Preferred citation style for this presentation

Erath A.L., P. Fourie,L. Sun, S Ordóñez, A. Chakirov (2015) From Big Data to Smart Data: developing a MATSim model that runs on Smart Card Data, IVT, ETH Zurich, Zurich, 7th August 2015

From Big Data to Smart Data: developing a MATSim model that runs on SmartCard Data

Seminar
IVT
ETH Zurich
7th August 2015

(FCL) FUTURE CITIES LABORATORY 未来城市实验室

(SEC) SINGAPORE-ETH CENTRE 新加坡－ETH研究中心

The potential of data driven transport planning

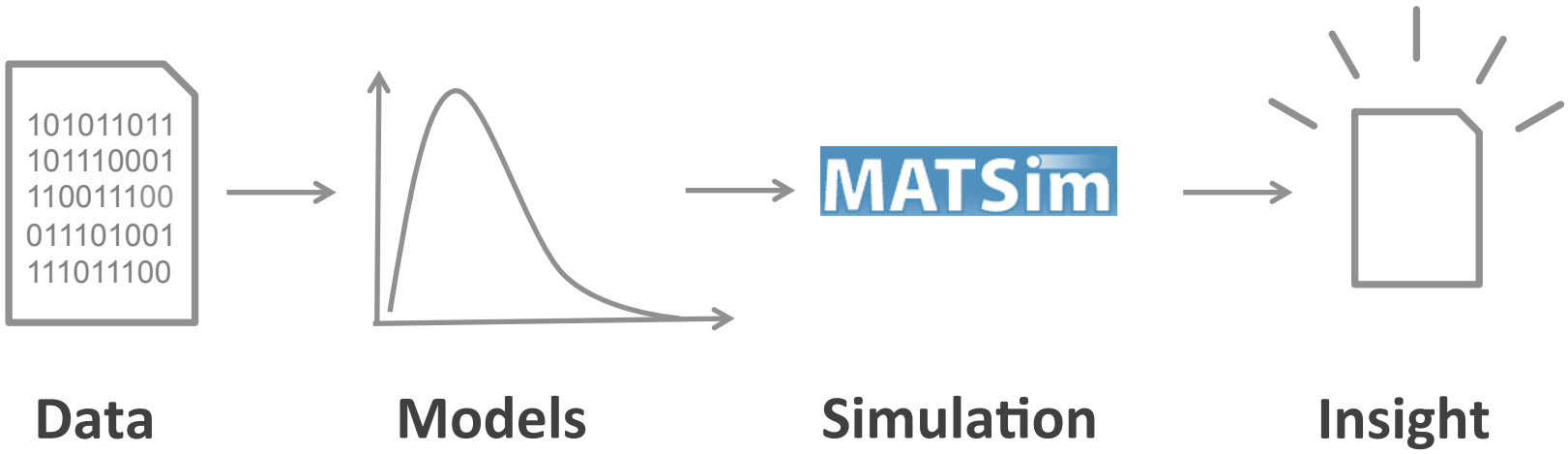
The potential of agent-based simulation

1. Traffic flows result from the movement of **single vehicles** and **individual people** and their interactions.

2. Accounts for dynamics that arise from **dwell processes, overcrowded vehicles** and congestion induced **delays** on links.

3. MATSim allows for big simulation scenarios with **millions of agents** and hundreds of public transport services.

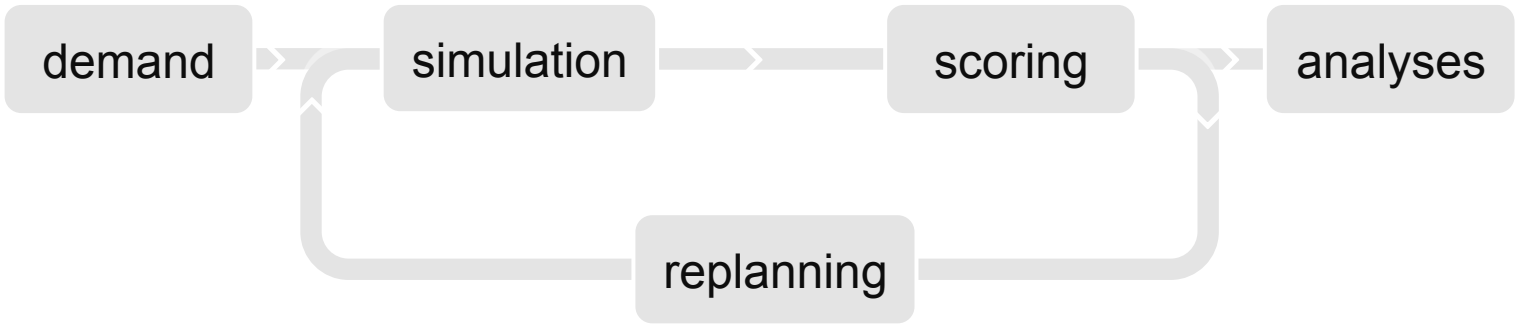
Turning Big Data into Smart Data



MATSim – Multi Agent Tansport SIMulation

- Open source software
- Under constant development
- well documented at **www.matsim.org**
- Can simulate scenarios with millions of agents

Engages a learning cycle to find (stochastic) equilibrium



Data

Singapore CEPAS system – quick facts & stats

Closed system on train and bus – distance-based charging

Adult, student, senior citizen cards

Senior citizens: 25% discount; students: 50% discount; both pay flat fare after 7.2km (so sometimes don't bother tapping out)

360+ bus services, 3000 buses, 5 heavy rail lines (153km, 104 stations), 3 light rail lines in new towns

10+ different bus types, information on deployment from bus spotter website.

Key statistics for 8 April 2013

- 4,138,780 journeys
- 5,675,986 stages (61% bus)
- 40,224,444 person-km (36% bus)
- 105,260 stages with no tap-out

Using real demand to simulate public transport

Travel demand directly derived from smart card transactions

- Transactions recorded on Tuesday, 22nd April 2011
- 4 Mio journeys, 5.7 Mio stages
- Adult, student, senior citizen cards
- Boarding stop (journey level)
- Boarding time
- Alighting stop (journey level)

Travel supply

- 362 bus lines
- 4 MRT lines
- 5 light rail serivces
- 1 Mono rail
- Each with detailed description of deployed vehicles

CEPAS smart card data (bus) to MATSim events

From this:

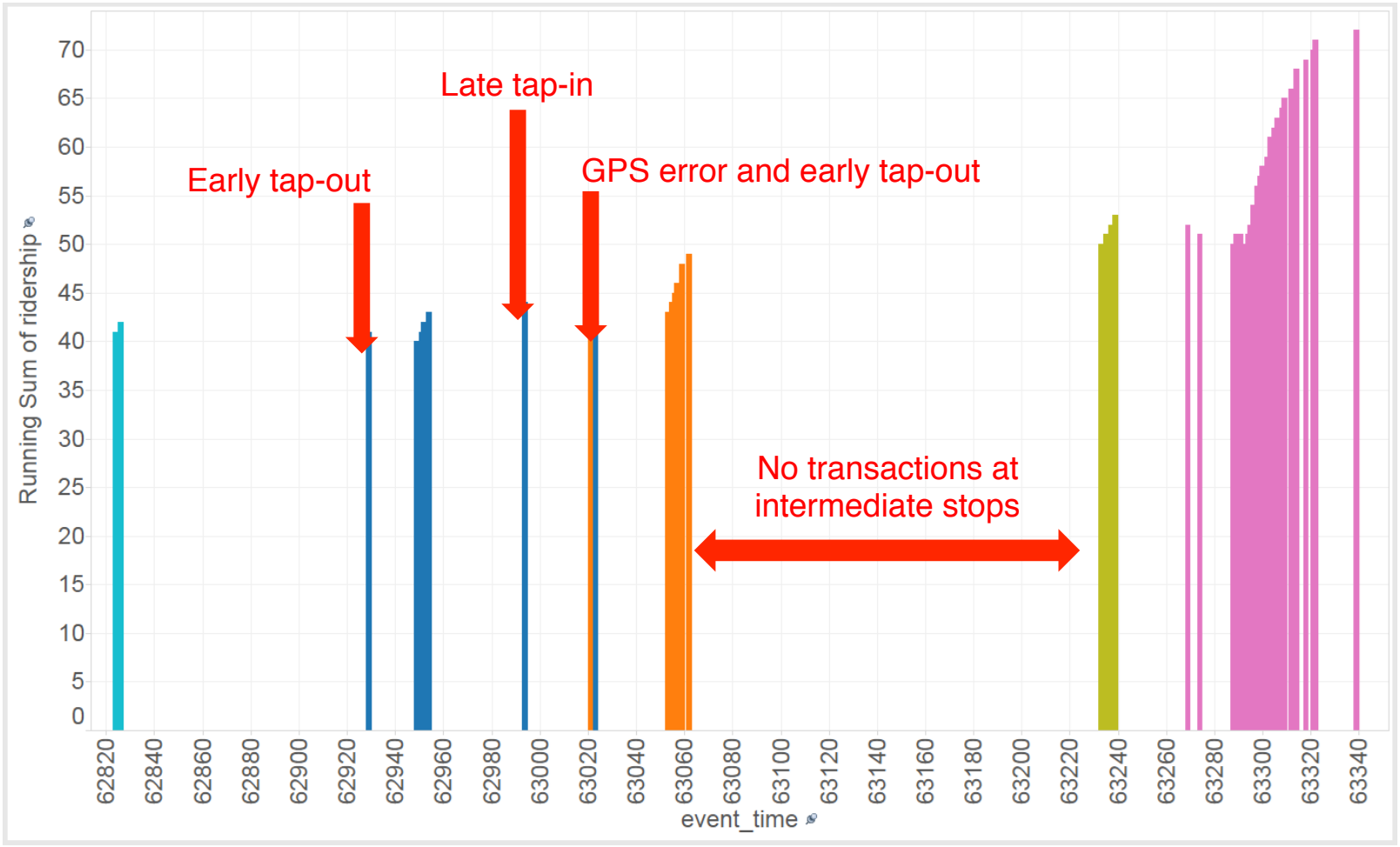
Column	Type
journey_id	bigint
card_id	bigint
passenger_type	character varying(14)
travel_mode	character varying(3)
srvc_number	character varying(5)
direction	character varying(1)
bus_reg_num	character varying(4)
boarding_stop_stn	character varying(20)
alighting_stop_stn	character varying(20)
ride_start_date	date
ride_start_time	time without time zone
ride_distance	double precision
ride_time	double precision
fare_paid	double precision
transfer_number	integer
tripid	integer

To this:

```
<?xml version="1.0" encoding="utf-8"?>
<events version="1.0">
    <event time="20407.996" type="TransitDriverStarts" driverId="pt_tr_2_0_2710" vehicleId="2_0_2710"
        transitLineId="2" transitRouteId="2_weekday_2-p" departureId="11468" />
    <event time="20407.997" type="departure" person="pt_tr_2_0_2710" link="67706" legMode="car" />
    <event time="20407.998" type="PersonEntersVehicle" person="pt_tr_2_0_2710" vehicle="2_0_2710" />
    <event time="20407.999" type="wait2link" person="pt_tr_2_0_2710" link="67706" vehicle="2_0_2710" />
    <event time="20408.0" type="VehicleArrivesAtFacility" vehicle="2_0_2710" facility="80051" delay="0.0" />
    <event time="20408.2" type="PersonEntersVehicle" person="9000123115170800" vehicle="2_0_2710" />
    <event time="20409.0" type="PersonEntersVehicle" person="2000063275680800" vehicle="2_0_2710" />
    <event time="20415.6" type="PersonEntersVehicle" person="1991035572811100" vehicle="2_0_2710" />
    <event time="20417.4" type="PersonEntersVehicle" person="9000055721680800" vehicle="2_0_2710" />
    <event time="20418.2" type="PersonEntersVehicle" person="2000032376700800" vehicle="2_0_2710" />
    <event time="20423.9" type="PersonEntersVehicle" person="9000033649531000" vehicle="2_0_2710" />
    <event time="20426.5" type="PersonEntersVehicle" person="9000226264400800" vehicle="2_0_2710" />
    <event time="20427.1" type="PersonEntersVehicle" person="2000032378970800" vehicle="2_0_2710" />
    <event time="20431.6" type="PersonEntersVehicle" person="2000001315531000" vehicle="2_0_2710" />
    <event time="20434.9" type="PersonEntersVehicle" person="2000145527100800" vehicle="2_0_2710" />
    <event time="20434.9" type="PersonEntersVehicle" person="8889105520747002" vehicle="2_0_2710" />
    <event time="20435.0" type="VehicleDepartsAtFacility" vehicle="2_0_2710" facility="80051" delay="0.0" />
    <event time="20435.0" type="left link" person="pt_tr_2_0_2710" link="67706" vehicle="2_0_2710" />
    <event time="20435.0" type="entered link" person="pt_tr_2_0_2710" link="67707" vehicle="2_0_2710" />
    <event time="20449.5" type="left link" person="pt_tr_2_0_2710" link="67707" vehicle="2_0_2710" />
    <event time="20449.5" type="entered link" person="pt_tr_2_0_2710" link="57601" vehicle="2_0_2710" />
    <event time="20458.68733191533" type="left link" person="pt_tr_2_0_2710" link="57601" vehicle="2_0_2710
    <event time="20458.68833191533" type="entered link" person="pt_tr_2_0_2710" link="51197" vehicle="2_0_2
    <event time="20468.029216544513" type="left link" person="pt_tr_2_0_2710" link="51197" vehicle="2_0_271
```

