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 corridor

Carlos Llorca [1]

Joanna Ji [2]

Joseph Molloy [3]

Rolf Moeckel [1]

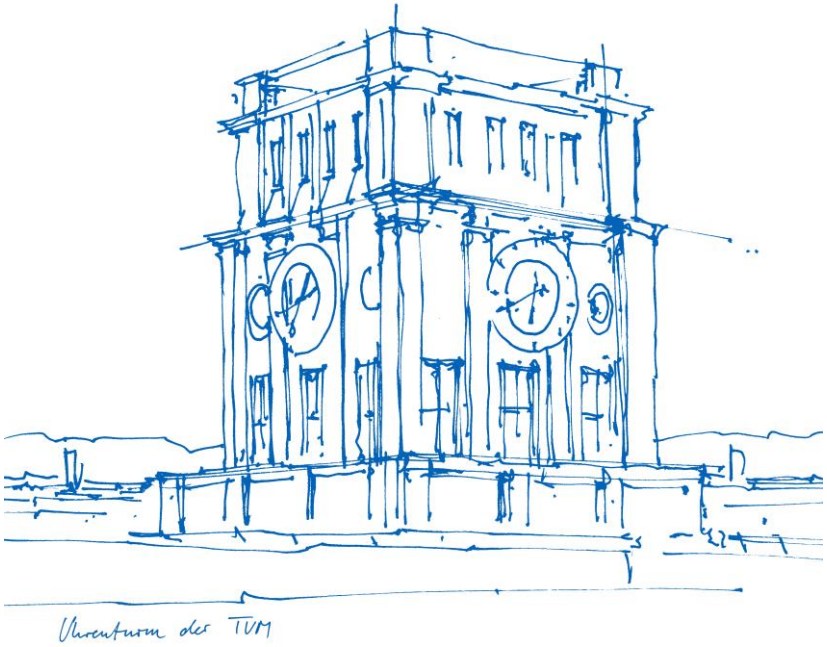
[1] Technical University of Munich

Department of Civil, Geo and Environmental Engineering

Research Group Modeling Spatial Mobility

[2] PTV Group

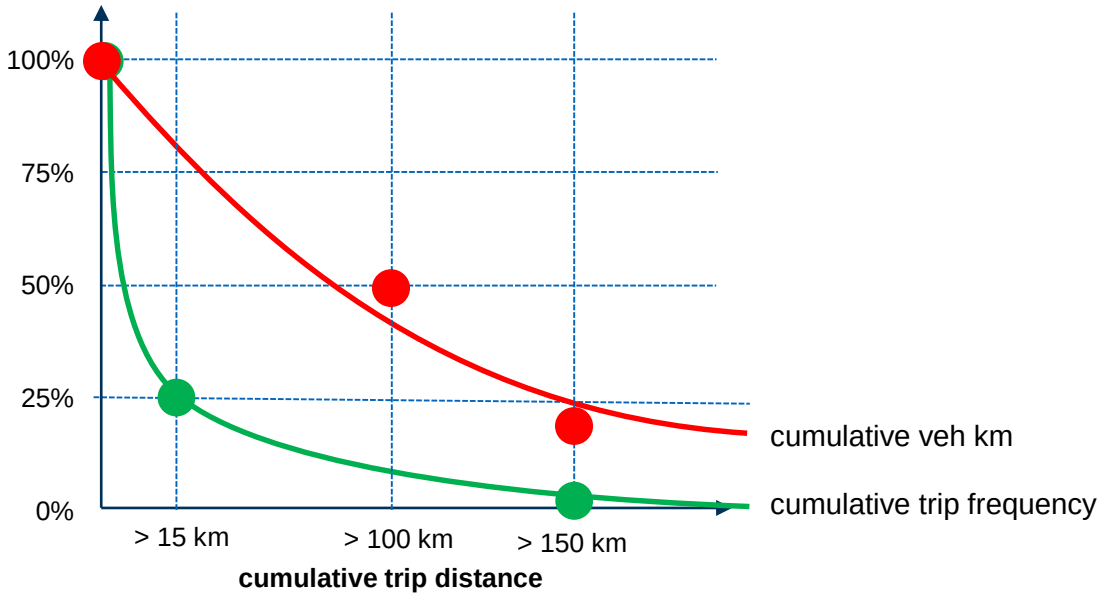
[3] IVT, ETH Zürich



Milan (Italy), September 2017

Introduction

Relative importance of long-distance travel demand:

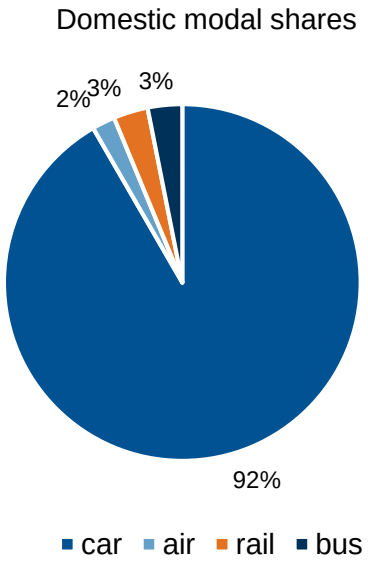


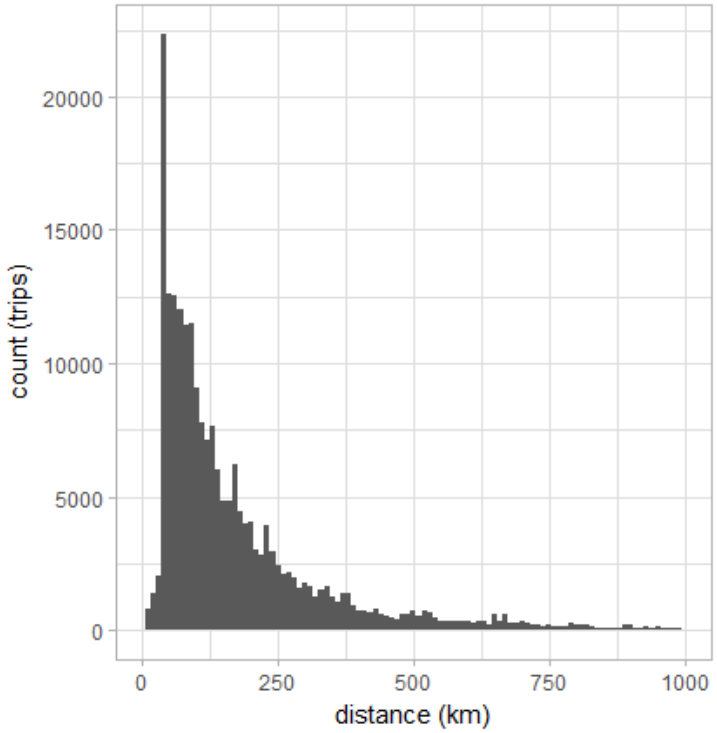
Less research compared to short-distance (urban) travel demand:

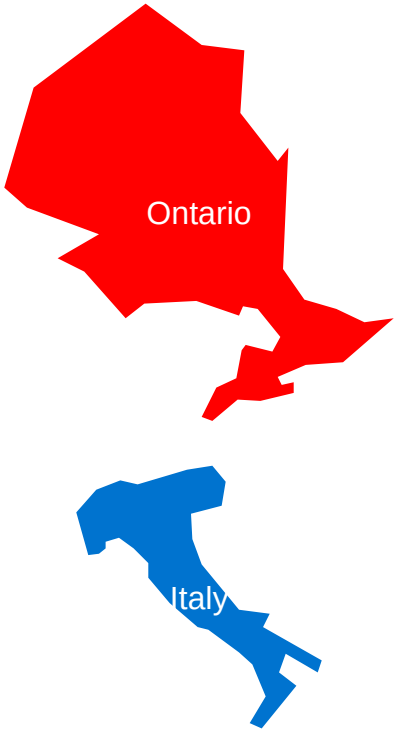
- Case specific or corridor specific
- Long-distance modules for large scale statewide/countrywide models

