Appendix A
Detailed Methodology and Data

Table of Contents

A1 Data Sources

This section details the data sources used in the Vehicle-for-Hire Transportation Impact Study.

A1.1 PTC Trip Records

The Municipal Licensing & Standards (ML&S) Division currently receives trip records, shown in Exhibit A1-1, for each trip performed by a PTC since September 7, 2016. This includes the start and end points of trips located to the nearest intersection, request times, pickup times, the type of service, trip status and a shared trip indicator. Starting April 1st, 2017, trip start times were truncated to the nearest hour and waiting times and trip status have been omitted. The data is provided by licensed PTCs to ML&S on a monthly basis.

Exhibit A1-1: Current PTC Trip data provided to ML&S

Trip Records	Description
origin	Tagged to the nearest intersection, or municipality if outside city
destination	Tagged to the nearest intersection, or municipality if outside city
request time	When trip was requested (only prior to April 2017)
pickup time	truncated to hour
end time/duration	combined with start hour (e.g. 7:20 = start between 7am and 8 am, 20 min duration)
distance in km	truncated/rounded to nearest 100 m
type of service	XL, WAV, X etc.
pooled trip ID	ID changes whenever vehicle is empty for pooled service
trip status	Whether the trip was completed, or driver or passenger cancelled, only until April 2017

A1.2 Pick-up/Drop-off Activity Data through Shared Streets

PTC trip data collected by ML&S is geo-referenced to the nearest intersection in order to protect individuals' privacy. While extremely valuable for understanding travel patterns and trends, the trip data does not provide the precise resolution to understand pick-up and drop-off hotspots, nor the interaction with curbside bylaws and regulation. Curb activity data was provided by Uber and Lyft using SharedStreets as a broker. An indicative sample is shown in Exhibit A1-2.

SharedStreets is a project of the Open Transport Partnership, a non-profit funded by Bloomberg Philanthropies and private companies including Uber, Lyft, and Ford Motor Co. Data was received for a total of nine weeks in 2018 from January to September and aggregated by hour of day to a 10m spatial resolution. The SharedStreets platform filters out any data if there was only one pick-up or drop-off in the requested time-period for any 10m segment of curb to avoid this data being personally identifiable.

Exhibit A1-2: SharedStreets Pickup/Drop-off Data

This curbside activity data collected through SharedStreets is exclusive to PTCs and represents a fraction of the curbside activity at these locations. This data can be an indicator of high activity locations for other curbside uses including other for-hire vehicle services, such as taxis, or commercial delivery vehicles.

Data limitations include:

- The side of street for pick-up and drop-off is based on the direction of travel of the vehicle prior to stopping.
- For one-way streets where vehicles could be stopping on either side of the street all pick-up and drop-off activity is aggregated to the right-hand side.

A1.3 Additional Data Provided by PTCs

Additional information was requested of both major PTCs. Uber provided the following data while Lyft declined to participate.

- **Aggregate Wait Times:** Average wait times by neighbourhood and time period (e.g. Weekday AM Peak, Friday & Saturday Evenings) for select weeks to provide information on the trends in wait times from April 2017 to September 2018.
- **Aggregate Proportion of Distance Travelled by Period:** For March 2017 and September 2018, the proportion of the total distance travelled by

drivers during each activity period – cruising while waiting for a request, en-route to a request, and in-service with a passenger. This was used to estimate and validate modelling the total amount of deadheading VKT.

- **Hourly Number of Active Vehicles:** The number of vehicles active on the platform by hour for select dates below:
  - Friday Dec 15th 2017
  - Thursday March 29th 2018
  - Thursday May 31st 2018
  - Thursday Sep 13th 2018
  - Friday Sep 14th 2018
  - Saturday Sep 15th 2018
  - Saturday June 23rd 2018

A1.4 Transportation Tomorrow Survey (TTS)

The Transportation Tomorrow Survey (TTS) is a regional household travel survey conducted by the University of Toronto in collaboration with local and provincial government agencies to collect information about urban travel trends and patterns in the Greater Golden Horseshoe area. The survey has been conducted every five years since 1986 and helps local and regional governments, as well as the province and its agencies make transportation planning and investment decisions. The most recent survey was conducted in the fall of 2016 and is used to understand the characteristics of PTC and taxi travelers.