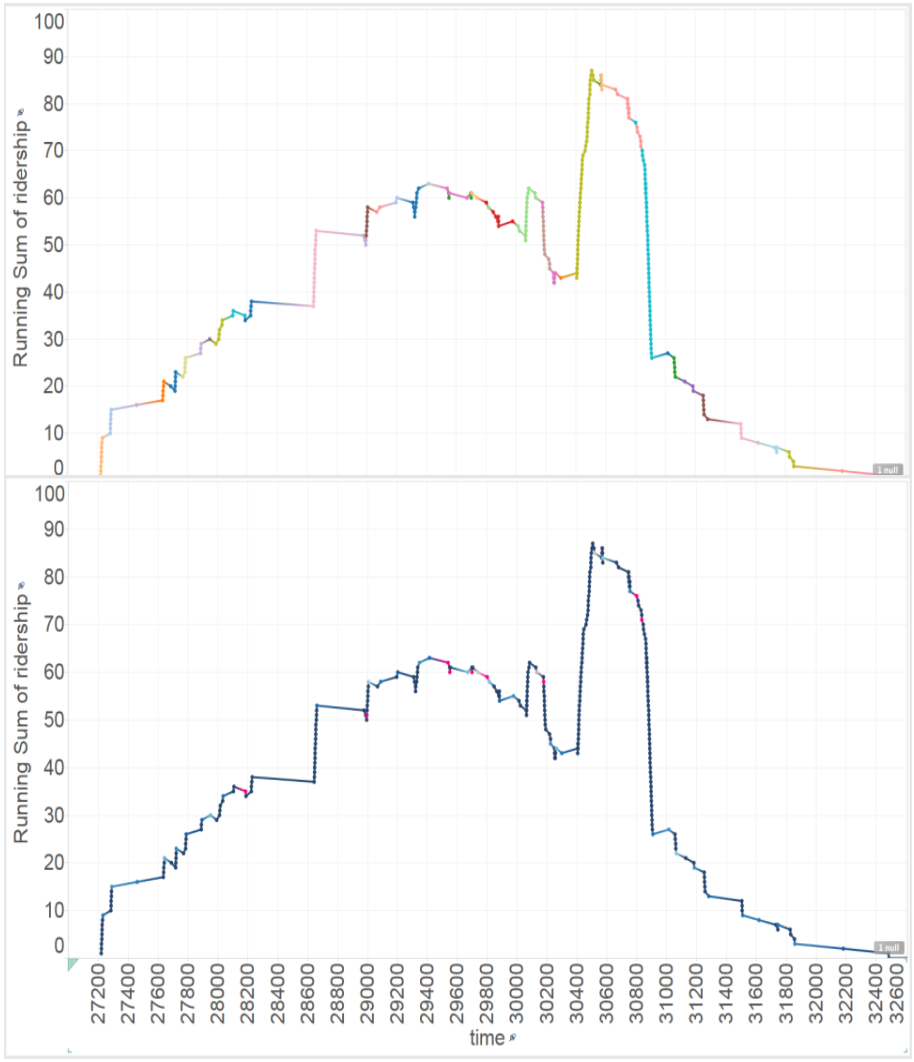
Extracting operational schedule from smart card data

Ridership vs transaction times, colored by stop ID



Fourie, Pieter Jacobus (2014). 'Reconstructing Bus Vehicle Trajectories from Transit Smart-Card Data', Working paper. Future Cities Laboratory.

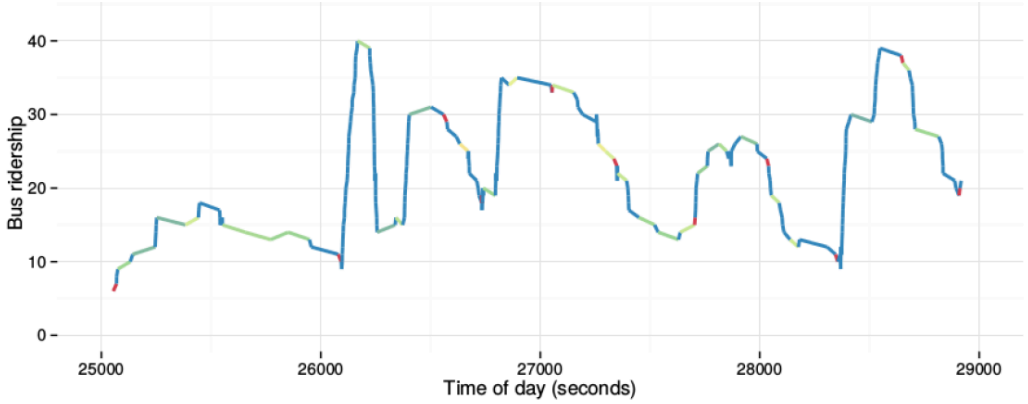
Removing GPS errors



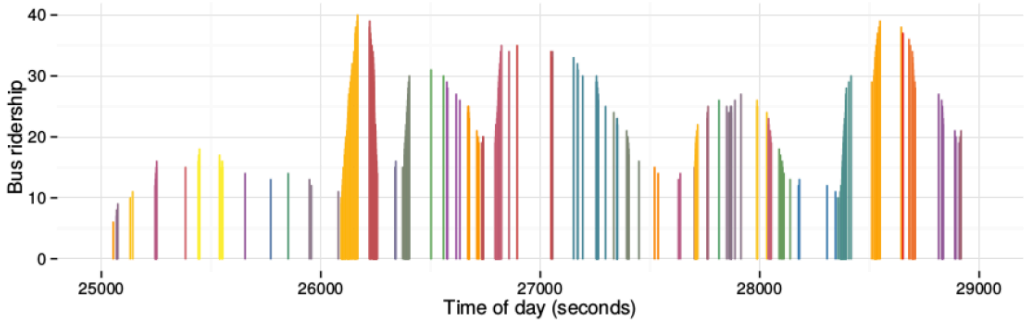
Two plots of ridership vs transaction times, colored by stop ID (top) and transaction ‘speed’ (bottom).

Note how high ‘speed’ transactions (pink to red) correspond to incorrect stop IDs in the top plot

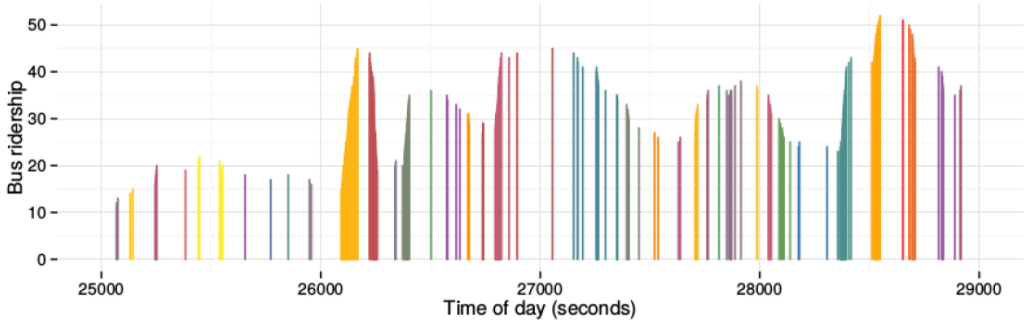
Removing GPS errors



(b) Transactions coloured by stop ID before filtering



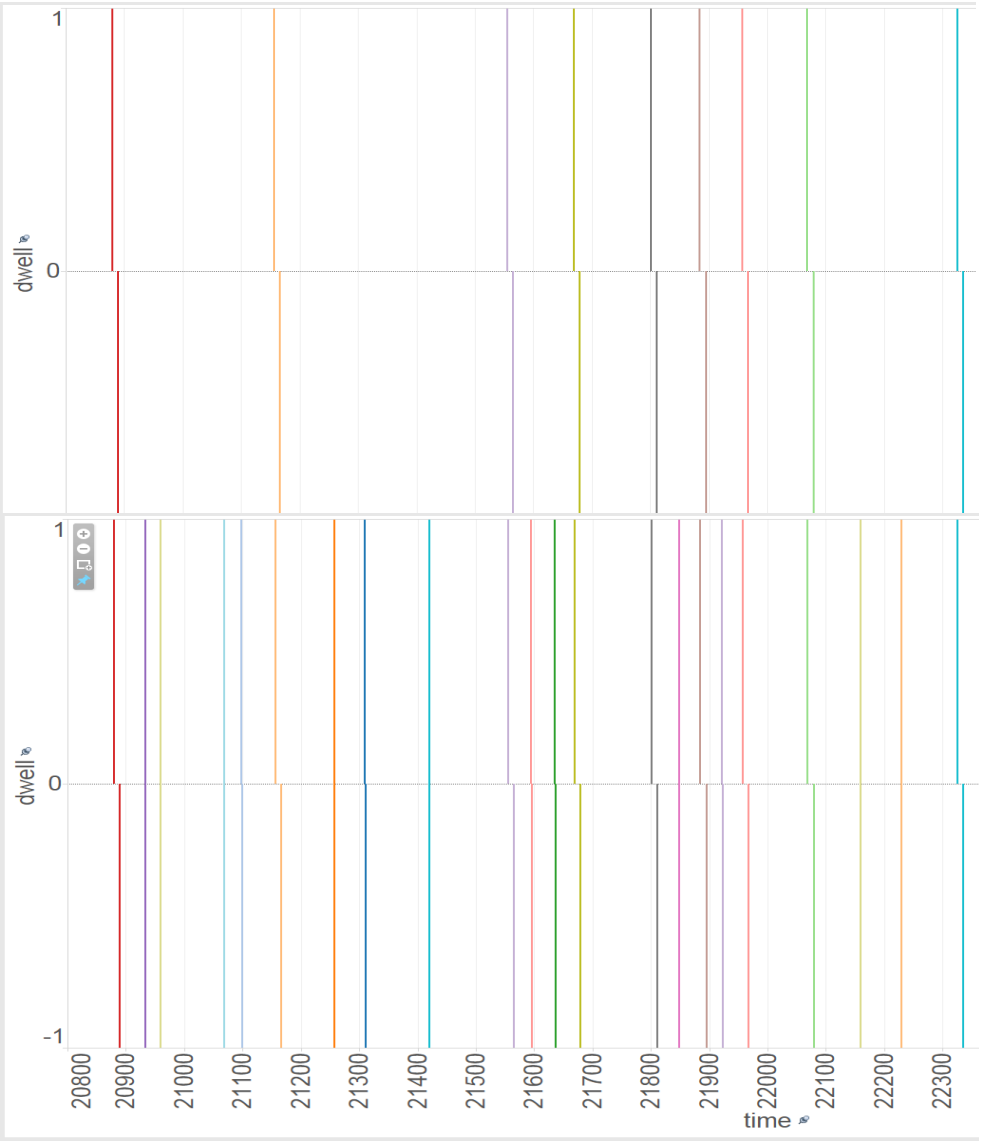
(c) Transactions coloured by stop ID after filtering.



Ridership vs transaction times, colored by stop ID (top) and transaction ‘speed’ (bottom).

Note how high ‘speed’ transactions (pink to red) correspond to incorrect stop IDs in the top plot

Interpolation of missing dwell events

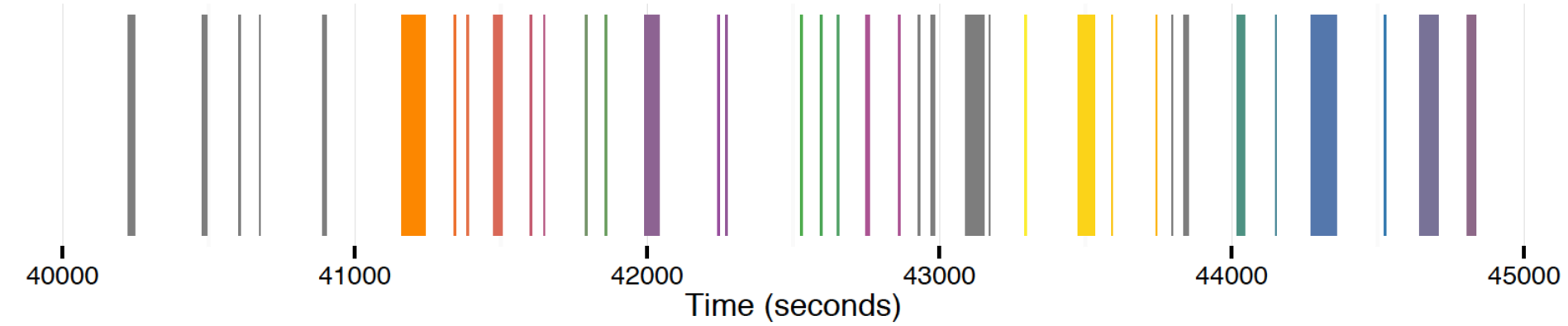


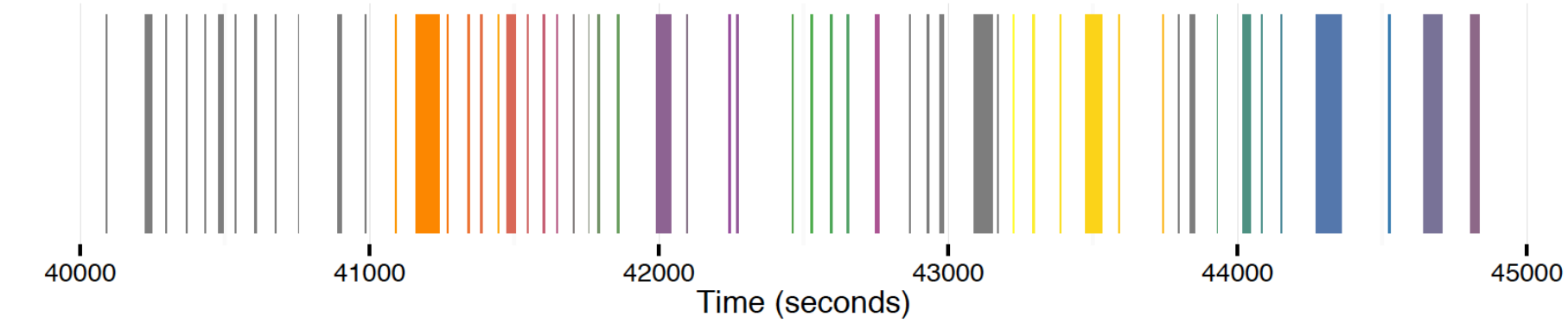
Two plots of bus arrivals (bars above zero) and departures (below zero), colored by stop ID.

