Research in Transportation Economics 72 (2018) 27–36

Contents lists available at ScienceDirect

Research in Transportation Economics

journal homepage: www.elsevier.com/locate/retrec

The usage of location based big data and trip planning services for the estimation of a long-distance travel demand model. Predicting the impacts of a new high speed rail corridor

Carlos Llorca[a,*], Joanna Ji[b], Joseph Molloy[c], Rolf Moeckel[a]

[a] Technical University of Munich, Arcisstr. 21, 80333, Munich, Germany
[b] PTV Group, Haid-und-Neu-Str. 15, 76131, Karlsruhe, Germany
[c] IVT, ETH Zürich, Ch-8093, Zürich, Switzerland

ARTICLE INFO

JEL classification:
R41
R42

Keywords:
Travel demand model
Long-distance travel
High-speed rail
Location-based social network
Online trip planning

ABSTRACT

Travel demand models are a useful tool to assess transportation projects. Within travel demand, long-distance trips represent a significant amount of the total vehicle-kilometers travelled, in contrast to commuting trips. Consequently, they pay a relevant role in the economic, social and environmental impacts of transportation. This paper describes the development of a microscopic long-distance travel demand model for the Province of Ontario (Canada) and analyzes the sensitivity to the implementation of a new high speed rail corridor.

Trip generation, destination choice and mode choice models were developed for this research. Multinomial logit models were estimated and calibrated using the Travel Survey for Residents in Canada (TSRC). It was complemented with location-based social network data from Foursquare, improving the description of activities and diverse land uses at the destinations. Level of service of the transit network was defined by downloading trip time, frequency and fare using the planning service Rome2rio.

New scenarios were generated to simulate the impacts of a new high speed rail corridor by varying rail travel times, frequencies and fares of the rail services. As a result, a significant increase of rail modal shares was measured, directly proportional to speed and frequency and inversely proportional to price.

1. Introduction

The planning and the design of transportation infrastructure depends on complex and interacting economic, social, environmental and technical factors. An accurate knowledge of current and future transportation demand is an important decision factor. With this purpose, travel demand models predict the number, destination and modal choice of trips. These characteristics depend, among other factors, on the frequency and length of the trips. The travel demand may be segmented into urban travel demand, such as short-distance daily commuting, and long-distance travel demand, such as overnight trips or touristic trips.

Traditionally, transport modelers paid most attention to daily short-distance urban traffic demand and their corresponding models, based on the fact that the number of trips is much larger compared to long-distance travel. The models for long-distance trips appeared later and usually transferred parameters from urban models. Some studies highlighted the importance of the long-distance travel demand market (sometimes mixed with the so-called intercity travel demand), based on their contribution to the vehicle kilometers travelled (Dargay & Clark, 2010), or motivated by the interest on analyzing high speed rail networks (Outwater et al., 2010). Data availability is reported generally as the major concern for long-distance models (Miller, 2004), since households travel surveys were typically designed for urban travel demand.

This paper shows the development of a long-distance passenger travel demand model for the province of Ontario (Canada) as part of a province-wide transportation model. The paper summarizes the model development, including the use of location-based big data and online trip planning services for its estimation. Additionally, the paper analyzes the impacts of a new high speed rail corridor by testing the sensitivity of the model to changes in travel time, price and frequency.

2. Literature review

Long-distance travel demand pays a significant role in the transportation system. While around 75% of trips in US are less than 15 km, they account for only 28.9% of vehicle-kilometers travelled (Schiffer,

* Corresponding author.
E-mail addresses: carlos.llorca@tum.de (C. Llorca), joanna.ji@ptvgroup.com (J. Ji), joseph.molloy@ivt.baug.ethz.ch (J. Molloy), rolf.moeckel@tum.de (R. Moeckel).

https://doi.org/10.1016/j.retrec.2018.06.004
Received 30 October 2017; Received in revised form 8 June 2018; Accepted 11 June 2018
Available online 14 June 2018

2012). On the contrary, trips over 150 km account for 1% of trips, but represent 15.5% vehicle-miles travelled. Similarly, 50% of all passenger kilometers in Europe corresponded to trips beyond 100 km (Rich & Mabbit, 2011).

Two main approaches were found when modeling long-distance travel demand. Some authors focused on the estimation of long-distance demand as a required component for large scale models (state wide, country wide or even larger scales) while others performed a wide variety of case-specific studies, concentrated solely on specific corridors (such as high speed rail lines) or certain demand or supply segments (such as vacation trips).

2.1. Large scale long-distance travel demand models

The integration of long-distance demand into large scale models started with very simple approaches. According to Giamo and Schiffer (Giamo & Schiffer, 2005, pp. 2–212), some statewide travel demand models did not even account for long-distance travel demand. Many others only forecasted future trips based on trip tables, which cannot capture sensibilities of person, environment or trip variables. However, the recent statewide models contributed to the development of long-distance travel demand sub-models (Miller, 2004).

An example in this direction was the long-distance model for Ohio (US), as part of the statewide model (Erhardt, Freedman, Stryker, Fujioka, & Anderson, 2007). Although the authors formulated a complete framework for trip generation and trip destination, mode choice modules could not be estimated because of the lack of air and transit observations in the survey. Therefore, the impact of transit frequency and schedules could not be considered. The TRANS-TOOLS project proposed a long-distance travel demand model for Europe (Rich & Mabbit, 2011). According to the authors, TRANS-TOOLS model was the largest model ever with respect of population and covered area. This resulted in a very coarse level of resolution, excessive complexity and long runtimes, making it unable to test transport planning policies. Lu, Zhu, Luo, and Lei (2015) developed a long-distance travel demand model considering multinomial logit models for trip generation, destination choice and mode choice, built on a multi-level nested structure. The model only applied to a small set of seven destination alternatives in the province of Guangdong (China), focusing explicitly on intercity travel and discarding rural areas. The Federal Highway Administration (2015) proposed a national long-distance passenger travel demand model combining different state long-distance travel surveys. The model used multinomial logit models to predict the number of trips, their duration and their travel party. Destination choice and mode choices were jointly estimated as a nested logit model. The California statewide model for High Speed Rail (Outwater et al., 2010) has tested the impact of high speed rail, but considering the whole long-distance transport demand. The high speed rail mode was defined as a new modal alternative. Then, it was required to estimate a mode choice model based on a stated preference survey. For the state of North Carolina, a long-distance person travel demand model was developed that explicitly accounted for the three closest transit stations to the origin zone and the three closest transit stations to the destination zones (Moeckel, Fussell, & Donnelly, 2015). The model was applied to test rail and regional bus network improvements.

2.2. Case-specific studies

Despite those large scale examples, most of studies about long-distance travel demand could be classified as case-specific or corridor specific, as they do not provide a framework for predicting the complete long-distance travel demand in a certain state or province. Limtanakool, Dijst, and Schwanen (2006) considered the effect of some land-use factors, such as density and use specialization, on long-distance trips in the Netherlands. They could only estimate a binary choice model considering the decision between car and train. Van Nostrand, Sivaraman, and Pinjari (2012) proposed a model only for vacation trips in US, without taking into account long-distance business demand. Other authors (LaMondia, Fagnant, Qu, Barrett, & Kockelman, 2016) tested the impact of automated vehicles in long-distance travel demand, reducing the share of conventional cars and airplanes for trips under 500 km.