A1.5 UTTRI Resident Survey

The University of Toronto Transportation Research Institute (UTTRI) undertook a survey of City of Toronto residents in May 2019 in order to analyze the factors that influence residents' choices of when or if they choose to travel by exclusive and/or shared PTC services in the City. The survey conducted was a specialized travel survey that uses a Stated Preference (SP) technique built on Revealed Preference (RP) information of daily travel.

The survey collected information from a random sample of residents selected from a market research panel of the City of Toronto. Respondents were asked a series of questions pertaining to personal and household characteristics, information on the extent to which respondents use PTC services, and their familiarity with and perceptions of PTC services. In addition, respondents were asked to complete a series of real (revealed) and hypothetical (stated) preference questions, which were used to understand the trade-offs that people make when choosing a mode of travel in the City. These trade-offs were structured around two types of trips: commute to work or school trips, and discretionary trips made for entertainment or other purposes.

The survey was conducted using a web-based questionnaire and was administered to the members of the Canadian Viewpoint ('CanView') consumer panel. Panel members were deemed to be eligible for the survey if their home address was within the City of Toronto. In total, 723 completed responses were obtained from a total of 913 participants.

A1.6 TTC Subway Delay Data

The TTC logs each Subway delay including the time, location and duration of the incident. This dataset is available on the City's Open Data Portal.

A1.7 HERE Traffic Speed Data

Transportation Services purchases traffic speed data from HERE, a navigation company, for real-time traffic operations and historical analyses. Data is provided in five-minute bins for all the city streets where data are available. This data was used for simulating the routing of PTC trips from origin to destination.

A1.8 Bluetooth Traffic Speed Data

The Transportation Services Division monitors travel times on a number of downtown arterial streets using Bluetooth readers, originally deployed for monitoring the King Street Transit Pilot and other downtown transportation initiatives. This data provides traffic speeds at a street-link and 5-minute resolution, where data is available, and is used to measure travel time trends.

A2 Methodology

The methodology was based on new approaches and best-practices from the academic literature developed in cooperation with the University of Toronto Transportation Research Institute. The methodology has been designed to build credible and conservative assessments of the volume of PTC vehicles on City streets in the absence of data about the volume of PTC vehicles on city streets and on deadheading activity.

A2.1 Trip Routing

In order to convert trip records into PTC vehicle volumes on the City's streets, it was necessary to model the likely path vehicles took from the recorded origin to the recorded destination. Routing was performed using pgRouting, a PostgreSQL implementation of Dijkstra's Shortest Path algorithm. Trips were routed through a network of historical traffic conditions at the time of pickup data using a snapshot of travel times sourced from travel speed data from HERE (see Appendix A1.7) in order to more accurately model the paths taken by PTC drivers. Gaps in traffic data were filled in by using data models provided by HERE for each street segment by time of week.