The top plot shows dwell events before interpolation, the bottom includes interpolated dwell events, with a zero duration.

The travel time allocated between interpolated dwell events is proportional to the free speed travel time between stops.

Interpolation of missing dwell events



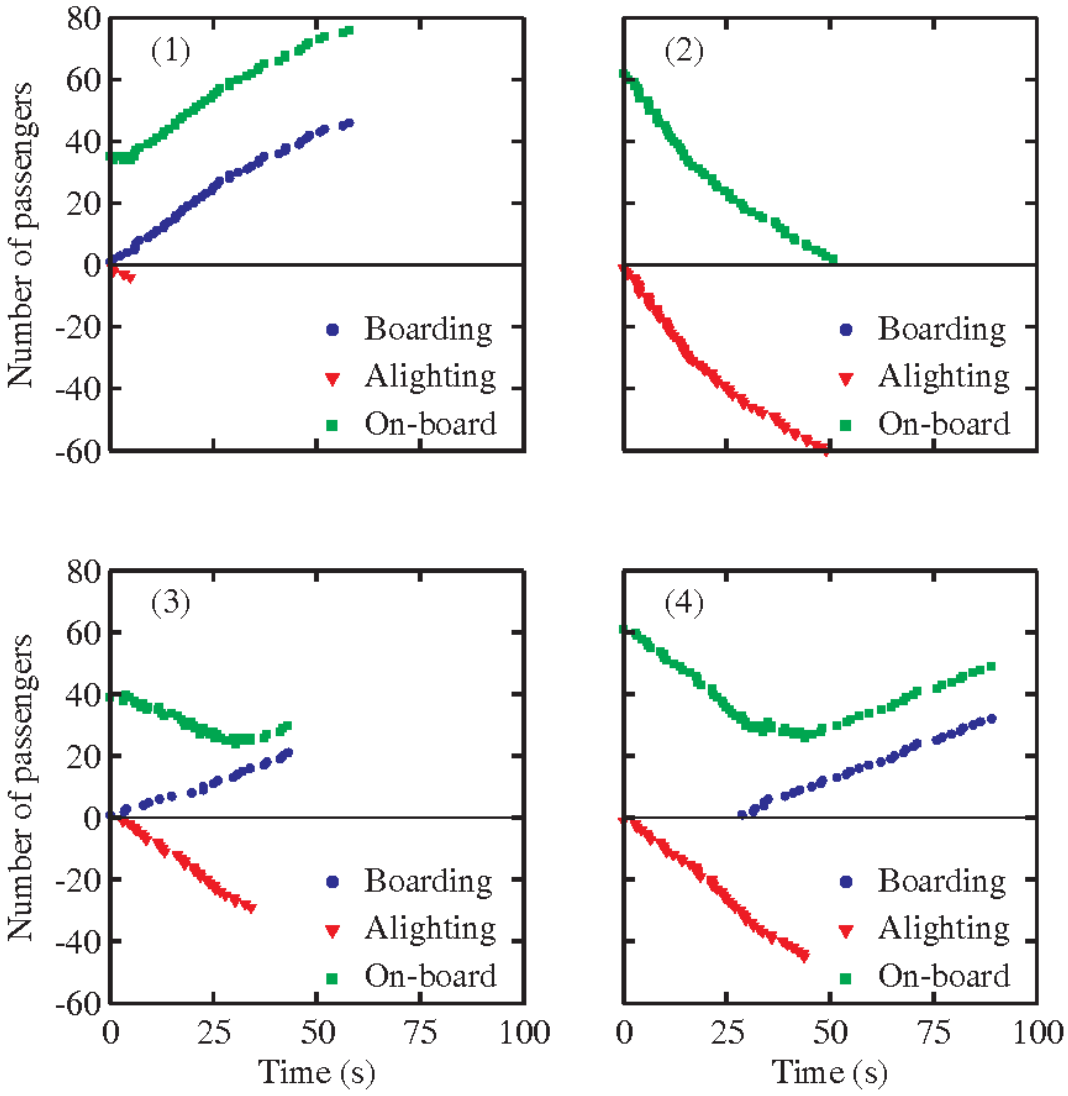


Error rates for 8th April 2013, bus only

Number of dwell events identified	1,143,619
Number of dwell events interpolated	466,723
Number of card transaction	6,775,855
Stop ID not part of route	156,437
Transaction dropped due to speed	87,352
Dwell event transaction dropped	148,162
Share of transactions dropped	5.8%

Models

Boarding and alighting process

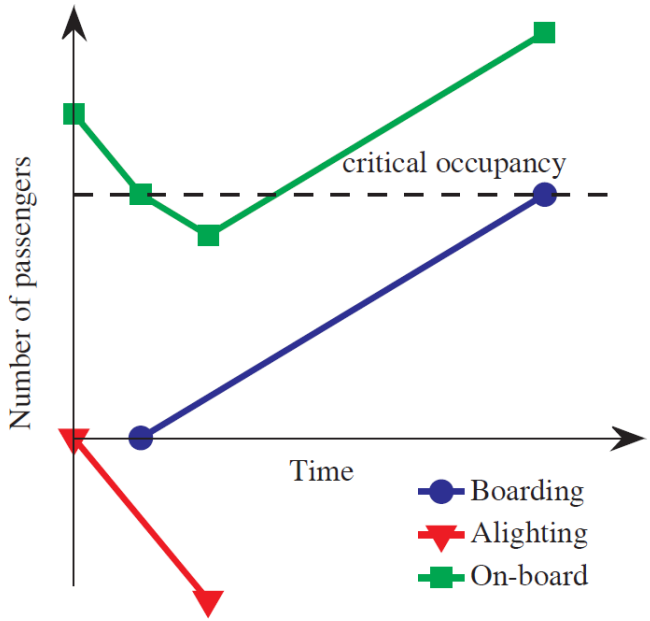


Boarding/alighting flow and instantaneous on-board passengers (single decker, low floor).

Sun, Lijun, Alejandro Tirachini, Kay W. Axhausen, Alexander Erath and Der-Horng Lee (2014). 'Models of Bus Boarding and Alighting Dynamics', *Transportation Research Part A: Policy and Practice* 69: 447–460.

Passenger activity time model

Boarding and alighting process



Statistical model

$$t = \{\beta B + a(\max(on - cr, 0)), aA\}$$

Sun, Lijun, Alejandro Tirachini, Kay W. Axhausen, Alexander Erath and Der-Horng Lee (2014). 'Models of Bus Boarding and Alighting Dynamics', *Transportation Research Part A: Policy and Practice* 69: 447–460.

Bus types in Singapore

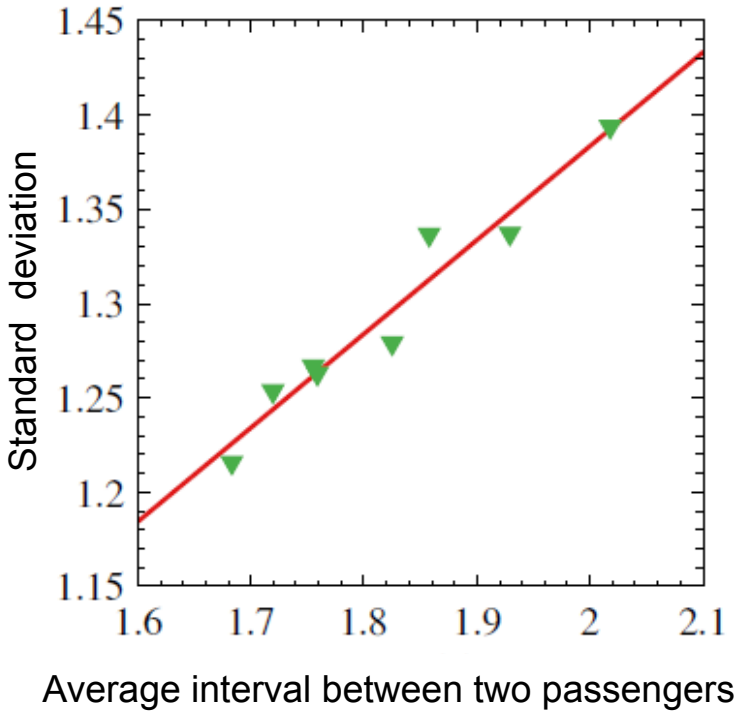
Standart type, single floor
Low floor or with step
1 boarding and alighting door each
83-85 pax capacity

Double decker
Low floor or with step
1 boarding and alighting door each
131 pax capacity

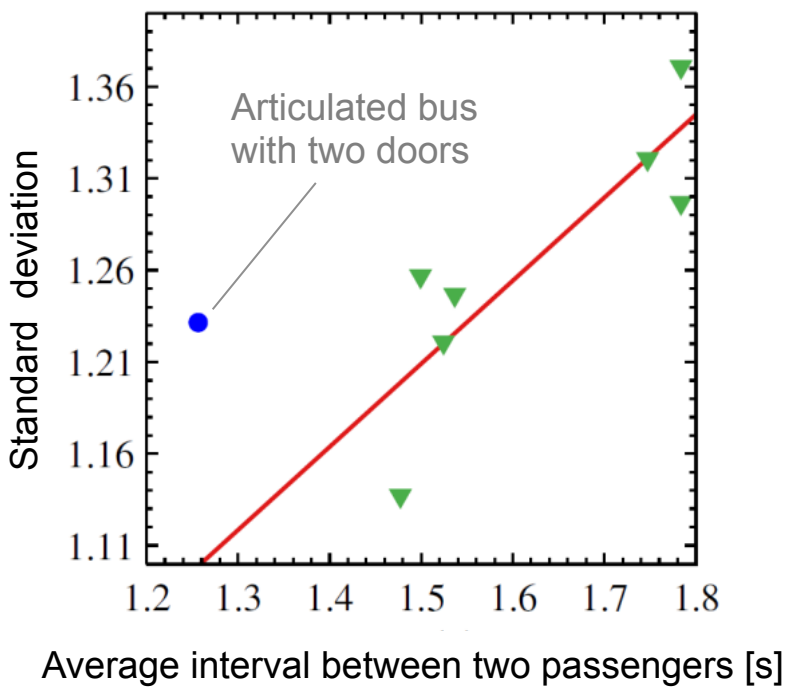
Articulated
With step
1 boarding and 2 alighting doors
132 pax capacity

