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

(FCL) FUTURE CITIES LABORATORY

CREATE
Campus for Research Excellence And Technological Enterprise

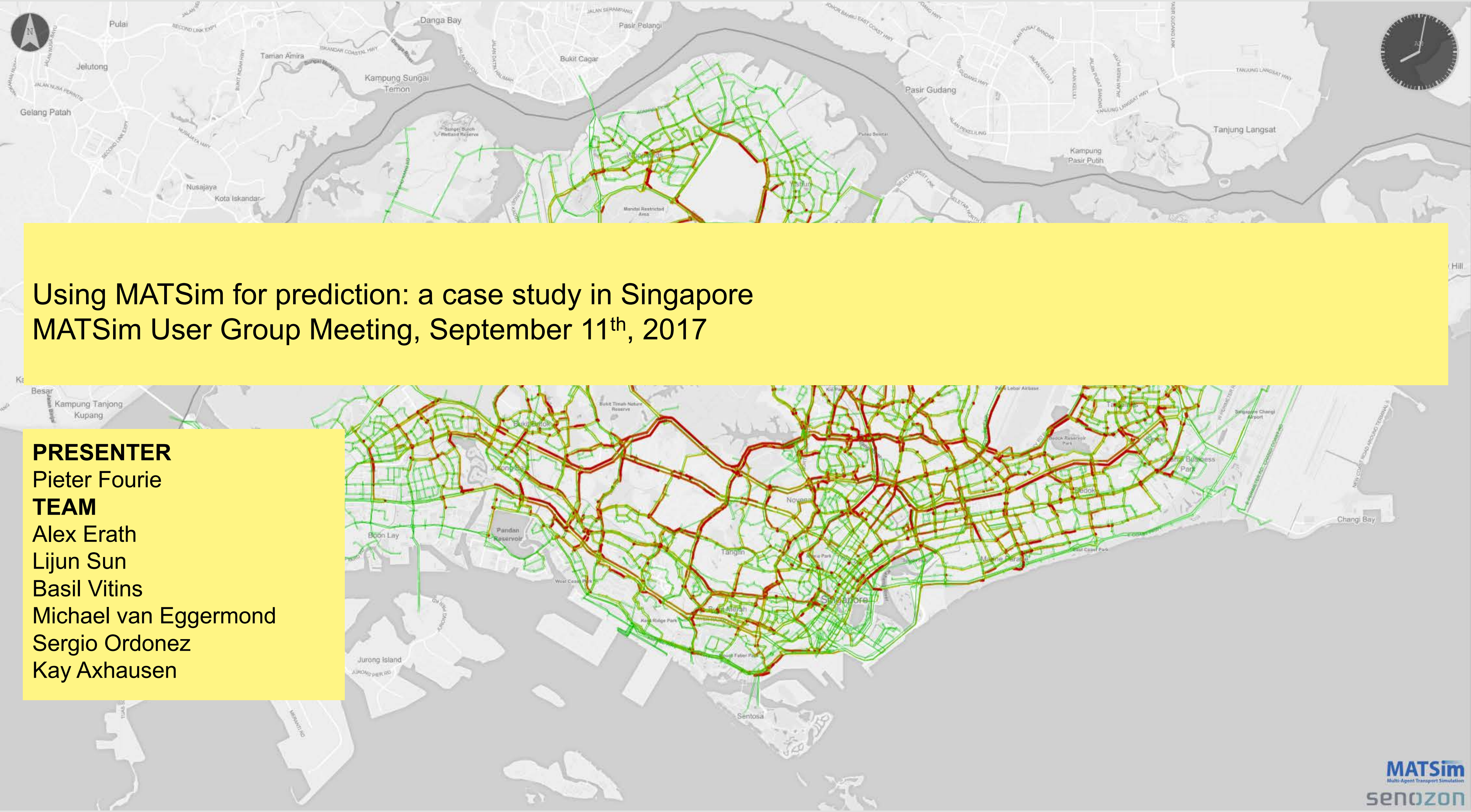
ETH zürich

EPFL

NANYANG TECHNOLOGICAL UNIVERSITY

NUS National University of Singapore

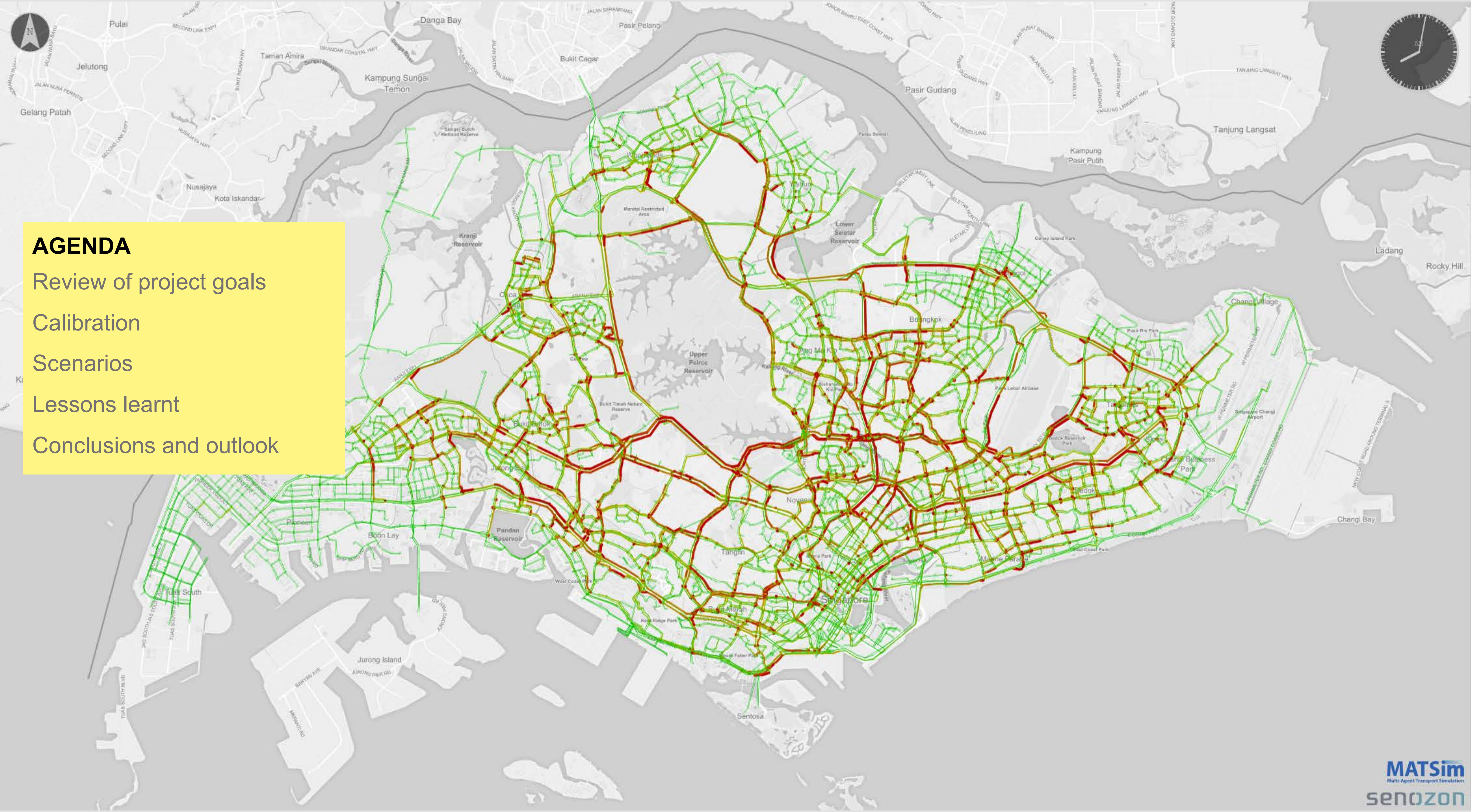
SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN



Using MATSim for prediction: a case study in Singapore
MATSim User Group Meeting, September 11th, 2017

PRESENTER
Pieter Fourie
TEAM
Alex Erath
Lijun Sun
Basil Vitins
Michael van Eggermond
Sergio Ordonez
Kay Axhausen

MATSim Multi-Agent Transport Simulation
senozon



AGENDA

Review of project goals

Calibration

Scenarios

Lessons learnt

Conclusions and outlook

PROJECT GOALS

Starting step for URA/LTA to **understand** how **agent-based transport modelling** can be applied in planning practice.

New datasets have become available, and some are constantly updated

Strengthen the **collaboration** between **research and practice** and accelerate knowledge spillover

Make MATSim data and results **available** for everyday planning tasks

Carefully **calibrate and validate** the model to quantify its accuracy and prediction capabilities.

Develop better understanding (and modelling) of **home-work relationship** in relation to demographics and locality

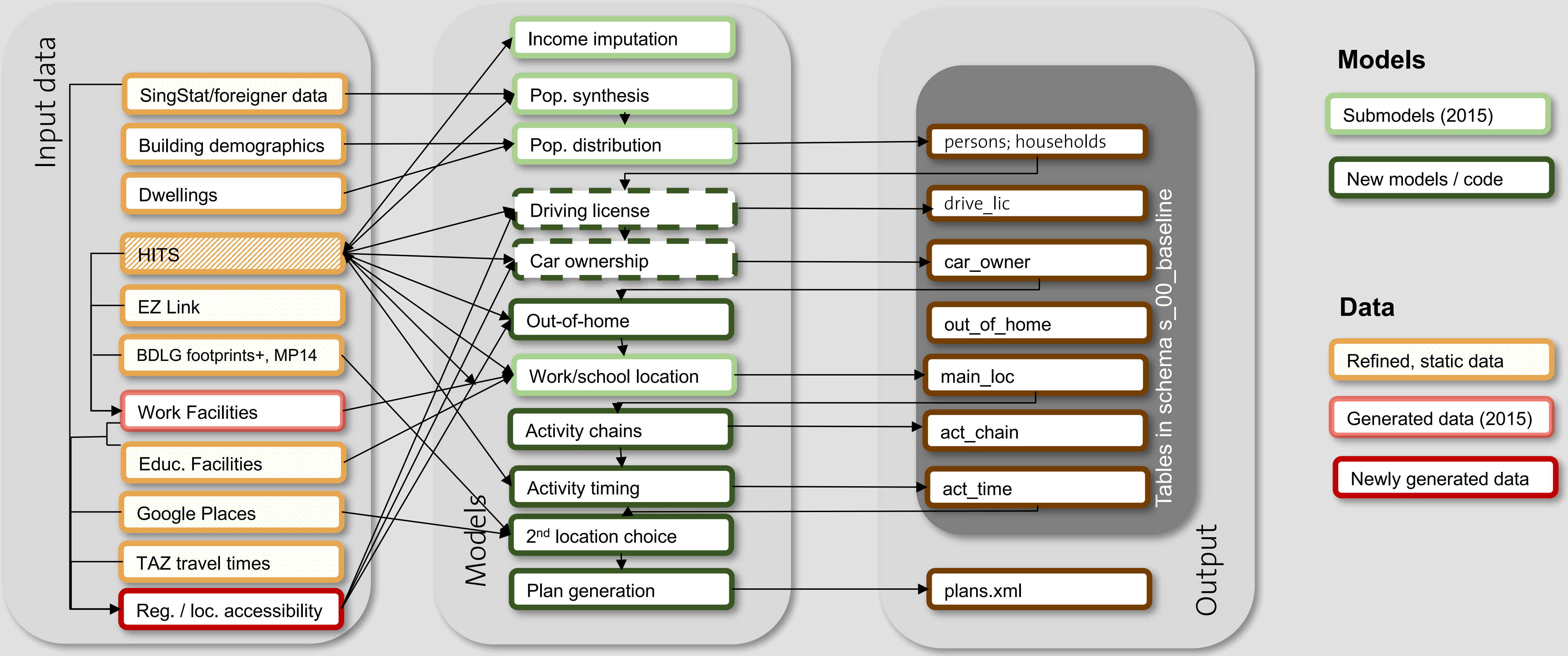
Collaborative project between FCL-URA-LTA to facilitate **knowledge transfer** through workshops, regular meetings and staff exchange.