The prediction of the effects of building high speed rail networks focused a significant number of case-specific models. Gutierrez (2001) analyzed the changes in accessibility because of the implementation of a new high speed rail in Spain. Different definitions of accessibility were considered (by testing different distance decay functions and potential attractors). The results identified significant impacts, not only in the geographical area where the new line was going to be built, but also in surrounding areas. However, the changes in accessibility increased the inequality in accessibility between areas. Wong and Habib (2015) explored the effects of accessibility to high speed rail stations in the Windsor-Quebec corridor (Canada) using a joint revealed preference and stated preference survey. The expected modal shares in this corridor were characterized by calibrating and validating a nested logit mode choice model, where auto, rail, high speed rail, bus and air were considered as choice alternatives. The effect of accessibility to stations was found to be relevant for the success of a new high speed rail system. Some studies have analyzed long-term passenger counts (and their temporal variation) of high speed rail in Spain (Gundelfinger-Casar & Coto-Millán, 2017), in Japan (Demizu, Li, Schmöcker, Nakamura, & Uno, 2017) or in China and Taiwan (Li & Schmöcker, 2017). However, these approaches did not provide an estimation of the potential usage of the projected high speed rail infrastructure and the subsequent effects over the rest of travel demand.

2.3. Research motivation

As a result of the abovementioned long-distance travel demand approaches, previous research identified further data requirements and research needs. Miller (2004) identified several unsolved issues regarding long-distance travel demand estimation. First, data collection could be improved by facilitating access to private-operator transit data. Secondly, Miller described the use of stated preference survey as complicated, expensive and potentially biased, making the test of new modes (such as high speed rail) extremely uncertain. Thirdly, the level of aggregation of long-distance models was in general too coarse in space, time and modes (especially regarding access and egress), thus a finer resolution was recommended. Miller proposed additional model capabilities that include non-resident visitors or long-distance through traffic.

Additionally, the datasets used in destination choice models need also further improvement. Van Nostrand et al. (2012) identified the need of improving destination attractiveness data. The California Statewide Model for High Speed Rail failed when estimating trips to certain touristic regions, as only population and employments at a coarse spatial resolution were considered as attractor (Outwater et al., 2010). The Federal Highway Administration (2015) recommended collecting better data of long-distance trips through household travel surveys or using smartphone-based collection methodologies. Regarding transport supply, the provision of transit network data at the country level was recommended, especially for long-distances buses. Although the amount of available data for the other modes was quite significant, its conversion to a usable transit network required very long processing times.

Based on the review of the existing literature, and motivated by the previously mentioned data availability issues, this paper has the objective of developing a long-distance travel demand model for the province of Ontario (Canada). The paper presents and discusses a methodology to enrich available travel survey datasets with location-based big data and trip planning services, as an additional data source of both selected and non-selected destination and mode alternatives. After implementing the model, the paper also explores its sensitivity to

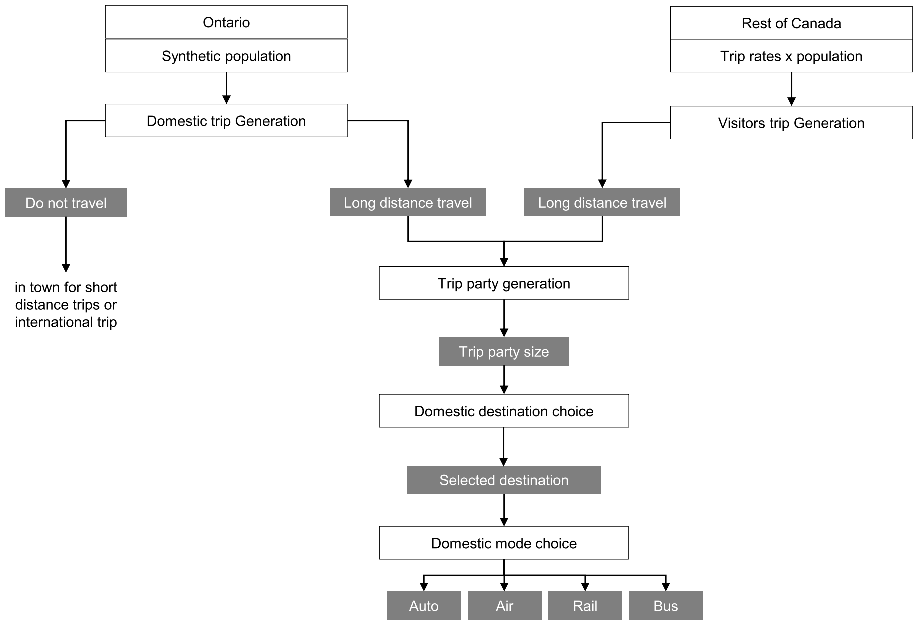

Fig. 1. Model framework.

travel time, price and frequency, based on a scenario of a new high speed rail corridor.

3. Materials and methods

This section describes the model development including its concept and theoretical framework (subsection 3.1), the data collection (subsection 3.2) and the definition of scenarios (subsection 3.3).

3.1. Model framework

The model is designed as part of a four-step travel demand model (although the last step, traffic assignment is not included here, since it is covered within a province-wide model where the long- and short-distance travel is assigned at the same time).

Fig. 1 shows the structure of the long-distance travel demand model for domestic trips. Ontario residents are represented by a synthetic population. For visitors arriving from the rest of Canada, the same structure was used, except socio-demography information about travelers was not known. Moreover, international trips are explicitly modelled using separate sub-models, which are not included in this paper.

The proposed approach is microscopic (individual persons, households and trips are simulated) and the different stages are modelled using a multinomial logit formulation (Erhardt et al., 2007; Rich & Mabbit, 2011). Previous approaches, however, have used survey data only and not the mix of data presented here. Logit models provide the probability of selecting an alternative within a set of independent choices, and are formulated as shown in equations (1) and (2) (Ben-Akiva, 1974; Train, 2009).

$$P_{pj} = \exp(U_{pj}) / \sum_{k=1}^{k=K} \exp(U_{pk}) \tag{1}$$

Where:

- P_{pj} is the probability of selecting the alternative j by the individual p.
- U_{pj} is the utility of selecting the alternative j for the individual p, described in equation (2).
- $k = 1, 2, \ldots K$ is the set of alternatives.

$$U_{pj} = \sum_{s=1}^{1=S} \beta_s x_s \tag{2}$$

Where:

- $s = 1, 2, \ldots S$ is a set of either individual-related, alternative-related or exogenous explanatory variables.
- β_s is the coefficient of the explanatory variable s.
- x_s is the value of the explanatory variable s.

With respect of the domestic trip generation, a multinomial logit model was proposed to predict individuals' decisions whether to travel or not. The alternative choice set is defined by four alternatives: "stay at home", "start a long-distance daytrip", "be away in an overnight long-distance trip" or "start or end an overnight long-distance trip". This way the model was able to select synthetic persons in Ontario to make a long-distance trips or to stay at home and be available for other trip types. The influence of person characteristics, household characteristics and potential accessibility of the origin zone (as described by equation (3)) was taken into account.

$$Accessibility_i = \sum_{j=1}^{j=N} Population_j^{\alpha} \cdot \exp(\beta \cdot tt_{ij}) \tag{3}$$

Where:

- $Accessibility_i$ is the potential accessibility of zone i.
- $j = 1, 2, \ldots N$ are zones.
- $Population_j$ is the population of zone j.
- tt_{ij} is the travel time (by car) between zone i and j.
- α: calibration parameter for population.
- β: calibration parameter for travel time.