Framework

- Development of a long-distance passenger model for the provincial model of Ontario (Canada)
- Ontario province:
  - Population = 13.5 M
  - Domestic trips starting in Ontario = 93 M/year



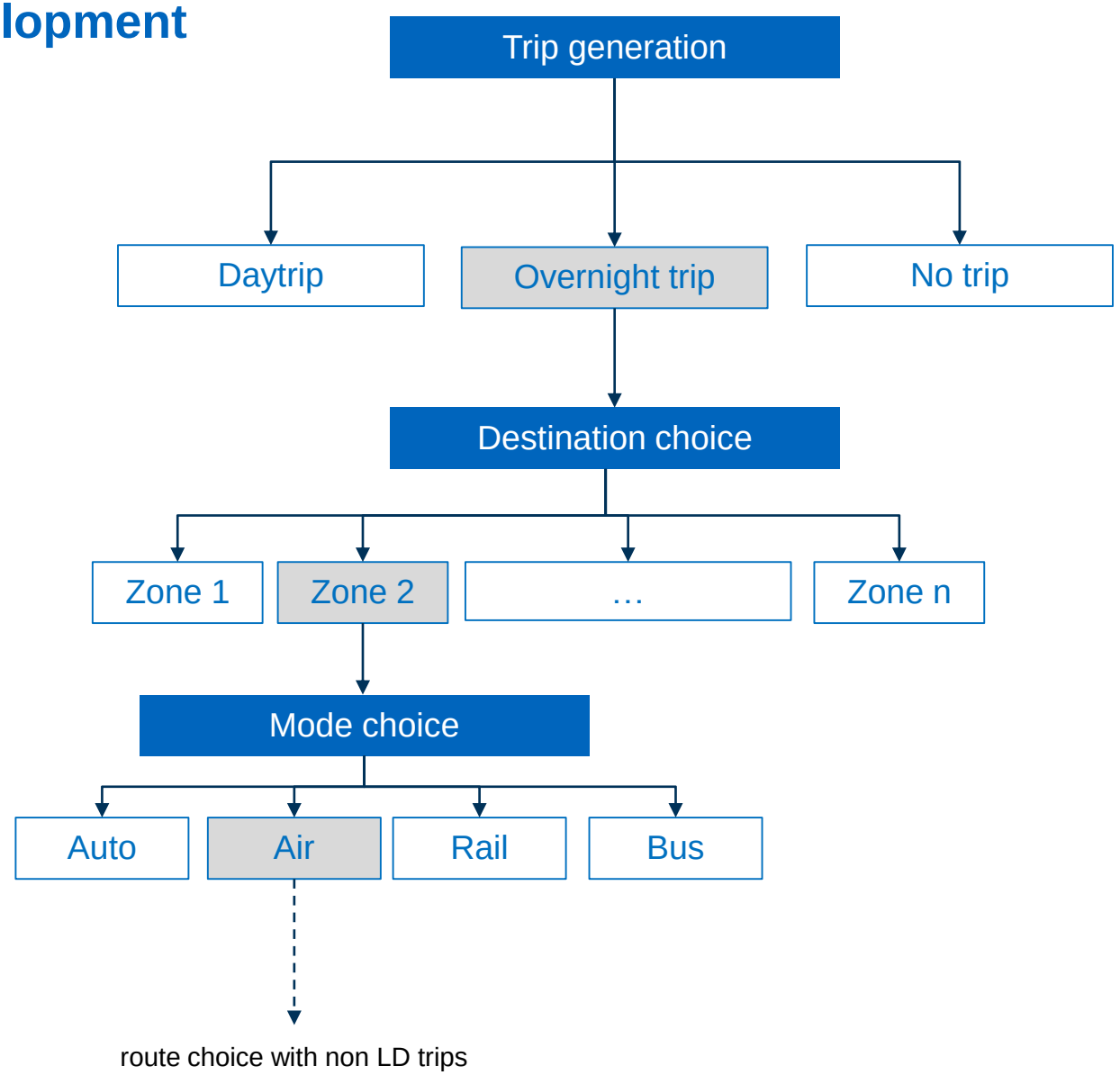




**according to the definition of long-distance trip in Canadian survey data: >40 km daytrip or overnight trip*

Model development

Model structure



Model development

Data collection (1):