Deliver urban and transport **data platform** that integrates behavioral models, MATSim i/o and various visualisation tools.

Develop and improve the **demand generation workflow** of MATSim Singapore

Create a **base model for further studies**

SYNTHETIC POPULATION AND TRAVEL DEMAND



PROGRAMMING LANGUAGES / MODEL TYPES

Models

- Income imputation
- Pop. synthesis
- Pop. distribution
- Driving license
- Car ownership
- Out-of-home
- Work/school location
- Activity chains
- Activity timing
- 2nd location choice
- Plan generation

Model types

- Sampling
- Bayesian network
- Random forrest
- Regression
- Time-space prism
- Logit model

Programming languages

- R
- Python
- Java

Paradigm

- KISS
- Make use of existing Open Source packages
- Restrict to R, Python and Java

Database

- Postgresql -> Open Source, many connectors
- PostGis extension

DATABASE REFACTORING

Before

- a_dos_population_data
- a_lta
- a_lta_emme
- a_lta_pt_supply
- a_sla_address_points
- a_sla_bldg
- a_ura_dgp
- a_ura_mp08
- baseline_simresults
- c_factual
- c_postal_codes
- c_transit_entrance_exit
- d_activity_plans
- d_hits_chains
- d_population_extended
- d_population_synthesis
- d_urbananalysis_sg
- d_workcapacities
- d_zonal_systems
- old_MATSim_d_work_capacities
- old_matsim_d_work_capacities_ii
- p_carlicensemodel
- p_destinationchoicemodel
- p_lta
- p_postalcodes
- p_ura_mp
- public
- r_lta_hits12
- r_lta_mtz_1092p
- r_lta_traveltimes
- r_osm_2013
- r_osm_2015
- r_sla_address_points
- r_streetdirectory_lines
- r_ura_bca_building
- r_ura_hotel
- r_ura_MOESchoolsEnrolment
- r_ura_moeschoolsenrolment
- r_ura_mp
- r_ura_officeretail
- ro_act_chain
- ro_freight_and_foreign_trips
- ro_main_act_timing
- ro_mainactfacilityassignment
- ro_mtz1169_traveltimes_lta
- ro_outofhomedecision
- ro_p_carownership
- ro_p_destinationchoice
- ro_p_drivinglic
- ro_p_facilities
- ro_p_incomemodel
- ro_p_schoollocationchoice
- ro_plansgeneration
- ro_population
- ro_postcodes
- ro_r_hits12_extended
- ro_r_lta_hits12
- ro_sec_activities
- ro_spatial
- ro_supply
- ro_workcapacities
- ro_zonal_systems
- topology
- u_artemc
- u_atizaza
- u_bvitins
- u_erathal
- u_lijunsun
- u_lta_fkw
- u_lta_kl
- u_lta_wll
- u_michaeva
- u_test_capstone
- u_ura_aa
- u_ura_cc
- u_ura_dl
- u_ura_dt
- u_ura_hkh
- u_ura_ht
- u_ura_hz
- u_ura_vp
- u_ura_whm
- urban_analysis
- w_1_1_intro
- w_1_2_synthpop
- w_1_3_spatial
- w_1_4_modelling
- w_attach_synthpop

After

- m_01_popsynth
- o_dos_population_data
- o_lta_hits12
- o_sla_bldg
- o_ura_mp
- o_zonal_systems
- p_controltotals
- p_inc
- p_workcapacities
- public
- s_00_baseline
- topology
- u_hits12_extended

“o_” : Original data
Refined dataset

“p_” : Preparatory data
Static data to inform travel demand models, e.g. income imputation, work capacity model, accessibility measure

“m_xx_” : Modelling data
_xx _> number of submodel
Contains intermediate model output

“s_xx” : Scenario data
”s_00” -> baseline
Each schema contains all necessary tables as produced when applying travel demand models in sequence

“u_” : Utility data
Transient data used in several modelling steps in earlier stages of the project.
Hard to maintain
Will be removed piece by piece

CALIBRATION AND VALIDATION

Travel demand models

All models calibrated against HITS 2012

Include spatial analysis indicators for enhanced sensitivity

Validation against hold-out samples

Simulation

Sub-population based calibration of behavioural parameters

Calibration is iterative systematic simulation parameter adjustment -> very computation & time intensive

No direct handles on OD flows compared with STM, as OD flows are done upstream in activity assignment models.

CALIBRATION AND VALIDATION

- Car: simulated with QSim
- Public transport: simulated with QSim
- Taxi: simulated in QSim like car with additional activity and leg
- Walk: teleported with beeline factor
- School bus: routed but not simulated in QSim
- Other: routed but not simulated in QSim
- Passenger: routed but not simulated in QSim
- Freight: simulated in QSim

CALIBRATION AND VALIDATION

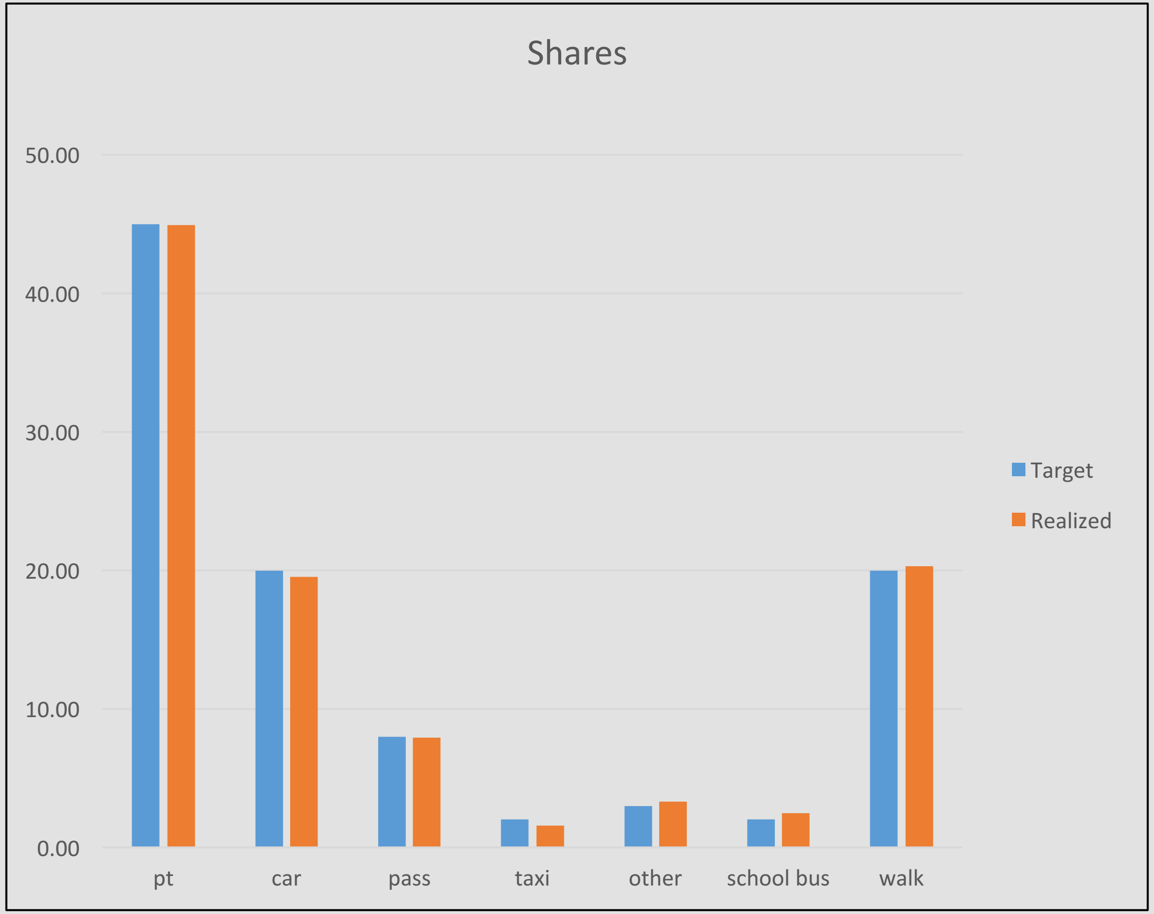
Calibration handles

3 household income-based sub-populations
(0-4k, 4k-9k, 9k+)

8 transport modes
(pt, car, psgr, taxi, other, school bus, walk, transit walk)

4 mode choice parameters each:
constant
marginalUtilityOfDistance_util_m
marginalUtilityOfTraveling_util_hr
monetaryDistanceRate