Variability of activity time for different bus types

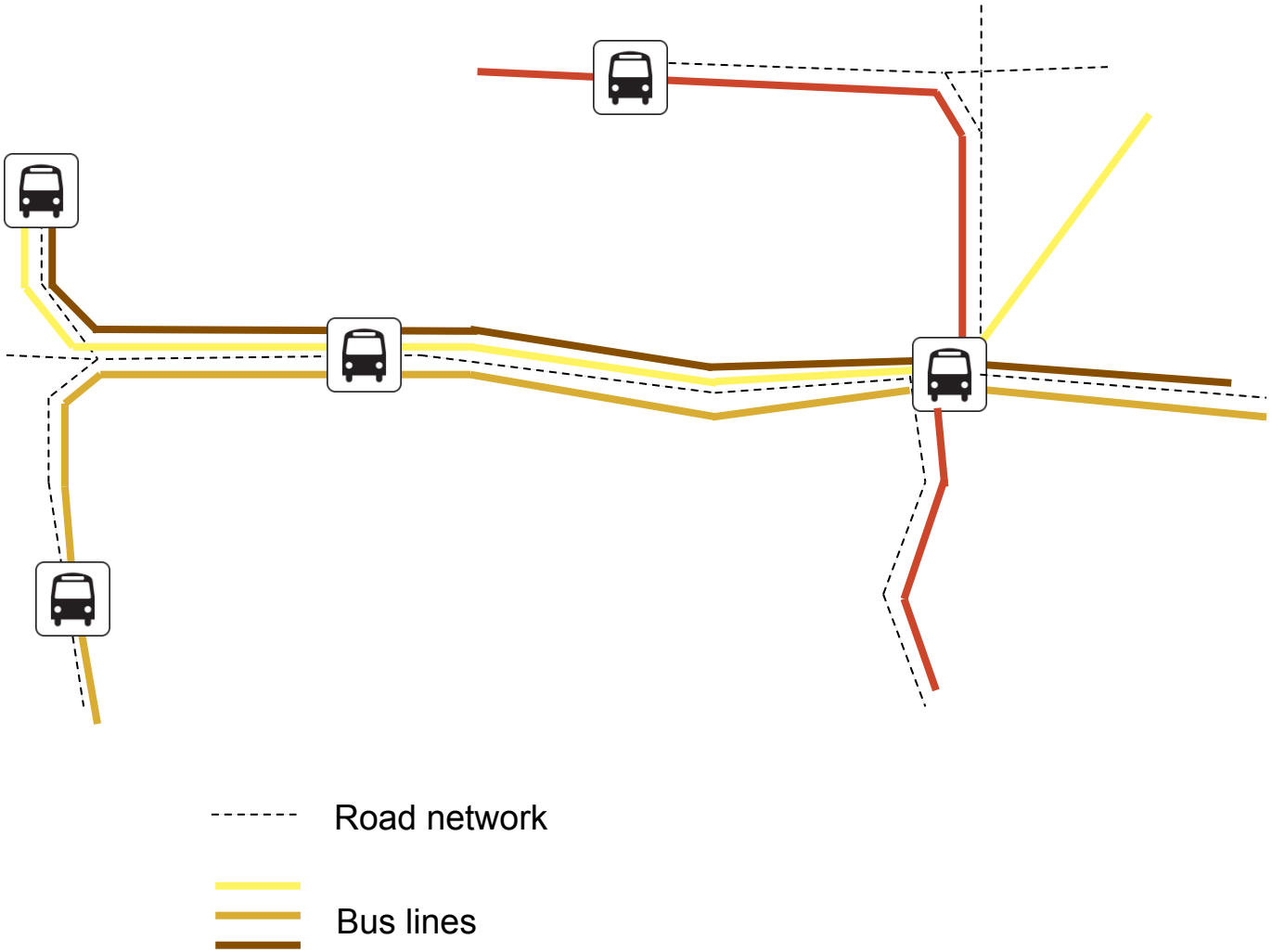
Boarding



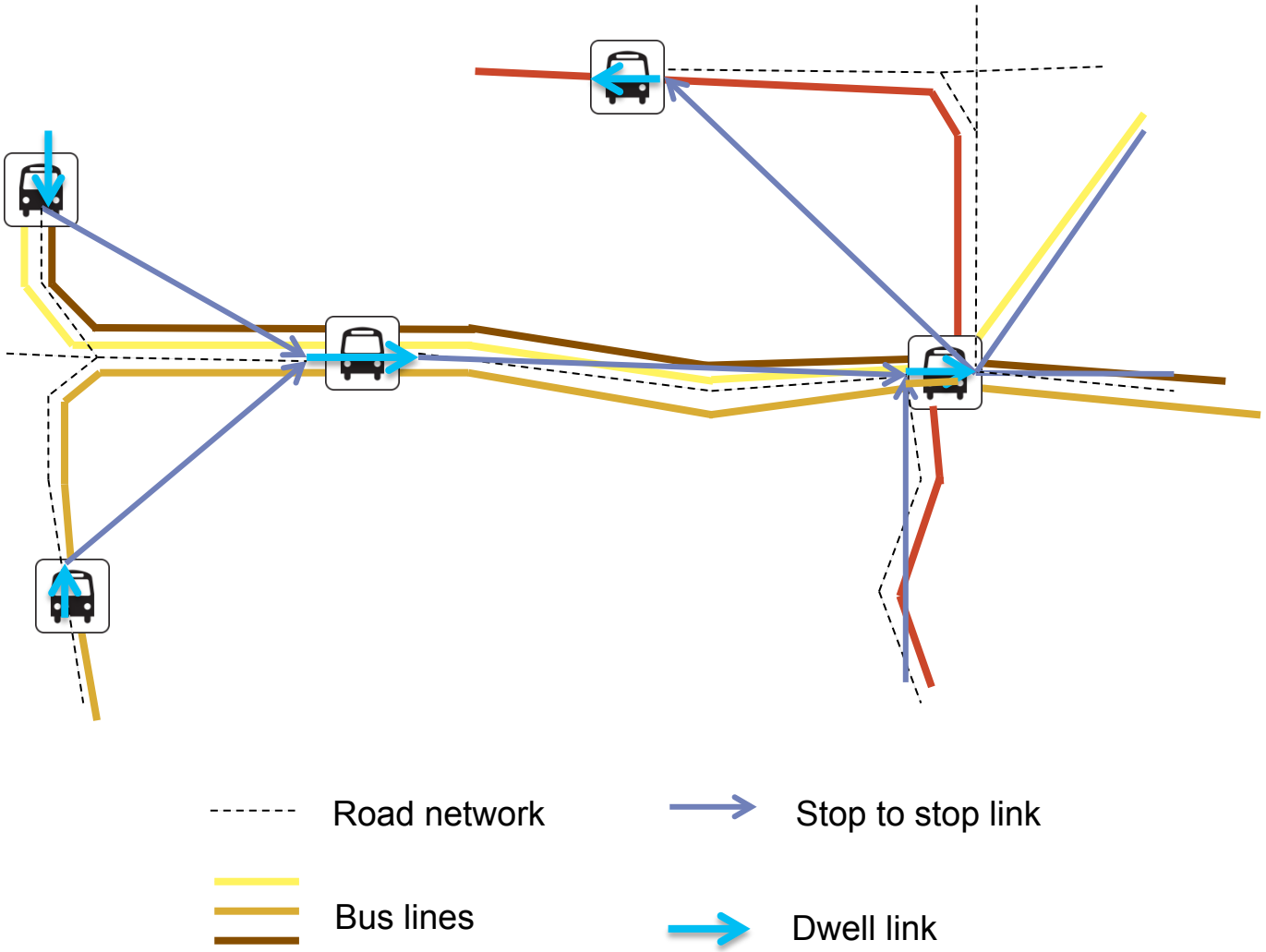
Alighting



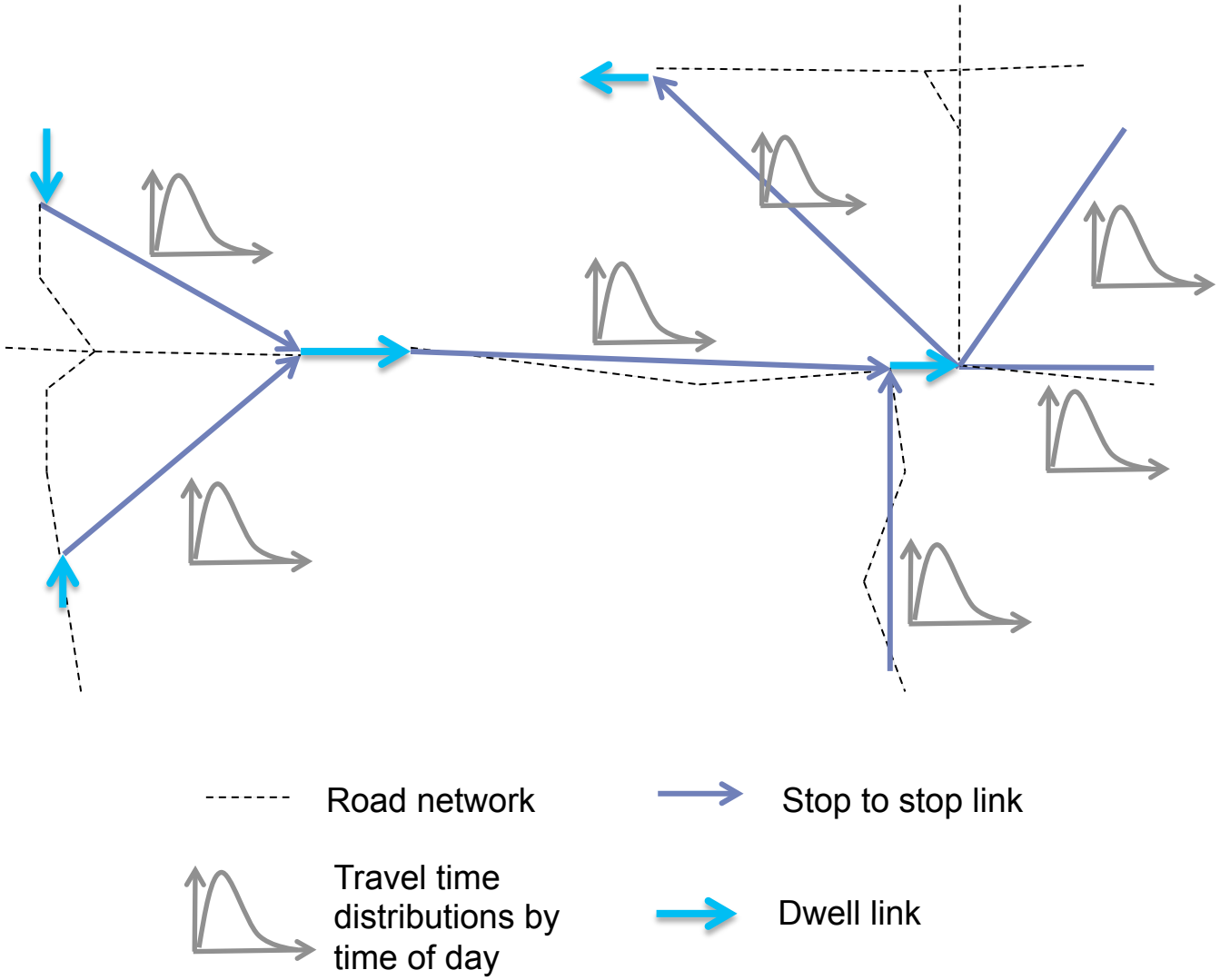
Variability of stop-to-stop travel time



Simplifying the network



Accounting for travel time variability



Modelling stop to stop travel times

Derive from Smart Card Data records travel times between stops

Each observed travel time between two subsequent stops contitutes one observation

Independent variables to be either derived from smart card data or GIS data, but do not require any other data source (e.g. traffic flow)

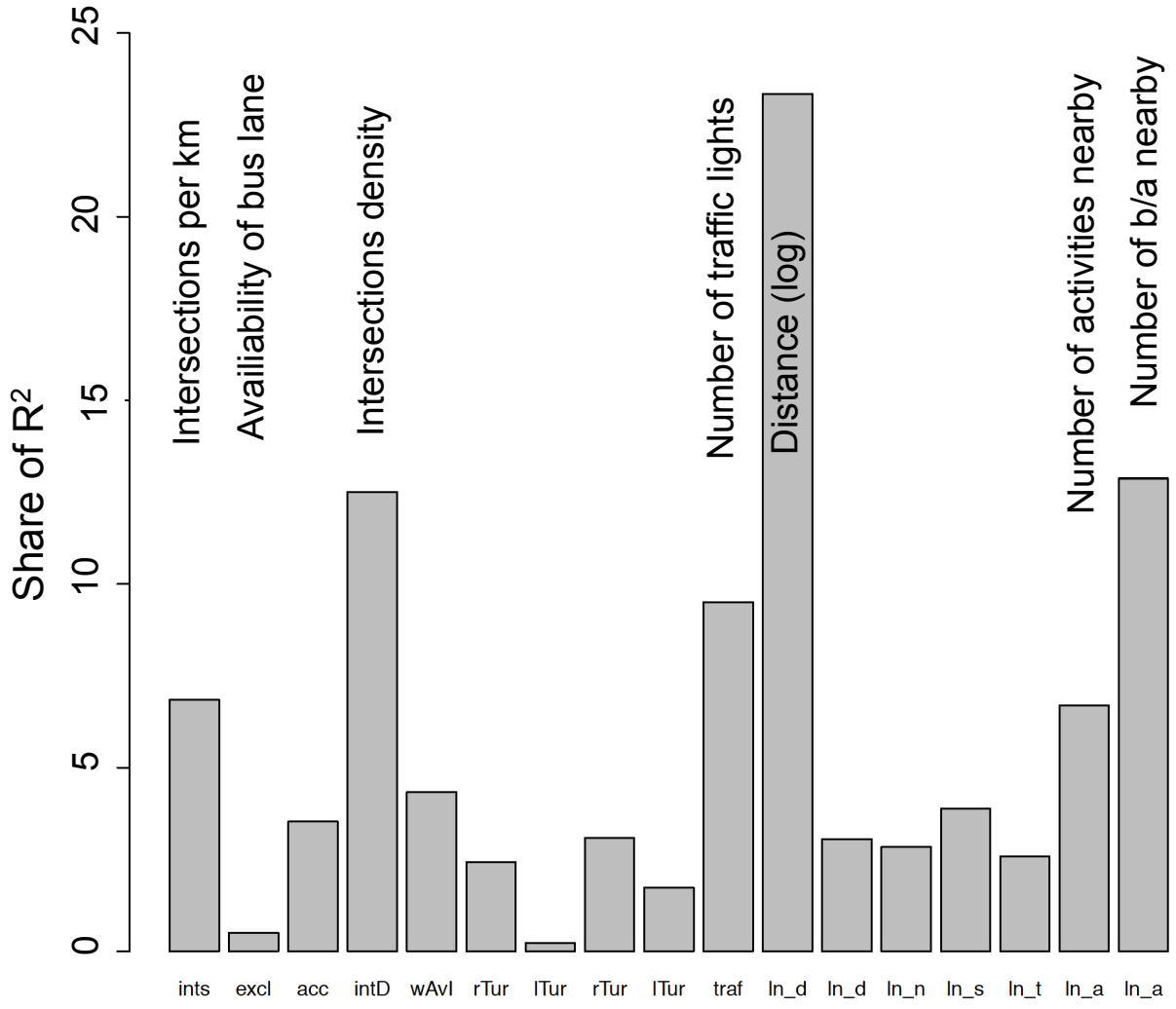
Static variables

- Availability of bus lane
- Number of intersections
- Number of left/right turns
- Curviness
- Deviation from crowfly distance
- Number of traffic lights
- Intersection density