The total number of trips made by visitors is obtained by multiplying observed trip rates by the population of the origin zone (see definition of zones in section 3.2.2). Once trips are generated, the trip party was randomly generated, based on observed trip party frequencies (from travel surveys).

Domestic destination choice models were estimated as logit models to predict the probability of selecting a destination zone, as a function of its attraction, as well as logsums obtained from the mode choice models (defined as shown in equation (4)). They described not only the distance between origin and destination, but also the availability and level of service of the travel modes. The destination choice model considered single destination trips, as shown by more than 96% of the survey records.

$$Logsum_{ij} = \sum_{m=1}^{m=M} \exp(U_{ij,m}) \tag{4}$$

Where:

- $Logsum_{ij}$ is the *logsum* between origin *i* and destination *j*.
- $U_{ij,m}$ is the utility of travelling from origin *i* to destination *j* using the mode *m*, obtained from an estimated mode choice model.
- $m = 1, 2, \ldots M$ is a set of alternative travel modes.

Lastly, mode choice models were defined as multinomial logit models, were the set of alternatives corresponded with the available modes: car, air, rail and bus. Apart from mode level of service measures, personal and household attributes were included in the estimation for domestic trips starting in Ontario. For travel behavior estimations for Canadians who live outside of Ontario, socio-demographic attributes could not be represented because those residents were not included in the synthetic population for Ontario.

This model framework required the three sub-models (trip generation, destination choice and mode choice) to be econometrically estimated. The software R and its packages mlogit (Croissant, 2017) and mnlogit (Hasan, Wang, & Mahani, 2016) were used. subsection 3.2. describes the process of collecting the data for such model estimation, while subsection 4.1. will provide the estimation results.

3.2. Data collection

The estimation of the long-distance travel demand model is primarily based on the use of two travel surveys. One was conducted for domestic long-distance trips within Canada (Travel Survey of Residents in Canada - TSRC) and the other one covered international to and from Canada (International Travel Survey - ITS). The use of additional and innovative data sources was required to complete the model estimation, as explained later in subsection 3.2.3.

3.2.1. Travel surveys

Both TSRC and ITS travel surveys are designed primarily to measure domestic and international tourism, including trip characteristics, activities at the destination and trip expenditures. The TSRC for domestic travel is conducted by telephone and includes information about all the non-recurrent daily trips over 40 km and all overnight trips made by the respondents during one month. The TSRC survey was the primary source of data used for the model estimation. Data for the years 2011–2014 were used. Table 1 shows the main variables included in the database. The ITS survey was used to estimate and calibrate the models for international trips, which is not presented in this paper.

Lastly, Table 2 summarizes the number of survey records available for model estimation, both in terms of survey records and expanded using the survey weights to the total amount of trips in the period 2011–2014.

3.2.2. Zone system

The development of the long-distance travel demand model inherited the zoning system in which the synthetic population of Ontario was georeferenced. This zoning system had more than 5000 raster cells. Although trip generation is made at the person level, and the origin of the trip was allocated in this fine zone system, the estimation of destination choice models required a different aggregated zone system. Two main reasons motivated the aggregation: first, as the model is estimated based on survey data, the finest resolution is the one provided by the survey; second, while using logit models for destination choice, the number of alternatives should not be too big (as behavioral models, the size should not exceed what a person can manage to compare), so an aggregation or subsampling of alternatives is recommended (Rich & Mabbit, 2011). Therefore, an aggregated zone system was defined from the topographical union between Census Divisions (CD, that cover the whole Ontario region) and Census Metropolitan Area (CMA, only in dense urban areas). This resulted in 69 zones to be used for destination choice and mode choice (Molloy & Moeckel, 2017a).

As a long-distance model, the locations outside Ontario needed to be included in the zone system. The aggregation level of the external zones increased with their distance from Ontario: The neighboring provinces Quebec and Manitoba were represented by 38 different zones; the rest of Canada was aggregated to provinces and territories (10 zones).

Table 1
Variables in TSRC survey datasets.

Table	Variable	Description, range and levels
Person	Person age (respondent)	0–99 years
	Person gender (respondent)	Male Female
	Person education level (respondent)	Under high-school High school Post-secondary University
	Person employment status	Employed Unemployed
	Household size	Number of adults Number of kids
	Household income	Low (≤50,000 CAD*) Medium low (50,000 CAD* < income ≤ 70,000 CAD*) Medium high (70,000 CAD* < income ≤ 100,000 CAD*) High (> 100,000 CAD*)
Trip	Trip purpose	Business Leisure Visit
	Trip date	Day, Month, Year
	Trip mode	Car Air Rail Bus
	Trip origin	Province, CD**, CMA***
	Trip destination	If domestic: Province, CD and CMA If to the U.S.: State If overseas: Continental region
	Trip duration	Number of nights
	Travel party	Travel party size

*CAD: Canadian Dollars.
** CD: Census Division.
*** CMA: Census Metropolitan Area.

3.2.3. Foursquare and Rome2rio data

Although the TSRC survey provided worthy data to characterize long-distance trips and travelers, the estimation of discrete choice models required information about all the sets of alternatives, included non-selected destinations, as well as the level of service of the alternative travel modes.

With respect of measures for destination attractiveness, population and employment of the analysis zones were considered initially. These variables provided a global approach of the importance of the destinations but do not describe differences in land-uses of them, nor do they consider special tourist or business attractors. To enrich the description of destination alternatives, location based social network data (LBSN) were obtained. Destination attractiveness data were obtained from the Foursquare network (Foursquare, 2017).

The users of the Foursquare check-in their visits to venues (locations

Table 2
Sample size by purpose (number of records and number of trips – expanded from weights).

Type of trip	Survey records for estimation			Expanded average number of trips per day		
	business	leisure	visit	Business	leisure	visit
Domestic from Ontario	4771	19,686	28,378	31,665	111,907	163,471
Domestic from rest of Canada*	1457	3223	5036	4165	8149	12,528

*trips starting in other Canadian provinces were only considered if they end in Ontario or cross Ontario.

such as airports, restaurants, parks, ski areas, etc.) with the goal of providing other users information and recommendations of places to go. The data stored in this LBSN was accessed through the Foursquare public venue API, which returns a list of geo-located venues, their category, and their number of check-ins.

Venues were classified into five main categories and the number of check-ins by category within each model zone. The five main categories (medical, ski area, hotel, outdoors and sightseeing) were defined by aggregating the original list of Foursquare venue categories and subcategories. A total of 34,041 venues and 7,981,458 check-ins were collected. A previous research study provides more details on data collection and processing (Molloy & Moeckel, 2017b).

Regarding the estimation of mode choice models, it was required to obtain the level of service (supply) of the different trip alternatives. Trip surveys do not include travel times, numbers of transfers, frequencies or prices of neither the selected nor the non-selected modes. Therefore, level of service of the available modes was obtained from the trip planning service Rome2rio (Rome2Rio, 2017). Rome2rio is a multimodal transport origin-to-destination search engine designed for long-distance and local journey planning.