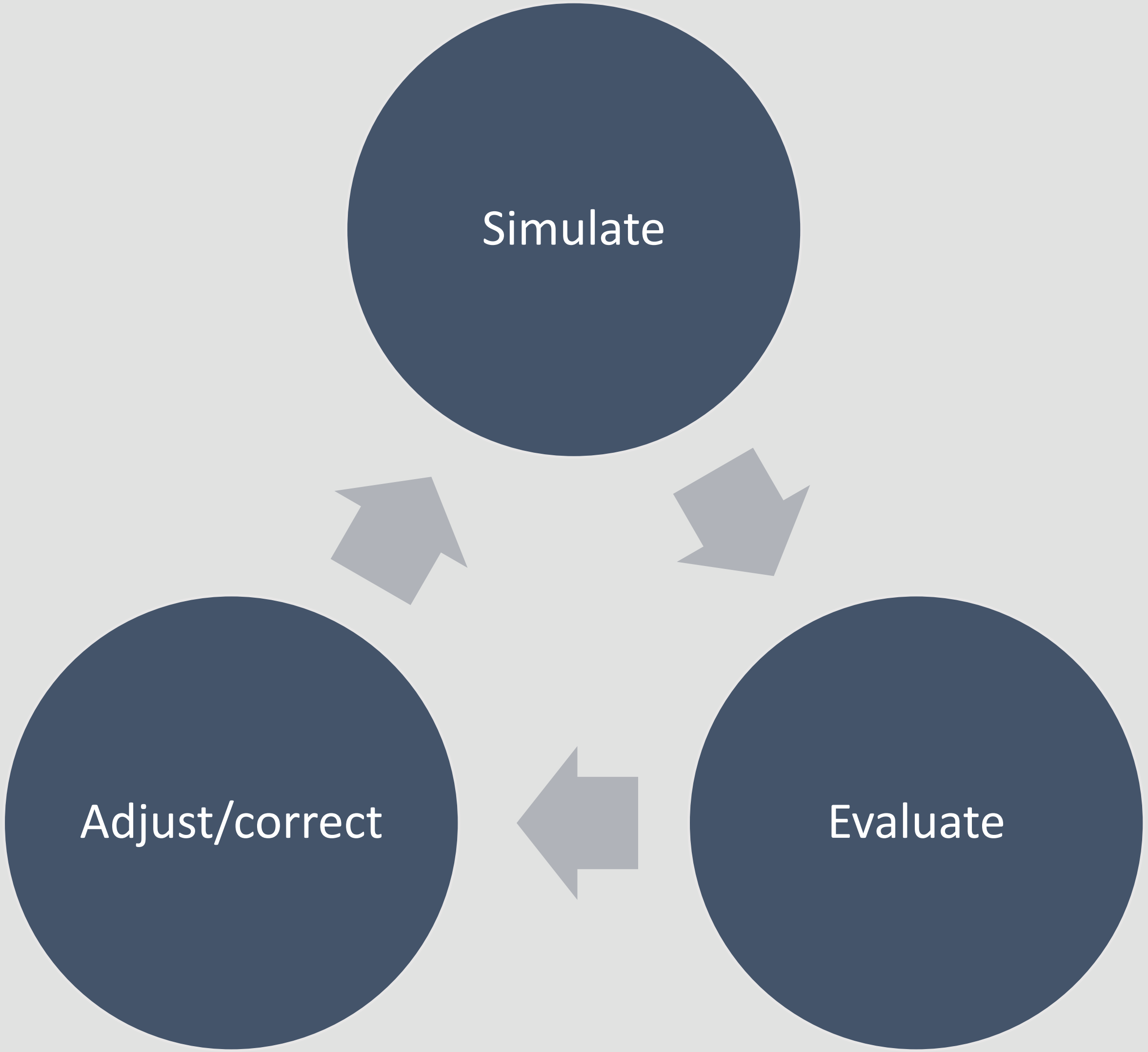
96 dimensions



CALIBRATION AND VALIDATION

Calibration runs exposed problems

Mapping of count stations
…
Bus-car interaction dynamics on highways
…
Missing plan attributes
…
Inconsistencies in plan file
…
Network capacity problems
…
Intersection capacity problems
…
Etc.



CALIBRATION ISSUES

The focus of calibration was mainly to

- Implement Singapore modes
- Achieve mode shares

Following calibration, a number of outstanding issues were identified:

Non-simulated modes

Cost parameters of non-simulated modes had to be set very high; this might be due to:

Access and egress, waiting times missing in sim

Detours in reality but fastest path in sim

Rides immediately available in sim, not in reality

Coordination effort (passenger)

