Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2023) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2023 Montreal 17-21 July 2023

Systemwide Variations and Factors Affecting Mixture Transit Travel Time Distributions

Yuxuan Wang[a,c,d,e,*], Catherine Morency[a,c,d,e], Martin Trépanier[b,c,d]

[a] *Polytechnique Montreal, Department of Civil, Geological and Mining Engineering, 2500 ch. de Polytechnique, Montreal, Quebec, H3T 1J4, Canada*

[b] *Polytechnique Montreal, Department of Mathematics and Industrial Engineering, 2500 ch. de Polytechnique, Montreal, Quebec, H3T 1J4, Canada*

[c] *Interuniversitary Research Centre on Enterprise Networks, Logistics, and Transportation (CIRRELT)*

[d] *Canada Research Chair in Transportation Transformation*

[e] *Canada Research Chair in Personal Mobility*

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

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

* Corresponding author Tel.: +1-514-340-4711.
E-mail address: yuxuan.wang@polymtl.ca

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., 2013). 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., 2022), transfer coordinations (Bhaskar 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.

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 (Kittelson and Associates et al., 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.

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.

2.2. Studies on transit travel time distributions

To reiterate, transit travel time distributions can provides 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 (Bhaskar 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.

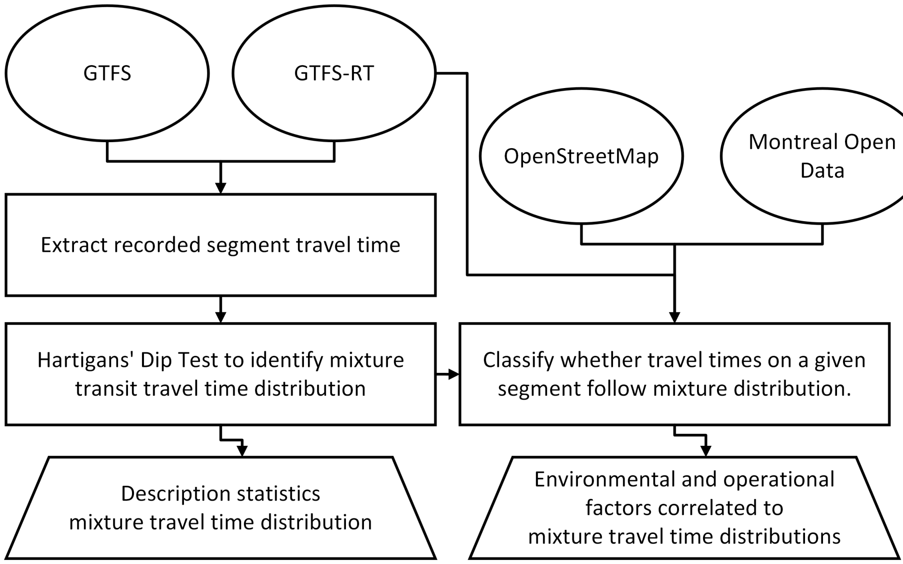

Fig. 1. Research Framework

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 regarding 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 1 and expained 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 timepoint 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.

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. However, 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.

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 (Mazloumi 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.

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. 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.