Using the Rome2rio API, data for all origin and destination zones in North America were queried (a total of 167 by 167 origin-destination pairs). Origin and destination coordinates were located in the centroids of the zones. Rome2rio provided different alternatives to travel from each origin to each destination, described by the following variables.

- Total travel time from origin to destination.
- Travel time by segment (segment is defined as a part of the trip using a single mode and vehicle; then, the number of transfers is equal to the number of segments minus one).
- Segment mode.
- Frequency as number of services per week by segment.
- Price range and average price from origin to destination.

The data were processed according to the following criteria:

- A “main mode” was assigned based on the following hierarchy: air, rail, bus and auto. Every alternative with a flying segment was then coded as “air”. If there is no air segment, every alternative with a rail segment was coded as rail, and so forth. Consecutive segments using the main mode belong to the main trip, while everything before or after main mode trip(s) was considered as “access trip” or “egress trip”, respectively.
- Accordingly, “main mode travel time”, “access time” and “egress time” were defined.
- If two alternatives using the same main mode are found, the faster one was selected.
- The frequency was defined as the minimum frequency among all the segments of the alternative, expressed in services per week.
- Driving costs were calculated by multiplying distance by fuel costs (0.072 CAD/km) instead of using Rome2rio data, as it assumed only taxi or shared vehicle costs.

Based on this post-processing, travel time, price, number of transfers and frequencies were expressed in form of origin-destination matrices, as detailed in previous work (Ji, 2017). Lastly, both the Foursquare and Rome2rio datasets were merged with the travel surveys in order to include them in the model estimation.

3.3. Scenarios

The developed model was applied to the demand prediction of a new high speed rail along the Toronto-Windsor corridor (Collenette, 2016), which is currently under planning. The sensitivity of the destination choice model and the mode choice model was tested by means of a set of scenarios that represented that high speed rail. A 300 km/h maximum speed rail line is planned to connect six stations in the south west area of Ontario, which concentrates 7 million people and 3.4 million jobs. The line is shown in the interactive map.

The implementation of the high speed rail line in the model was defined by the modification in the rail mode level of service variables (as defined from the Rome2rio datasets). Specifically, travel times were reduced, while prices and service frequency were increased. Consequently, and in contrast to previous examples, no new mode alternative was implemented for high speed rail (Outwater et al., 2010; Wong & Habib, 2015). The reason for this approach was that the primary goal was to test sensitivities of the estimated model, and the fact of not having any information on stated preferences of travelers for the use of high speed rail. Different scenarios were defined by varying the above-mentioned level of service variables with different intensities.

Travel times by rail among the six planed new rail stations were obtained from Collenette (2016). These six stations were georeferenced to six model zones, and access and egress times to rail stations were assumed. For these 30 origin-destination pairs (6 origins by 5 destination stations) the travel time using high speed rail was set as the new travel time by rail.

The reduction in travel times from and to other locations near the planned high speed rail was defined by using a simplified route choice model that provided the minimum travel time by rail for every origin-destination pair, as shown by equation (5). The approach found the shortest path considering all the available high speed rail stations, deducting access and egress times in the zones were travelers would only need to transfer from conventional to high speed rail. This process was repeated assuming different maximum speeds of 200, 300 and 400 km/h, and their corresponding average speeds and travel times, as summarized in Table 3.

With respect of price, a fare increase of 50, 100 and 150% in the high speed rail segments compared to conventional rail services were considered. For origin-destination pairs that started or ended in zones not served by the high speed rail, the increase in fare was applied only to the proportion of travel time on the high speed rail. Additionally, the increase in frequency was applied to the origin-destination pairs among the six high speed rail served zones. Increases of 25% and 50% were considered together with the case of not increasing frequency. The changes in frequency were not applied to trips that only used high speed rail for a part of the segment, as their maximum frequency would be constrained by other modes they use. Table 3 summarizes the complete set of scenarios.

$$tt'_{o-d} = \min(\, tt_{o-d},\, tt_{o-i} + \, tt_{i-j} + tt_{j-d} - egress_i - access_i - egress_j - access_j) \;\; \forall \;\; i, j \tag{5}$$

Table 3
Scenarios for high speed rail.

Scenario	Travel time scenario		Frequency scenario (% of increase*)	Price scenario (% of increase*)
	Maximum speed (km/h)	Average speed in the high speed segments (km/h)		
0	Unchanged			
1a	200	130	Unchanged	unchanged
1b			+25%	unchanged
1c			+50%	unchanged
2a	300	200	Unchanged	unchanged
2b			+25%	unchanged
2c			+50%	unchanged
3a	400	260	Unchanged	unchanged
3b			+25%	unchanged
3c			+50%	unchanged
3c.1				+50%
3c.2				+100%
3c.2				+150%

*Legend: the number represents an increase respect the original variable. I.e. "+50%" of frequency means that the original frequency is increased by a 50% or multiplied by 1.5.

Where:

- $tt'_{o\text{-}d}$ is the travel time in the high speed rail scenario between o and d.
- $tt_{o\text{-}d}$ is the travel time in the base scenario between o and d.
- o and d are origin and destination zones (not served by high speed rail).
- $tt_{\text{-}i\text{-}j}$ is the travel time in the high speed rail scenario between i and j.
- i and j are origin and destination zones (served by high speed rail).
- *access* and *egress* are access and egress times to rail stations at the considered zone (i or j).

4. Calculation

This section describes firstly the estimation process of the trip generation, destination choice and mode choice models, based on the already collected data. Later, subsection 4.2. describes the implementation of the model and its application to the scenarios.

4.1. Model estimation

Model estimation results for domestic trip generation, destination choice and mode choice are presented in this section. Table 4 shows the estimation results of the trip generation model. The columns show the coefficients of the utility functions for the three trip alternatives, where the base case (utility equal zero) is to "stay in town". As can be deducted from the negative values of the intercepts, starting a long-distance trip has a rather low probability. The impacts of person household and location attributes varied across purposes, although in general a higher income, a higher education level and a lower short-distance accessibility was related with increased trip probabilities.

Table 5 shows the results of the estimation of domestic destination choice models by purpose. Domestic destination alternatives are the aggregated zones (as explained above, 69 zones in Ontario and 48 in the rest of Canada). Table 5 shows the impact of distance between origin and destination as a function of the mode choice logsums (note that mode choice models were estimated before, although applied later) plus an added term of the logarithm of travel distance by car. The effect of logsums was different depending on whether the trip is a daytrip or an overnight trip (as mentioned by the if-clause in the table). The higher coefficient of logsums for daytrips explains that daytrips are more likely to select a closer destination, while the very long trips are in almost all cases, overnight trips.

Additionally, the impact of the type of origin and destination zone (urban or rural) was captured using three additional dummy variables, defined as follows:

- Intra-metro: the trip starts and ends at the same urban zone.
- Intra-rural: the trip starts and ends at the same rural zone.
- Inter-metro: the trip starts and ends in different urban zones.

Accordingly, it was found that intra-metro leisure and visit trips were less common. On the contrary, intra-rural trips were likely for all the purposes. As expected, trips between different urban zones (inter-metro) were more likely for business purposes, but less likely for leisure trips.