No cost of crowding, capacity

No surcharges

Counts

Model calibration mainly for mode share

Further calibration runs required after various interventions

Routing parameters need to be part of calibration

Intersection friction needs investigation

Comparison of STM vs navigation network

Other issues

Hard facility timings

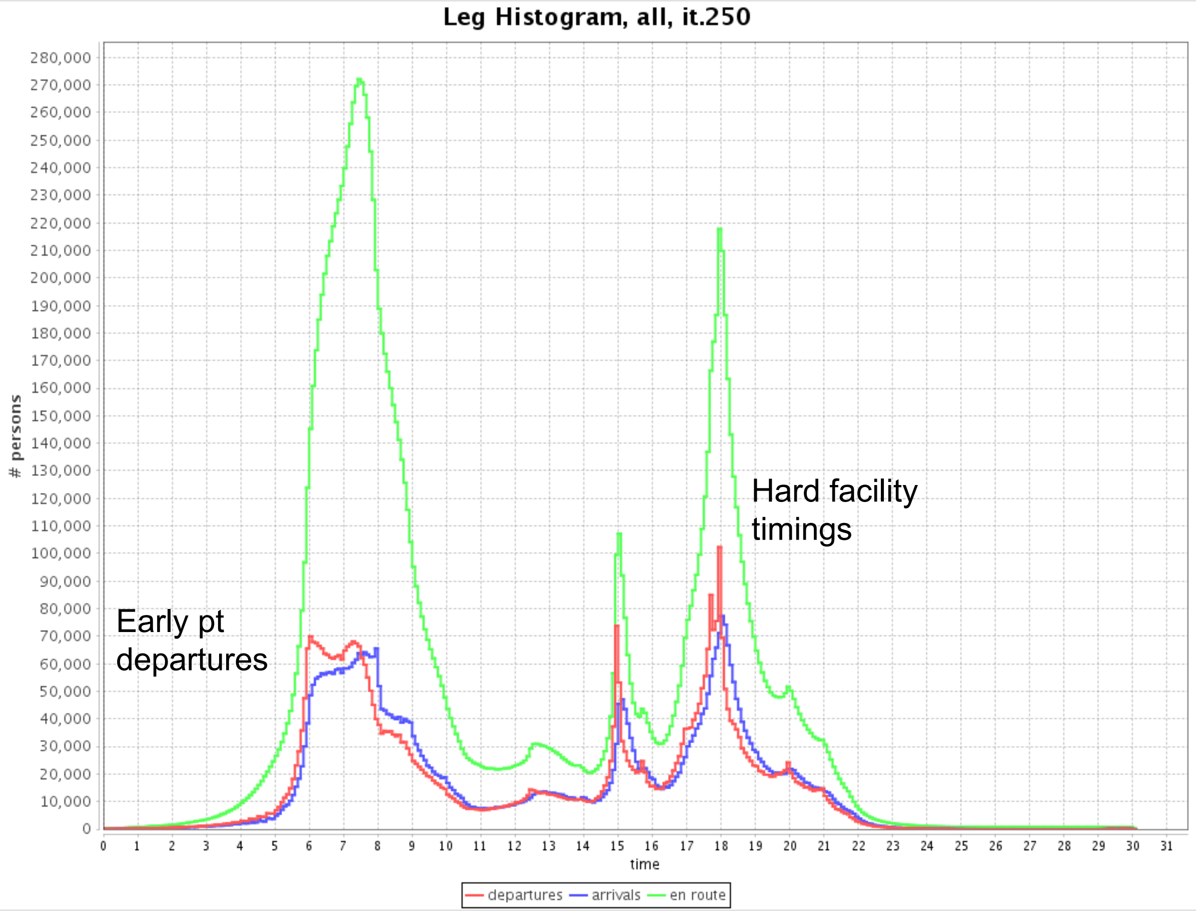
Pickup/drop-off timing

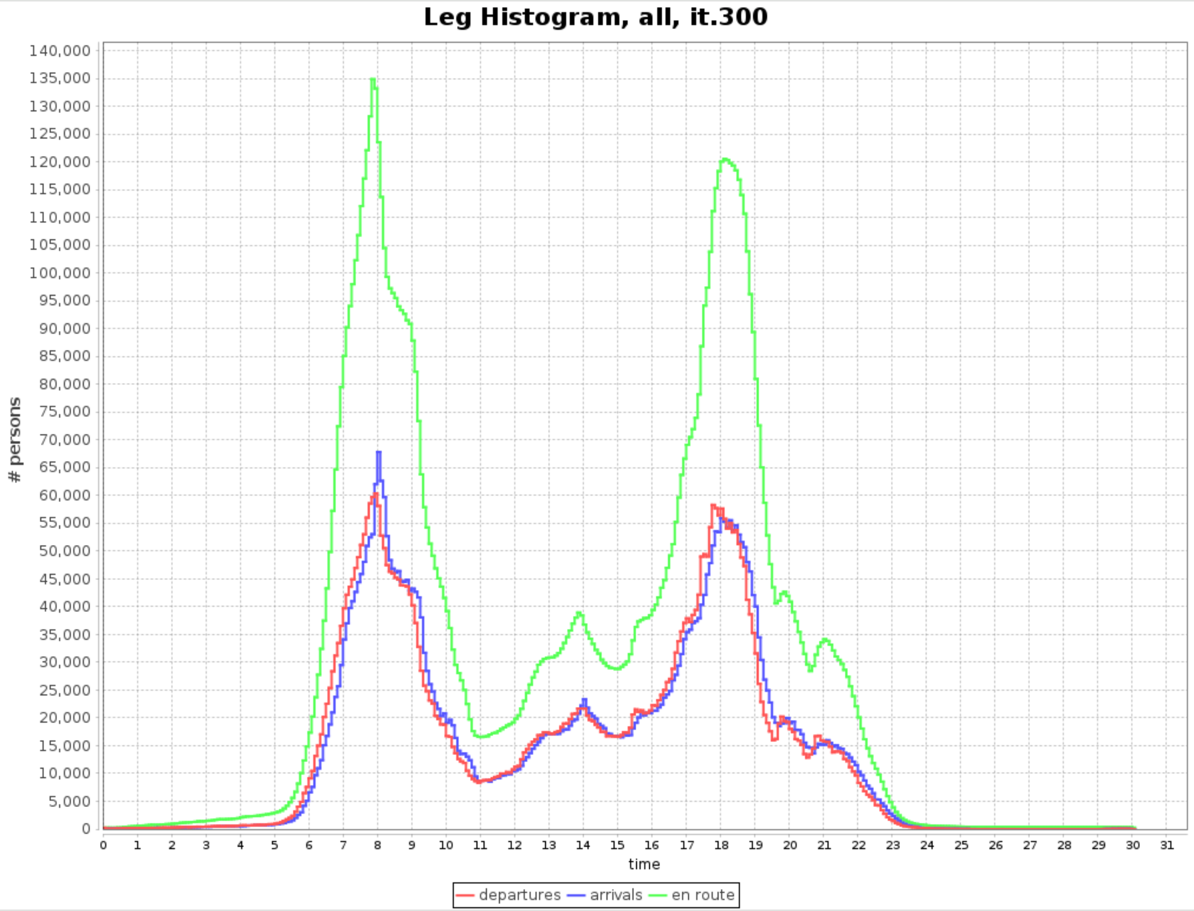
Transit waiting time not penalised

Single home activity

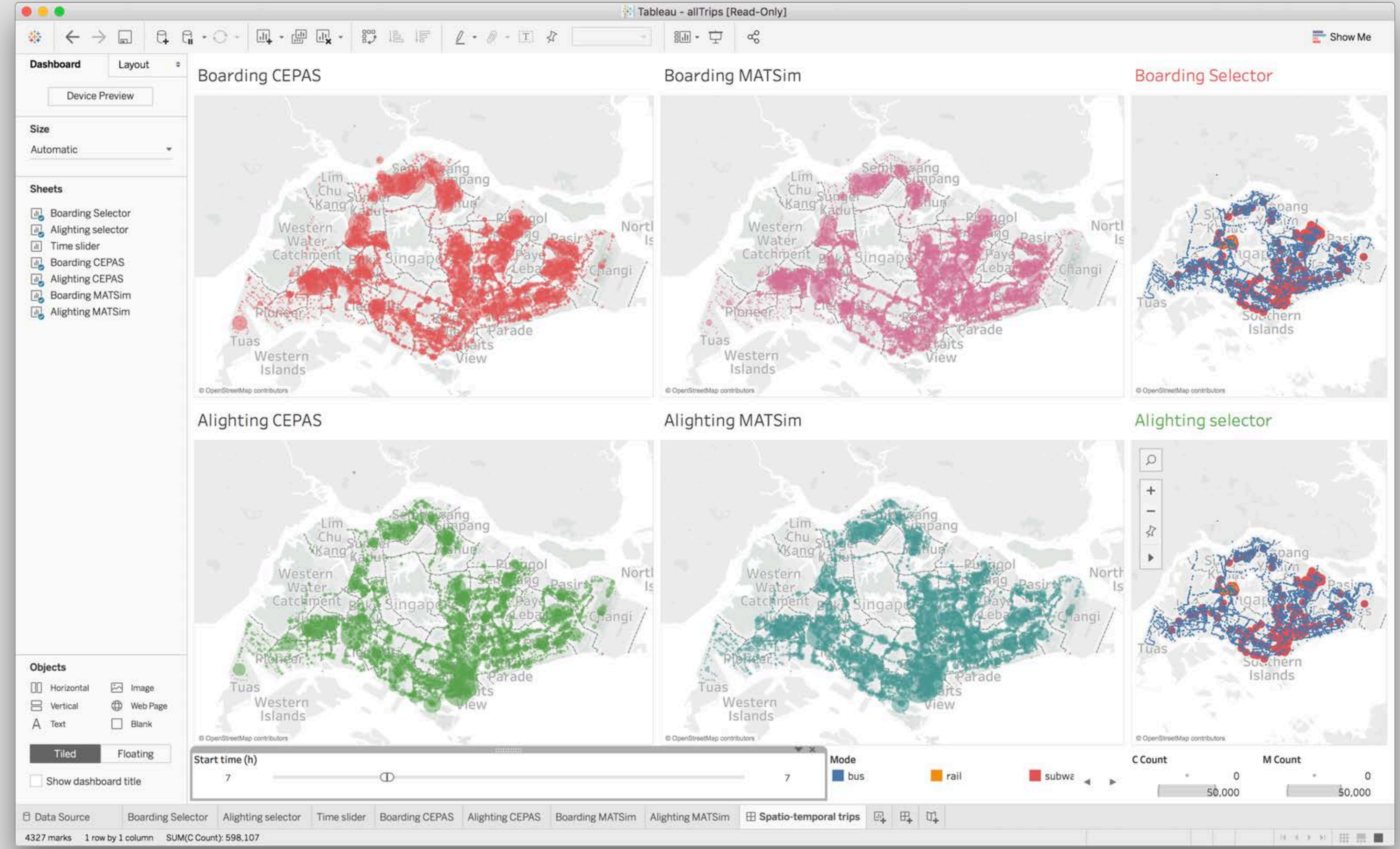
Home activity end time

DYNAMICS: BEFORE AND AFTER

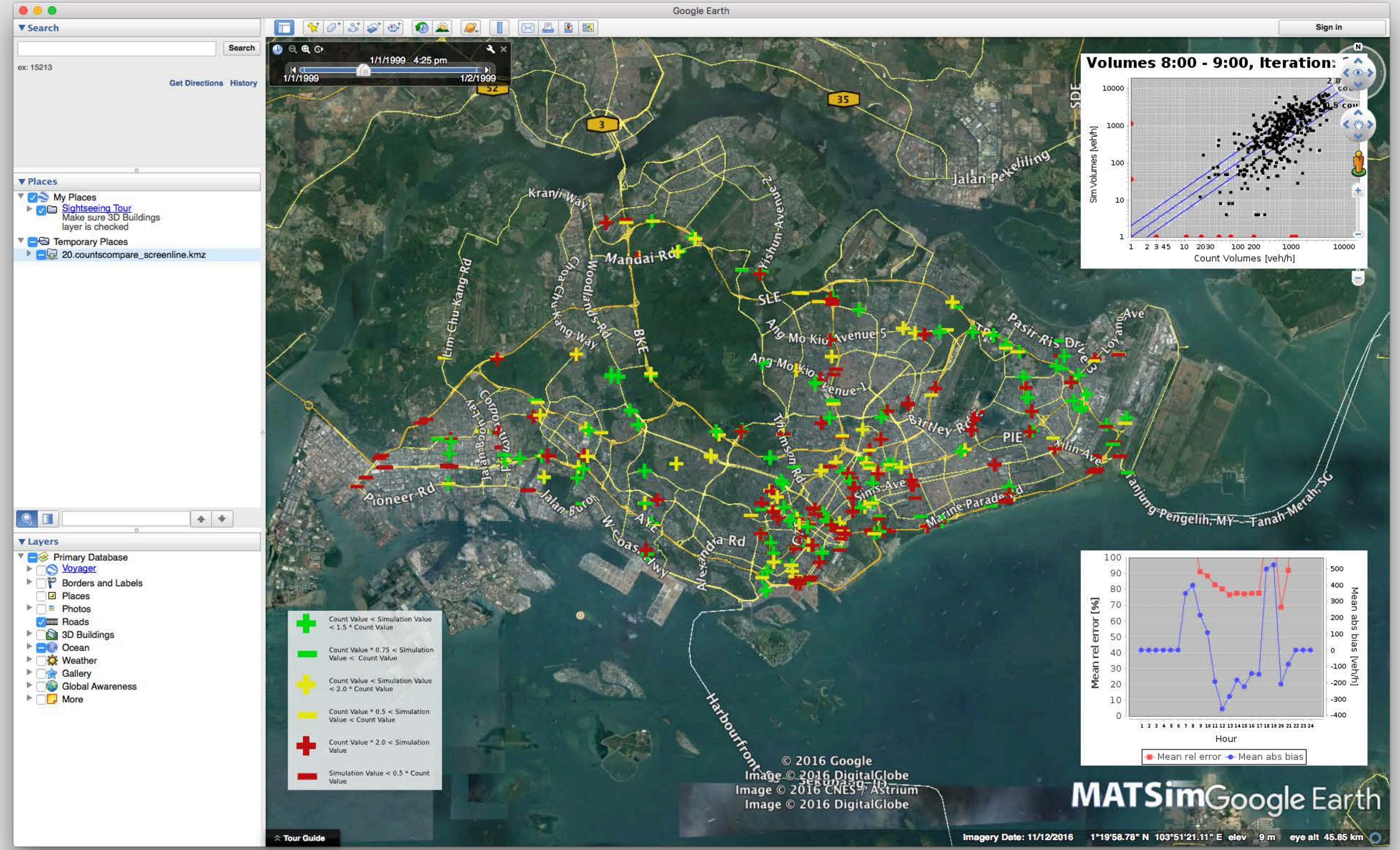




PT VALIDATION



CAR VALIDATION

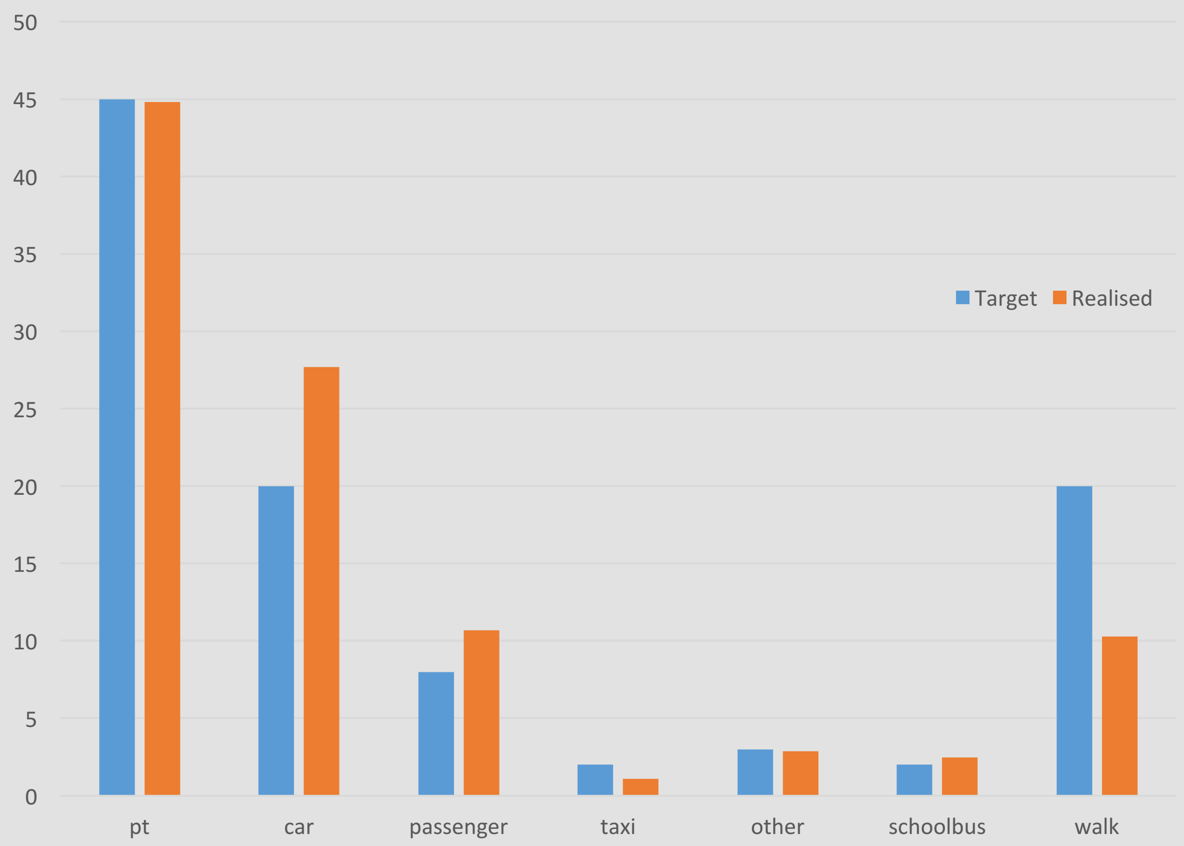