Time-dependent variables

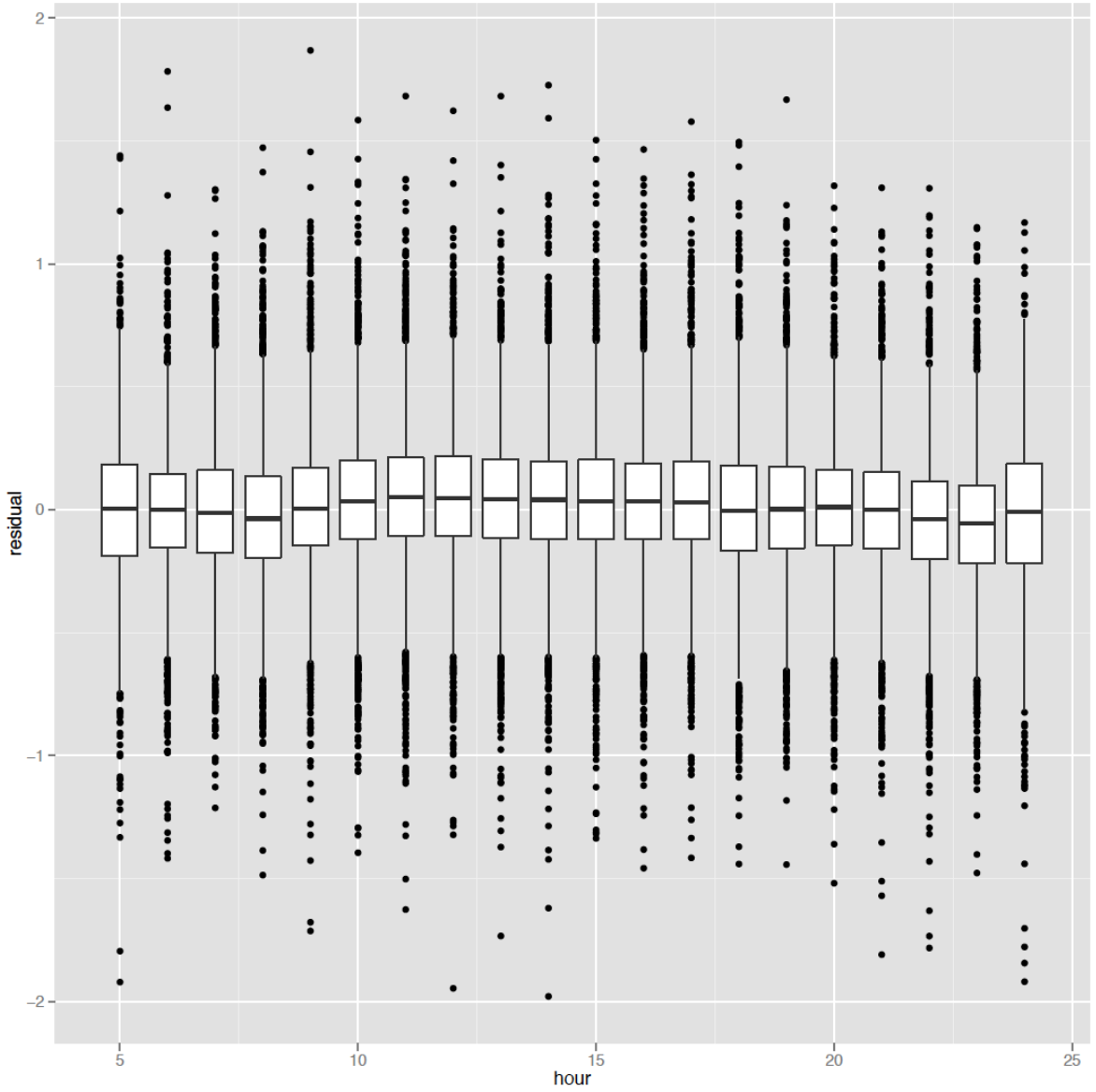
- Boarding/alighting activites in 500m radius

Modelling stop to stop travel times



R^2 = 19.75%, metrics are normalised to sum 100%

Residuals by time of the day



Simulation

Key assumptions

Fixed demand

Demand concentrated at the bus stop

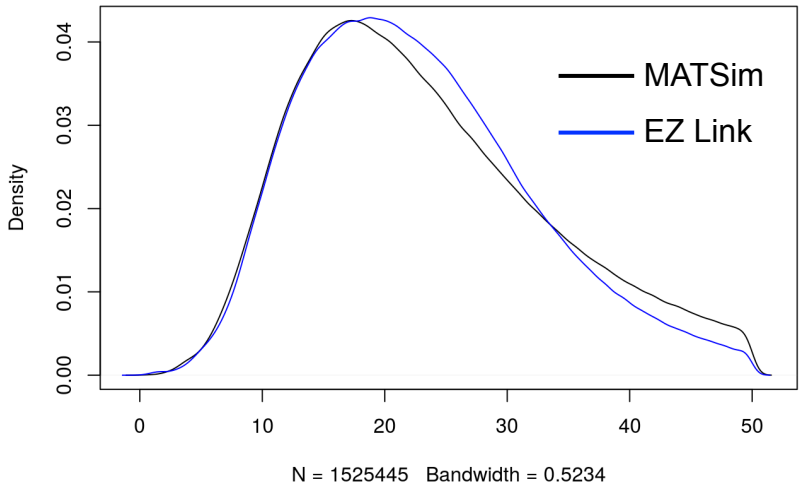
Simulation of public transport only

Simplistic description of transfer characteristics

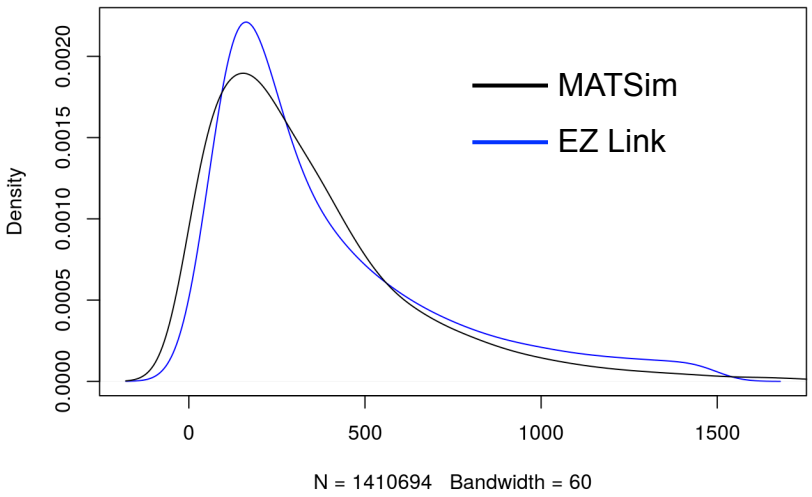
Routing according to parameters of stated preference survey

Validation

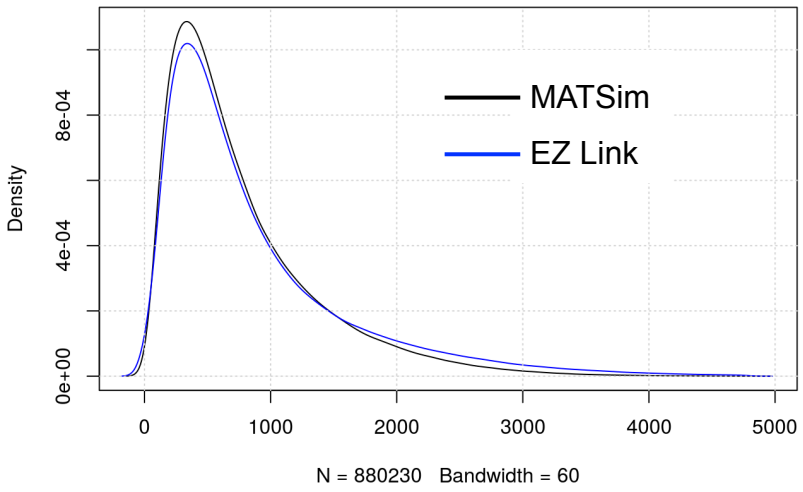
Bus speed



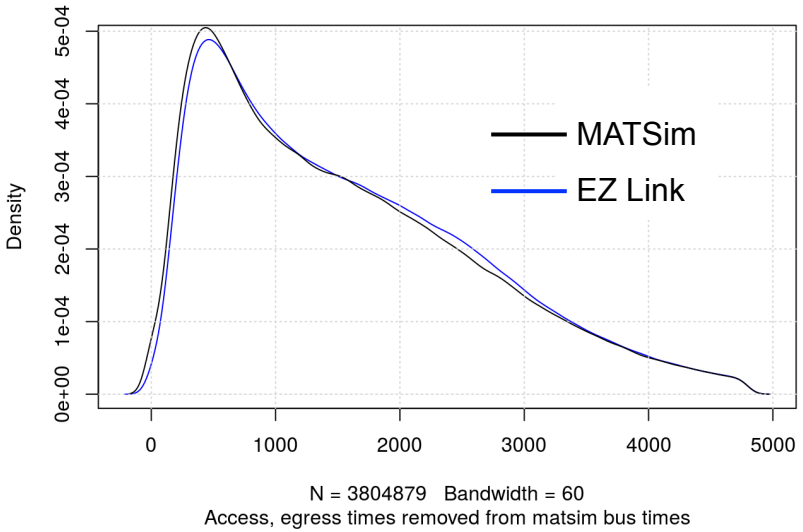
Transfer times



Trip duration (Bus)



Journey duration all modes



Case studies

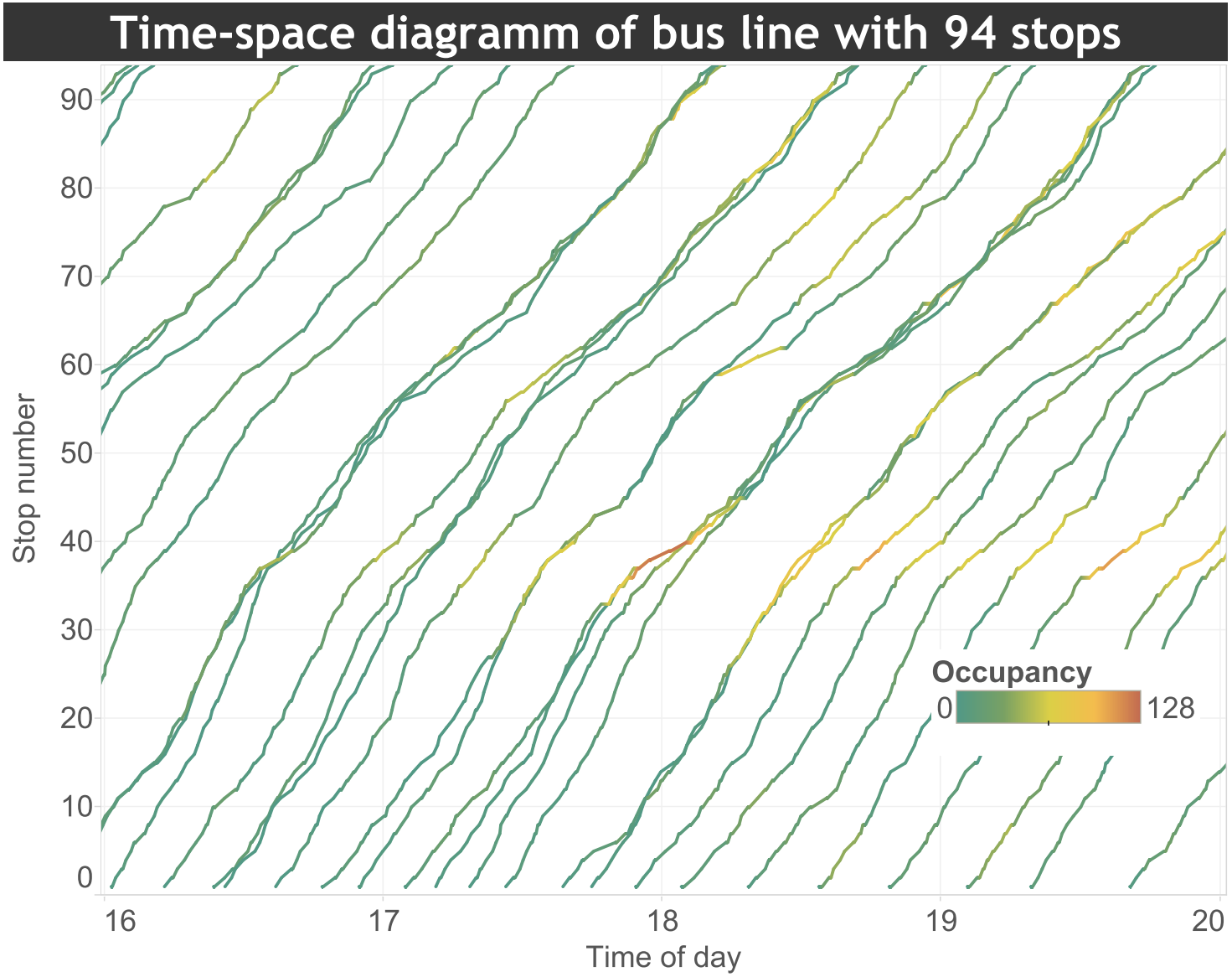
Evaluation of new services and routes:

- How can new network designs improve reliability and tackle overcrowding?
- How many passengers will be attracted by a new service?