A2.1.1 Routing Methodology before April 2017

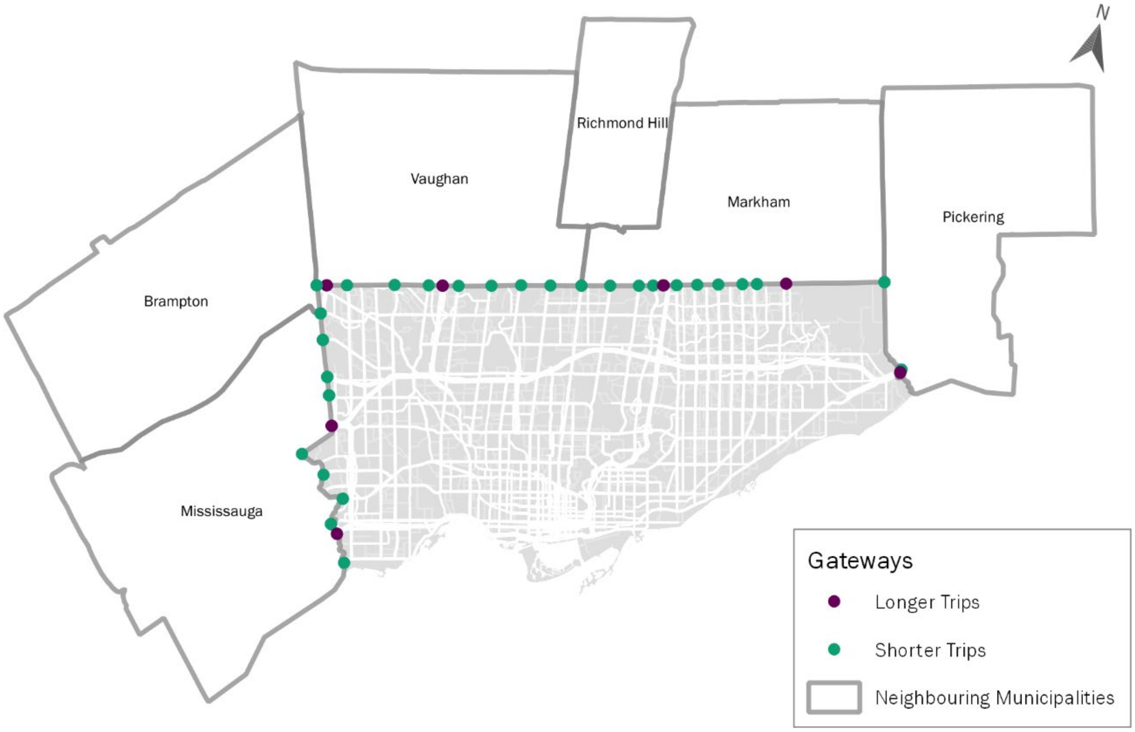
Prior to April 2017, timestamps for trip requests, pick-ups and drop-offs are accurate to the minute or second. HERE traffic data is available in five-minute increments.

The following methodology was implemented to route origins and destinations using this data:

1. **Generate a routing network:** For each five-minute bin, historical traffic data for that time was joined with models for that day of week, 15-minute period and link provided by HERE. Link IDs were duplicated for bidirectional streets and are re-drawn in the direction of travel. Source and target nodes for each link were also corrected to the direction of travel. The network mostly accounts for access restrictions and differences in road elevation but does not account for turn restrictions at intersections.
   a. **Trips within Toronto:** For each trip record, the nearest node was found in the routable HERE network. These were typically the exact same intersections. Multi-level intersections were not dealt with explicitly.
   b. **Trips from/to the six nearest municipalities outside of Toronto:** For trip records where the origin or destination was outside the city but within the six nearest municipalities, the node was assigned to be a “gateway”, an intersection on that municipality’s border representative of a major arterial or highway. Exhibit A2-1 shows the map of gateways used. If a trip started or ended more than three kilometers outside of the City of Toronto, it was assumed the PTC driver took a highway; otherwise, major arterials became candidate intersections.
   c. **Trips from/to beyond the six nearest municipalities outside of Toronto:** These trips were excluded.
   d. **Pooled trips:** Passenger segments were re-ordered to represent driver segments: trips were not routed from origin to destination, but from stop to stop, in the order the driver would have logically conducted these journeys.
2. **Route the trips:** Five-minute batches were sent to a many-many Dijkstra routing engine with the network for that time period in batches of 250 unique origins and their corresponding destinations in order to avoid memory issues. The routing engine was implemented within the pgRouting extension of the PostgreSQL database, and returned the shortest path for each origin-destination pair given traffic conditions at that time.
3. **Combine route and Origin-Destination (OD) data:** Routing results were joined back to trip (or pooled segment) ODs to link them with their trip record.

To produce neighborhood-level vehicle-kilometres travelled (VKT), the number of distinct trips for each link is multiplied by that link’s length and then aggregated by neighbourhood.

Exhibit A2-1 Map of Gateways to External Municipalities



A2.1.2 Routing Validation

Routed trips were validated by comparing routed VKT with the network distance for trips summed over the individual trip records provided to the City by PTCs.

VKT for routed trips was calculated by summing up the total length of routed street network using HERE's street geometry. Outliers – trips whose routed and recorded distances were significantly different – were then mapped to investigate potential errors. This method was used to resolve bugs in the routing process. Once any major variances were resolved, the average differences between routed and recorded distances were mapped by neighborhood to ensure no significant bias was present.

A2.1.3 Routing Methodology after April 2017

All timestamps for dates after March 30, 2017 were shifted to the start of the hour (for example '2018-09-13 07:47:30' becomes '2018-09-13 07:00:00'). Uber provided additional timestamps accurate to the minute; Lyft, did not.

