VALIDATING AND CALIBRATING A DESTINATION ESTIMATION ALGORITHM FOR PUBLIC TRANSPORT SMART CARD FARE COLLECTION SYSTEMS

Li He[1,2], Research Associate

Neema Nassir[3], Postdoctoral Research Fellow

Martin Trépanier[1,2], Professor

Mark Hickman[3], ASTRA chair & Professor

1 POLYTECHNIQUE MONTRÉAL
LE GÉNIE EN PREMIÈRE CLASSE

2 CIRRELT
Interuniversity Research Center on Entreprise Networks, Logistics and Transportation

3 THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

In this presentation

- **Introduction**
  - Why do we need to estimate a destination?
- **Background**
  - Smart card data in public transport planning
- **Methodology**
  - Data source
  - Destination estimation algorithm(s)
- **Results**
  - Accuracy, estimation phase, temporal analysis
  - Tolerance distance calibration
- **Conclusion**

Introduction

« Tap-in » smart card systems

- Smart card data is very useful to transport planners because it is a **continuous source of data** on ridership and travelers' habits
- Many smart card automated fare collection systems **only validate the transactions at the entrance** of vehicles/ stations (« tap-in » only systems)
- For some studies (ie. models fed by OD matrices), there is a need to **estimate the destination** for each boarding transaction

Introduction

Aim of the research

- Through the years, a **destination estimation algorithm** has been developed to add « tap-out » information to the Gatineau, Canada, smart card dataset.
- Many studies use destination estimation algorithms based on the sequence of transactions during the day, but **none really validated the results** → Munizaga et al. tried to match smart card data and household surveys, mainly to validate survey responses
- The aim of this study is to apply to Australian « tap-in / tap-out » data the algorithms developed for Canadian datasets
  - To **validate** the algorithm
  - To help to **calibrate** the algorithm

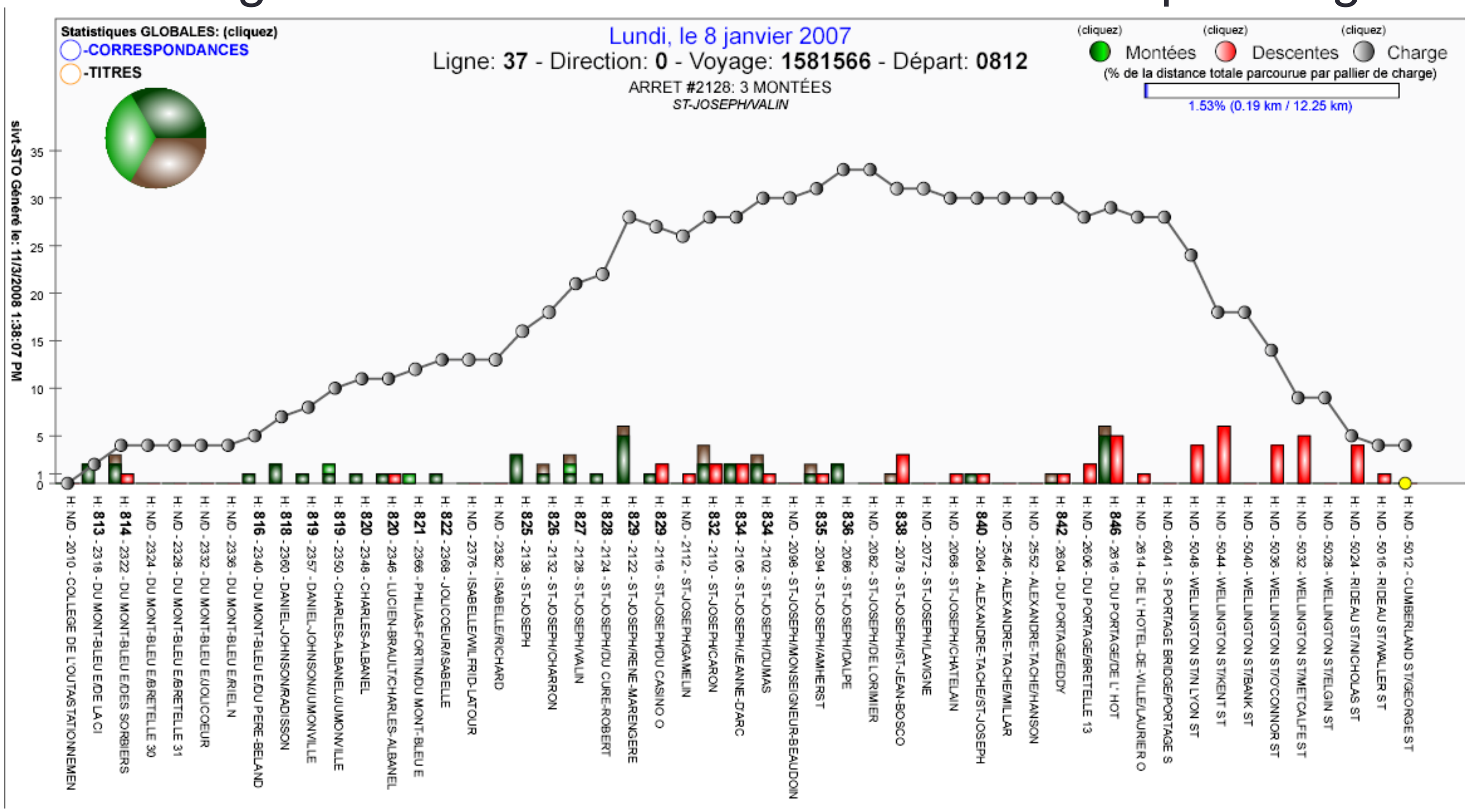
Background
Smart card data in public transport planning