4.1. 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

Fig. 2. Map for Route 51

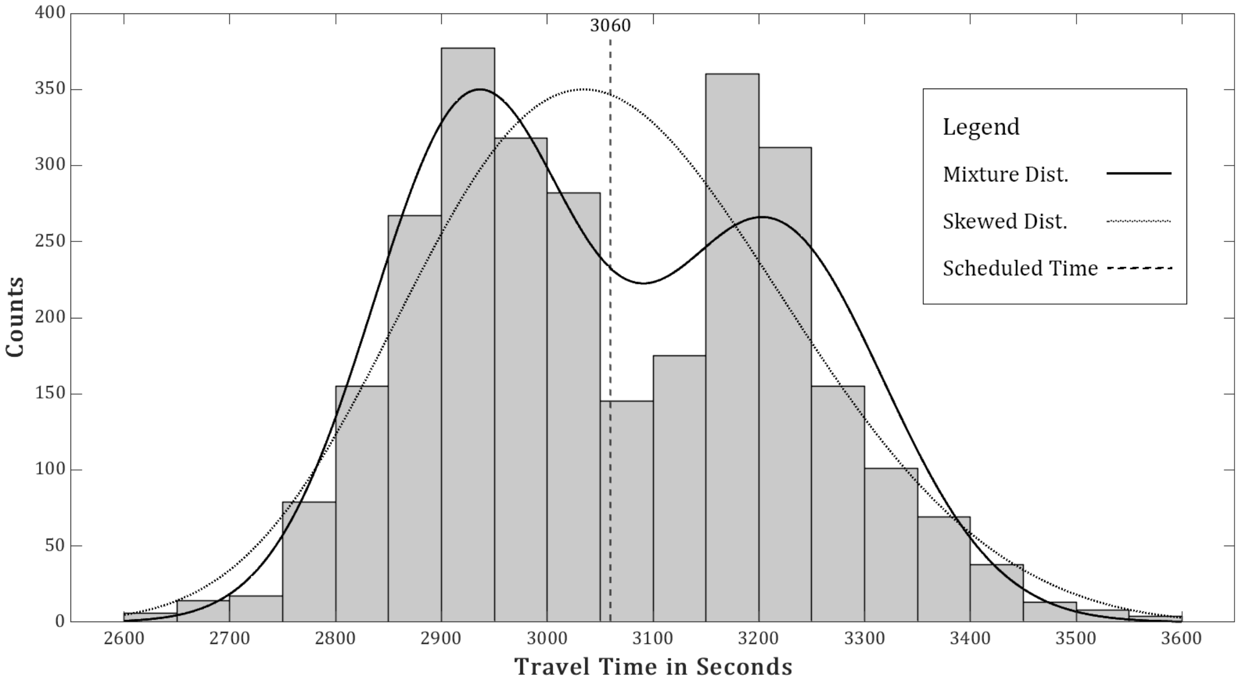

Fig. 3. Example of Mixture Transit Travel Time Distribution at Service Pattern Level (Route 51, Eastbound, Weekdays 11 AM to 12 PM, STM data between May 1, 2021 and Oct 24, 2022)

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 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 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 3 is an example taken from all trips on weekdays eastbound Line 51 between 11 AM and 12 PM. On the route map (Figure 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 pan-

