# Analysis of SimMobility-Aimsun iterations

# Estimation of equilibrium search space in integrated demand-supply transportation model

Aalto University
Aalto Science Institute

## 1 Theoretical and technical concepts

### What the poster is about

The poster is devoted to the integration of supply and demand models in transportation modelling. We use the integration of supply and demand models because they mutually complement each other: the demand model simulates mobility of the population but it does not take into account traffic load, while supply model provides information concerning how efficiently urban infrastructure may manage movement needs but it does not simulate mobility by itself.

### Technical concept

Our study area was Tallinn, Estonia. From the side of demand SimMobility was used, a model taking a synthetic population as an input. This dataset is completely anonymized, since information about each individual may cause privacy concerns. SimMobility generates a list of trips for each person assigning mode of transport and origin-destination zones

Based on these lists of trips an origin-destination matrix was created with total number of trips for each O-D pair and it was fed to another model (Aimsun) as an input. This program simulates traffic in the city and outputs travel times for each O-D couple. Afterwards received travel times were inserted in SimMobility building a looping cycle between the 2 programs that is shown in **Figure 1**.

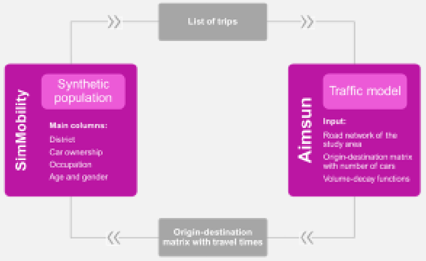
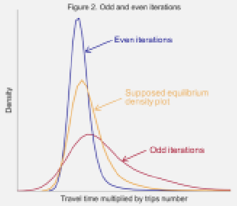



Figure 1. SimMobility-Aimsun loop iterations

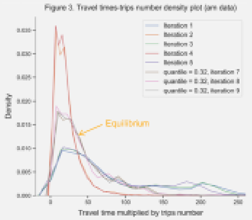
## 2 Search of equilibrium

In order to analyse the difference between each iteration in the designed loop it was decided to consider simultaneous change of trips number received from SimMobiliry and travel times from Aimsun by their multiplication. Only morning time period was reported (7:00 – 10:00) because pm results followed the same trends.




Initially it was supposed that after certain number of iterations equilibrium between supply and demand would be reached. However, the loop got stuck between odd and even iterations instead (**Figure 2**).

This may be explained by the fact that only simulated data was used which contained incorporated inaccuracy.




As a result, in odd iteration we got higher number of trips but lower travel times and vice versa in case of even iterations.

To reach equilibrium we caused perturbation calculating the quantile of the first five iterations' travel time and putting it in SimMobility to run a new loop. **Figure 3** demonstrates that quantile q=0.32 gives us practical equilibrium.

## 3 Relationship between quantile and trips number

To identify which quantile gives us equilibrium we defined a search space between quantile = 0.2 and quantile = 0.8. Then we ran 3 iterations for each quantile separately.

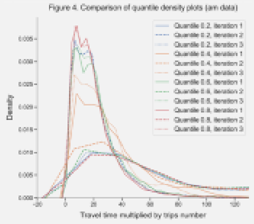
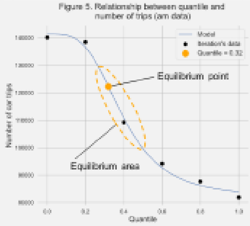



**Figure 4** displays density plots of travel times multiplied by number of trips for each quantile. It is well seen that fluctuations between iterations that are closer to quantile = 0.4 become smaller.

Trying to reduce number of iterations to figure out quantile which gives us equilibrium, we modeled a relationship between number of trips and quantile. From the previous plot it is apparent that equilibrium area is around quantile 0.4. Relationship has a form of inverse logistic function as **Figure 5** shows, thus to fit this curve, a 4-parameter logistic regression was used:

$$y = a + \frac{d - a}{1 + \frac{x^b}{c}}$$

Where a, b, c, d are model parameters

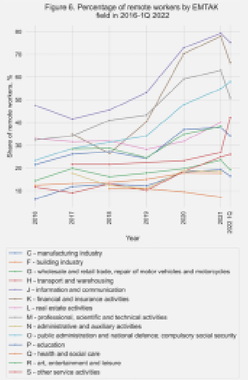



## 4 Future application and conclusions

### Future application

Received methodology and results will allow us to analyse different research problems such as: mobility on demand, automated vehicles or ride sharing. This is made possible by integrating demand and supply models.