- A smart card system collects data on **every transaction** aboard vehicles or stations
  - Date and timestamp, card number, fare type, route, location, etc.
  - Data is usually collected asynchronously (2-3 days delay)
- Data is **useful** for **planning**
  - Universal and **continuous source of data** on ridership, evolution, by fare type, etc.
  - **Classification** of passengers with data mining based on daily, weekly, monthly behaviour → *welcome to big data community*
  - Calculation of **performance indicators** for both demand and supply
  - **Loyalty** to service, **turnover** rates

Background

Smart card data in public transport planning → destinations

- For each trip on public transit, **having the destination** is essential to:
  - Obtain **load profile** of the route, for each run, vehicle, stop, etc.
  - Obtain Origin-Destination **matrices** to be used in planning tools



Methodology
Data source

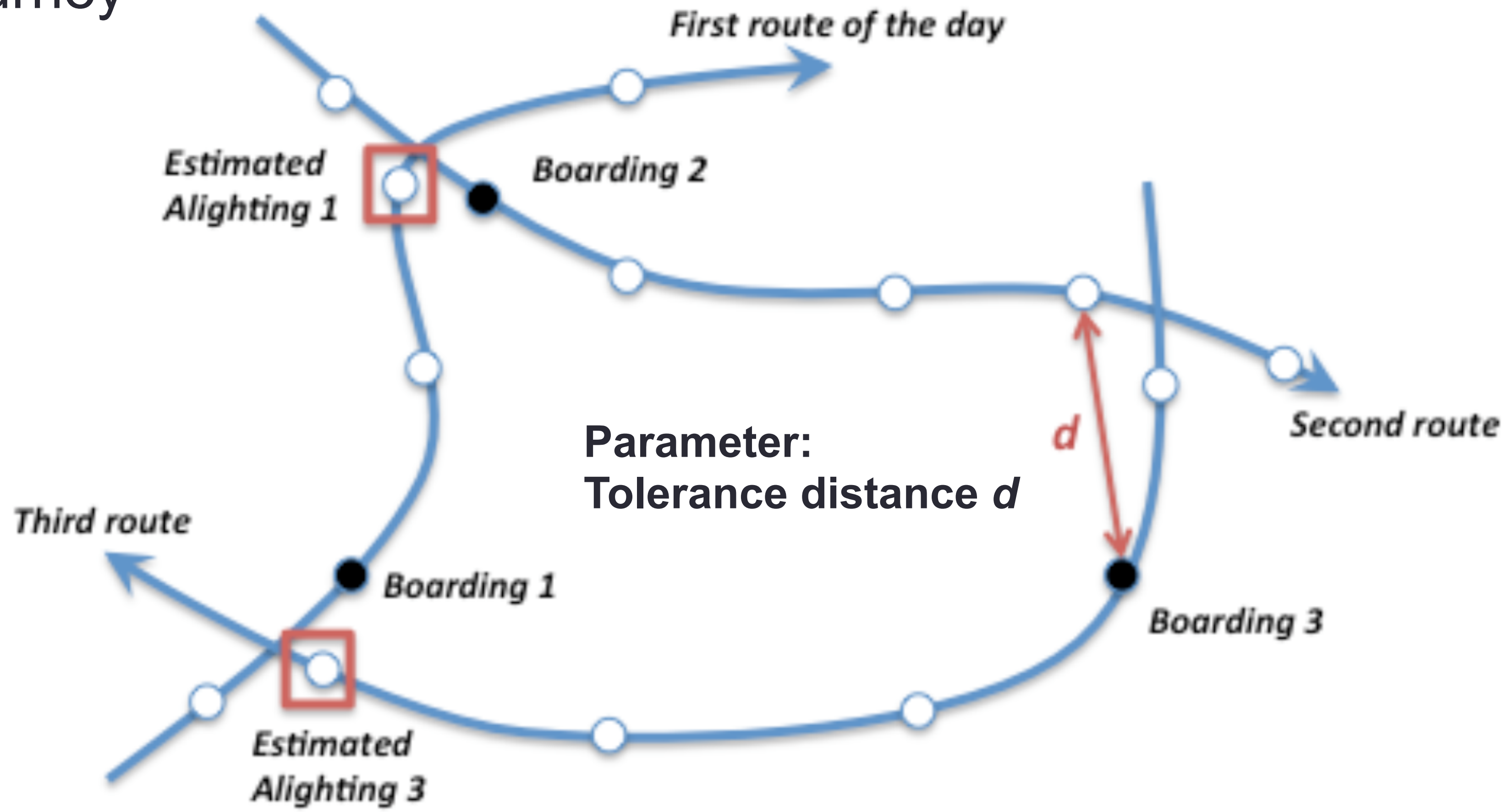
- Go Card from **Brisbane**, Australia
- Used by approximately **85%** of the travelers
- 40,341 trips made in March 2013 by a **random set** of card users

- Tap-in & tap-off information available: location of boarding and alighting stops