Simulation and analysis:

- A full day simulated to steady-stated conditionsion in just about 40 minutes.
- Leverage on off-the-shelf business analytic software for interactive analysis (plus Senozon Via, of course)

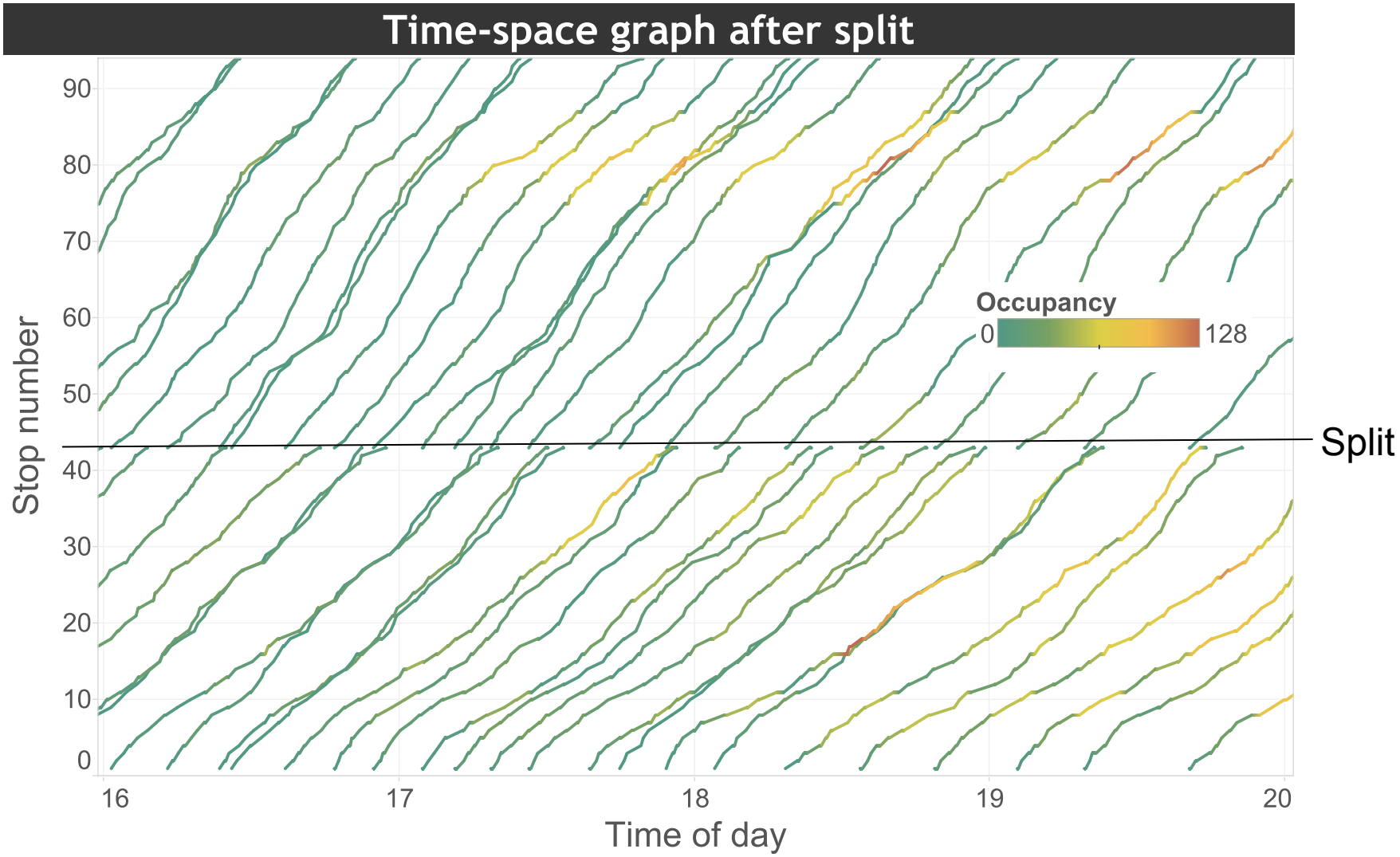
The reliability of a long bus line



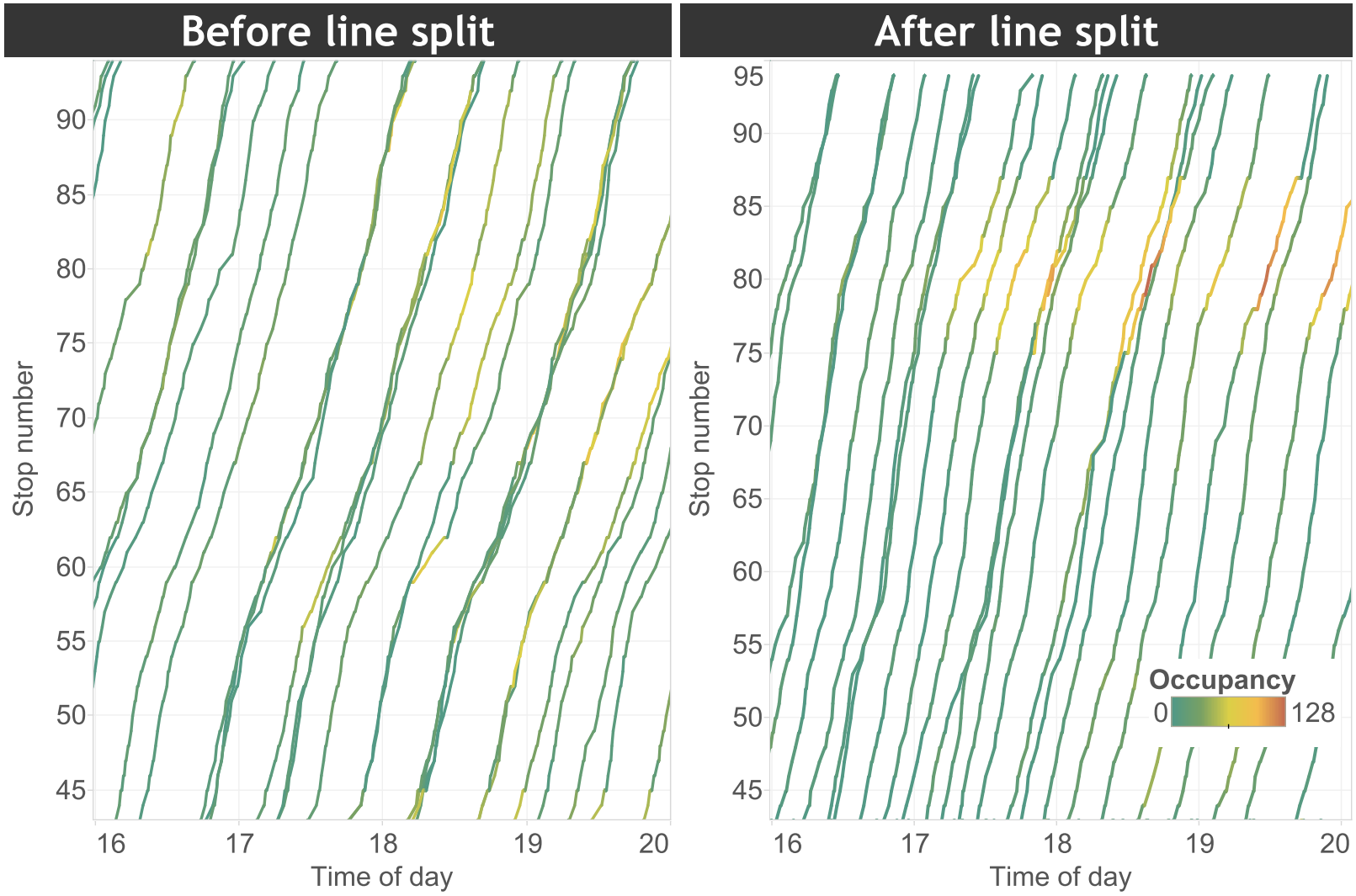
Simulating and evaluating a line split

https://vimeo.com/117976373

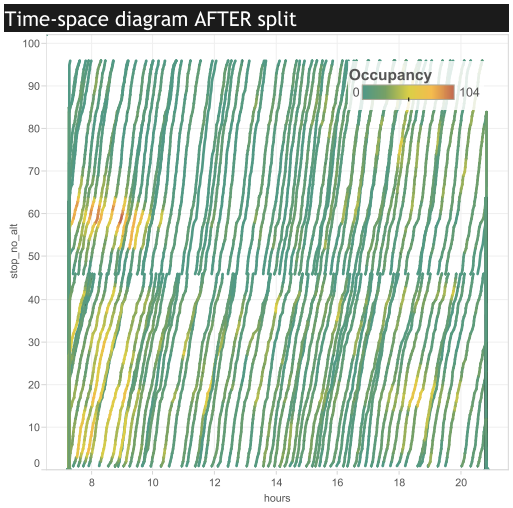
The effect of splitting the line

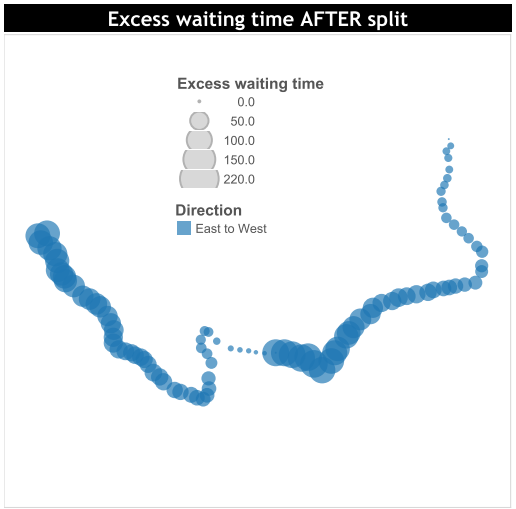


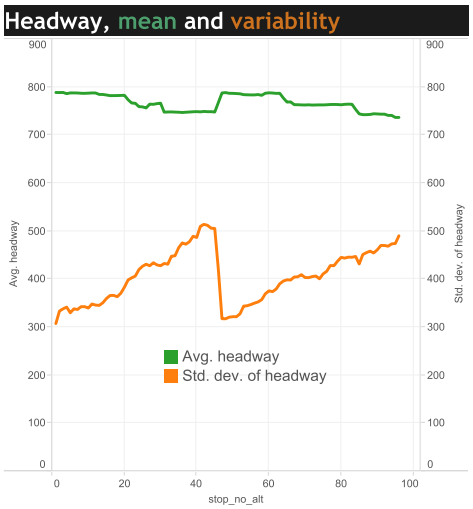
Reliability before and after line split



Some data visualisation fancyness



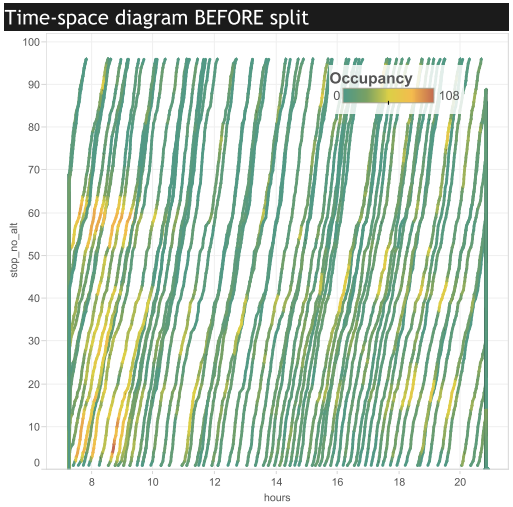


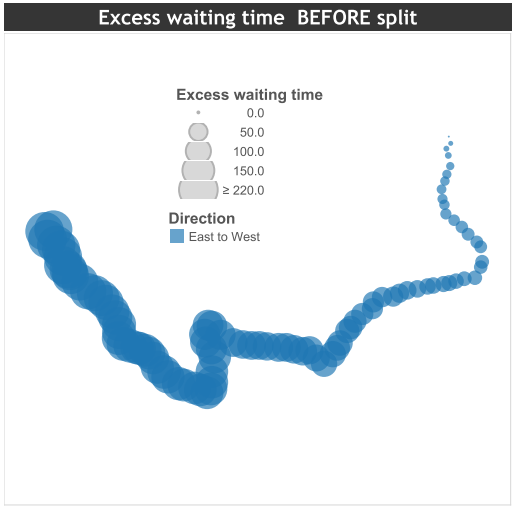


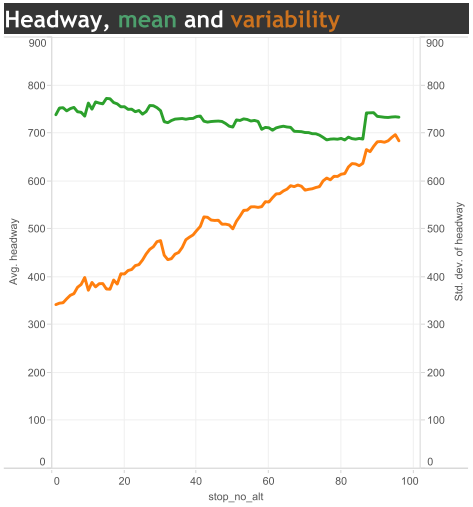
Hour min
7.25862931465733

Hour max
20.8314157078539

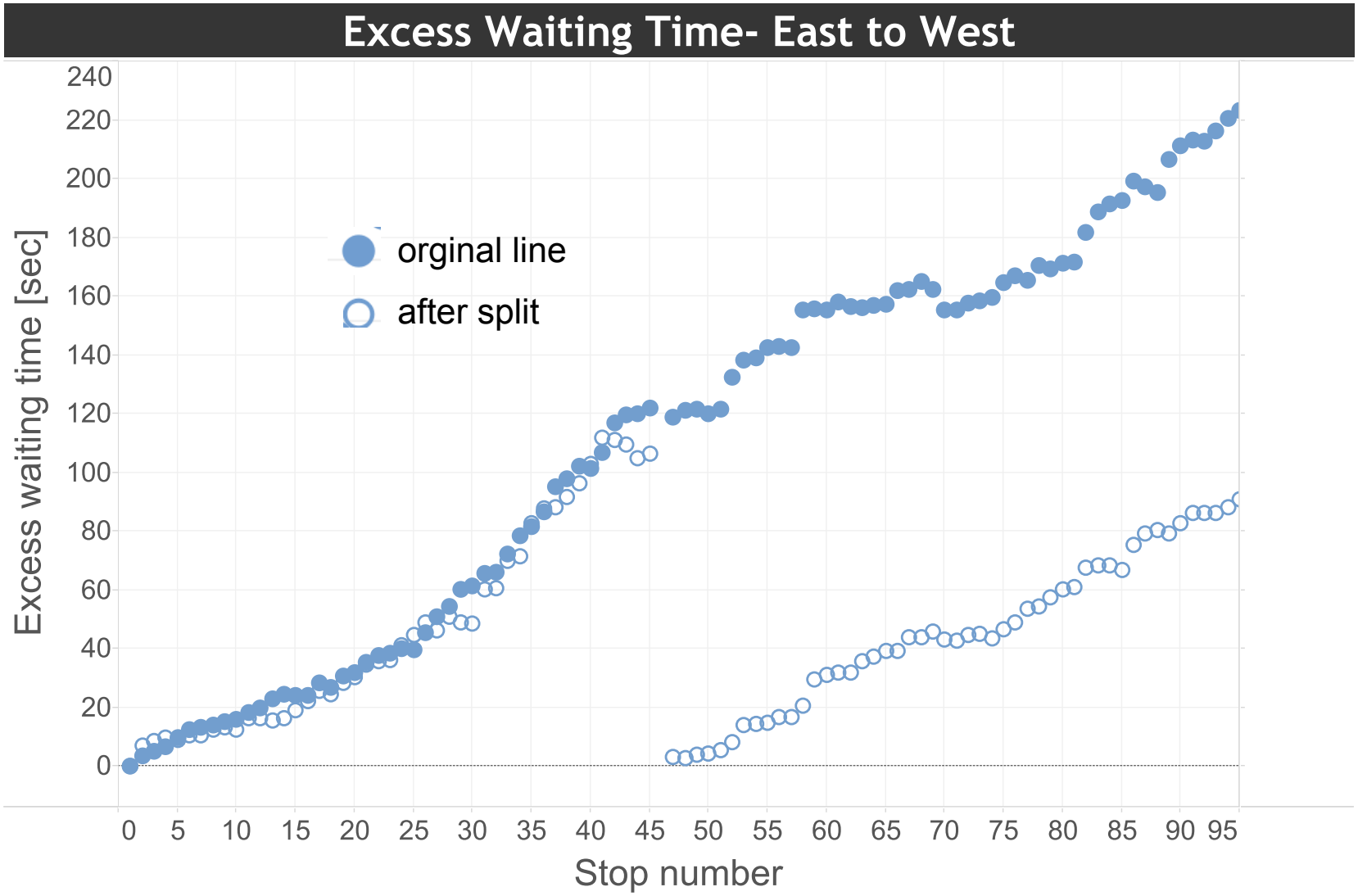
Direction
1 East to West
2 West to East







Reliability: Excess waiting time along line



Conclusion

From Big Data to Smart Data

- Use Big Data to understrand the underlying operational patterns of public transport operations
- Agent-based simulation to model the inherent dynamics of public transport operations

A tool to evaluate alternative service provison

- How do different fare collection alternatives impact service quality?
- Which bus type for which line?
- How to adjust service provision in case on road works or congestion?
- How is the perfomance of alternative network designs?

Further research

- Inferring activity purposes and locations
- Integration of induced demand

The team to make it happen

Pieter Fourie
PhD student
Operations Research

Dr. Alex Erath
Deputy PI,
project manager

Lijun Sun
PhD student
Data scientist

Sergio Ordonez
PhD student
Computer Scientist

Artem Chakirov
PhD student
Electric Engineer