The next planned application is remote working which have seen huge rise due to the pandemic. For that purposes, it is planned to utilize the share of remote workers by occupation in Tallinn (**Figure 6**). A forecast for each occupational group will be built in order to further assign remote working class to each individual in synthetic population and analyse the difference between real and future predicted data.

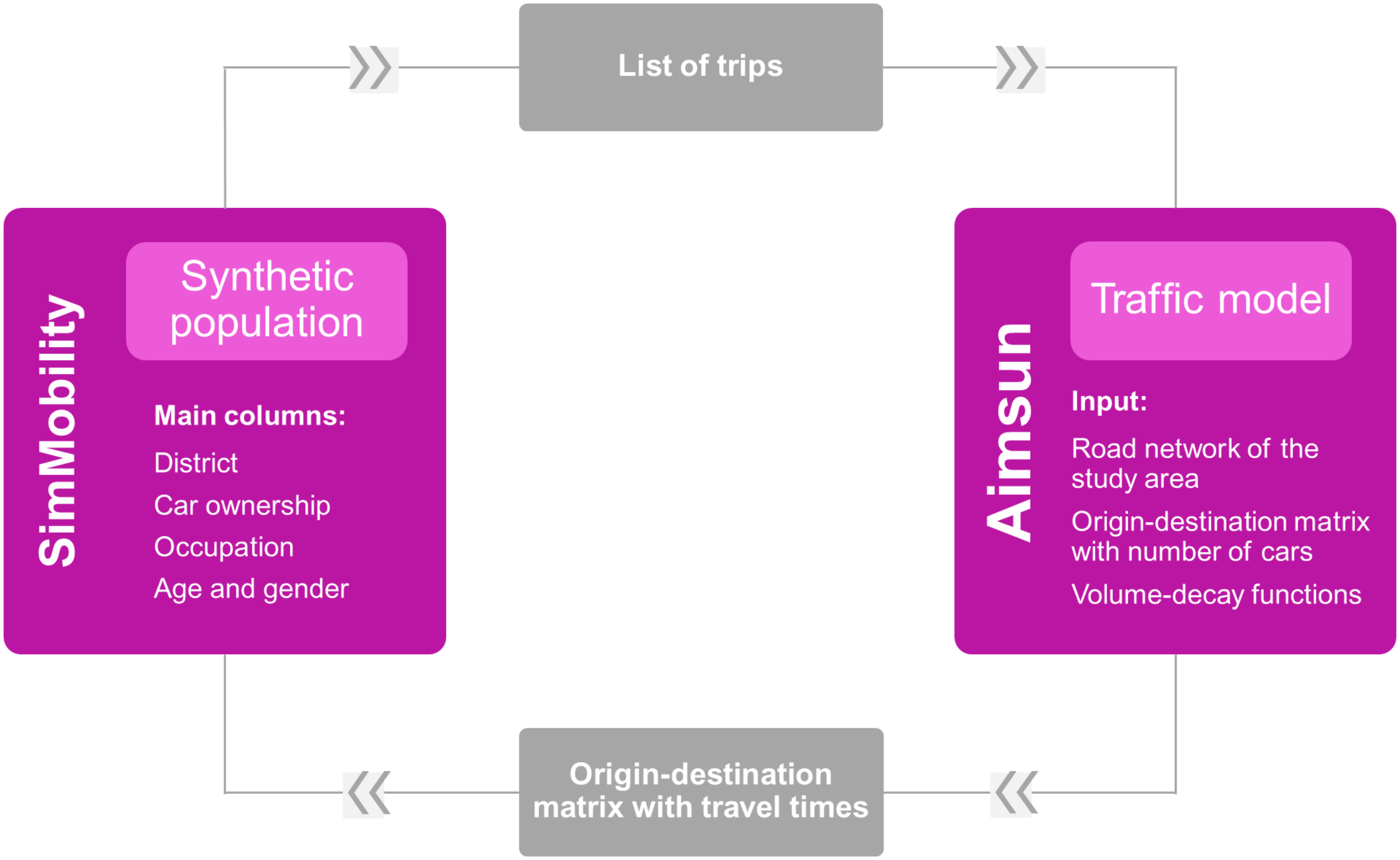



### Conclusion

In the research a **methodology** that allows to integrate supply and demand models, which usually are observed separately, was developed. The proposed methodology is able to identify the **equilibrium point** in the **search space** without the need of iterating multiple loops for each quantile (through the 4-parameter logistic regression). This results in **reduced computational effort** and **increased applicability**. Once the equilibrium is reached, different research direction may be investigated.