+ GTFS data of March 2013 for the transit network

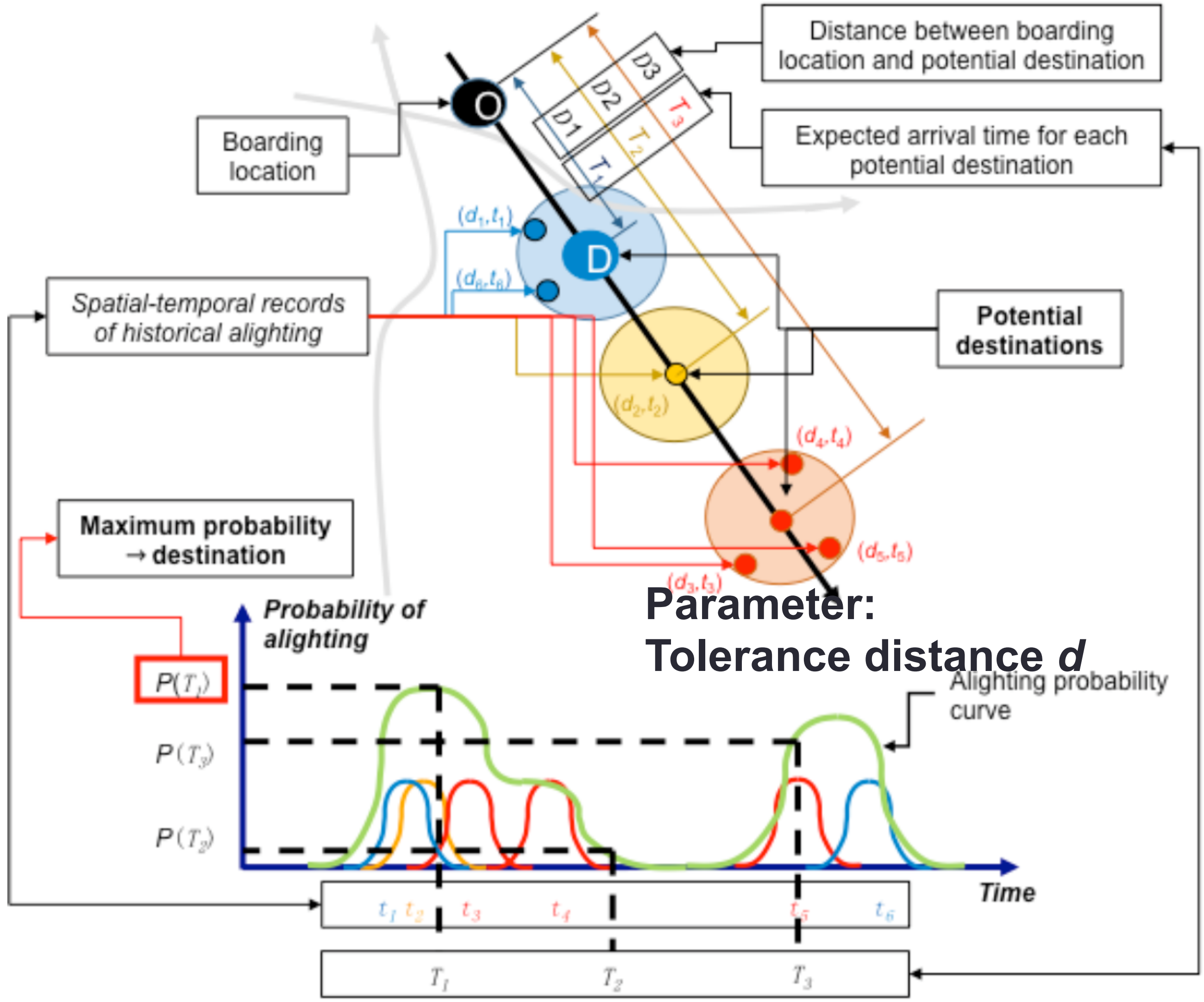
Methodology
Destination estimation algorithm (part I)

- This part is based on the **sequence** of trips made during a journey



Methodology
Destination estimation algorithm (part II)

- This part is used to process « **unlinked** » trips by looking at the history of the cards
- Probability from a kernel density method



Parameter: Tolerance distance *d*

Methodology
Validation

- The destination stop is **estimated** with the algorithms
- The estimated stop is **compared** to the real « tap-off » observation

- We use a **distance threshold** for the accuracy:
  - Estimated stop can be the same as the tap-off (distance of 0 metre)
  - Estimated stop can be near the tap-off (distance > 0 metre)

- We also try to **calibrate** the tolerance distance parameter of the parts I & II of the algorithm