MODE SHARE AFTER POST-CALIBRATION FIXES

Constant for walk mode would need to be lowered

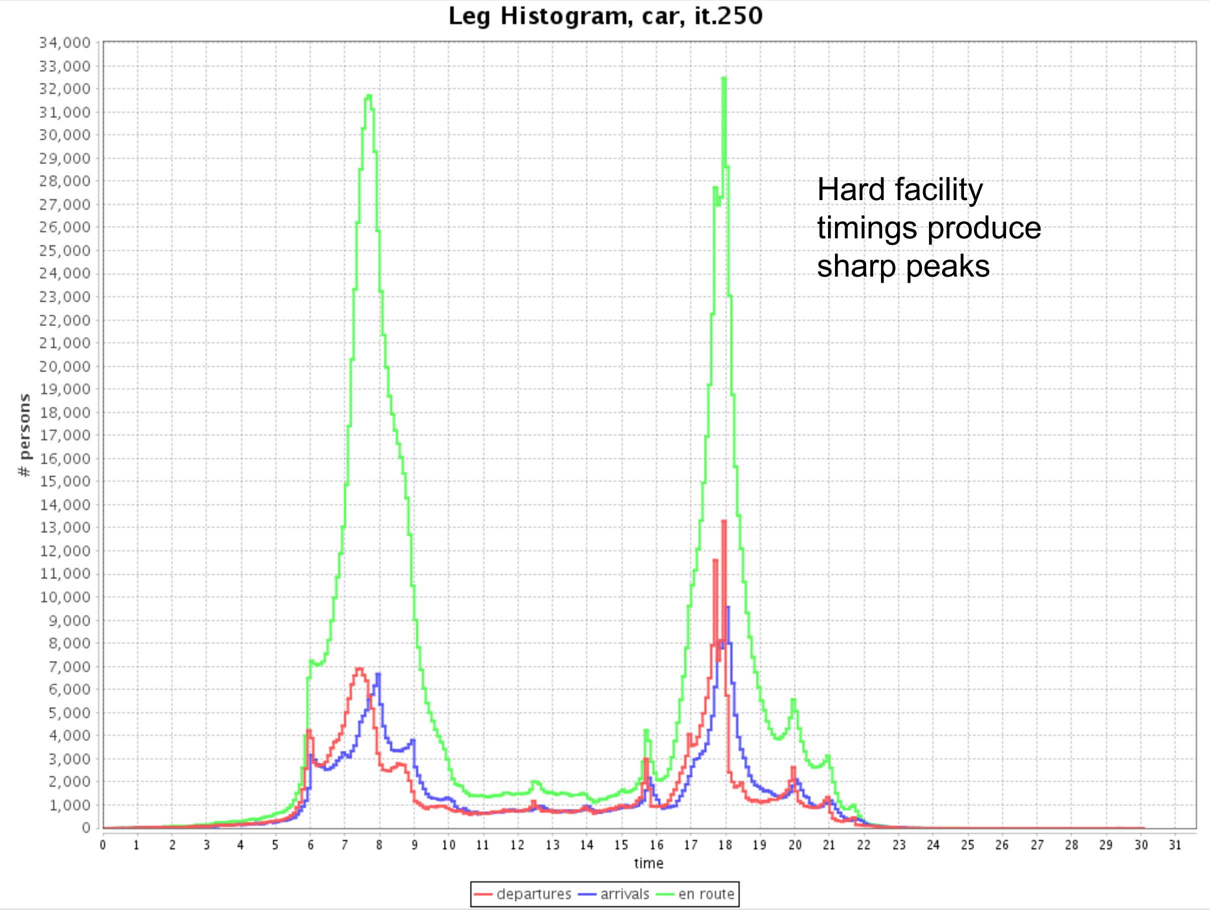
More short trips made by car in updated model

Previously, walking took place during peak hour (congestion avoidance)

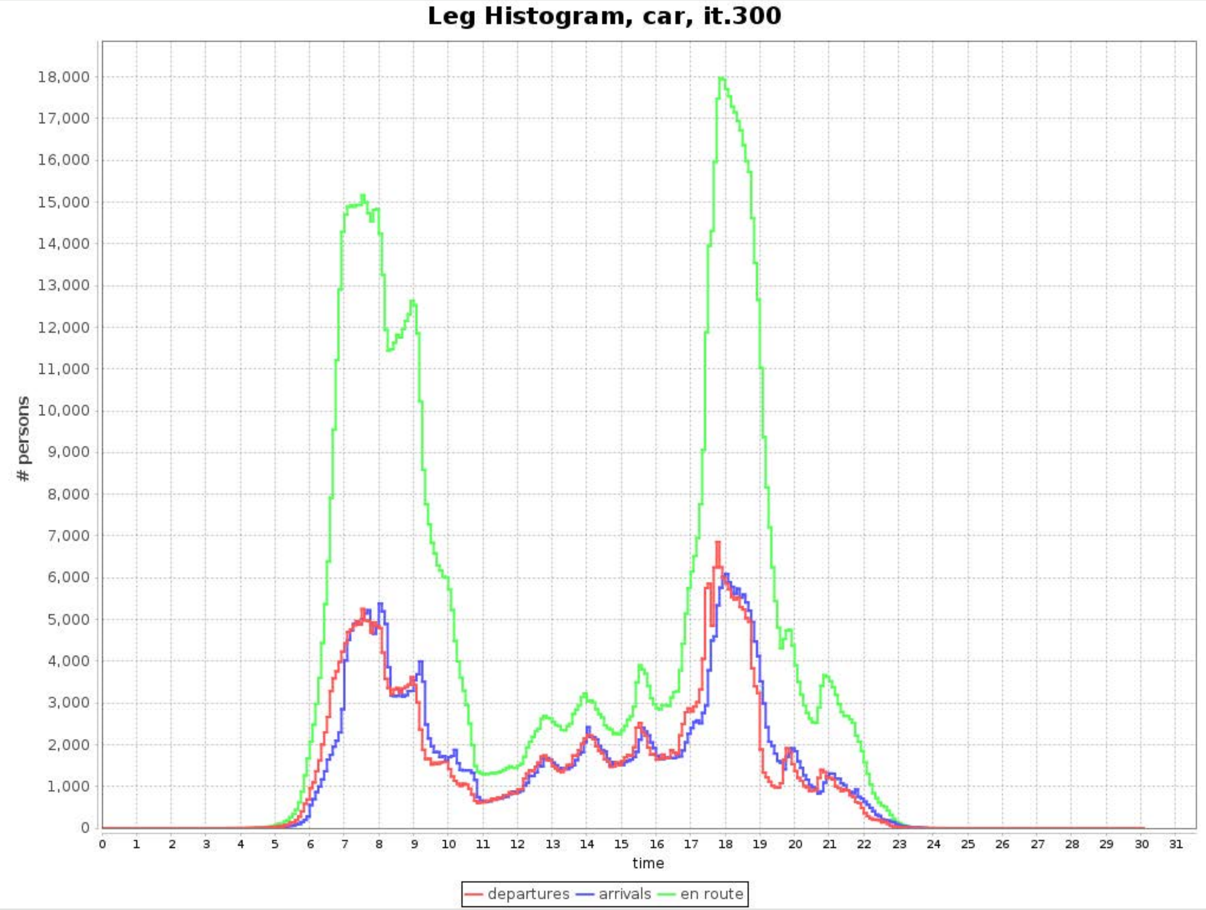


MODE SHARE AFTER POST-CALIBRATION FIXES

Car before

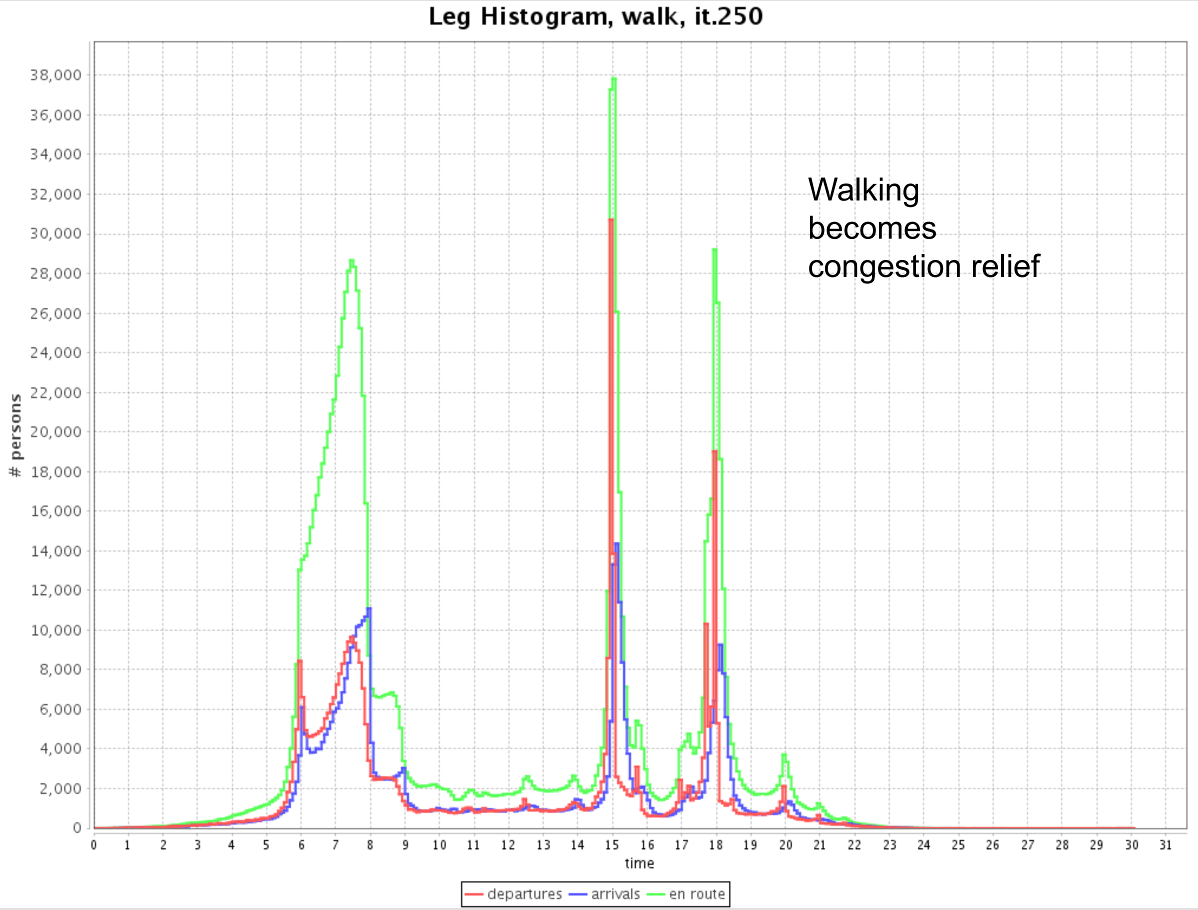


Car after



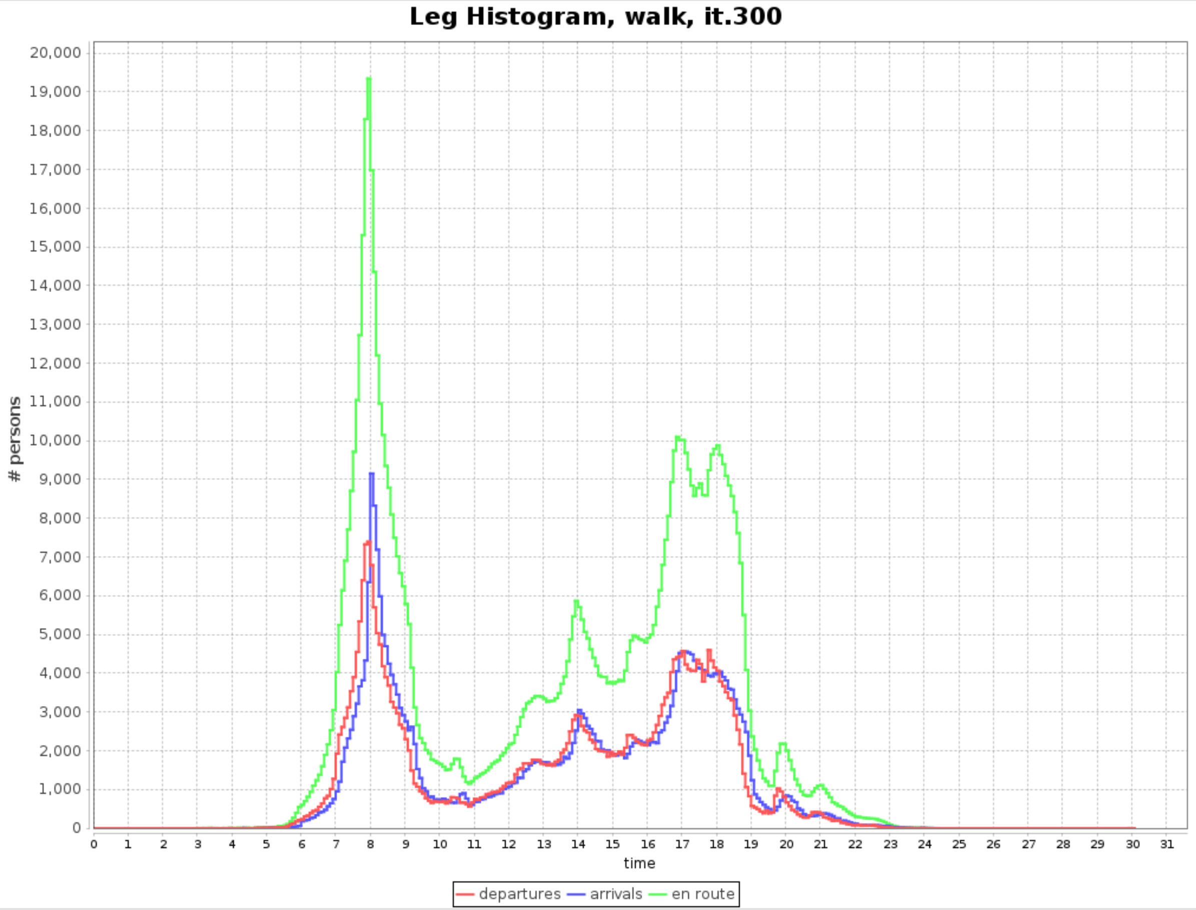
MODE SHARE AFTER POST-CALIBRATION FIXES

Walk before



Walking becomes congestion relief

Walk after



CALIBRATION CONCLUSIONS

Sources of deviation:

- Input data:
  - foreigners
  - **job locations**
  - facility-to-link-assignment
- Route choice parameters
- (Pace of urban development?)