With respect of the destination attractiveness, the use of Foursquare data increased the explanatory power. The number of hotel check-ins was significant for all purposes, and sightseeing, outdoors and skiing check-ins were significant for leisure trips. After comparing the models with and without Foursquare data, an increase of Log-Likelihood of 1% for business, 2% for visit, and 8% for leisure was found. Thus, adding the dataset was especially relevant for leisure trips. Lastly, it was necessary to add a dummy variable for the Niagara destination, as previously the model had underestimated the number of trips to that particular destination.

Lastly, mode choice models were estimated. During model estimation, a strong positive correlation between mode price and mode travel time (generally, longer trips are also more expensive) was observed. It was decided to combined both variables into one *generalized time* by using values of time, as shown in equation (6). Then, the generalized time variable was converted to an exponential function, namely impedance, as shown in equation (7). After testing different combinations, it was decided to use one generic coefficient for *impedance* (one coefficient for all the modes, but distinguished by purpose). The value of time was equal to 32 CAD/h for visit and leisure trips and 65 CAD/$ for business trips, as selected for the provincial model. The values of the parameter α in equation (7) were selected through trial-and-error to produce the better model in terms of log-likelihood and AIC (Akaike Information Criterion). The chosen values for γ are shown in Table 6.

$$generalized\ time = travel\ time + travel\ cost/value\ of\ time \tag{6}$$

$$impedance = \exp(\gamma \cdot generalized\ time) \tag{7}$$

Where:

- γ is a weighting, negative parameter

As shown in Table 6, the impact of the impedance was found to be always positive, decreasing the utility with increasing distance. The frequency, defined as number of services per week, was found to be significant in the case of domestic trips, too. Additional trip characteristics were significant as well. Higher travel parties are related with trips by car, while overnight trips increase the probability of selecting air, which is correlated with the selection of further destinations. In the case of domestic trips, the model is able to capture some person-related and household-related preferences, such as the lower probability of selecting bus and rail for high income groups, the higher likelihood of young people to select bus or rail modes, or the smaller likelihood of travelling by car for women.

The values of Log-Likelihood for the estimate with constants only is shown in the table in row LL(interc. only). The comparison with the Log-Likelihood of the full model LL(full model) showed a considerable improvement, justifying the addition of mode level of service, person attributes and trip characteristics. Furthermore, the sensitivity of the model to changes in time, cost and frequency was captured by adding those mode-specific variables. An alternative model without person related attributes was used for those trips made by visitors (results are similar and not shown here due to space limitation).

Table 4
Domestic trip generation of Ontario residents by purpose.

Variable	Business			Leisure			Visit		
	away	daytrip	in/out	daytrip	away	in/out	Away	daytrip	in/out
(intercept)	−7.929***	−6.265***	−7.707***	−5.462***	−5.236***	−5.524***	−4.691***	−4.401***	−4.448***
isYoung (age ≤ 25)	–	–	–	–	–	–	1.033***	0.356***	1.215***
isRetired (age > 65)	–	–	–	–	–	–	–	–	–
isFemale	−0.610***	−0.945***	−0.628***	–	–	–	0.338***	−0.107*	0.158***
#adultsInHousehold	–	–	–	−0.171***	0.127***	−0.060*	−0.307***	−0.111***	−0.228***
#kidsInHousehold	–	–	–	–	–	–	−0.222***	−0.119***	−0.205***
IsHighSchool	–	–	–	0.540***	0.273**	0.542***	–	–	–
isPostSecondary	–	–	–	0.520***	0.323***	0.495***	0.147*	0.212***	0.221***
isUniversity	0.783***	0.348**	0.874***	0.575***	0.322**	0.490***	0.459***	0.176**	0.456***
isEmployed	1.369***	0.923***	1.399***	–	–	–	–	–	–
hasIncomeMedLow	–	–	–	0.598***	0.392***	0.535***	0.211*	0.374***	0.231**
hasIncomeMedHigh	–	–	–	0.701***	0.561***	0.7820***	0.345***	0.467***	0.464***
hasIncomeHigh	0.641***	0.452***	0.807***	1.209***	0.701***	1.146***	0.408***	0.521***	0.443***
accesibility	−0.012**	−0.014***	−0.014**	−0.005***	−0.012***	−0.004***	−0.008***	−0.009***	−0.009***
LL	−3736			−20,438			−20,398		
AIC	7557			40,960			40,880		

LL: Log-Likelihood, AIC: Akaike Information Criterion.
Significance codes: ***: p-value < 0.001; **: p-value < 0.01; *: p-value < 0.05 .: p-value < 0.1, -: not significant.
Legend for variables: "isCondition" or "hasCondition" is a dummy variable equal to 1 if the condition is fulfilled, zero otherwise; "#countVariable" is the count of the variable, "variable" is a continuous variable.

4.2. Model implementation, calibration and application

After the completing estimation, the model was implemented in a JAVA program and calibrated to represent properly the observed trip length frequency distribution and modal shares. Logsum coefficients of destination choice and mode specific constants of mode choice were adjusted (multiplied by calibration parameters) to match average trip distances and modal shares with an error under 5% with respect of observations.

The developed and implemented model was applied to the set of scenarios, by modifying the input datasets according to the definition of scenarios (modified rail travel time, price and frequency). Every scenario was replicated 10 times to average out stochastic variations inherent to the microsimulation approach.

5. Results

This section analyzes the sensitivity of the model to the implementation of a high speed rail line, focusing on the variation of modal shares and destination choices. By definition, the trip generation model is not sensitive to changes in the transport supply, therefore, no induced travel demand could be captured.

The presented results correspond to an average simulated day and are describe separately for two different analysis areas: a) all domestic trips in Ontario and b) domestic trips among the six zones served by high speed rail.

Fig. 2 shows the changes in rail modal shares with the implementation of the high speed rail scenarios varying travel time and service frequency (without modifying the fare). The left side (Fig. 2a) corresponds to the changes of all domestic trips, while the right (Fig. 2b) corresponds to the trips among the zones which are served by

Table 5
Domestic destination choice by purpose.

Type of variable	Variable	Business	Leisure	Visit
Attraction	Log(#population + #employment)	0.276***	−0.193***	0.302***
	Log(#Foursquare_hotel)	0.147***	0.228***	0.138***
	Log(#Foursquare_sightseeing)	0.125***	0.112***	0.03***
	Log(#Foursquare_outdoors)	–	0.132***	–
	Log(#Fousquare_skiing)	–	0.073***	–
	isNiagara	–	1.444***	–
Origin-destination types	isIntra-metro	–	−3.053***	−1.843***
	isIntra-rural	0.98***	0.812***	0.293***
	isInter-metro	0.169***	−1.223***	–
Distance origin - destination	[if daytrip] Logsum	3.615***	9.671***	4.384***
	[if overnight trip] Logsum	0.899***	2.461***	0.909***
	Log(tripDistance)	–	−0.267***	−0.231***
LL		−17831	−73649	−110309
AIC		35,663	147,297	220,617

LL: Log-Likelihood, AIC: Akaike Information Criterion.
Significance codes: ***: p-value < 0.001; **: p-value < 0.01; *: p-value < 0.05 .: p-value < 0.1, -: not significant.
Legend for variables: "isCondition" or "hasCondition" is a dummy variable equal to 1 if the condition is fulfilled, zero otherwise; "#countVariable" is the count of the variable, "variable" is a continuous variable.