Petr Anashin, Assistant Researcher. Supervisors: Claudio Roncoli, Assistant Professor; Serio Agriesti, PhD Student
School of Engineering, Department of Built Environment, Transportation Engineering
Contact details:
anashinpetya@gmail.com

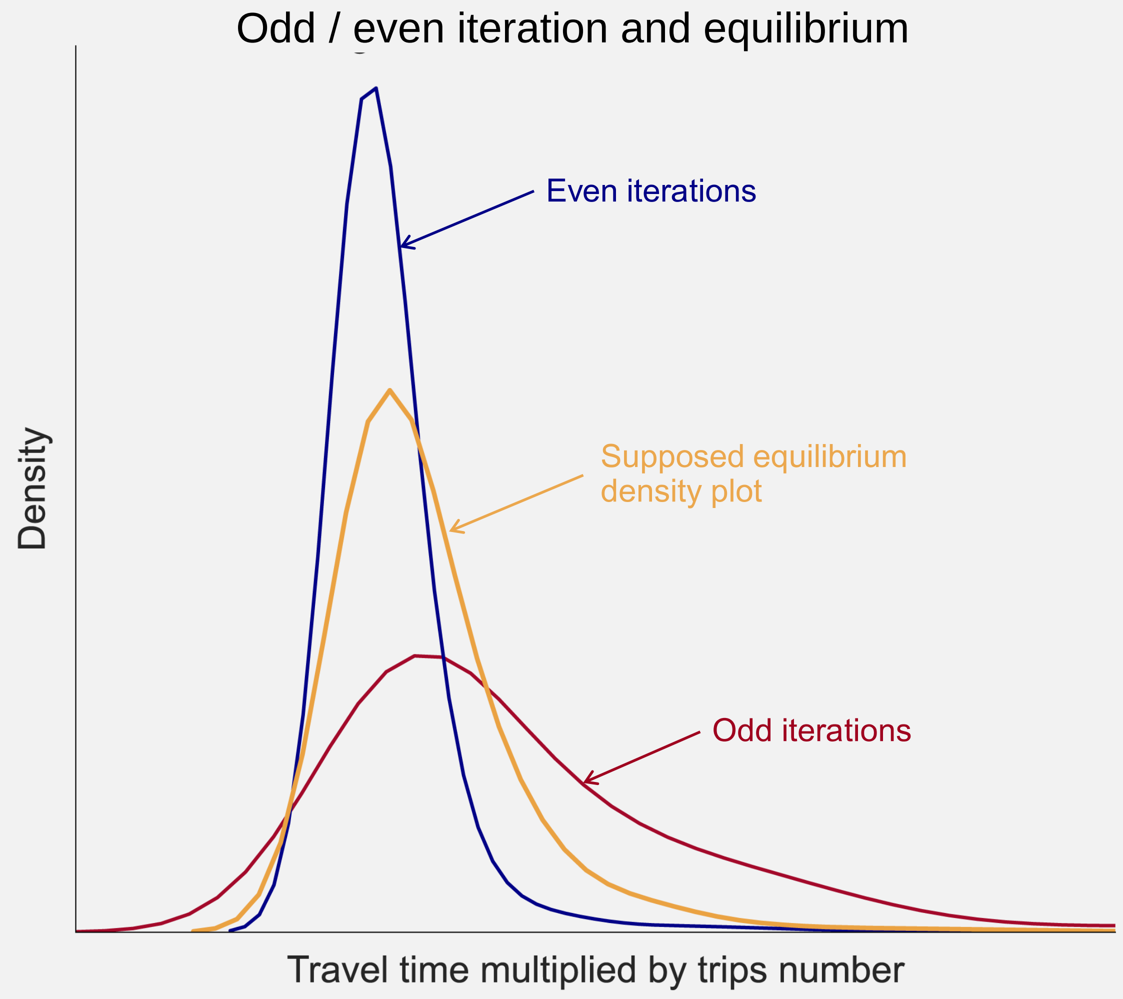