Methodology

Estimation codes

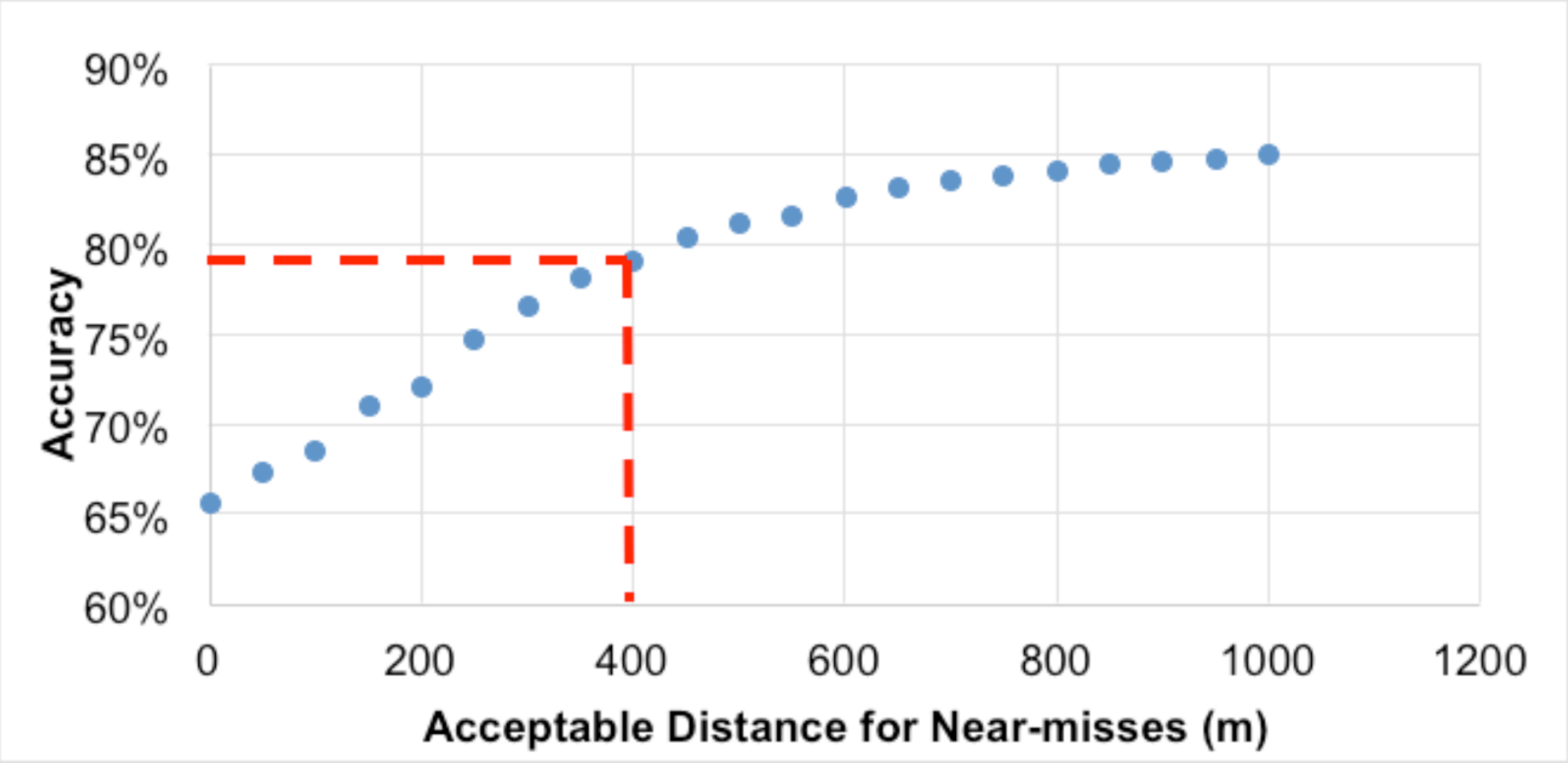
Given to every stop based on the step of the algorithms used to find destination

- **11** – Part I, trip following another
- **12** – Part I, destination is found using the first trip of the day (return to home)
- **13** –Part I, destination is found using the first trip of the next day
- **21** –Part II, destination found with the kernel density method, many choices
- **22** –Part II, destination found with the kernel density method, only one choice

Results

Overall accuracy

- The **accuracy** of the algos varies from 65% at 0m to 80% at 400m distance threshold