Table 6
Mode choice by purpose (for residents of Ontario).

Variable	Business				Leisure				Visit			
	auto	air	bus	rail	auto	air	bus	rail	auto	air	bus	rail
intercept	0	-7.235^{***}	-5.088^{***}	-6.646^{***}	0	-7.773^{***}	-3.93^{***}	-4.457^{***}	0	-5.15^{***}	-5.796^{***}	-5.824^{***}
impedance	2.696^{***}	2.696^{***}	2.696^{***}	2.696^{***}	6.659^{***}	6.659^{***}	6.659^{***}	6.659^{***}	2.601^{***}	2.601^{***}	2.601^{***}	2.601^{***}
#frequency	–	0.004^{***}	0.004^{***}	0.004^{***}	–	0.005^{***}	0.005^{***}	0.005^{***}	–	0.007^{***}	0.007^{***}	0.007^{***}
isOvernight	–	2.909^{***}	0.387^{***}	0.387^{***}	–	3.896^{***}	1.356^{***}	0.794^{***}	–	2.766^{***}	1.072^{***}	0.517^{***}
isInter_metro	–	1.724^{***}	1.199^{***}	2.007^{***}	–	1.559^{***}	1.49^{***}	0.901^{***}	–	1.107^{***}	0.842^{***}	0.842^{***}
#tripParty	–	-0.153^{***}	0.032^{***}	0.032^{***}	–	−0.055.	-0.966^{***}	-0.494^{***}	–	–	–	–
isIncomeLow	–	–	–	–	–	–	–	–	–	-0.878^{***}	–	–
isIncomeHigh	–	–	-0.931^{***}	-0.931^{***}	–	0.272^{***}	-0.944^{***}	-0.993^{***}	–	–	–	–
isYoung	–	–	0.797^{***}	0.797^{***}	–	-0.378^{***}	1.466^{***}	1.682^{***}	–	-1.39^{***}	1.96^{***}	1.05^{***}
isFemale	-0.489^{***}	–	–	–	–	–	–	–	-0.493^{***}	–	–	–
isUniversity	–	–	–	–	–	–	–	–	–	0.47^{***}	0.361.	1.072^{***}
γ	−0.0015 (not estimated)				−0.0004 (not estimated)				−0.0004 (not estimated)			
LL(interc. only)	−3817				−5499				−12,880			
LL(full model)	−2568				−4481				−9152			
AIC(full model)	5172				9018				18,346			

LL: Log-Likelihood, AIC: Akaike Information Criterion.
Significance codes: ***: p-value < 0.001; **: p-value < 0.01; *: p-value < 0.05 .: p-value < 0.1, -: not significant.
Legend for variables: "isCondition" or "hasCondition" is a dummy variable equal to 1 if the condition is fulfilled, zero otherwise; "#countVariable" is the count of the variable, "variable" is a continuous variable.

the high speed rail. Obviously, the impact was much stronger in the second case, as all trips by rail benefited from the existence of high speed rail. Moreover, the higher reduction of travel time (from scenarios at 200 km/h to 400 km/h), the higher the increase in rail modal share.

However, the reduction of travel time was found to be less significant compared with the increase in service frequency. In fact, with no increase in frequency, the modal share at zones served by high speed rail increased between 0.4 and 0.6 percent points (around 2500 to 3500 additional trips by train per day – note that trip is a different measure that may be smaller than passengers, due to multi-passenger trips), but if the frequency was increased by 50%, the modal shares increased between 3.5 and 4.3 percent points.

A deeper analysis of the limited impact of travel time revealed the relative importance of access and egress times in the trip duration, which reduced drastically the benefit achieved in the main (high speed rail) segment of the trip for certain origin-destination pairs.

As expected, the changes in modal level of service had an impact on destination choice (represented by the aggregated logsum terms in the utility equations). The analysis detected an increase of 43% in the number of trips between zones served by high speed rail.

For the scenarios where the frequency increased by 50% (represented by the dashed line in Fig. 2) and maximum speed was equal to 400 km/h, the sensitivity of the model to changes in fares was tested. Fig. 3 shows the impact of different relative increases of the high speed rail fare (with respect to the conventional rail fare). This comparison has been also segmented by purpose (shown by line types in Fig. 3).

Fig. 3 shows the general trend of a decreasing rail share with an increase in fares (while keeping frequency and travel times constant). Although this effect was uniform, the decrease was stronger for business trips, where the rail share was generally larger.

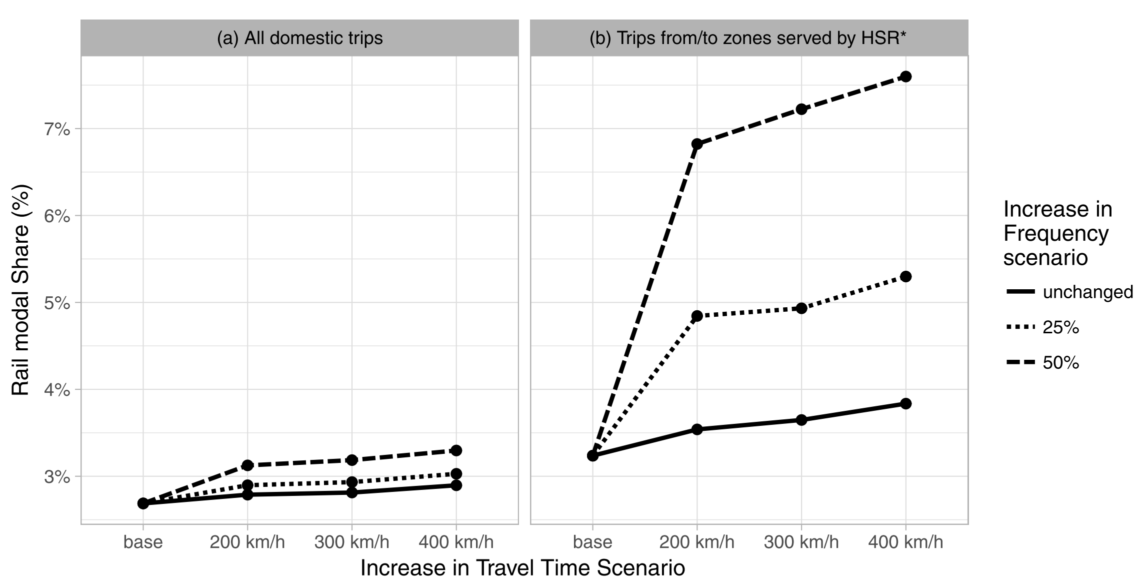

Fig. 2. Modal shares by scenario (varying travel time reduction and frequency) (*HSR: High Speed Rail).