Implications

- Relative differences, e.g. before and after, can provide insight, but
- Until issues are addressed, cannot use in policy planning

What is required:

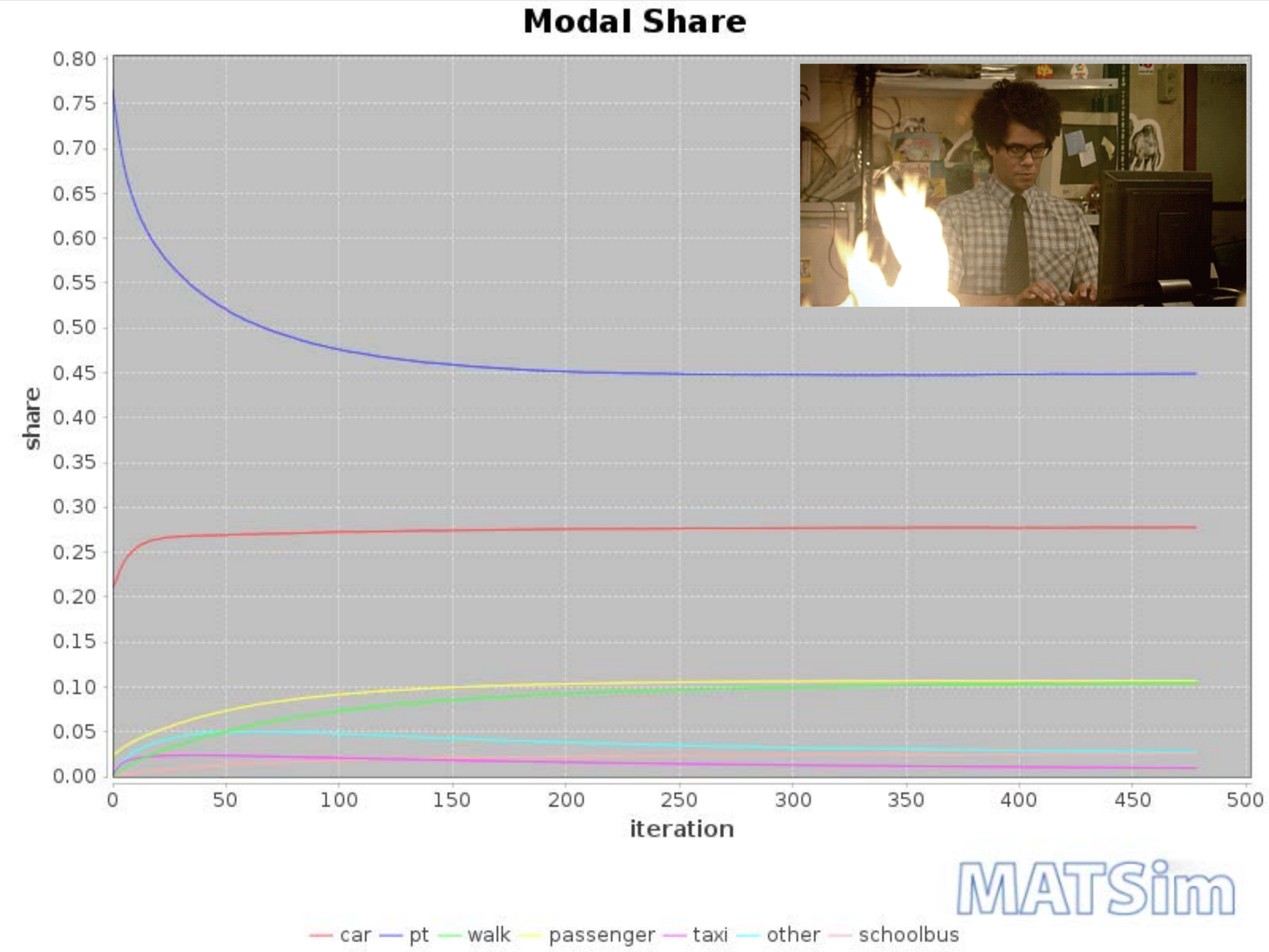
- **Improved demand information, e.g. cell phones**
- Continued investment in model; e.g. Switzerland model is revised every year
- Calibration should be done **on-site;** and be **massively parallel, re-run after major changes**
- **Reducing number of modes** should make things more manageable

Complexity & performance

Calibration took 67 CPU months on high-end Xeon servers

Mode share changes very slowly over iterations, so the graph on the right took 8 days to construct…

While performance was not a focus for this project, it will have to be addressed in future



SCENARIOS

BASELINE

DOWNTOWN LINE II

No changes in population

New transit schedule, network extension only

JPR

Baseline population, network

30% jobs moved closer to home locations

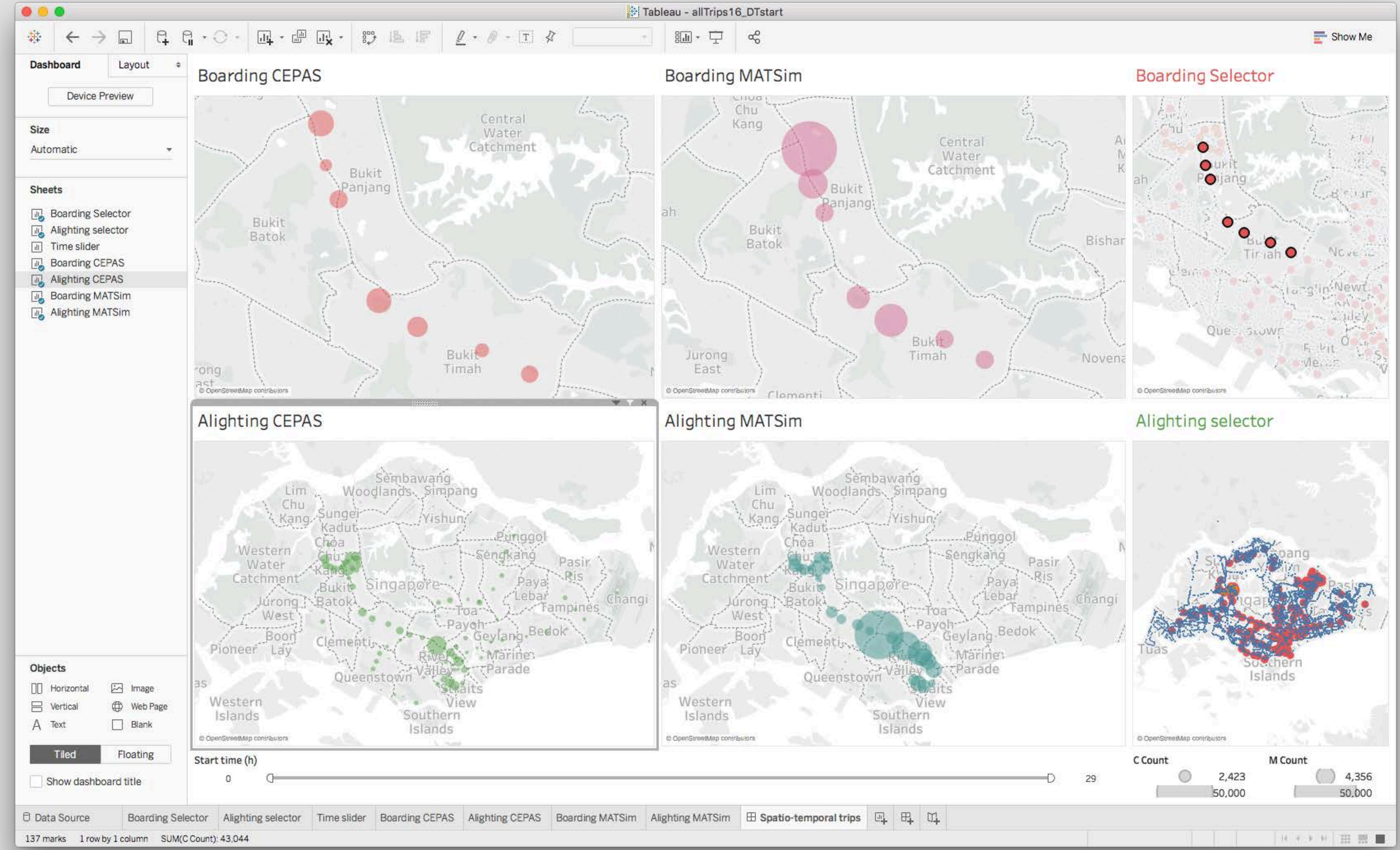
Re-run model stack from work location choice forward