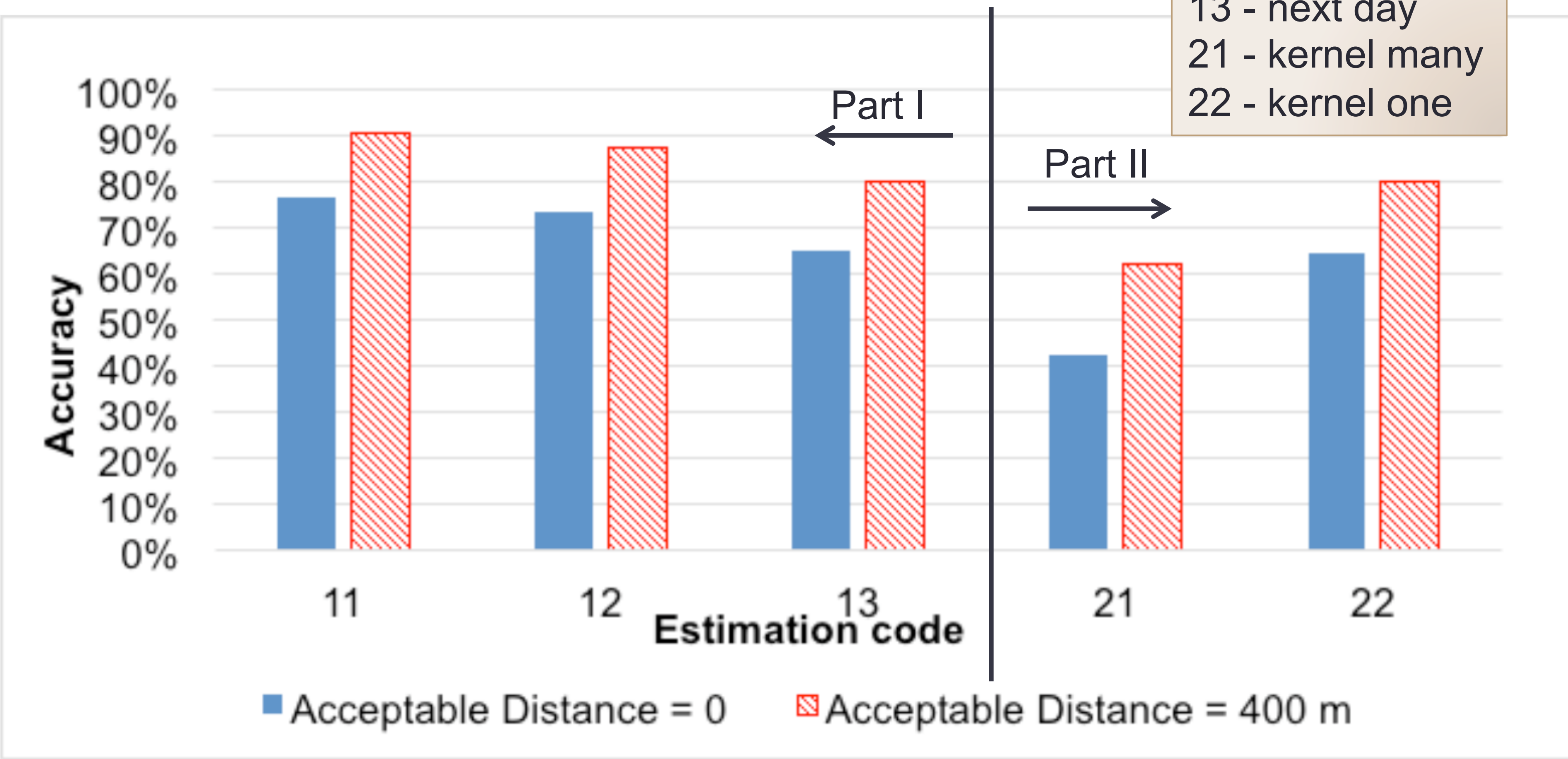


Results

Accuracy per estimation code

- As expected, accuracy is higher for **part I**

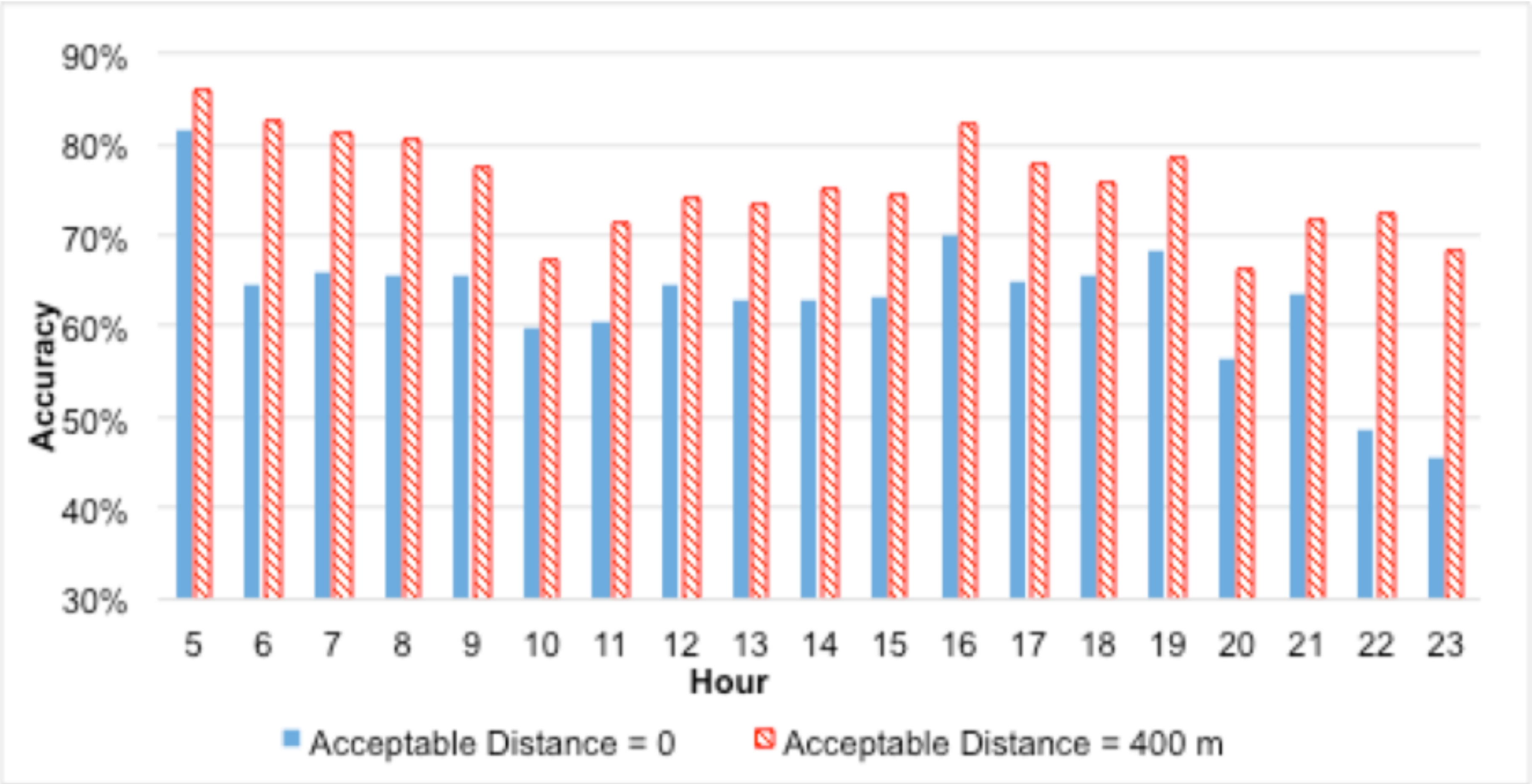
11 - sequence
12 - first trip
13 - next day
21 - kernel many
22 - kernel one