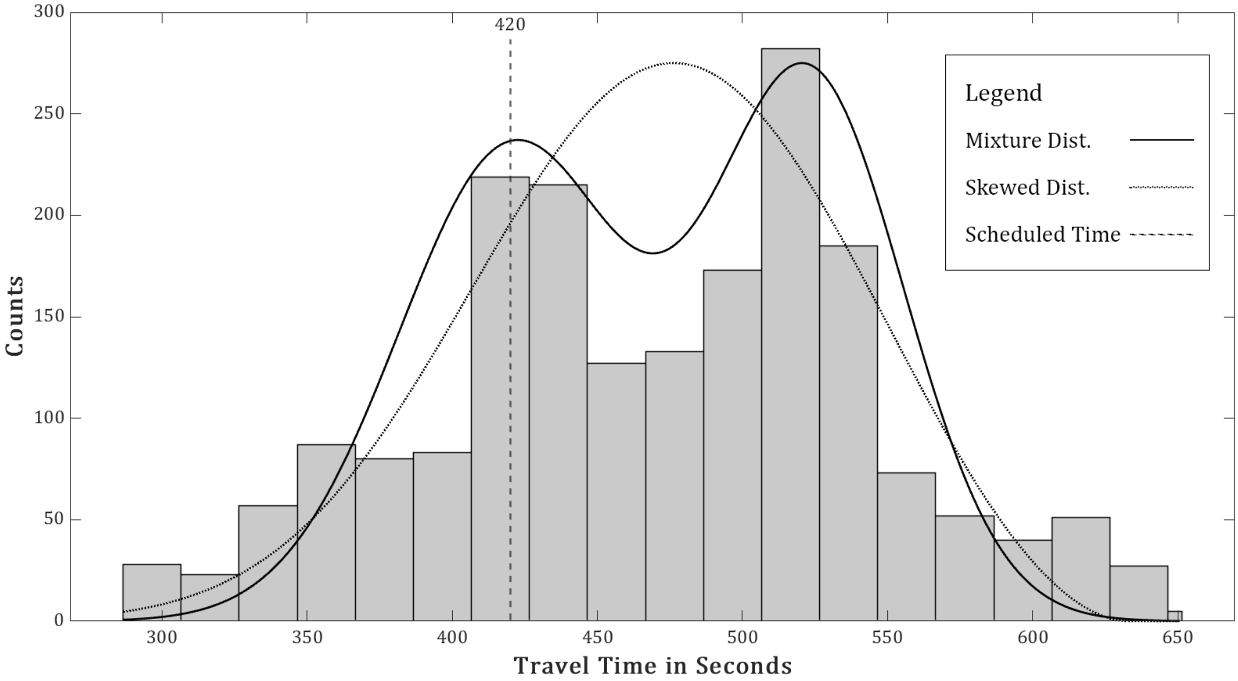

Fig. 4. Example of Mixture Transit Travel Time Distribution at Timepoint Level (Route 51, Eastbound, from Decelles / Jean-Brillant to Station Édouard-Montpetit, Weekdays Between 5 PM and 6 PM, STM data between May 1, 2021 and Oct 24, 2022)

demic, 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 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 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.

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 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 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 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.