DOWNTOWN LINE II

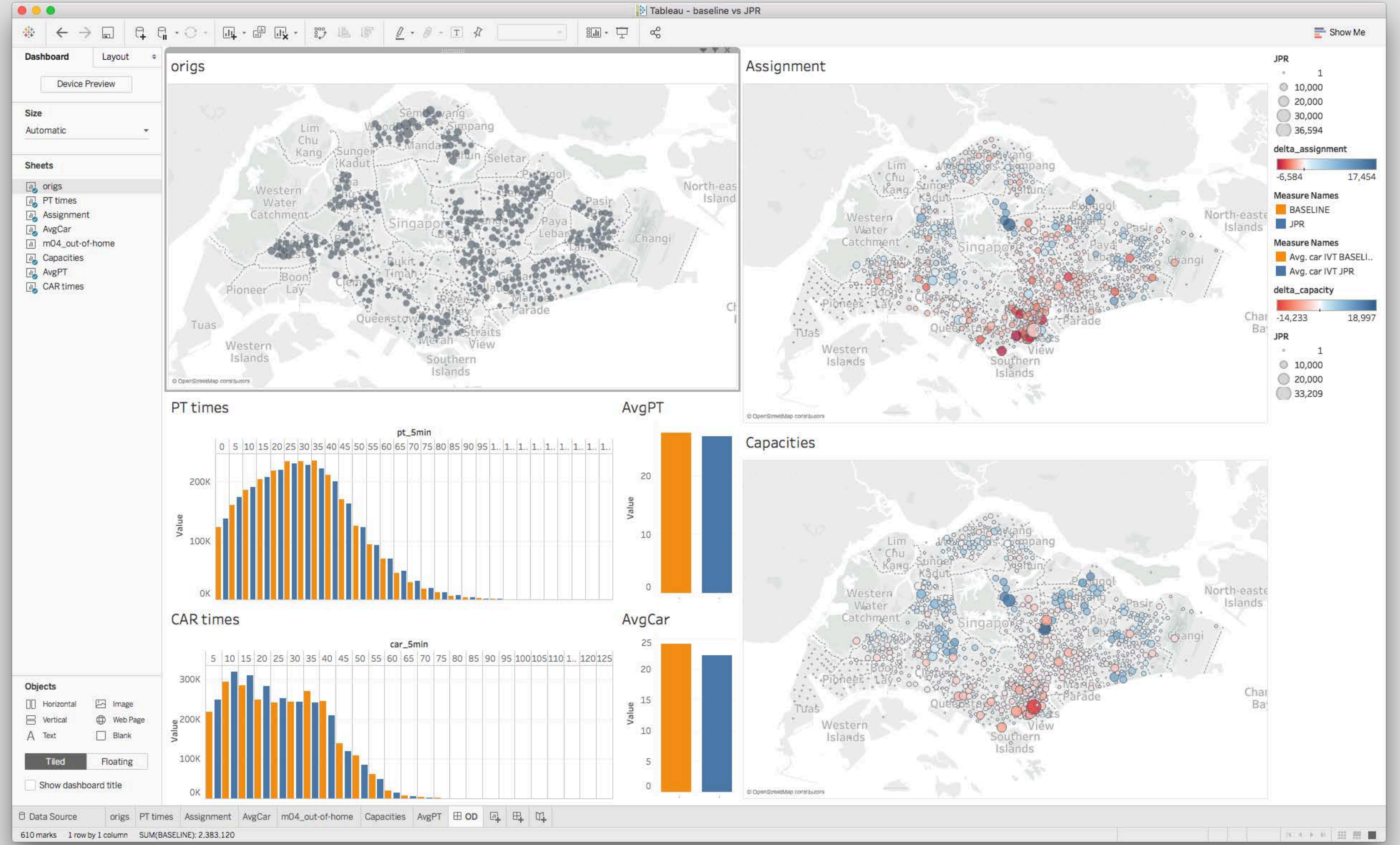
Affected data sets
Transit network

Assumptions
No change in base population work location choice
No change in accessibility
Only affecting mode choice and route choice decisions

DOWNTOWN LINE II



SCENARIOS: REACH-BASED JOB PROVISION RATIO OPTIMIZATION



Affected data sets
Workplace capacities
Facilities

Assumptions
30% of jobs moved closer to workers
No change in local accessibility and diversity following move

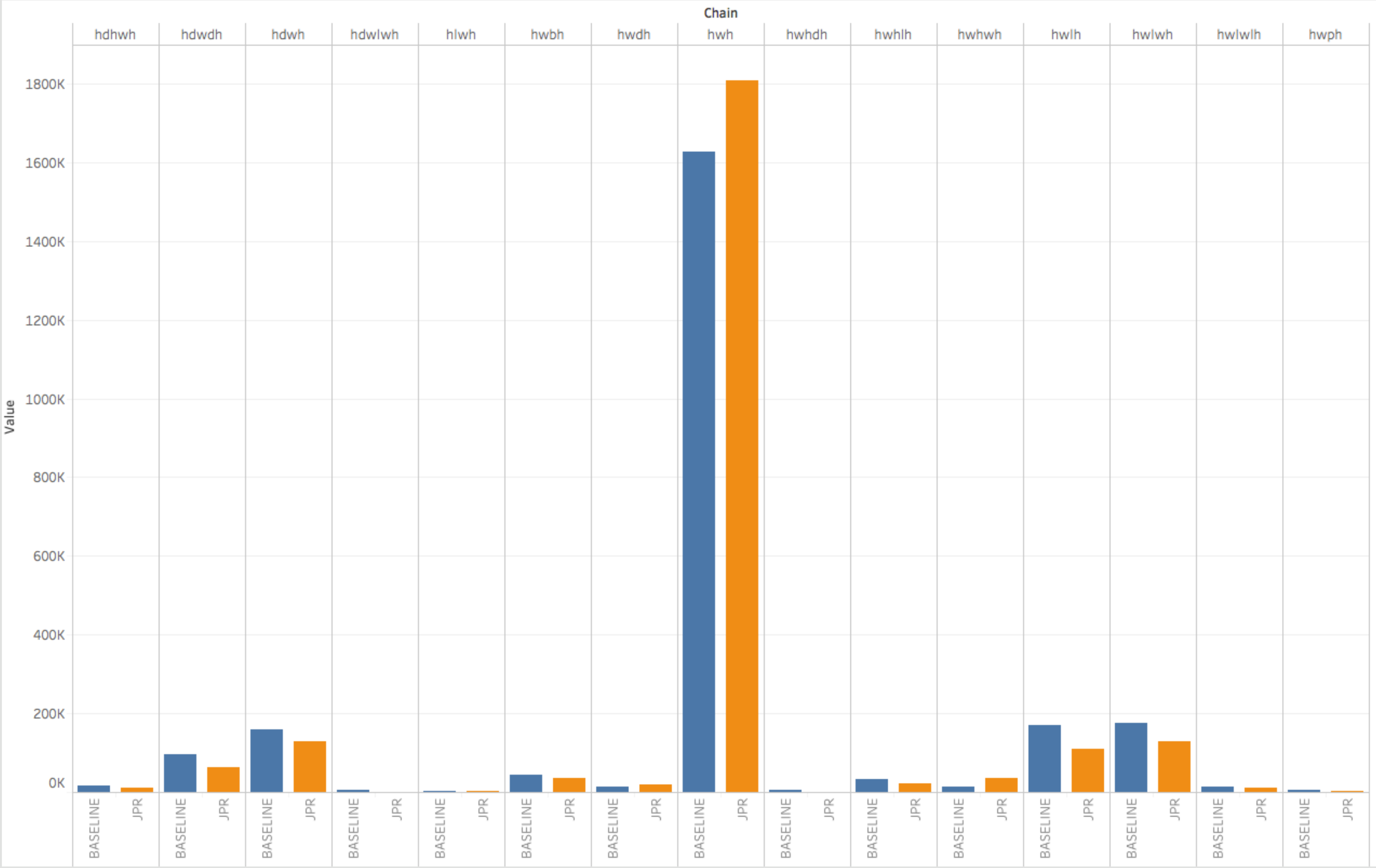
SCENARIOS: REACH-BASED JOB PROVISION RATIO OPTIMIZATION

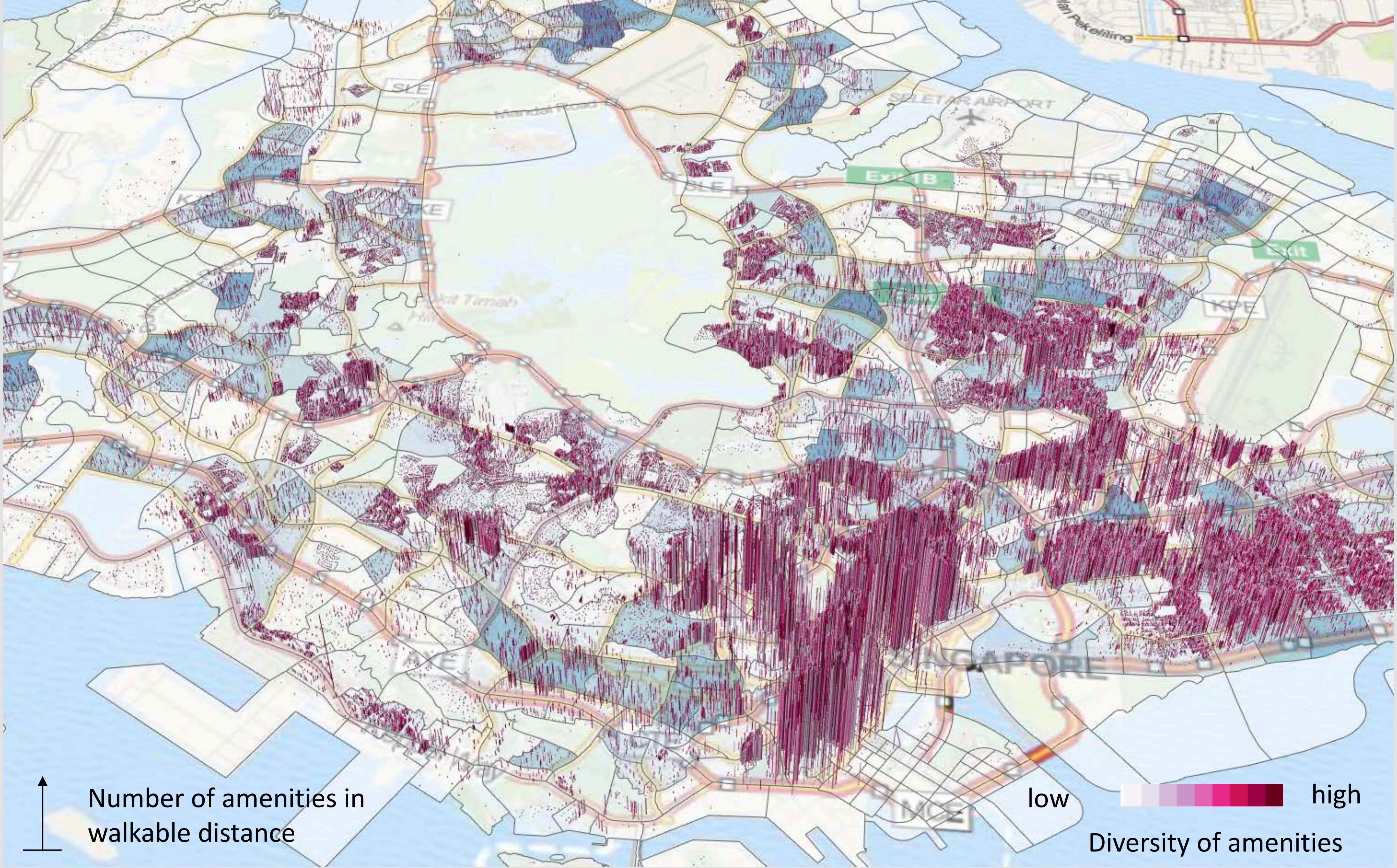
Activity chains

Lower diversity around work location favours shorter activity chains

Only exception: hwhwh, up from 12k to ~ 34k

Activity chain model doesn't consider tour length




SLE
SELETAR AIRPORT
Exit 1B
TPE
KPE
Exit
AYE
SINGAPORE
MCE
Number of amenities in walkable distance
low
high
Diversity of amenities