# Technical concept



# Stucked iterations

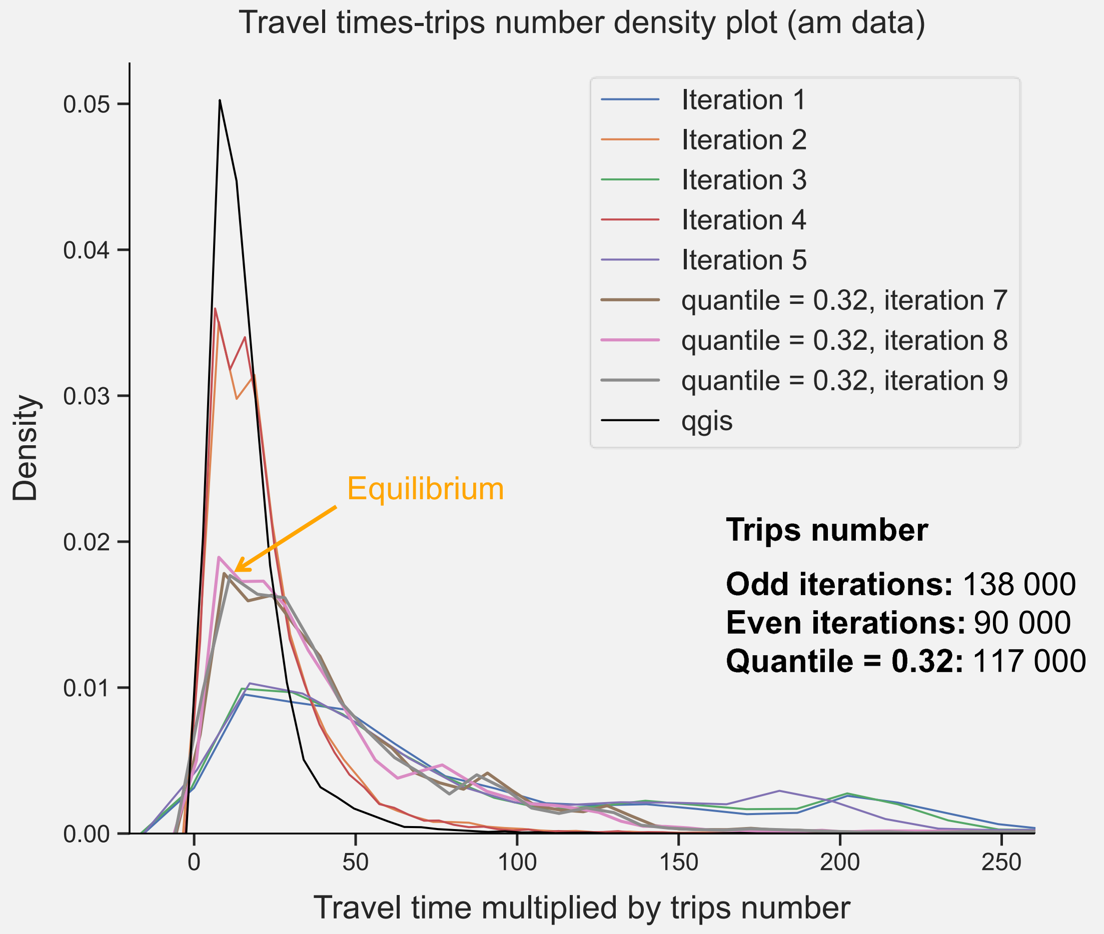
## Stucked iterations

1

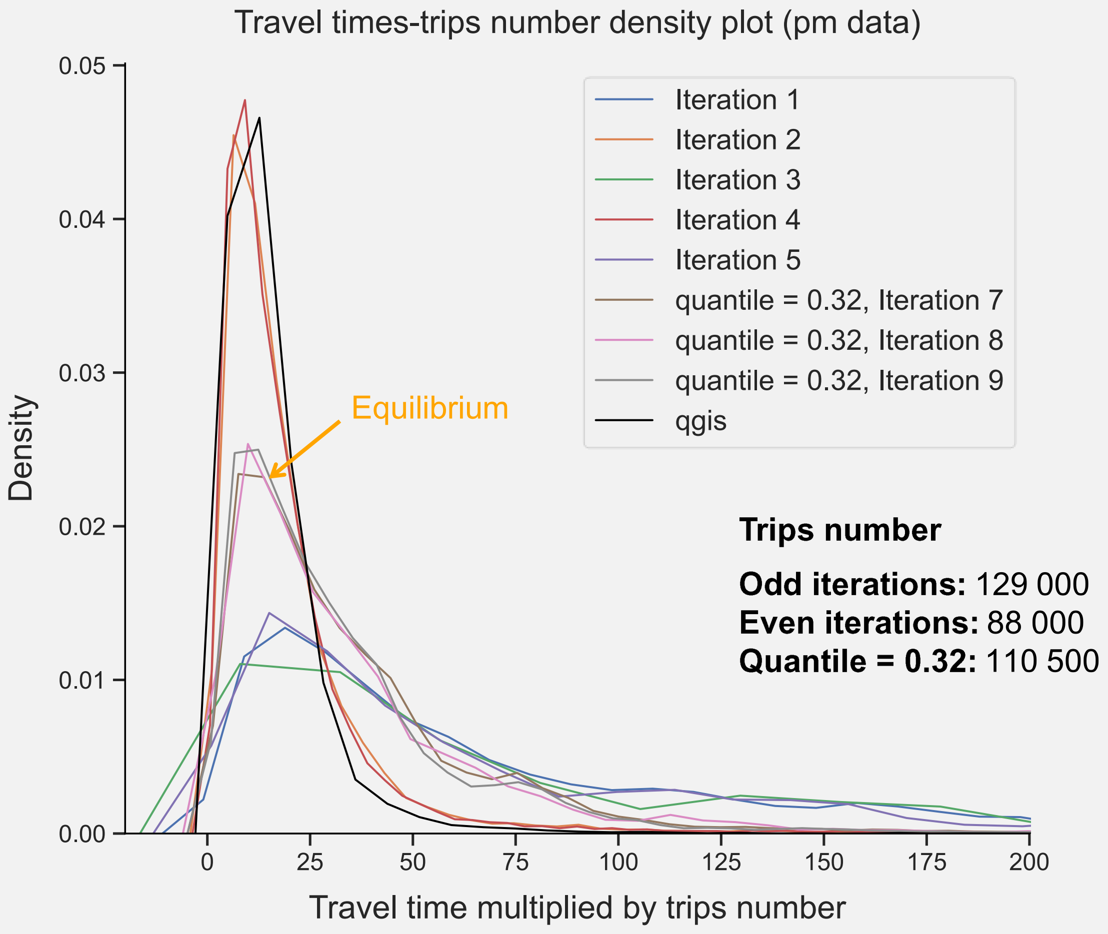