Zoning system: 69 zones in Ontario + 175 external zones (38 in Canada and 137 abroad)

Model development

Data collection (2)

Travel survey: Travel Survey of Residents in Canada (TSRC)

Person:
- Id
- Age
- Gender
- Education level
- Employment status
- Household size
- Household income
- Province, census division, census metropolitan area

Trips:
- Traveler id
- Purpose
- Date
- Mode
- Origin
- Destination
- Duration
- Party size

52K records (expanded to 948M trips) during years 2011 to 2014

Informs partially about destination and mode alternatives **selected** by respondents

Model development

Data collection (3):

Destination alternatives – zone characteristics

- Population
- Employment – by industry categories
- **Foursquare: location based social network**
  - Users can:
    - Register places (=venues)
    - Check-in visits to venues
    - Write reviews of venues

Access: foursquare.com

Model development

Foursquare → looking for venues (desktop version)

Example query:

https://foursquare.com/explore?mode=url&near=Milan%2C%20Lombardy%2C%20Italy&nearGeoId=72057594041101371&q=shopping

Model development

Foursquare data processing:

- Define a raster search grid of 1x1 degrees
- Access to the number of check-ins by category by raster cell trough Foursquare-API
- Aggregate to the desired geographical resolution and desired venue categories

- Venue categories (after aggregation for this analysis):
  - Medical
  - Ski area
  - Hotel
  - Outdoors
  - Sightseeing

Summary of data collection for Canada:

- 34,041 venues
- 7,981,458 check-ins

Result → better characterization of diverse land uses at the alternative destinations

Model development

Data collection (4):

Modal level of service (transport supply) → **Rome2Rio**

- Online trip planning service → users search Point to Point trip alternatives

Access in https://www.rome2rio.com/

Model development

Data collection:

4. Modal level of service (transport supply) → Rome2Rio

Example query:

https://www.rome2rio.com/s/Munich/Milan

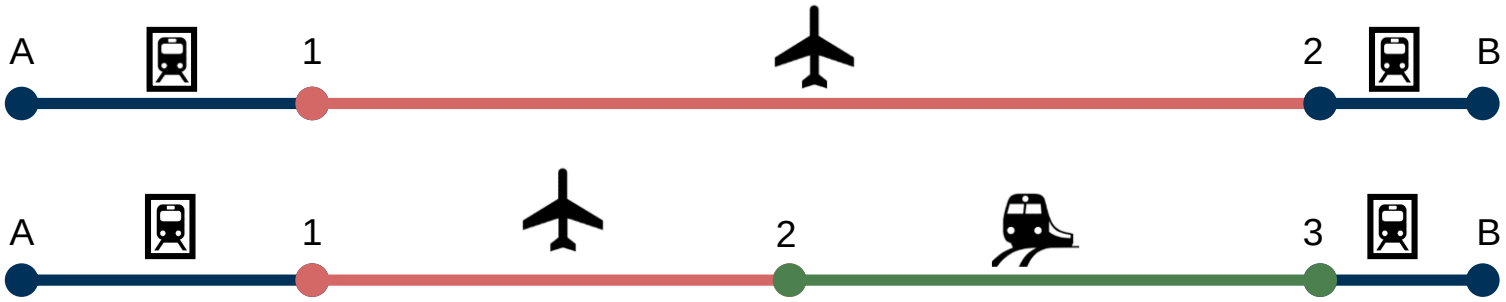
Model development

Data collection:

4. Modal level of service (transport supply) → Rome2Rio

Data processing:

- Access via Rome2rio API
- For each mode alternative:
  - Total travel time, access and egress times
  - Average price
  - Frequency
  - Main mode hierarchy: air > rail > bus > auto
- Example: 2 alternatives with air as main mode:



Result → travel time, travel cost, frequency and number of transfers zone-to-zone matrices

Model development

Model estimation:

- Multinomial logit models → random choices of individuals
- Trip generation
  - Four alternatives: stay at home, daytrip, being away in overnight trip, start or end an overnight trip
  - $Utility(travel) = f(person\ and\ household\ attributes, accessibility\ of\ origin\ zone)$
- Destination choice
  - 69 Alternatives (= number of zones in Ontario)
  - $Utility(destination) = f(population, employment, Foursquare\ attractors, mode\ choice\ logsum, daytrip/overnight)$
- Mode choice
  - Four alternatives: auto, air, rail, bus
  - $Utility(mode) = f\ (person\ and\ household\ attributes, mode\ level\ of\ service)$

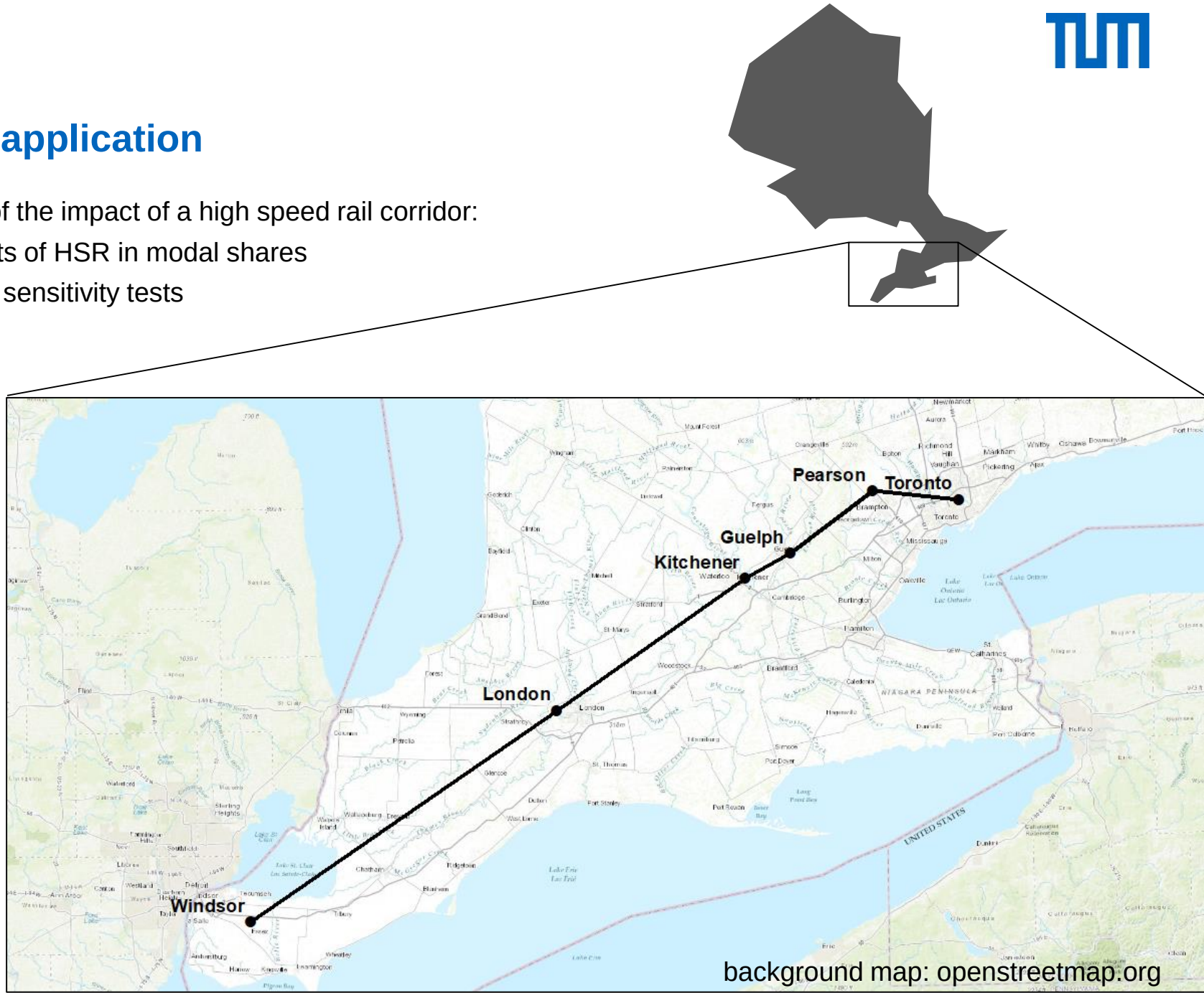
Model calibration:

- Adjustment of logsum coefficients and mode specific constants to match survey average trip length and modal shares (errors < 5%)

Model application

Analysis of the impact of a high speed rail corridor:

- Impacts of HSR in modal shares
- Model sensitivity tests



Model application

Scenarios:

- Travel time* reduction:
  - Base scenario: conventional rail
  - HSR scenarios: 200, 300** and 400 km/h maximum speed (130, 200** and 260 km/h commercial speed)
- Travel cost* increase:
  - Base scenario: current travel cost
  - HSR scenarios: increase by 50%, increase by 100%, increase by 150%
- Service frequency* increase:
  - Base scenario: current service frequency
  - HSR scenarios: no increase, increase by 25%, increase by 50%

- No route choice or assignment models → modification of zone-to-zone matrices:
  - Between zones served by HSR (6 stations) → apply directly the improvement
  - Between zones not served by HSR → applies only to the HSR segment

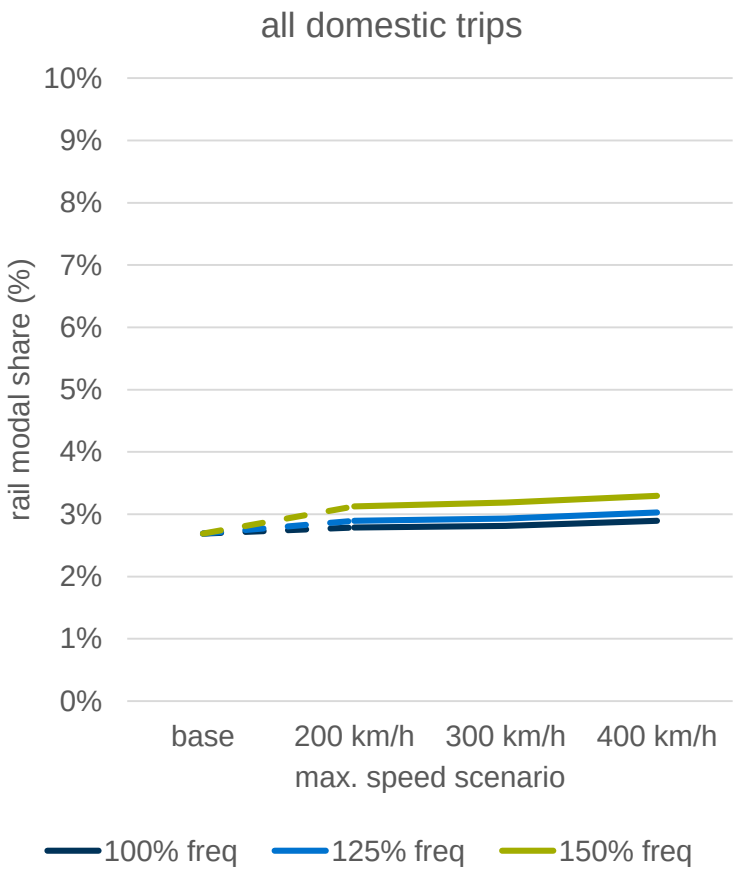
**variables significant at the 99% confidence level in mode choice MNL models*