Since accurate timestamps are critical to routing and to linking trips together (Section A2.2), imputed timestamps for Lyft trips were generated by bootstrapping from Uber trips. This is done for each Lyft pick-up timestamp by randomly sampling it from an Uber pick-up within a one kilometer radius, on the same day and hour. The drop-off timestamp was then self-consistently calculated from the duration of the trip. Pooled Lyft trips are treated as a single trip from the first pick-up to the final drop-off, with the timestamp of the first pick-up imputed using the methodology outlined above.

A2.2 Routing Deadheading

The previous section described the process used to estimate PTC vehicle volumes when PTC drivers are traveling with a passenger. The travel between the destination of one trip and the origin of the next (i.e. deadheading) is also a critical component of the total VKT generated by PTCs.

Since the trip data available to the City does not contain an explicit driver identifier to link individual trips together, it is necessary to model the behaviour of drivers to estimate their behaviour between passenger trips. Fortunately, the assignment of drivers to passengers is governed by the driving applications of PTC companies, and emulating these applications allows us to make educated guesses of how drivers move from one trip to the next. The process of connecting drivers with passengers is referred to as "trip linking" in this report.

A2.2.1 Linking Methodology

A PTC driver serving multiple trips over the duration of their work period will cycle between three distinct periods:

- Cruising while waiting for a passenger request (Period 1)
- Driving en-route to a request (Period 2), and
- Driving in-service of the request (Period 3).

Cruising and driving en-route to a request collectively constitutes deadheading. At the beginning and end of the work period, the driver may also commute from and to another location. The trip linking in this report only estimates the time taken and distance travelled by drivers en-route to a request. Using the data available, it is virtually impossible to reconstruct the exact service history of individual drivers. It is possible, however, to produce a set of trip linkages that, in the aggregate, resemble the real-life distribution of en-route times and distances.

The methodology used to link individual trips together is as follows:

1. **Generate feasible links:** Which trips can feasibly be linked together was determined over the course of a day in five-minute increments. For each increment, the drop-off locations of all trips ending in the increment were collected. For each drop-off, the closest 30 pick-up points of trips beginning within the next 20 minutes were found (these values were selected to make the problem computationally tractable). The set of drop-off points was then routed to the set of pick-up points using the methodology detailed in Section A2.1.1. All routes that take longer to travel than the time difference between the drop-off and pick-up were discarded. The remaining routes comprise the choice set of feasible links between drop-offs and pick-ups.

2. **Transform the feasible links into a graph:** The set of feasible links was then transformed into a directed graph. The nodes of the graph represent trips, and directed edges represent the feasible links.
3. **Determine the linking solution:** One of several graph algorithms was utilized to determine which of the feasible links were actually taken by drivers. These graph algorithms find a "matching", or set of links such that every trip's pick-up and drop-off are each joined to at most one link. Differences between algorithms are discussed below. For this report, the batched fleet minimizing algorithm is used, as the distribution of trip wait times it produces most closely matches the actual distribution found in the data (Section A2.2.2).
4. **Convert the solution to volumes:** The linking solution can then be treated as a set of trips, and their paths converted to neighbourhood VKTs as detailed in Section A2.1.1. A set of trips linked together can then also be treated as the path taken by a hypothetical driver over the course of their work period. This can be used to produce a variety of interesting measures including, for example, the total amount of time drivers spent between servicing trips. Drivers are assumed to begin their work period at the pick-up of their first trip, and end at the drop-off of their last trip.

The linking algorithms are simplified versions of driver-to-passenger matching algorithms used by PTC companies and autonomous vehicle simulations[1]. They include:

- **Greedy**: Connect each drop-off with the feasible pickup with the shortest travel time, handling the drop-offs in order of time. This is similar to Uber's driver-passenger matching algorithm[2], though when there are multiple drivers and passengers Uber has been reported to make additional corrections to minimize wait times.
- **Fleet-minimizing**: Link as many trips together as possible (without connecting multiple drivers to the same trip), thereby minimizing the number of drivers (more precisely the number of driver work periods) needed to satisfy all trips. This algorithm is outlined by Vazifeh et al. 2018[3], who used it to determine the minimum size of a hypothetical automated vehicle fleet to service New York City's taxi demand. In practice, it forces drivers on average to wait and travel for longer between trips than for greedy linking (conversely, greedy linking requires more drivers to operate). It also acts on the entire graph at once (unlike the greedy algorithm, which handles each trip in order of time), and so is realistic only in cases where a PTC company predicts trip demand several hours into the future.
- **Batched fleet-minimizing**: Divide up the feasible links graph into time increments of t_{bin}. Then, use the fleet-minimizing algorithm to link the

---

[1] Hanna, J. P., Albert, M., Chen, D., & Stone, P. (2016). Minimum cost matching for autonomous carsharing. *IFAC-PapersOnLine, 49*(15), 254-259.
[2] Stanford University School of Engineering. (2018, April 3). Dawn Woodward: How Uber matches riders and drivers to reduce waiting time [Video file]. Retrieved from https://youtu.be/GyPq2joHZv4
[3] Vazifeh, M. M., Santi, P., Resta, G., Strogatz, S. H., & Ratti, C. (2018). Addressing the minimum fleet problem in on-demand urban mobility. *Nature, 557*(7706), 534.

trips in each increment (potentially to unlinked trips from past increments) in order of time. This is the algorithm used for the main report, with t_{bin} = 1 minute.

Linking was performed for October 20, 2016 and September 13, 2018, representative days near the beginning and near the end of the study period, respectively.

As with routed trips outlined in Section A2.1, this methodology uses the shortest-time route connecting trips, without incorporating turn restrictions or accounting for driver behaviour. It additionally does not model drivers as agents with, for example, limits on how much they wish to drive, or preferred geographic regions for servicing trips, as data on these are not available. This results in some work periods that are unrealistically long, but are in the minority (e.g. about 10% of the work periods from the 2016 trip linking solution are longer than 4.6 hours, but the median work period length is 1.4 hours). Trip linking alone is also unable to constrain driver behaviour while cruising – since drivers may pause or continue driving during this time – or commuting to and from other locations.

A2.2.2 Linking Optimization and Validation

For the batched fleet-minimizing algorithm, t_{bin} is a tunable free parameter. The maximum feasible deadheading time, t_{max}, can also be decreased from 20 minutes. To tune these two values, a large set of link solutions are calculated using t_{bin} and t_{max} values selected by a Bayesian hyperparameter optimizer. Once each solution is available, the passenger wait time can be derived from the driver en-route time for the selected link, and the distribution of linked passenger wait times can be compared to the distribution of true passenger wait times by their Jensen-Shannon divergence. This optimization was tried on October 20, 2016 and March 30, 2017, and for both days a t_{bin} of approximately one minute and a t_{max} of 20 minutes produces linking results that best reproduce the distribution of true wait times. For October 20, the first quartile, median and third quartile of the linked wait times are 3.3, 5.2, and 7.7 minutes, respectively, while the true wait times are 3.7, 5.5, and 7.8 minutes. The same t_{bin} and t_{max} were also used for September 13, 2018 (since accurate trip request timestamp data was not available for after March 2017, preventing optimization).

The VKT was compared to aggregate distance (by period) data provided by Uber. It shows that 35 to 40% of total VKT is spent cruising, 5 to 10% en-route, and 55% in-service, meaning the ratio of en-route to in-service VKT is about 10 to 20%. This ratio is around 15% for both the October 20 and September 13 linking solutions, consistent with the data.

Uber also provided the hourly number of unique active (in-service or deadheading) vehicles for dates throughout 2018. A linear fit of the number of active vehicles against the total number of trips led to the following estimated model (adjusted R^2 = 0.964):

$$N_{Vehicles} = 0.48N_{Trips} + 274$$

This is equivalent to about two trips per vehicle (somewhat of an overestimate, since the fit is to Uber vehicles and trips only, and some Uber drivers simultaneously drive for Lyft). The linking solution for October 20 predicts 30% fewer vehicles, and for September 13 predicts 15% fewer trips. This indicates that trip linking is more efficient than real PTC operations, and introducing driver work constraints may lead to a more accurate estimate.