# Iterations comparison

## 1 Iterations density plots and trips number, a.m.



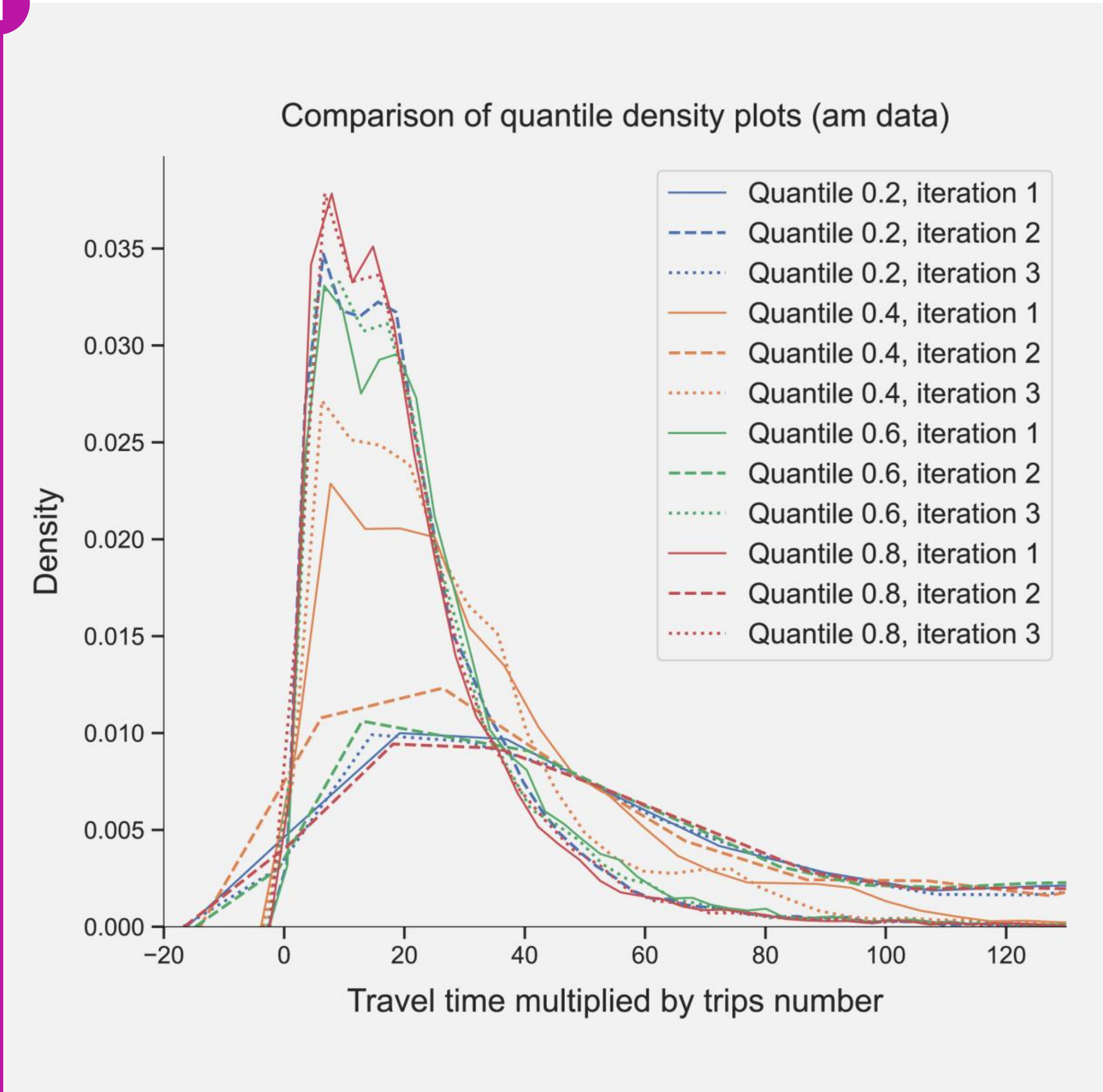

## 2 Iterations density