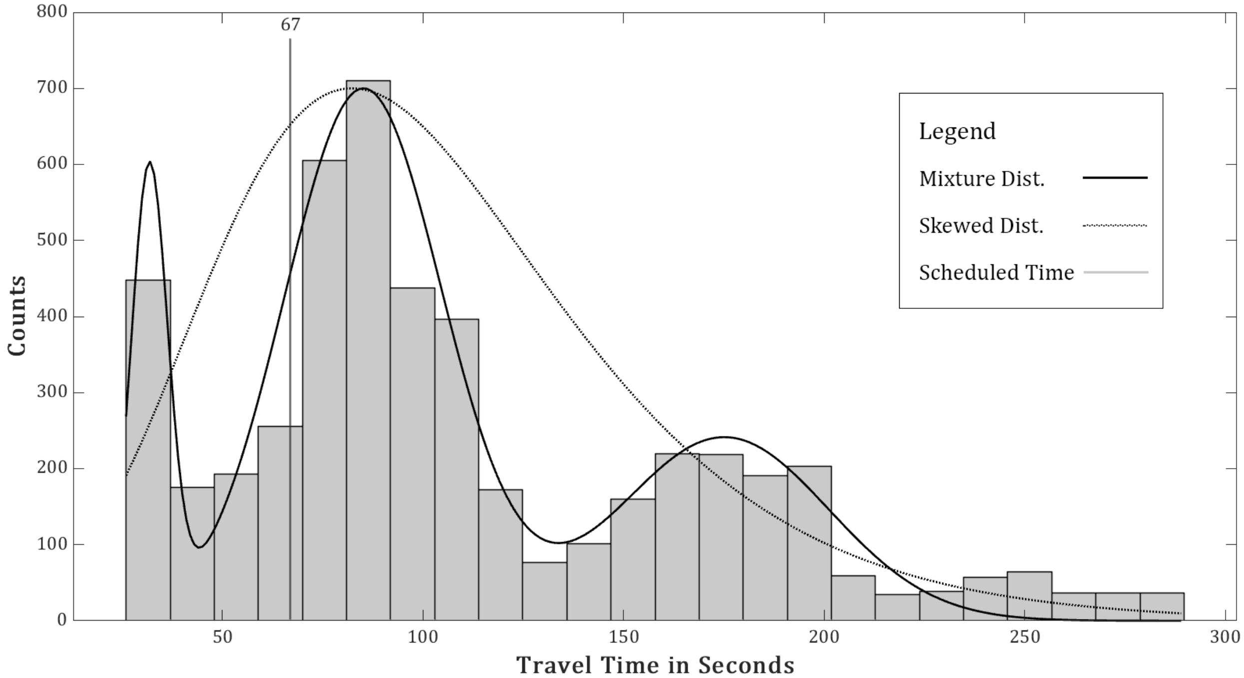

Fig. 5. Example of Mixture Transit Travel Time Distribution at Stop Level (Route 51, Eastbound, , from Decelles / Jean-Brillant to Decelles / Édouard-Montpetit, Weekdays Between 5 PM and 6 PM, STM data between May 1, 2021 and Oct 24, 2022)

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 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.

4.2. *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 chaonge could lead to a large variation in terms of percentages.

The service pattern level variations are shown in Figure 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 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.

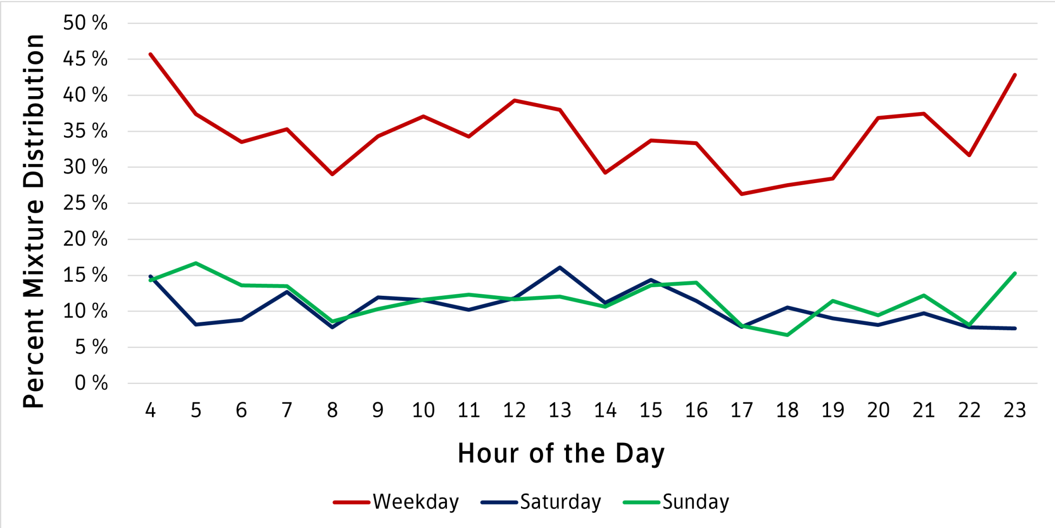

Fig. 6. Temporal variations of Mixture Transit Travel Time Distribution at Service Pattern Level (STM data between May 1, 2021 and Oct 24, 2022)

