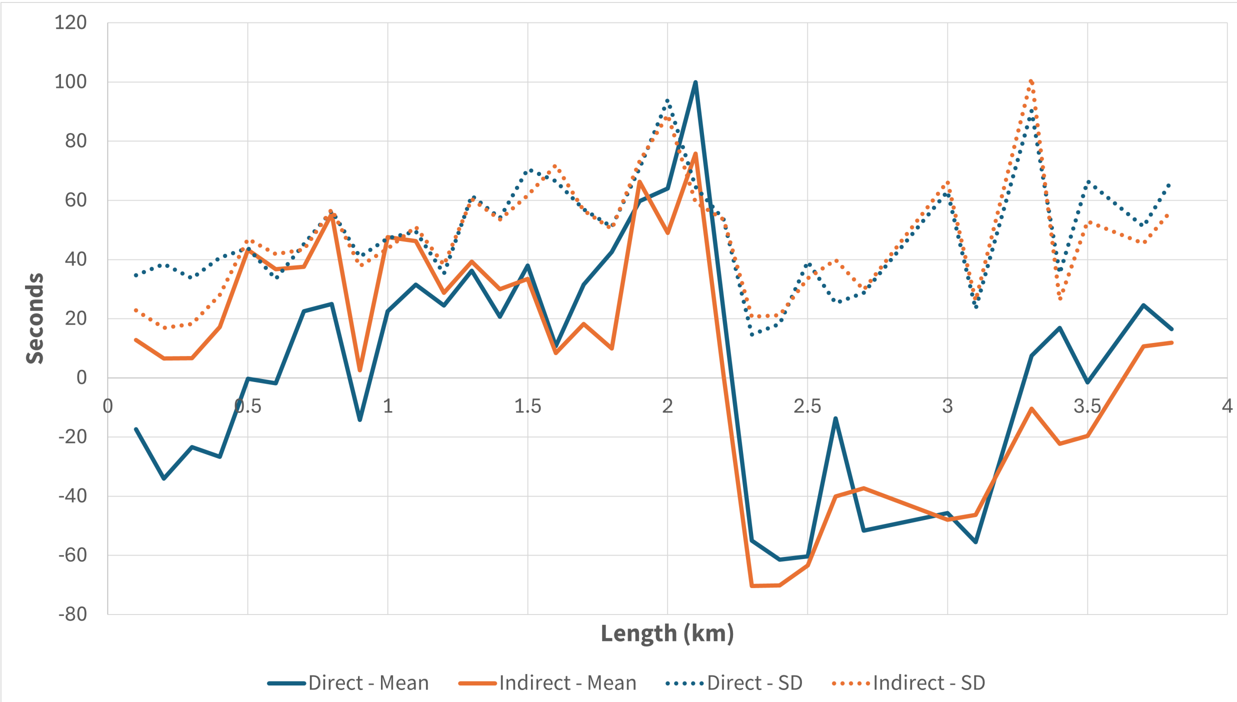


(a) Error

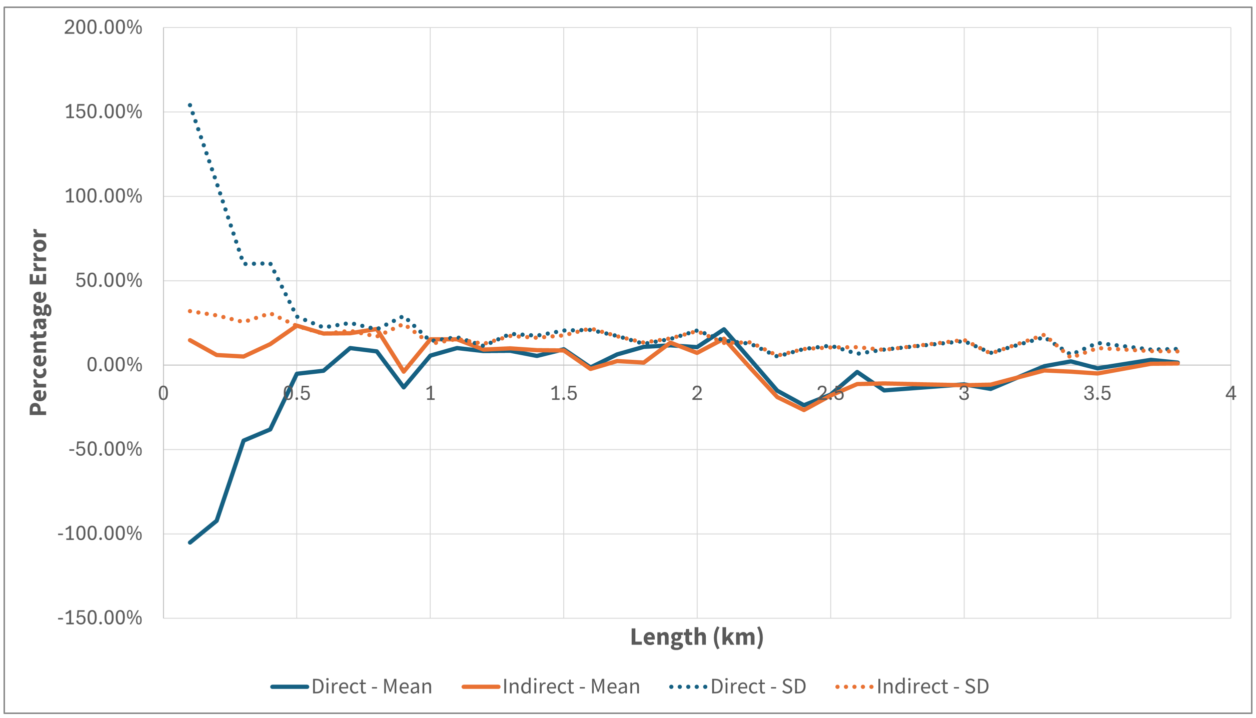


(b) Percentage Error

FIGURE 3: Errors Aggregated by Observed Average Segment Speed



(a) Error



(b) Percentage Error

FIGURE 4: Errors Aggregated by Segment Length

whether we would prefer certain biases in the model or if we are willing to accept the larger relative errors on short segments. Again, since different measures evaluate the results "from different angles" (*17*), transit planners and future researchers need to think more about which measure is more suitable given a specific context.

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 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, faster speed will result in shorter travel times, and slower speed will result in longer travel times. Hence, it is possible to use both as inputs for planning purposes.

Most of the previous literature focuses on travel times, and travel speeds 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 plan for or model a transit system?

We hypothesized that speed 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 don't necessarily travel between timepoints. We also consider 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 or speed 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 the modelling results are in line with our expectations. Our evaluations show that the non-linear models perform slightly better. Transit travel times and speeds 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 uses are not significant for travel time models. The coefficients in travel time models are paces, the inverse of speed, 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 measures 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 further study.

Segments with existing observations can help greatly when trying to model them in the new service hour scenario since they have a segment-specific intercept to take account of the differences between segments. However, the indirect results from speed models typically win for the new route scenarios, given the lack of segment-specific intercepts. Planners could refer to similar existing segments based on local knowledge when planning for a new route. 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, future researchers could also test a hybrid model using detailed inter-stop speed, traffic light timing, dwell times, etc.

We also highlight the shortcomings of using specific aggregated measures in previous literature, since different measures evaluate the results "from different angles" (*17*). 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. Both models perform similarly in terms of the average errors. However, the speed 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 talked about how transit vehicles may get stuck in traffic. However, for passengers, another potential way to get stuck is when missing a transfer. Thus, operational measures like the ones we compared may not necessarily reflect passenger experiences. Therefore, future researchers could also introduce additional measures regarding passenger experiences.

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, the Canada Research Chair in Mobility, 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.