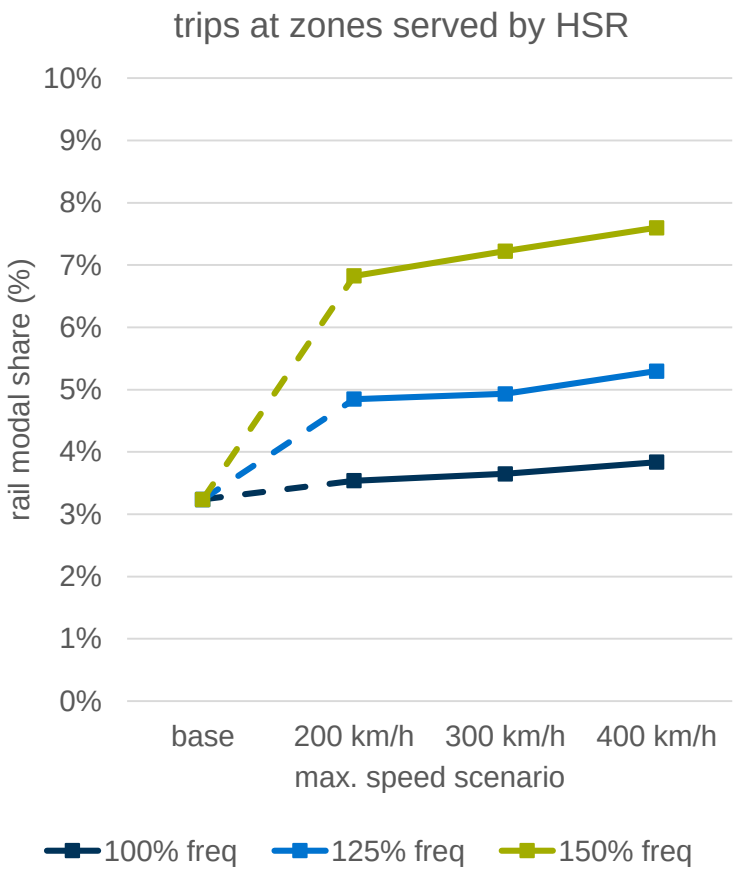
**planned by MTO maximum speed and commercial speed*

Model application

Results:

Modal share by travel time and frequency (without price change):