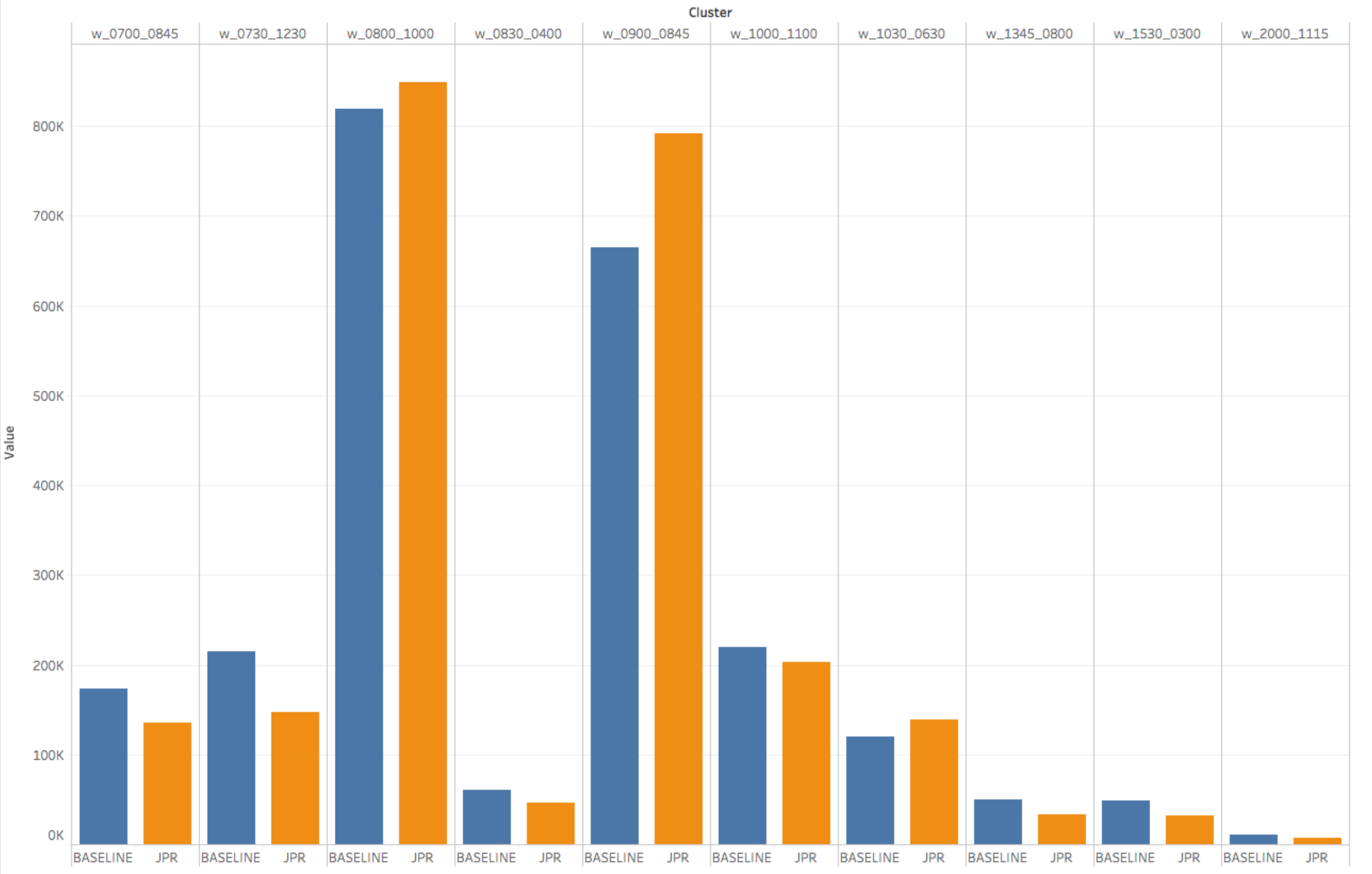
SCENARIOS: REACH-BASED JOB PROVISION RATIO OPTIMIZATION

Work start time and duration

Currently, two linear regression models run in sequence, then assigned to one of 10 clusters

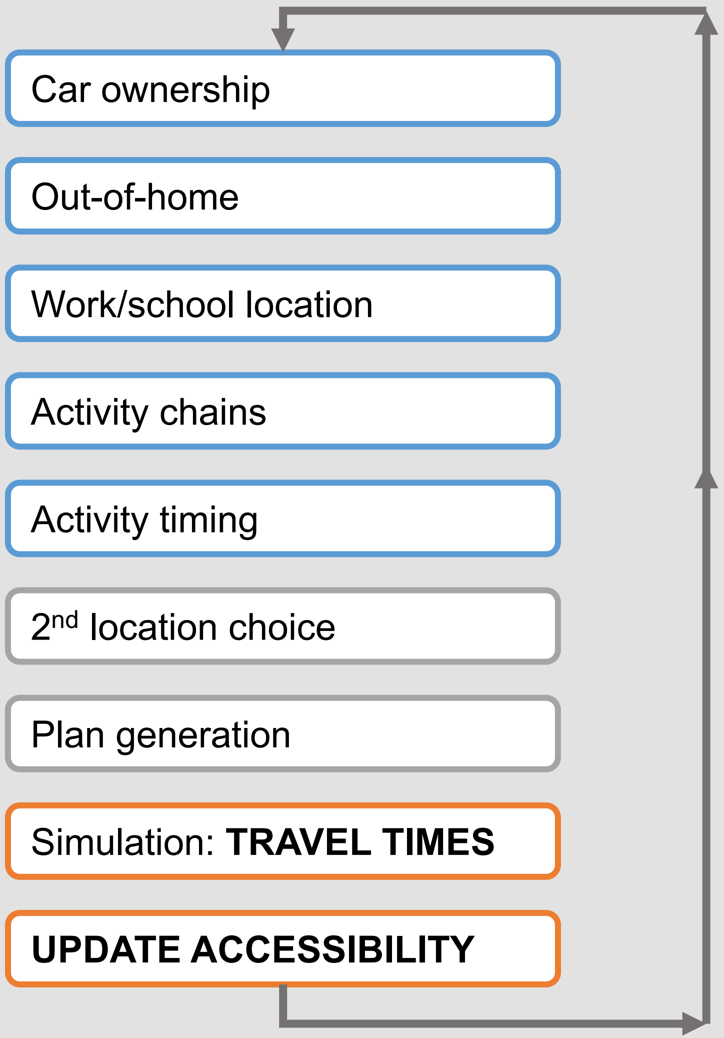
Activity chain, home and work diversity are strong predictors in those models

Less activity chain diversity -> less diversity in work start time and duration



SCENARIOS: CONCLUSIONS

1. Evaluating network interventions simpler than land use intervention
2. Initial implementation of JPR shows no strong effect
3. Need to have a cycle where all **accessibility variables are updated** to capture induced demand effects
4. Land-use intervention does not consider the effect of **local activity diversity and secondary activity options increasing** due to the work opportunities created there
5. JPR scenario should also incorporate **distance decay function** in activity location choice
6. Need **integrated activity chain/timing/location** assignment, with cognizance of tour length
7. Working through the land-use intervention is useful in **exposing the dimensions** of the problem, even if no conclusive answer could be provided



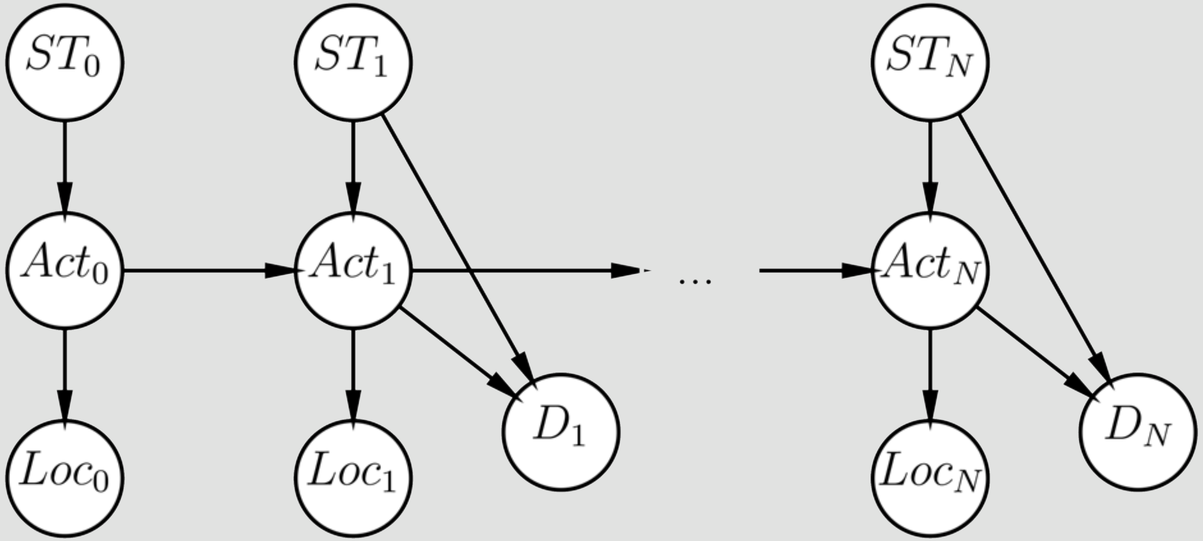
SCENARIOS: CONCLUSIONS

Progress on simultaneous activity and location assignment has been made.

The IO-HMM produces a very good fit when estimated against HITS (Anda & Ordonez, 2017)

Will be extended to work with cellular phone data in order to impute activity purpose

Input-Output Hidden Markov Model for Activity Scheduling



LESSONS LEARNT

WORKFLOW, DATABASE

Clear conventions for workflow emerged later in the project

Output-driven development in future, with tests

As far as possible, maintain a single programming language

Incorporate report into scripts, e.g. Bayesian network scripts

Remake package in R as a possible solution

CALIBRATION

Calibration focused on modes and mode share, routing parameters lacked

Input data of especially work locations affect results dramatically

Repetitive cycle of development and calibration needed, as changes have far-reaching effects

CONTACT SESSIONS

Walking through the demand modelling steps raised awareness of interacting processes

Identifying shortcomings and assumptions in modelling process helpful in raising awareness

Platform gradually became a common framework of understanding between line depts

AGENDA FOR IMPROVEMENT

Problem	Solution	Implication	Importance
Work locations are currently estimated from CEPAS and reported mode shares	Mobile phone data	Realistic demand produces realistic network loadings, e.g. MATSim SF Bay Area	*****
Accessibility effects of scenario changes don't affect demand generation	Repeated loop of simulation > accessibilty calculation > demand gen > simulation…	Better capture induced demand effects	****
Junctions in MATSim have too little impedance; traffic lights are modelled as a change in capacity only	Improved junction dynamics with realistic traffic signals	Improved network loading. Ability to test new junction dynamics, e.g. AVs	****
Some agents should be more flexible than others in deviating from prescribed activity timing	Based on household and personal demographics, assign a 'flexi-time' factor to an agent	More realistic activity timing	**
Currently no coordination between household members	Intra-household coordination model	More realistic mode choice, activity timing	*
Calibration is currently a manual, serial process	Semi-automatic, massively parallel calibration	Larger number of parameters evaluated in a shorter time	***

STAY IN TOUCH

Pieter Fourie
fourie@ivt.baug.ethz.ch

Web
http://www.fcl.ethz.ch/research/responsive-cities/engaging-mobility.html