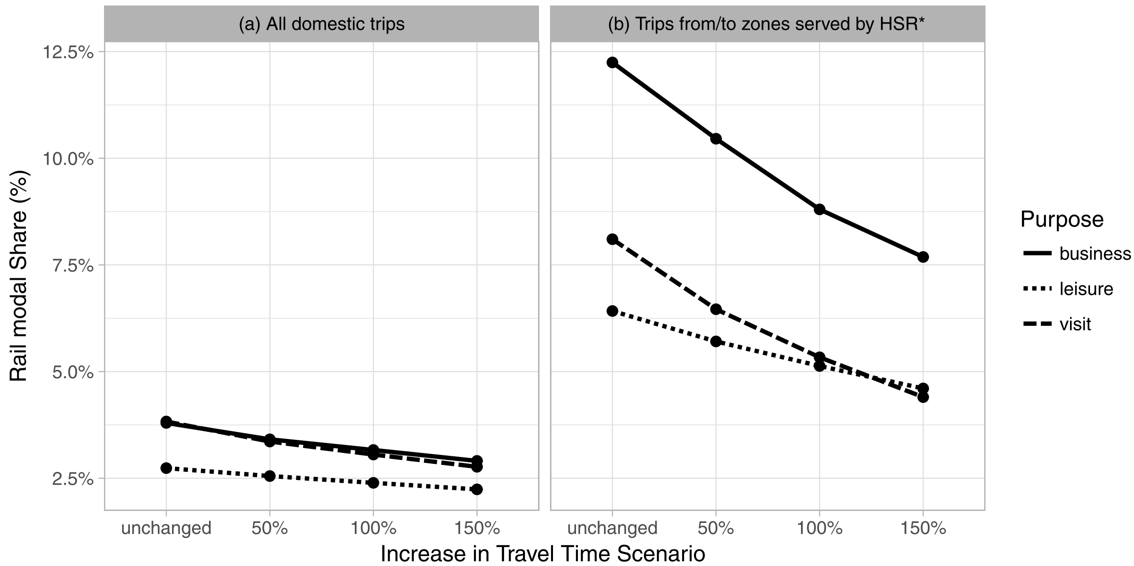

Fig. 3. Modal shares by scenario (varying price) by trip purpose (*HSR: High Speed Rail).

6. Discussion

This section describes firstly the implications of the use of alternative data sources for the model development. Secondly, the results of sensitivity analyses for the assessment of a high speed rail project are analyzed.

The estimation of discrete choice decisions required the availability of alternative specific data, for both selected and non-selected choices. Surveys are used as primary sources for the model estimation but rarely contain data on non-selected choices.

For destination alternatives, the number of check-ins at the Foursquare location-based social network was used to characterize specific and diverse land uses across zones. Then, the goodness of fit of the models (especially for leisure purposes) increased significantly compared to models that only used population and employment. This suggests that these types of data are suitable to trace leisure activities. Population and employment are probably enough to predict visit or business destinations. Although a significant improvement of model was achieved, the use of location-based social network data also presents the challenge of being used for forecasts. Further work is needed to link these data with land use, demographic or economic changes for long-term predictions.

Regarding modal level of service, the use of trip planning services, such as Rome2rio, was found to be a fast and reliable source of data. Among other advantages, a service such as Rome2rio replicates the way users plan their trips. Consequently, it seems to be an adequate tool to estimate a discrete choice model as a representation of human decisions, having the same information that the traveler would have. The collection of Rome2rio data allowed the estimation and implementation of the model without performing a traffic assignment (or route choice). On the one hand, this simplified the model development. On the other hand, this method cannot account for congestion issues, especially in transit modes, as Rome2rio data are mainly based on schedules instead of actual time travel times.

With respect of the sensitivity analysis of the developed model, and in contrast to the use of stated preferences for a new mode, high speed rail was implemented as an improvement of the current rail supply.

The sensitivity analysis showed reasonable positive impacts (i.e., an increase of the rail mode shares) with a reduction of travel time and an increase of service frequency. Negative impacts on rail shares due to fare increases were found to be consistent as well. All these effects were stronger for zones directly served by the new high speed rail, but could be appreciated in its surroundings, too.

The model increased the number of trips to zones served by high speed rail, extending the impact of mode choice decisions into destination choice. In addition, the model created additional trips by car, plane or bus among locations with high speed rail. Theoretically, this effect can be explained by the feedback from mode choice to destination choice through aggregated logsum terms. Such a behavior might seem counter-intuitive but can be observed in reality by a small degree. Once a high speed rail line opens between zone A and zone B, additional travelers may choose this origin-destination even if they do not travel by rail. This might be due to two reasons. For one, having high speed rail as a second alternative may serve as a back-up mode in case the originally planned mode fails (i.e., flights could be cancelled due to weather, which would case some travelers to switch to rail). Secondly, origin-destination pairs with high speed rail may be better known than other destinations, because travelers might have taken the train on this trip before or because advertisements for the train might make people more aware of that destination.

The methodology of this paper can be used by policy makers to understand the changes of planned projects at a very early planning stage. The model is able to define the influence area of the project and identify its potential users within the global long-distance passenger travel demand. In contrast to case-specific or corridor-based analyses, and despite its coarser level of detail, this method does not ignore the rest of the modes, and does not delimitate the influence area prior to analyzing the impacts of the planned infrastructure.

The simulation of high speed rail as an enhanced rail service was found to be simpler and relied on fewer assumptions than the stated preferences survey method. This is supported by several reasons. Firstly, to estimate the model, the mode-specific variables for rail (i.e. time and price) were chosen as the fastest rail trip alternative. Consequently, if high speed rail is considered as a rail service, the new mode-specific variables had to be the high speed rail time and price, for the origin-destination pairs where it is the fastest trip alternative. Secondly, this assumption would be reasonable if the operator replaces the current conventional services by high speed rail services. Lastly, it implies that users' perception of high speed rail is not very different from the current feeling against conventional rail. On the contrary, if some competition between conventional and high speed rail is expected, this approach might be failing at the prediction of modal shares. Some studies also reported deeper changes in the demand, when the new high speed rail is more than a new service but creates a new "brand" (Cartenì, Pariota, & Henke, 2017). To overcome this issue, and even without a stated preference survey, an incremental logit approach could help introducing a new mode, the coefficients of which cannot be estimated based on the existing survey (Ameen & Kamga, 2013). However, the need of additional assumptions regarding the new coefficients may increase the uncertainty of the model. Whether the first approach (substitution of conventional rail) or the second (add a new mode) works better for the Toronto-Windsor corridor depends on the new service characteristics and the reaction of users, which is still unknown to the authors.

Although the developed model is a provincial model, this paper tested their application to a corridor-specific case. At this point, the results should be interpreted with caution. Firstly, the level of resolution at the destination choice and at the mode choice model was too coarse to account for sufficient detail with access and egress times to high speed rail stations (only aggregate values per zone were used). Secondly, the estimation of the models was based on country-wide data. It reflected properly the total amount of domestic trips in Canada, but country-wide data may not be representative for this particular corridor. Thirdly, the model was unable to capture induced travel demand because of the improvement of the transit network level of service. Lastly, the obtained sensitivities were found to be reasonable, but cannot be validated until real observations of such scenarios are available.

7. Conclusions

This paper described the development of a long-distance travel demand model for the province of Ontario (Canada). The major contributions of the paper are the integration of alternative data sources for destination alternatives and modal level-of service, as well as a sensitivity analysis of the impact of high speed rail scenarios.

With respect of the first contribution, this research provided a method to acquire additional data that complemented travel surveys from both location-based social network and online trip planning services. These data were useful to increase the goodness of fit of destination choice models (especially for leisure purposes) and to facilitate and simplify the characterization of multimodal transport supply.