While trip linking does not estimate cruising VKT, it does estimate the cruising *time*. The aggregate cruising time varies considerably depending on the linking algorithm used, however, ranging from only 15% of the aggregate in-service time for the batched fleet-minimizing algorithm to 33% for the day-long fleet-minimizing algorithm. If it is assumed that driving speeds during cruising are not very different than those in-service, the data from Uber indicates cruising time is closer to 60 to 70% of the in-service time. It is unclear, however, whether this data includes drivers making trips they would have completed otherwise while having the Uber app open, which would inflate the cruising VKT fraction. Since there is very little data available on driver behaviour during cruising, it is difficult to determine why trip linking underestimates it.

A2.3 Example Routing and Linking

Exhibit A2-1: Example of a Sequence of Routed and Linked Trips

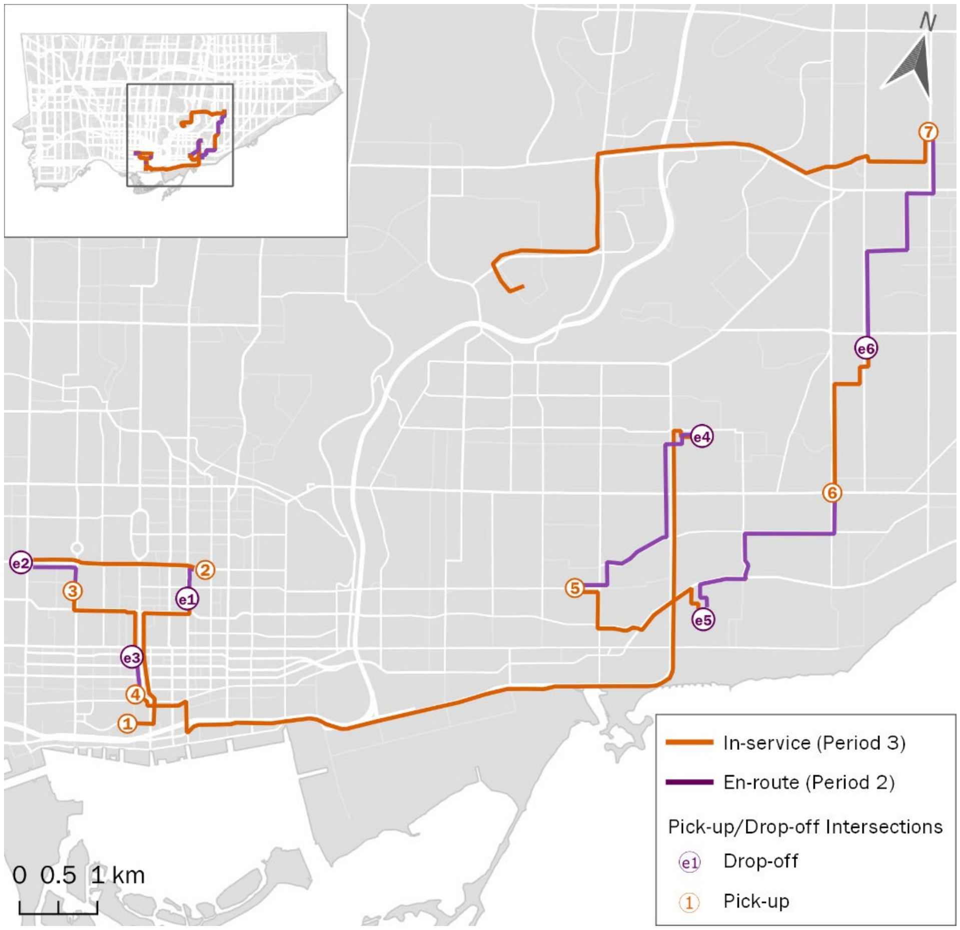


Exhibit A2-3 shows a set of origin/destination points for the evening of October 20, 2016 that have been routed and linked together into the path taken by a hypothetical driver using the methodology described in this Section. Each trip, or trip en-route to the next passenger, represents the shortest travel-time path between the origin and destination given traffic patterns at the time (see Section A2.1.1). Connections from one trip to the next are the result of the batched fleet-minimizing algorithm (see Section A2.2.1), which tries to connect as many trips to available and close-by drivers as possible. The order in which the trips and en-route trips were taken is labelled on the map.