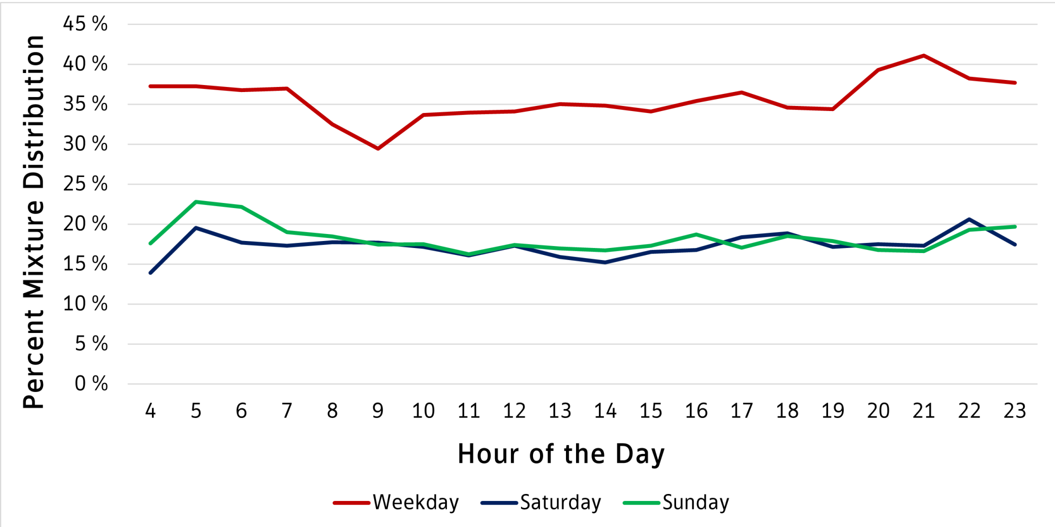

Fig. 7. Temporal variations of Mixture Transit Travel Time Distribution at Route Timepoint Level (STM data between May 1, 2021 and Oct 24, 2022)

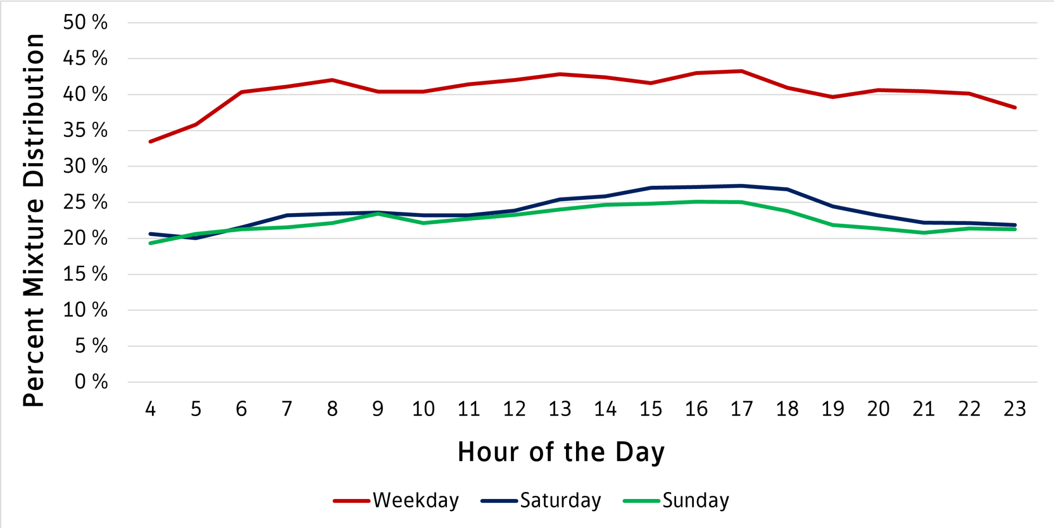

Fig. 8. Temporal variations of Mixture Transit Travel Time Distribution at Stop Pairs Level (STM data between May 1, 2021 and Oct 24, 2022)

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 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 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.

4.3. 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 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 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