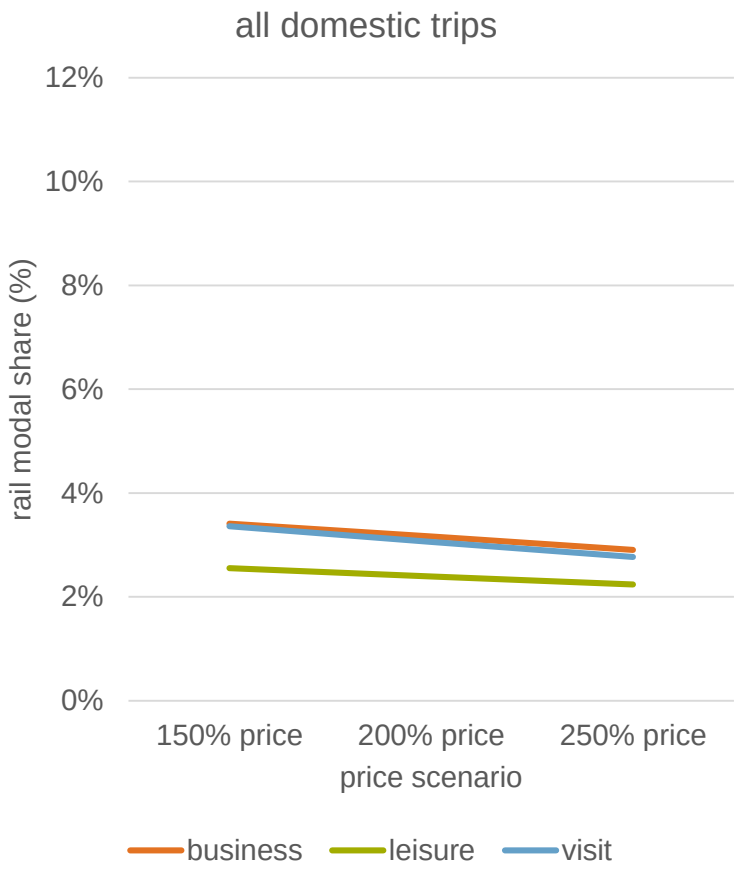


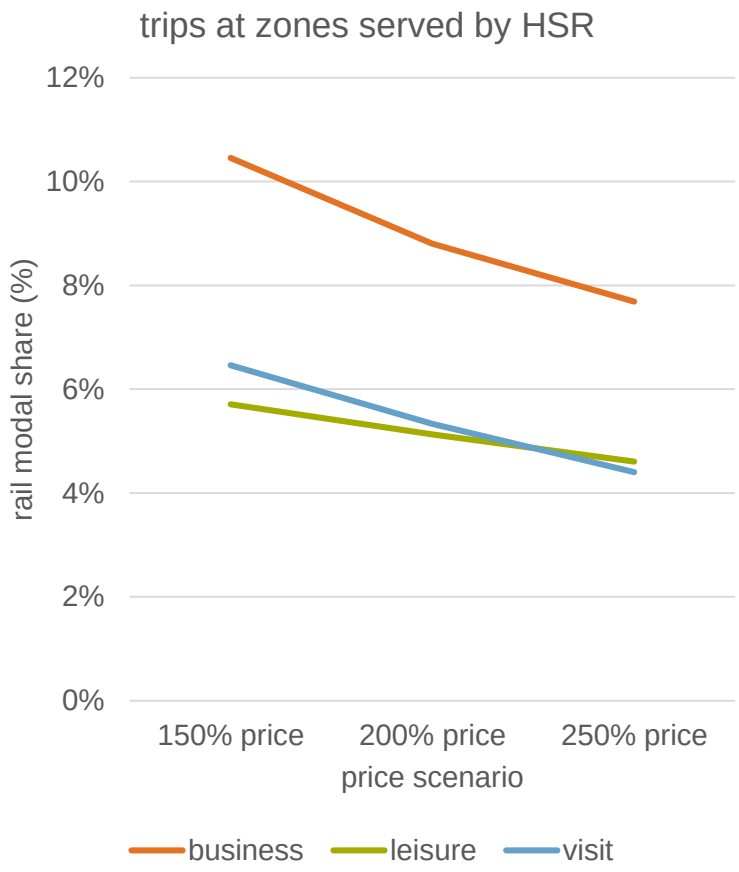


Model application

Results:

Modal share by travel cost and purpose (with max. speed = 400 km/h and frequency increases by 50%)





Discussion and conclusions

Methodological contributions:

- Usage of location-based social network data (Foursquare):
  - Pro: improved goodness of fit of destination choice models – especially leisure (+8% in LL)
  - Cons: prediction of such data in long-term estimations, bias
- Usage of trip-planning services (Rome2rio):
  - Pro: complete, quick and accessible, one single source, no assignment needed, close to users' decision criteria
  - Cons: congestion of modes cannot be directly considered – no assignment performed

Discussion and conclusions

Sensitivity analysis of HSR scenarios:

- Improvement of level of service of existing alternatives (vs. stated preference survey):
  - Complete substitution of rail supply instead of alternative train services
  - No further assumptions on mode perception → travel behavior remains as today
- Impacts are reasonable and seem plausible – but lack of validation:
  - Positive impacts of travel time reduction and increase of frequency
  - Negative impacts of increase in price
- Provincial model for a corridor-specific study:
  - Models are representative of the provincial demand –on average most of trips made by car
  - Suitable to delimitate the area of influence of the corridor measure – planning level

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.