The second contribution of the paper included a sensitivity analysis of the developed model and the assessment of a planned highs speed rail project for the Toronto-Windsor corridor. The results provide a quantitative prediction of the expected demand and a delineation of the area of influence of the proposed project. The effects of the high speed rail on the demand of other modes were quantified as well.

Consequently, and taking into account the limitations mentioned in section 7, the large scale long-distance travel demand model developed in this paper provides an adequate tool for planning agencies to rapidly quantify the impacts of different project alternatives in the entire transport system and to easily identify scenarios that warrant further and more detailed examination.

Acknowledgements

The research was completed with the support of the Technische Universität München – Institute for Advanced Study, funded by the German Excellence Initiative and the European Union Seventh Framework Programme under grant agreement n° 291763. The authors would like to thank Sundar Damodaran, Shan Sureshan, Arthur Tai and Rob Tardif from the Ministry of Transportation Ontario as well as Rick Donnelly, Mausam Duggal, Bryce Sharman and Peter Kucirek for their continued support and brainstorming for the development of the presented models.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.retrec.2018.06.004.

References

Ameen, N., & Kamga, C. (2013). Forecast of airport ground access mode choice with the incremental logit model. *Transportation Research Record: Journal of the Transportation Research Board, 2336*(2336), 97–104. https://doi.org/10.3141/2336-12.

Ben-Akiva, M. E. (1974). *Structure of passenger travel demand models.* Massachusetts Institute of Technology.

Cartenì, A., Pariota, L., & Henke, I. (2017). Hedonic value of high-speed rail services: Quantitative analysis of the students' domestic tourist attractiveness of the main Italian cities. *Transportation Research Part a: Policy and Practice, 100*, 348–365. https://doi.org/10.1016/j.tra.2017.04.018.

Collenette, D. (2016). High Speed Rail in Ontario: Transforming mobility, connecting communities, integrating centres of innovation and fostering regional economic growth and development. Special Advisor for High Speed Rail: Final Report. Retrieved from http://www.mto.gov.on.ca/english/publications/high-speed-rail-in-ontario-final-report/.

Croissant, Y. (2017). Estimation of multinomial logit models in R: The mlogit Packages. https://cran.r-project.org/package=mlogit.

Dargay, J., & Clark, S. (2010). The determinants of long distance travel: An analysis based on the British national travel survey. *World conference on transport research, lisbon, Portugal*.

Demizu, F., Li, Y. T., Schmöcker, J. D., Nakamura, T., & Uno, N. (2017). Long-term impact of the shinkansen on rail and air demand: Analysis with data from northeast Japan. *Transportation Planning and Technology, 40*(7), 741–756. https://doi.org/10.1080/03081060.2017.1340023.

Erhardt, G. D., Freedman, J., Stryker, A., Fujioka, H., & Anderson, R. (2007). Ohio long distance travel model. *Transportation Research Record: Journal of the Transportation Research Board, 2003*, 130–138. https://doi.org/10.3141/2003-16.

Federal Highway Administration (2015). *Foundational knowledge to support a long-distance passenger travel demand modeling framework implementation report*.

Foursquare (2017). *foursquare.com*.

Giamo, G. T., & Schiffer, R. G. (2005). *Statewide travel demand modeling. A peer exchange.* Transportation Research Circular, E-C075.

Gundelfinger-Casar, J., & Coto-Millán, P. (2017). *Intermodal competition between high-speed rail and air transport in Spain*, *vol. 47*, 12–17. https://doi.org/10.1016/j.jup.2017.06.001.

Gutierrez, J. (2001). Location, economic potential and daily accessibility: An analysis of the accessibility impact of the high-speed line madrid-barcelona-French border. *Journal of Transport Geography, 9*, 229–242. https://doi.org/10.1016/S0966-6923(01)00017-5.

Hasan, A., Wang, Z., & Mahani, A. S. (2016). Fast estimation of multinomial logit models: *R* package mnlogit. *Journal of Statistical Software, 75*(3)https://doi.org/10.18637/jss.v075.i03.

Ji, J. Y. (2017). *Long-distance mode choice modeling of Ontario province*M.Sc. Thesis. Technical University of Munich.

LaMondia, J. J., Fagnant, D. J., Qu, H., Barrett, J., & Kockelman, K. (2016). *Shifts in long-distance travel mode due to automated vehicles: Statewide mode-shift simulation experiment and travel survey analysis*, *vol. 2566*, 1–10. https://doi.org/10.3141/2566-01.

Limtanakool, N., Dijst, M., & Schwanen, T. (2006). The influence of socioeconomic characteristics, land use and travel time considerations on mode choice for medium- and longer-distance trips. *Journal of Transport Geography, 14*(5), 327–341. https://doi.org/10.1016/j.jtrangeo.2005.06.004.

Li, Y. T., & Schmöcker, J. D. (2017). Adaptation patterns to high speed rail usage in Taiwan and China. *Transportation, 44*(4), 807–830. https://doi.org/10.1007/s11116-016-9679-5.

Lu, M., Zhu, H., Luo, X., & Lei, L. (2015). Intercity travel demand analysis model. *Advances in Mechanical Engineering, 6*(0), 108180. https://doi.org/10.1155/2014/108180.

Miller, E. (2004). The trouble with intercity travel demand models. *Transportation research board annual meeting. Washington DC (US)*(pp. 1–8).

Moeckel, R., Fussell, R., & Donnelly, R. (2015). Mode choice modeling for long-distance travel. *Transportation Letters, 7*(1), 35–46. https://doi.org/10.1179/1942787514Y.0000000031.

Molloy, J., & Moeckel, R. (2017a). Automated design of gradual zone systems. *Open Geospatial Data, Software and Standards, 2*(1), 19. https://doi.org/10.1186/s40965-017-0032-5.

Molloy, J., & Moeckel, R. (2017b). Improving destination choice modeling using location-based big data. *International Journal of Geo-Information, 6*(9), 291. https://doi.org/10.3390/ijgi6090291.

Outwater, M., Tierney, K., Bradley, M., Sall, E., Kuppam, A., & Modugula, V. (2010). California statewide model for high-speed rail. *Journal of Choice Modelling, 3*(1), 58–83. https://doi.org/10.1016/s1755-5345(13)70029-0.

Rich, J., & Mabbit, S. L. (2011). A long-distance travel demand model for Europe. *European Journal of Transport and Infrastructure Research, 11*(1), 1–20.

Rome2Rio (2017). *rome2rio.com*.

Schiffer, R. G. (2012). *Long-distance and rural travel transferable parameters for statewide travel forecasting models*NCHRP Report 735https://doi.org/10.17226/22661.

Train, K. E. (2009). *Discrete choice methods with simulation.* Cambridge University Press.

Van Nostrand, C., Sivaraman, V., & Pinjari, A. R. (2012). Analysis of long-distance vacation travel demand in the United States: A multiple discrete–continuous choice framework. *Transportation, 40*(1), 151–171. https://doi.org/10.1007/s11116-012-9397-6.

Wong, B., & Habib, K. M. N. (2015). Effects of accessibility to the transit stations on intercity travel mode choices in contexts of high speed rail in the Windsor–Quebec corridor in Canada. *Canadian Journal of Civil Engineering, 42*(11), 930–939. https://doi.org/10.1139/cjce-2014-0493.