Results

Accuracy per hour of the day

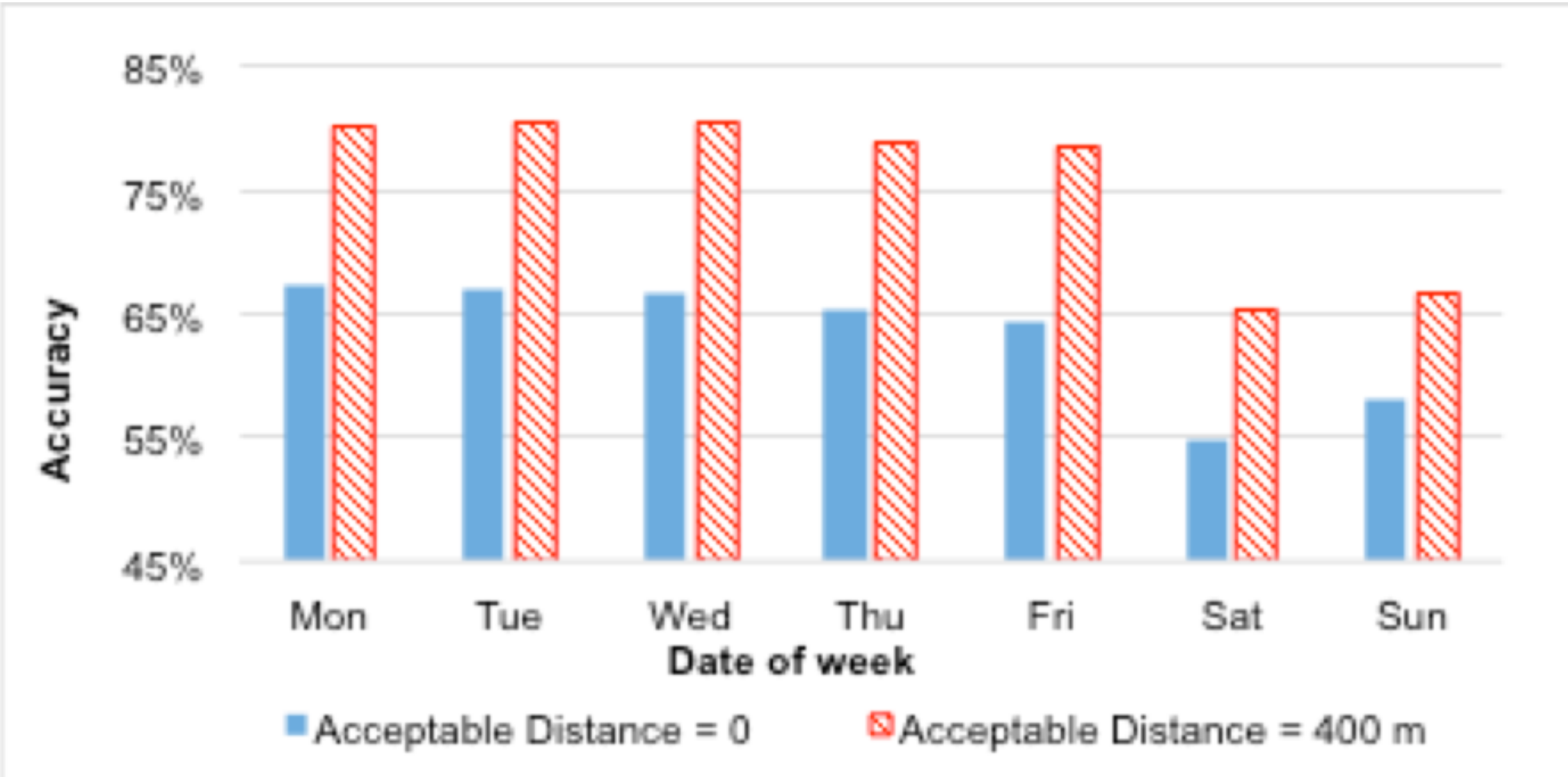
- Accuracy is higher at **peak hours** (regular trips)