This example also shows that drivers within the downtown core can freely be diverted by trip linking to service other areas of the city. This may or may not be realistic depending on typical driver behaviour and how the PTC driver-passenger matching service functions. Additional data on these could potentially make trip linking considerably more accurate.

A2.4 Estimating Transit Alternatives to PTC Trips

Estimated travel attributes of the fastest transit alternatives to PTC trips were determined using OpenTripPlanner (OTP), an open-source software suite that provides transportation network analysis services given general transit feed specifications (GTFS) and OpenStreetMap (OSM) data.

GTFS data for the TTC were downloaded from the Transitland Feed Registry[4], and OSM data from OSM Extracts by Interline[5]. OTP was run locally, and transit alternatives for a given PTC trip were estimated by passing its origin location, destination location, and time as inputs. OTP outputs multiple trip itineraries for each set of inputs; the one with the fastest travel time was selected for this analysis.

A3 Transportation Network Impacts Studies in Other Jurisdictions

To date, there have been a number of congestion studies that have been completed by municipalities, academics, and consultancies across North America, varying in scope and overall approach. A selection of these are summarized in Exhibit A3-1. Most of these studies are in agreement that the introduction of PTCs are resulting in additional vehicle-kilometers to the street networks on which they're operating, but the connection to resulting changes in congestion is less certain.

The most comprehensive study to date was published in October 2018 by the San Francisco County Transportation Authority (SFCTA). It attempted to isolate the total congestion that PTCs were adding to its street network, using a combination of its local long-term travel demand forecasting model, the application of traditional volume-delay functions to convert observed speeds to volumes, and estimated PTC trip volumes. This model was also used to estimate traffic volumes in the absence of PTCs, in order to provide an alternative scenario against which to compare current conditions.

[4] Transitland Feed Registry. Retrieved from https://transit.land/feed-registry/.
[5] OSM Extracts by Interline. Retrieved from: https://www.interline.io/osm/extracts/

Exhibit A3-1: Summary of Congestion Studies in Other Jurisdictions

Jurisdiction (Author)	Publication Date	Summary of Findings
Denver (Alejandro Henao, PhD Disseration)[6]	May 2017	**Approach** • Uses data collected personally as an Uber/Lyft driver to directly measure VMT for each trip, compared to the trip it replaced (based on a passenger survey). **Findings** • On average, the VMT for PTC trips was found to be 84.6% greater than the trip it replaced. • This increase would correspond to an additional 5.5 billion miles travelled in the USA in 2016 if the findings in Denver were transferrable to the rest of the country.
United States (Schaller Consulting)[7]	July 2018	**Approach** • Uses a variety of simplified scenarios varying the amount of total PTC trips that are shared, and the modes from which these trips are replacing, to estimate the total additional miles added to the transportation network. **Findings** • On average, each additional PTC trip is associated with a 41 to 180% increase in kilometers travelled relative to the mode it's replacing, on average. • In the USA's nine largest metropolitan areas, PTCs are estimated to have added 5.74 million miles in 2017.
San Francisco (San Francisco County Transportation Authority)[8]	October 2018	**Approach** • Two separate analysis methods: one using historical INRIX probe-based data, and the use of volume-delay functions to isolate the impact of PTC volumes from overall volume changes, and the second using a travel demand model to estimate the congestion impacts estimated over a series of scenarios. **Findings** • An overall decrease in arterial speeds between 2009 and 2016 of 26% and 27% in the AM and PM peak periods. • An estimated 55-65% of the overall changes in speed due to the contribution of PTCs. • An estimated 44-47% of the overall increase in vehicles-miles travelled due to the contribution of PTCs.

[6] Henao, Alejandro. (2017). Impacts of Ridesourcing – Lyft and Uber – on Transportation Including VMT, Mode Replacement, Parking, and Travel Behavior. Retrieved from https://pqdtopen.proquest.com/pubnum/10265243.html?FMT=AI

[7] Schaller Consulting. (2018). The New Automobility: Lyft, Uber and the Future of American Cities. Retrieved from http://www.schallerconsult.com/rideservices/automobility.pdf

[8] San Francisco County Transportation Authority. (2018). TNCs & Congestion. Retrieved from: https://www.sfcta.org/sites/default/files/2019-02/TNCs_Congestion_Report_181015_Final.pdf