Prof. Dr. Kay Axhausen
PI

Thank you!

www.futurecities.ethz.ch

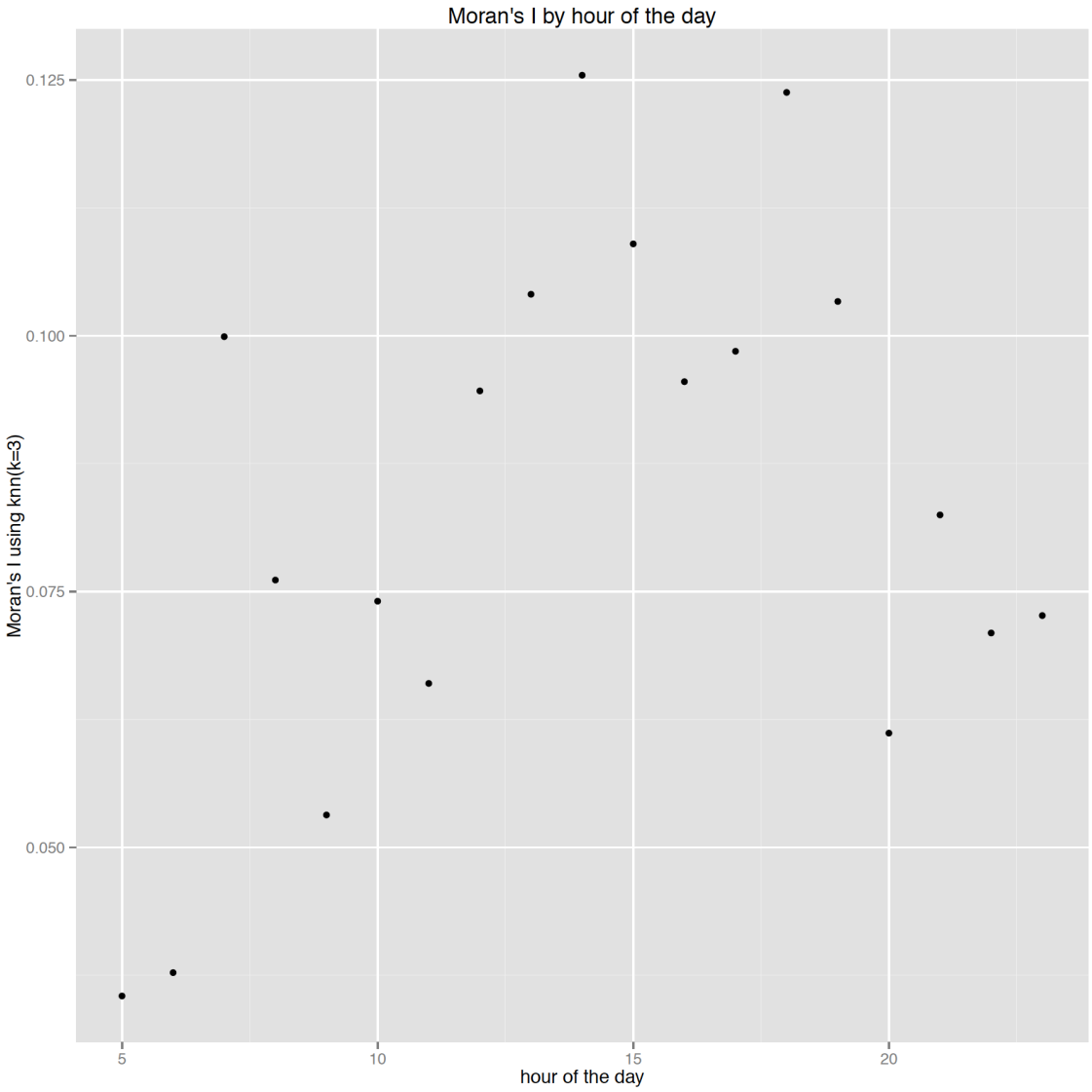
www.ivt.baug.ethz.ch

www.matsim.org

@alex_erath

Appendix

The (non-) issue with spatial autocorrelation



Other smart card data research at FCL

Understanding Metropolitan Collective Encounter Patterns (Lijun Sun *et al.*)

- The familiar stranger phenomenon on the bus
- Published in PNAS http://www.pnas.org/content/early/2013/07/31/1306440110

Efficient detection of contagious outbreaks in massive metropolitan encounter networks (Lijun Sun *et al.*)

- Using smart card data to develop efficient detection of disease spreaders
- Published in Science Reports
  http://www.nature.com/srep/2014/140606/srep05099/full/srep05099.html

Study of bus service reliability in Singapore using fare card data (Lijun Sun *et al.*)

- http://www.futurecities.ethz.ch/wp-content/plugins/zotpress/lib/request/rss.file.php?api_user_id=151611&download=XM4VQFC8

Models of Bus Boarding/Alighting Dynamics and Dwell Time Variability

- Sun, Lijun, Alejandro Tirachini, Kay W. Axhausen, Alexander Erath and Der-Horng Lee (2014). 'Models of Bus Boarding and Alighting Dynamics', Transportation Research Part A: Policy and Practice 69: 447–460.

Estimation of revealed preference route choice models to account for crowdedness (Tirachini, Sun, Erath)

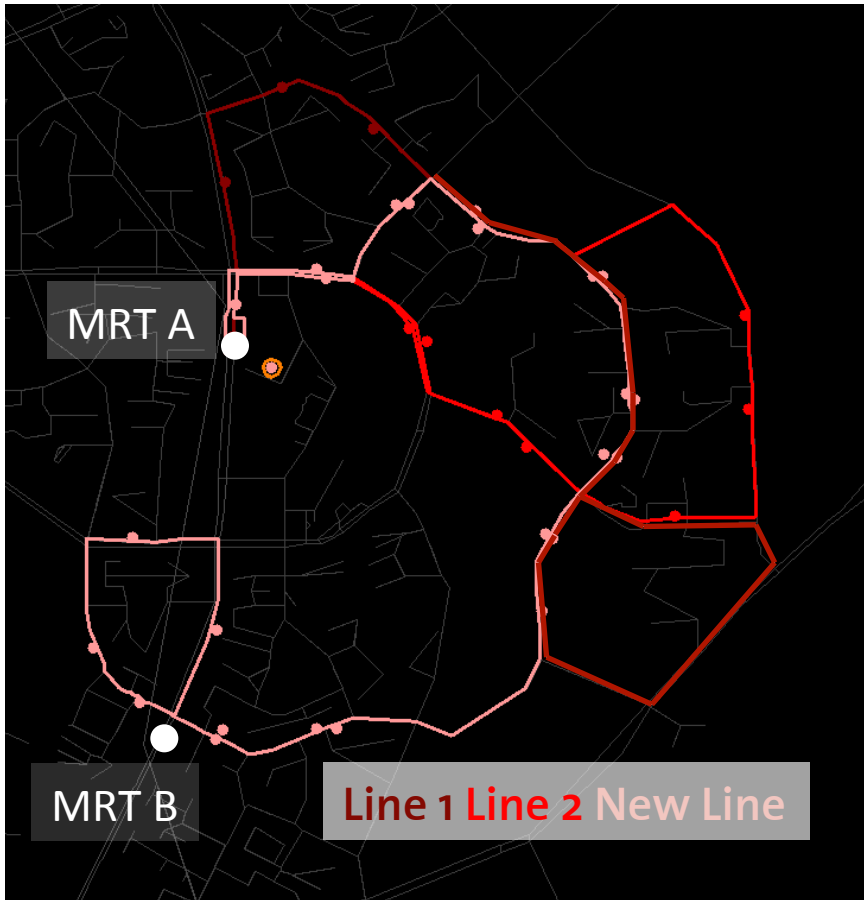
- People travel in the wrong direction to secure a seat
- Paper to be submitted the coming days

Activity Identification and Primary Location Modelling based on Smart Card Payment Data for Public Transport (Chakirov, Erath)

- http://www.futurecities.ethz.ch/wp-content/plugins/zotpress/lib/request/rss.file.php?api_user_id=151611&download=RQZWFXBZ

Stay tunes for more at http://www.futurecities.ethz.ch/module/mobility-and-transportation-planning/