plots and trips number, p.m.



# Quantiles comparison

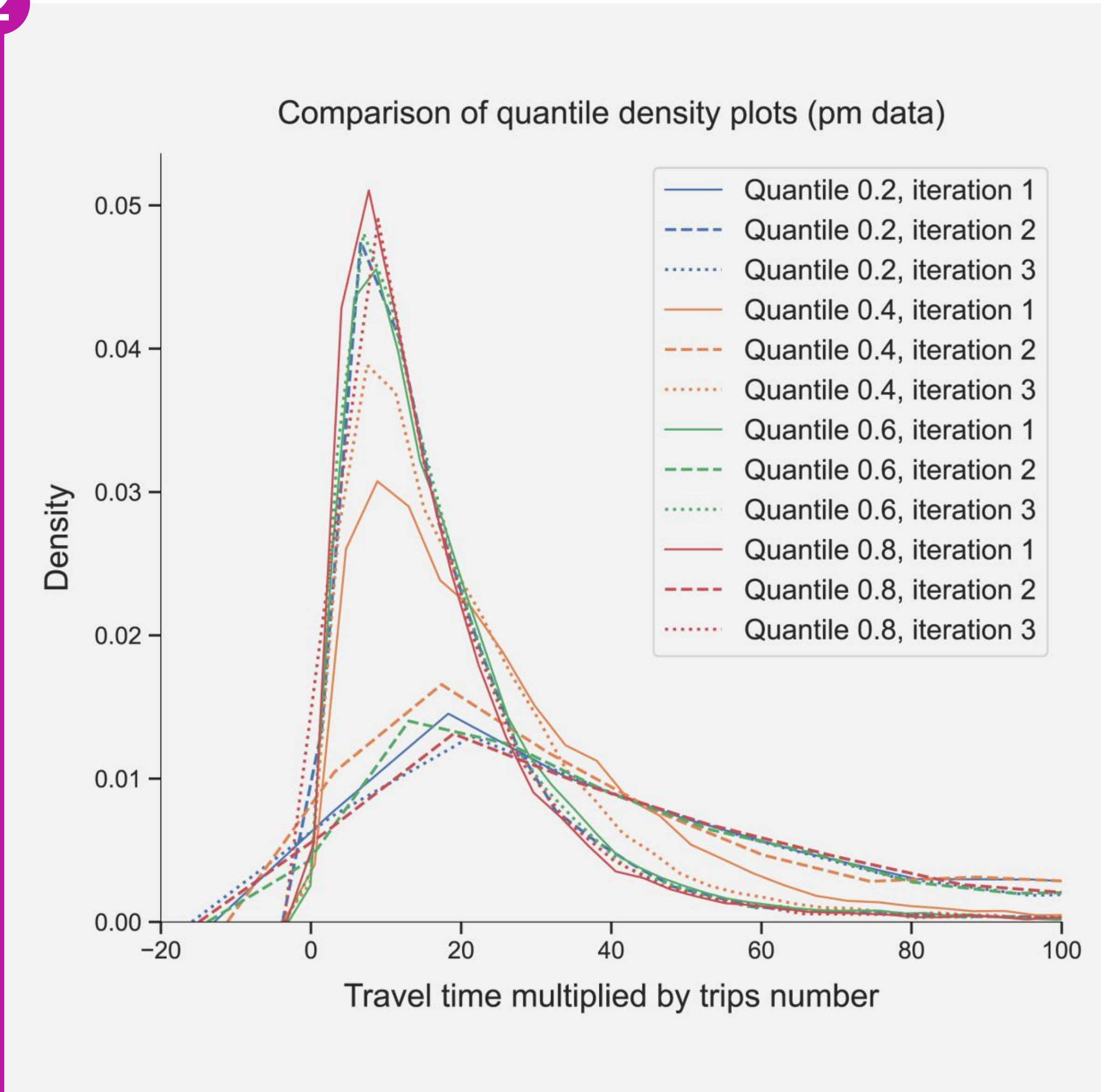
## Quantile search space, am

1