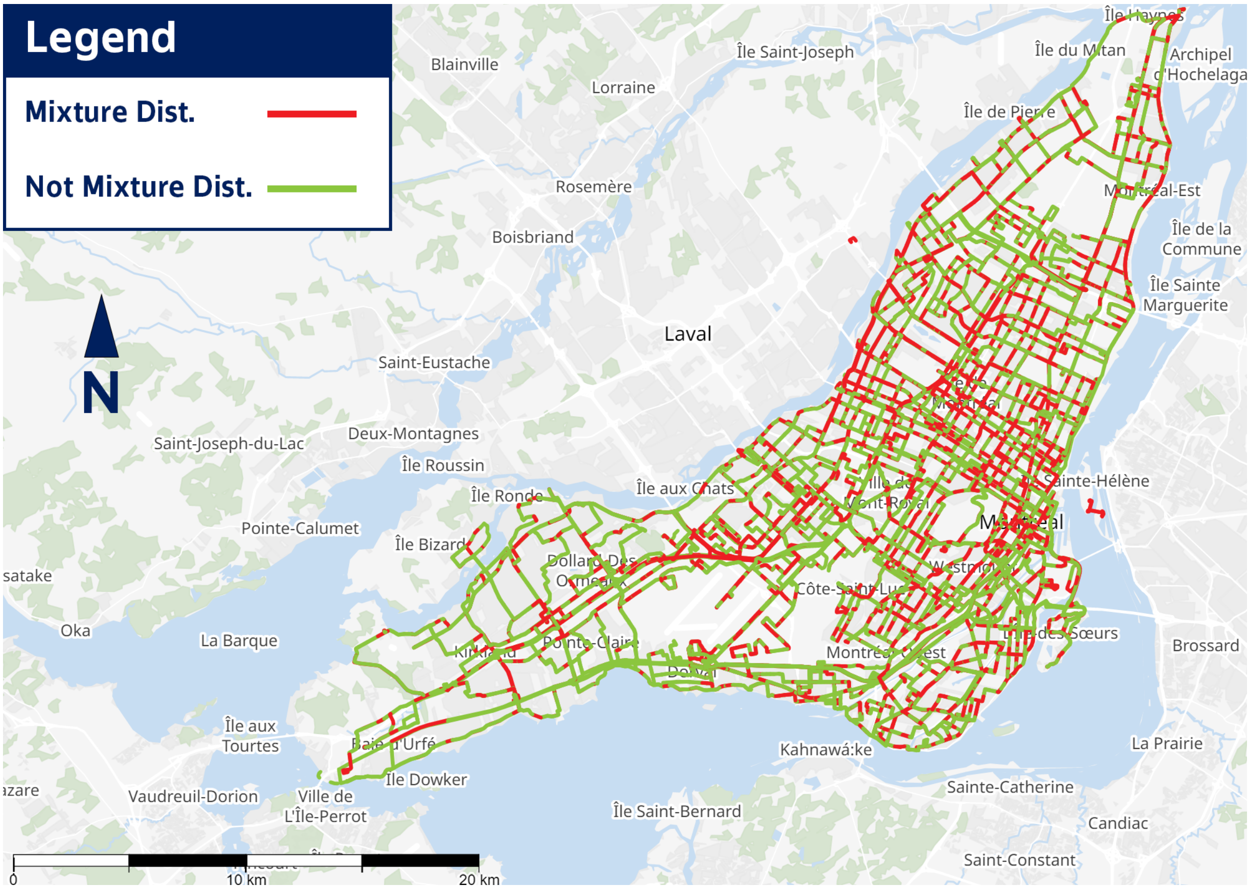

Fig. 9. Locations of Stop Level Mixture Transit Travel Time for Weekdays at 7 AM

mixture distributions, 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 the same observations for all three levels in this section.