Case study 2: adding a new bus line



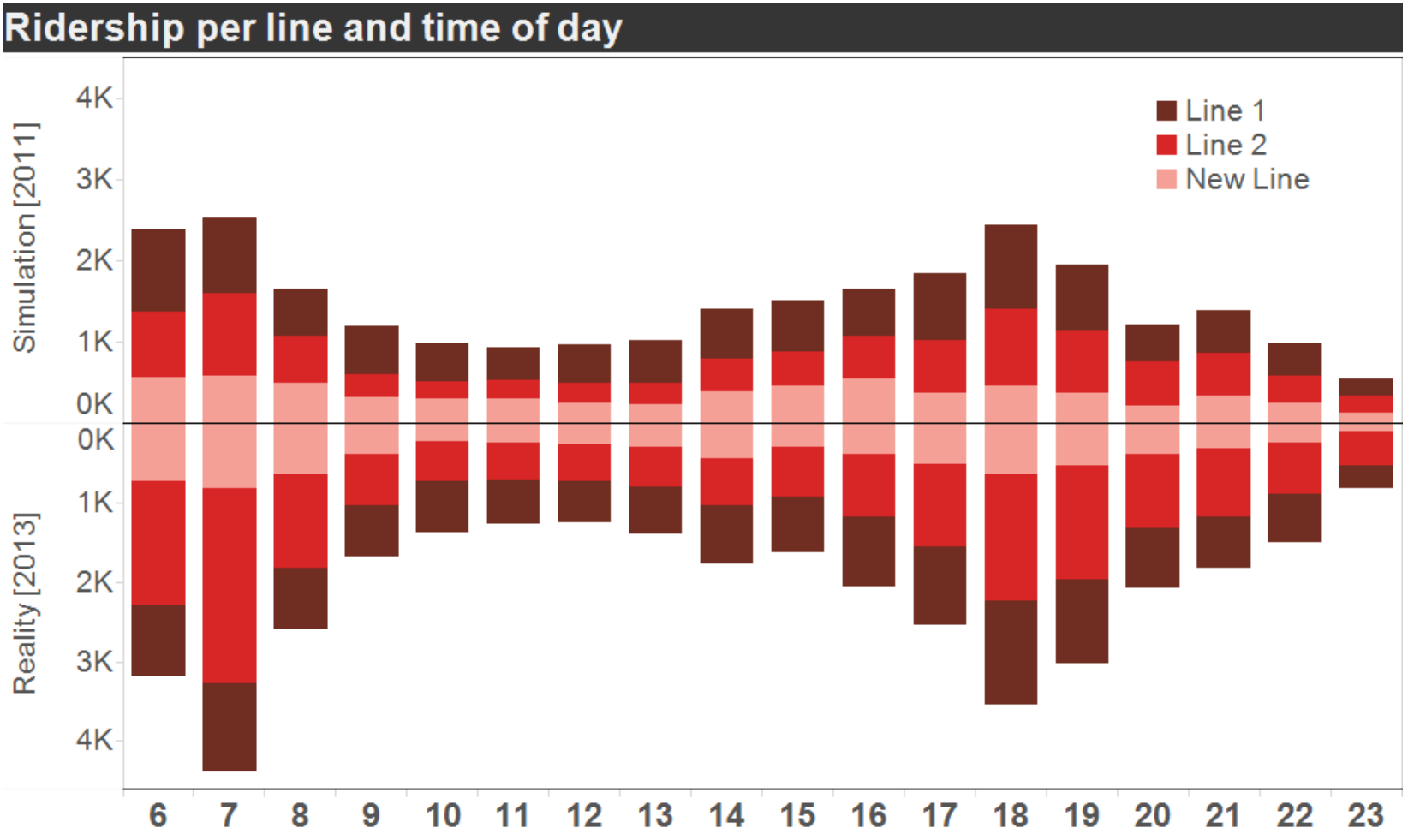
Residential new town

- Tidal demand patterns
- Issues with overcrowding during peak hours

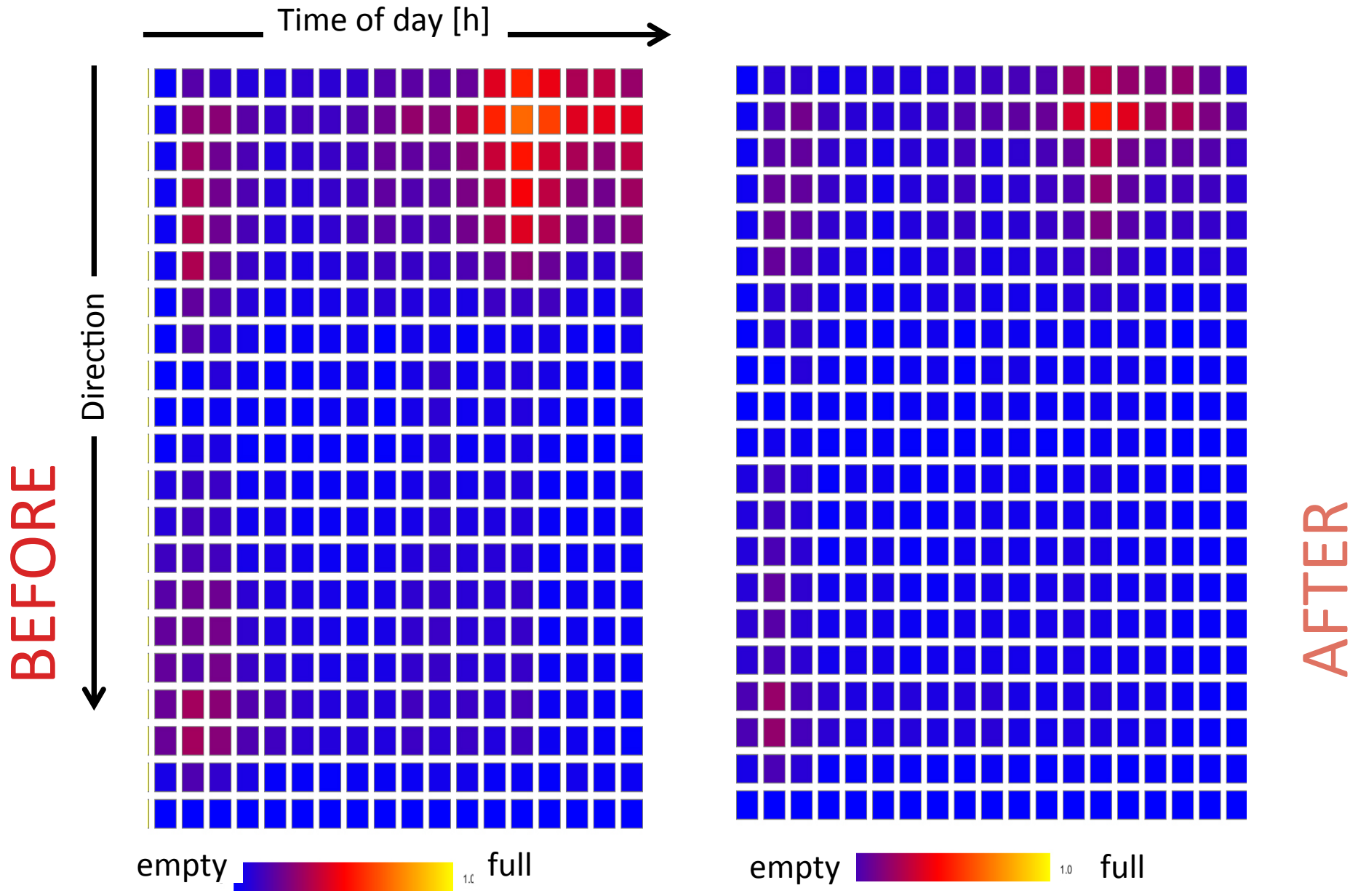
New bus line:

- 26 stops
- 10km
- Loop from MRT A to MRT B and back

Predicted vs actual ridership



Crowding heatmap: Line 1



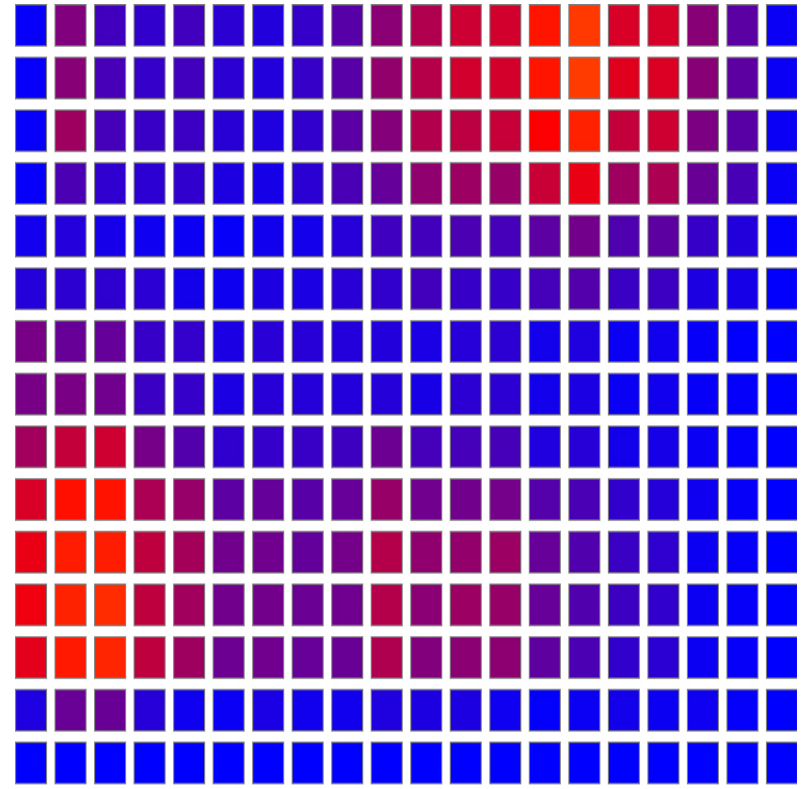
Crowding heat map: Line 2

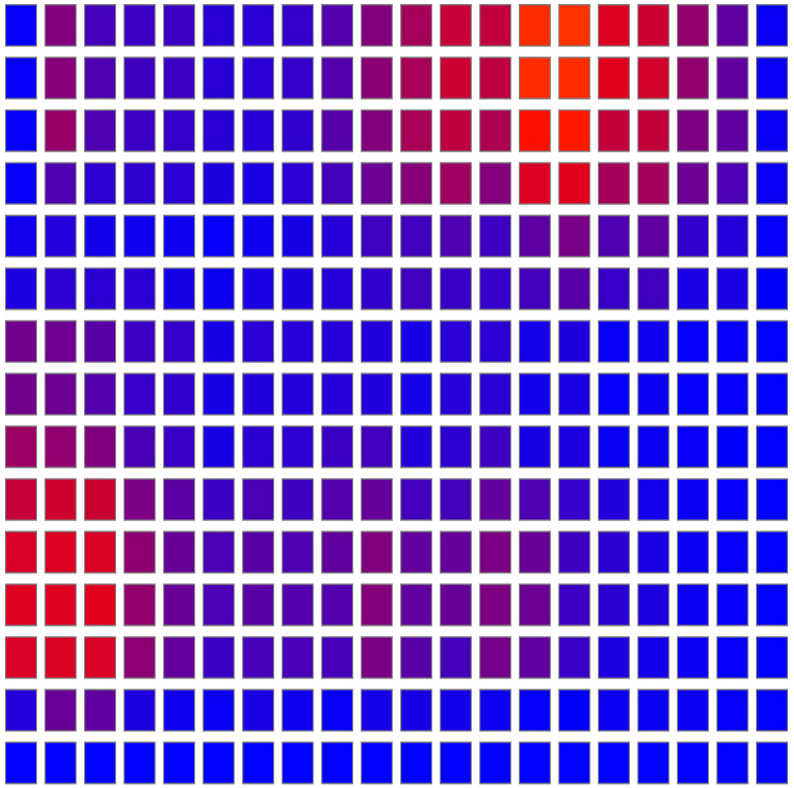
BEFORE

AFTER

Time of day [h]

Direction

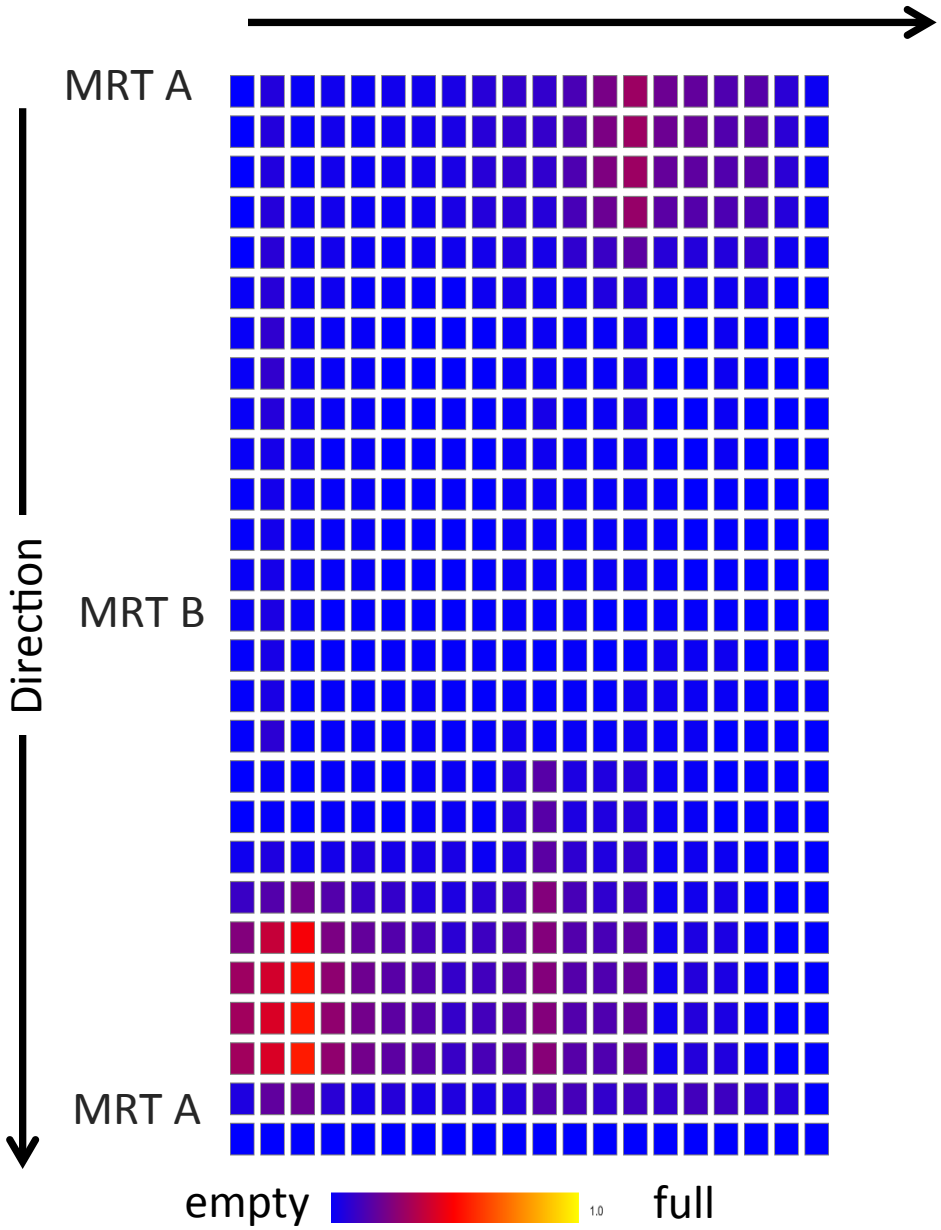




empty full

empty full

Crowding heat map: New line



AFTER