Results

Accuracy per day of the week

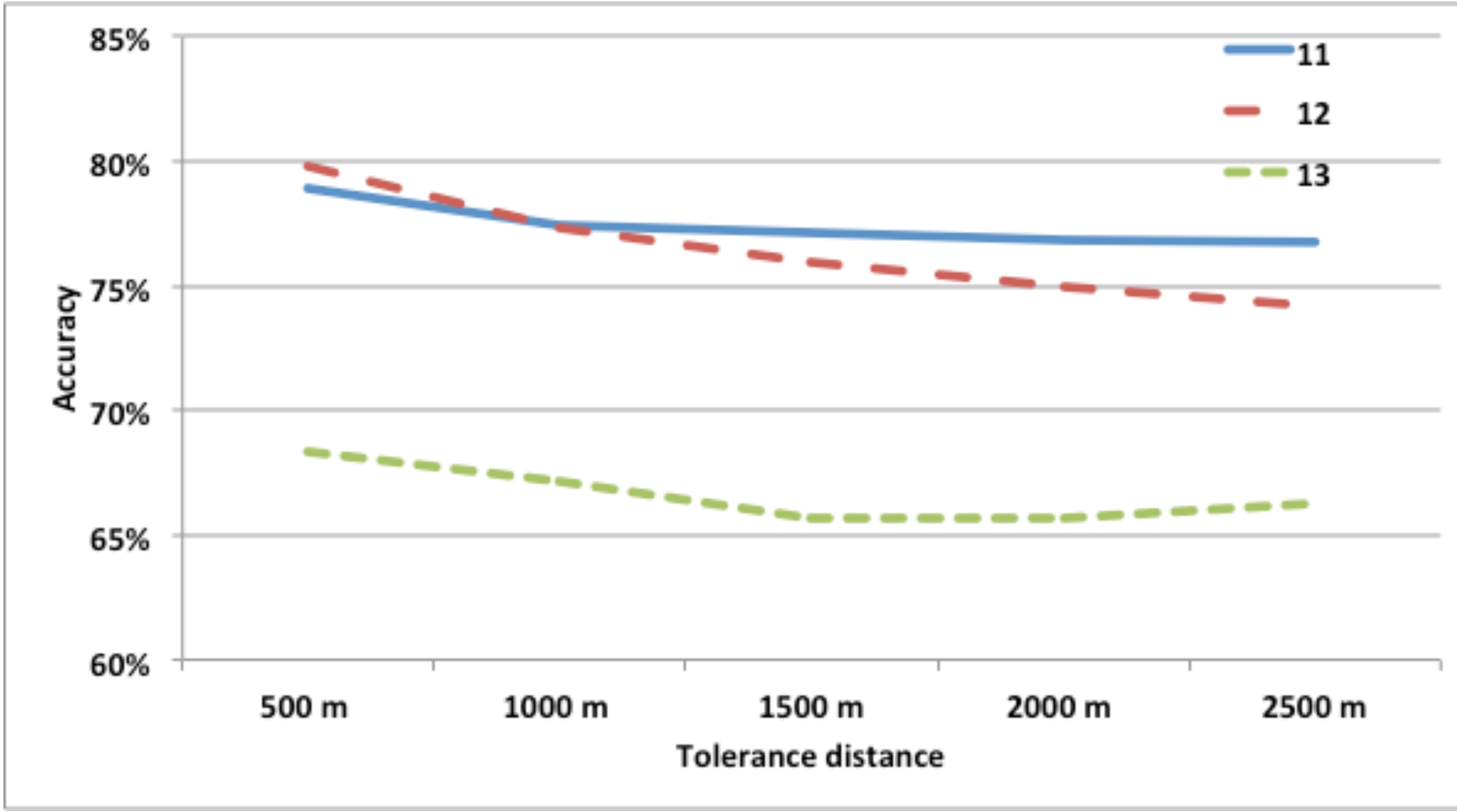
- Accuracy is higher on **weekdays**



Results

Calibration of the tolerance distance (pt. I)

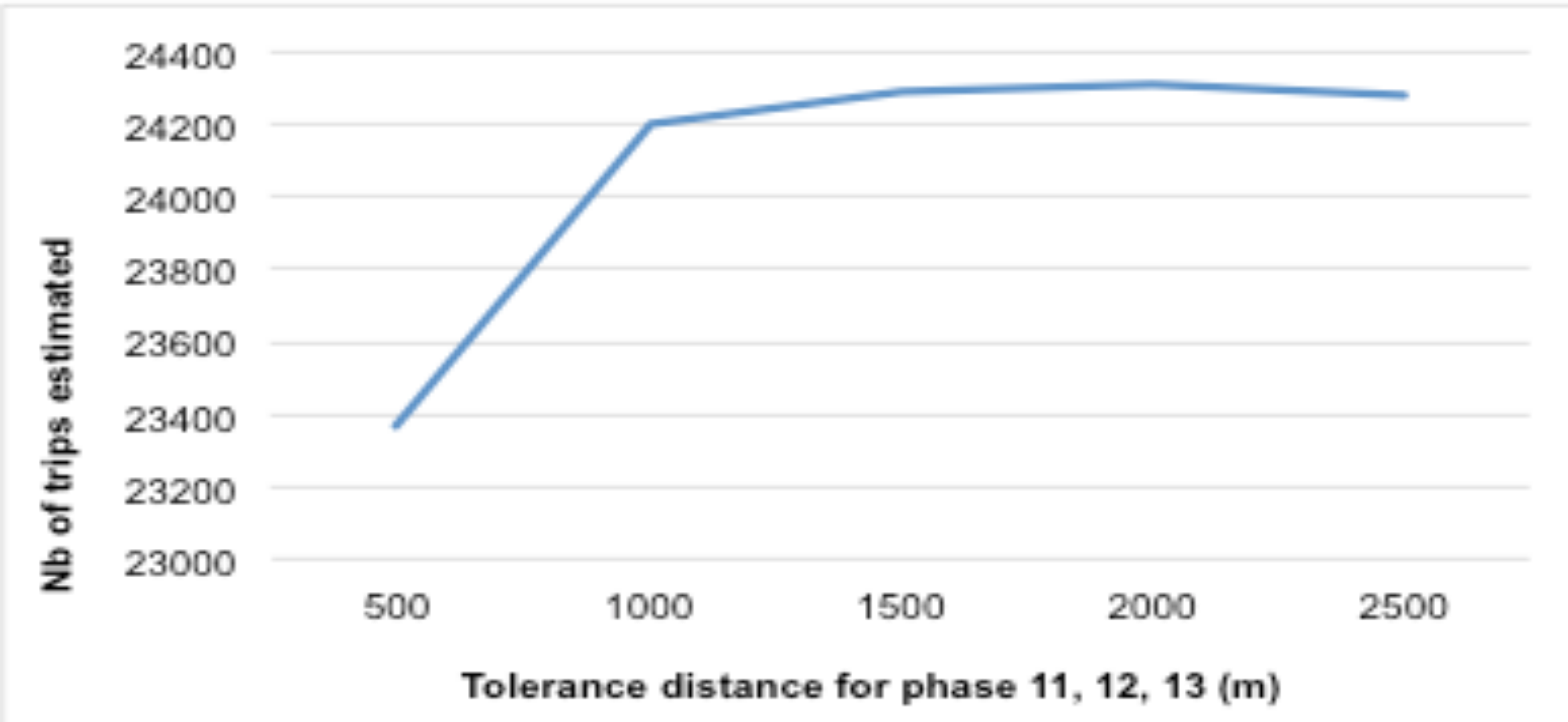
- Accuracy decreases with higher distance