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,

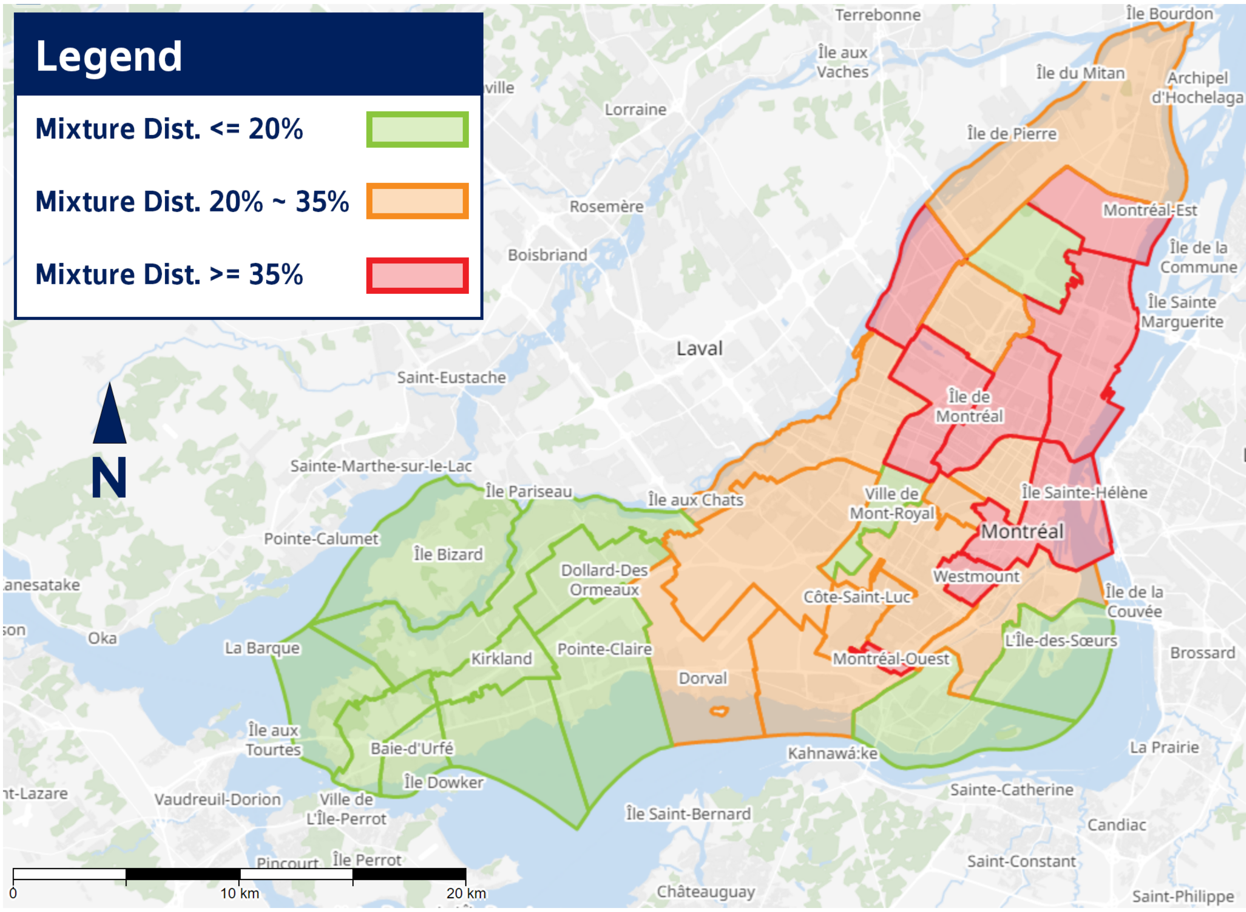

Fig. 10. Percent Timepoint Pair Segment Length Following Mixture Distributions in each Neighbourhood for Weekdays at 7 AM

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.

5.1. Variables Used in Classification Models

Based on the literature review and the observation from the descriptive statistics, we selects 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 request 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 1) are included here for readers' reference.

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 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.

Table 1. Descriptive Statistics for the Variables

	Stop Pairs				Route Timepoint Pairs				Service Pattern			
Variable	Min	Max	Avg	SD	Min	Max	Avg	SD	Min	Max	Avg	SD
Is Mixture Distribution	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

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.

Table 2. Confusion Matrix for Classification Methods at Different Analysis Levels

	Stop Pairs			Route Timepoint Pairs			Service Pattern		
K-Nearest-Neighbour		False	True		False	True		False	True
	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	True		False	True		False	True
	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
Random Forest		False	True		False	True		False	True
	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 Regression		False	True		False	True		False	True
	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

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.

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.

5.3. Identified Factors Related to Mixture Transit Travel Time Distribution

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

Table 3. Logistics Regression Coefficients

	Stop Pairs				Route Timepoint Pairs				Service Pattern			
Variable	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

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.

Finally, we try to rank the variables by determining their average effects to the results. Table 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.

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

Table 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

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.

References

Abkowitz, M., 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.

Ameijeiras-Alonso, J., Crujeiras, R.M., Rodríguez-Casal, A., 2019. Mode testing, critical bandwidth and excess mass. Test 28, 900–919.