REFERENCES

1. Wessel, N. and M. J. Widener, Discovering the space–time dimensions of schedule padding and delay from GTFS and real-time transit data. *Journal of Geographical Systems*, Vol. 19, 2017, pp. 93–107.
2. Danaher, A., J. Wensley, A. Dunham, T. Orosz, R. Avery, K. Cobb, K. Watkins, C. Queen, S. Berrebi, M. Connor, et al., *Minutes Matter: A Bus Transit Service Reliability Guidebook*. Washington, DC: The National Academies Press, 2020.
3. Carrel, A., A. Halvorsen, and J. L. Walker, Passengers' perception of and behavioral adaptation to unreliability in public transportation. *Transportation Research Record*, Vol. 2351, No. 1, 2013, pp. 153–162.
4. Coleman, M., L. Tarte, S. Chau, B. Levine, and A. Reddy, A data-driven approach to prioritizing bus schedule revisions at New York City transit. *Transportation Research Record*, Vol. 2672, No. 8, 2018, pp. 86–95.
5. Trépanier, M., C. Morency, and B. Agard, Calculation of transit performance measures using smartcard data. *Journal of Public Transportation*, Vol. 12, No. 1, 2009, p. 5.
6. Bauer, D. and M. Tulic, Travel time predictions: should one model speeds or travel times? *European Transport Research Review*, Vol. 10, 2018, pp. 1–12.
7. Furth, P. G., *Data analysis for bus planning and monitoring*. 34, Transportation Research Board, 2000.
8. Mandelzys, M., B. Hellinga, and P. Eng, Automatically identifying the causes of bus transit schedule adherence performance issues using AVL/APC archived data. *Transportation Research Board of the National Academies*, 2010.
9. Gurmu, Z. K. and W. D. Fan, Artificial neural network travel time prediction model for buses using only GPS data. *Journal of Public Transportation*, Vol. 17, No. 2, 2014, pp. 45–65.
10. Chen, X., Z. Cheng, J. G. Jin, M. Trepanier, and L. Sun, Probabilistic forecasting of bus travel time with a Bayesian Gaussian mixture model. *arXiv preprint arXiv:2206.06915*, 2022.
11. Ma, J., J. Chan, G. Ristanoski, S. Rajasegarar, and C. Leckie, Bus travel time prediction with real-time traffic information. *Transportation Research Part C: Emerging Technologies*, Vol. 105, 2019, pp. 536–549.
12. Wessel, N., J. Allen, and S. Farber, Constructing a routable retrospective transit timetable from a real-time vehicle location feed and GTFS. *Journal of Transport Geography*, Vol. 62, 2017, pp. 92–97.
13. Graves, E., S. Zheng, L. Tarte, B. Levine, and A. Reddy, Customer journey time metrics for New York city bus service using big data. *Transportation Research Record*, Vol. 2673, No. 9, 2019, pp. 1–10.
14. Cortés, C. E., J. Gibson, A. Gschwender, M. Munizaga, and M. Zúñiga, Commercial bus speed diagnosis based on GPS-monitored data. *Transportation Research Part C: Emerging Technologies*, Vol. 19, No. 4, 2011, pp. 695–707.
15. Aemmer, Z., A. Ranjbari, and D. MacKenzie, Measurement and classification of transit delays using GTFS-RT data. *Public Transport*, Vol. 14, No. 2, 2022, pp. 263–285.
16. Zhang, L., J. Weng, and Z. Chen, Characteristic Analysis of Bus Travel Speed on Commuting Corridors Based on GPS Data. In *CICTP 2014: Safe, Smart, and Sustainable Multimodal Transportation Systems*, 2014, pp. 1443–1453.

17. Kolassa, S., Why the “best” point forecast depends on the error or accuracy measure. *International Journal of Forecasting*, Vol. 36, No. 1, 2020, pp. 208–211.
18. Ester, M., H.-P. Kriegel, J. Sander, X. Xu, et al., A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, 1996, Vol. 96, pp. 226–231.
19. National Academies of Sciences, Engineering, and Medicine and others, *Transit Capacity and Quality of Service Manual, Third Edition*. Washington, DC: The National Academies Press, 2013.
20. Yang, J., N. A. Zaitlen, M. E. Goddard, P. M. Visscher, and A. L. Price, Advantages and pitfalls in the application of mixed-model association methods. *Nature genetics*, Vol. 46, No. 2, 2014, pp. 100–106.
21. Laird, N. M. and J. H. Ware, Random-effects models for longitudinal data. *Biometrics*, 1982, pp. 963–974.
22. Ho, T. K., Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, IEEE, 1995, Vol. 1, pp. 278–282.
23. Hajjem, A., F. Bellavance, and D. Larocque, Mixed-effects random forest for clustered data. *Journal of Statistical Computation and Simulation*, Vol. 84, No. 6, 2014, pp. 1313–1328.
24. Hox, J., M. Moerbeek, and R. Van de Schoot, *Multilevel analysis: Techniques and applications*. Routledge, 2017.
25. Goodwin, P. and R. Lawton, On the asymmetry of the symmetric MAPE. *International journal of forecasting*, Vol. 15, No. 4, 1999, pp. 405–408.