Results

Calibration of the tolerance distance (pt. I)

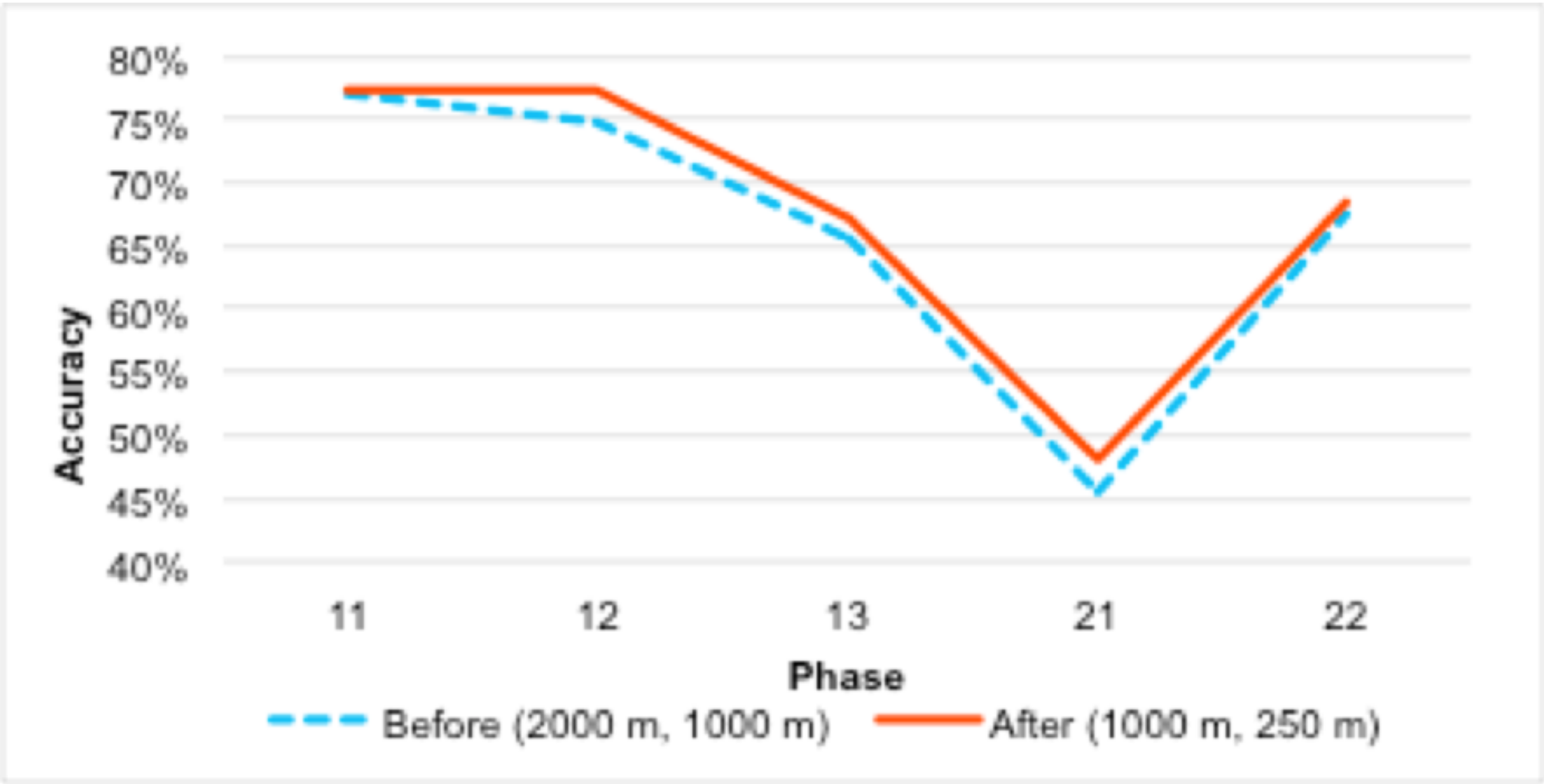
- However, the number of destinations increases, so there is a trade-off to set between the tolerance distance and the accuracy



Results

Calibration results

- There is a **slight improvement** after calibration process (+ 1 to 2%)



Conclusion

- We proposed a **validation** of the destination estimation algorithm with tap-in/tap-off data from Brisbane, Australia
- The results are: 65% accuracy at 0m distance threshold, **80% at 400m** (+ 1 to 2% after calibration)
- Results may show that:
  - Many transit users walk or use other modes between transit trips, making it difficult to find true destination
  - Irregularities of trips make it difficult to estimate
- However, accuracy of 80% on almost 85% of the trips is a very good start to **estimate an OD matrix** for each route, zone, etc. → better than survey!
- Many indicators (pass-km, pass-hr) do not need full accuracy

Acknowledgements

STO Société de transport de l'Outaouais

Queensland Government
Department of Transport and Main Roads

THALES

CRSNG
NSERC