Balcombe, R., Mackett, R., Paulley, N., Preston, J., Shires, J., Titheridge, H., Wardman, M., White, P., 2004. The demand for public transport: a practical guide. Transportation Research Laboratory.

Bhaskar, A., Cools, M., Chung, E., et al., 2017. An investigation of timed transfer coordination using event-based multi agent simulation. Transportation Research Part C: Emerging Technologies 81, 363–378.

Carrel, A., Halvorsen, A., Walker, J.L., 2013. Passengers' perception of and behavioral adaptation to unreliability in public transportation. Transportation Research Record 2351, 153–162.

Chen, C., Skabardonis, A., Varaiya, P., 2003. Travel-time reliability as a measure of service. Transportation Research Record 1855, 74–79.

Chen, X., Cheng, Z., Jin, J.G., Trepanier, M., Sun, L., 2022. Probabilistic forecasting of bus travel time with a bayesian gaussian mixture model. arXiv preprint arXiv:2206.06915 .

Coleman, M., Tarte, L., Chau, S., Levine, B., Reddy, A., 2018. A data-driven approach to prioritizing bus schedule revisions at new york city transit. Transportation Research Record 2672, 86–95.

Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. IEEE transactions on information theory 13, 21–27.

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. doi:https://doi.org/10.17226/25727.

Daskalakis, N.G., Stathopoulos, A., 2008. Users' perceptive evaluation of bus arrival time deviations in stochastic networks. Journal of Public Transportation 11, 2.

Du, Y., Deng, F., Liao, F., Ji, Y., 2017. Understanding the distribution characteristics of bus speed based on geocoded data. Transportation Research Part C: Emerging Technologies 82, 337–357.

Hartigan, J.A., Hartigan, P.M., 1985. The dip test of unimodality. The annals of Statistics , 70–84.

Ho, T.K., 1995. Random decision forests, in: Proceedings of 3rd international conference on document analysis and recognition, IEEE. pp. 278–282.

Hosmer Jr, D.W., Lemeshow, S., Sturdivant, R.X., 2013. Applied logistic regression. volume 398. John Wiley & Sons.

Kieu, L.M., Bhaskar, A., Chung, E., 2015. Public transport travel time variability definitions and monitoring. Journal of Transportation Engineering 141, Article–number.

Kittelson and Associates, Parsons Brinckerhoff, Inc., KFH Group, Inc., Texam A&M Transportation Institute, Arup, 2013. Transit Capacity and Quality of Service Manual, Third Edition. Washington, DC: The National Academies Press. doi:https://dx.doi.org/10.17226/24766.

Ma, Z., Ferreira, L., Mesbah, M., 2014. Measuring service reliability using automatic vehicle location data. Mathematical Problems in Engineering 2014.

Mazloumi, E., Currie, G., Rose, G., 2010. Using gps data to gain insight into public transport travel time variability. Journal of transportation engineering 136, 623–631.

Muller, T.H., Furth, P.G., 2001. Trip time analyzers: key to transit service quality. Transportation Research Record 1760, 10–19.

Noland, R.B., Polak, J.W., 2002. Travel time variability: a review of theoretical and empirical issues. Transport reviews 22, 39–54.

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.R., 1987. Simplifying decision trees. International journal of man-machine studies 27, 221–234.
Strathman, J.G., Dueker, K.J., Kimpel, T., Gerhart, R., Turner, K., Taylor, P., Callas, S., Griffin, D., Hopper, J., 1999. Automated bus dispatching, operations control, and service reliability: Baseline analysis. Transportation Research Record 1666, 28–36.
Trépanier, M., Morency, C., Agard, B., 2009. Calculation of transit performance measures using smartcard data. Journal of Public Transportation 12, 5.
Turnquist, M.A., Blume, S.W., et al., 1980. Evaluating potential effectiveness of headway control strategies for transit systems. Transportation Research Record 746, 25–29.
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. Master's thesis. The Ohio State University.