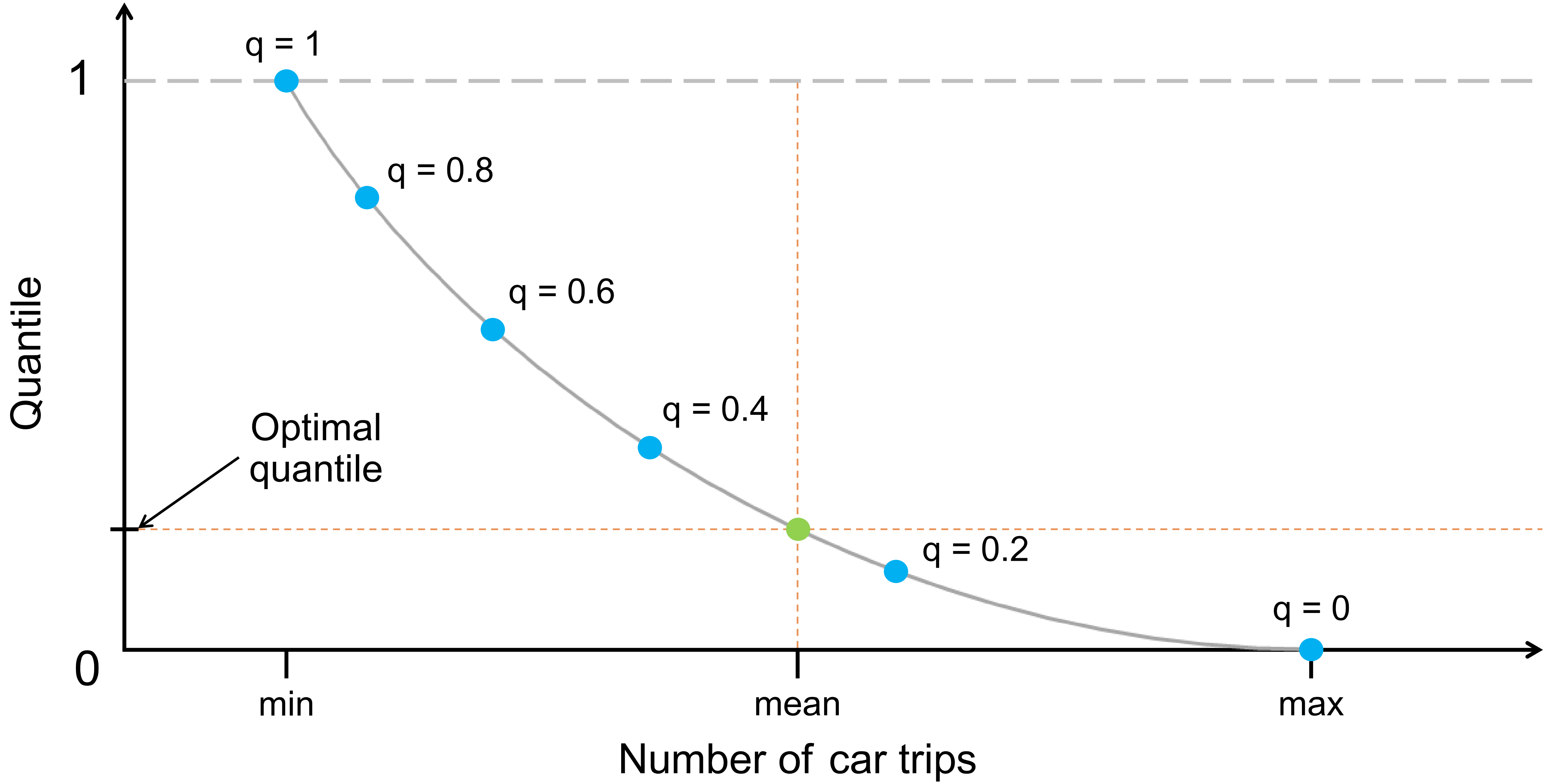



## Quantile search space, pm

2

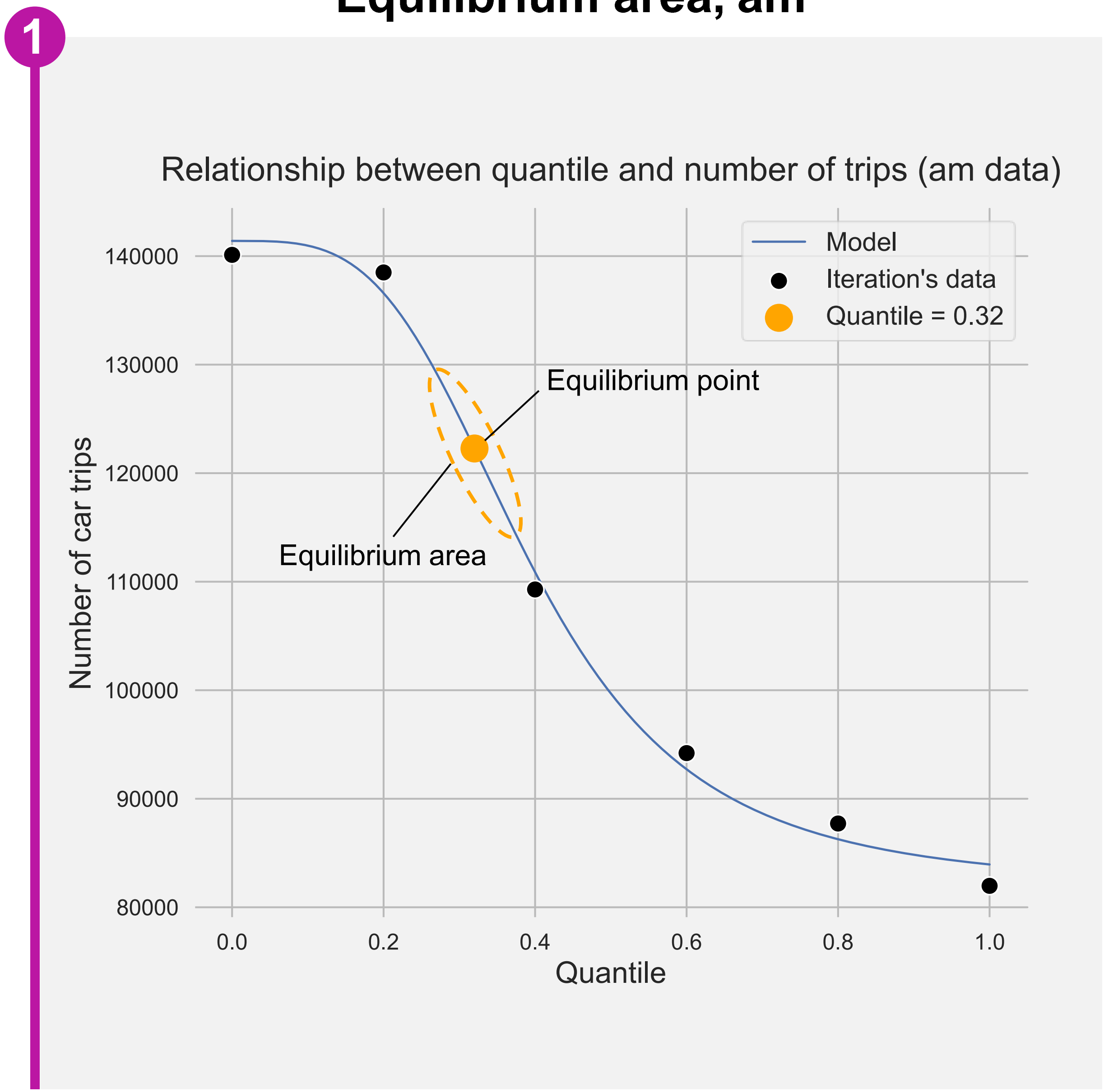


## Identification of optimal quantile

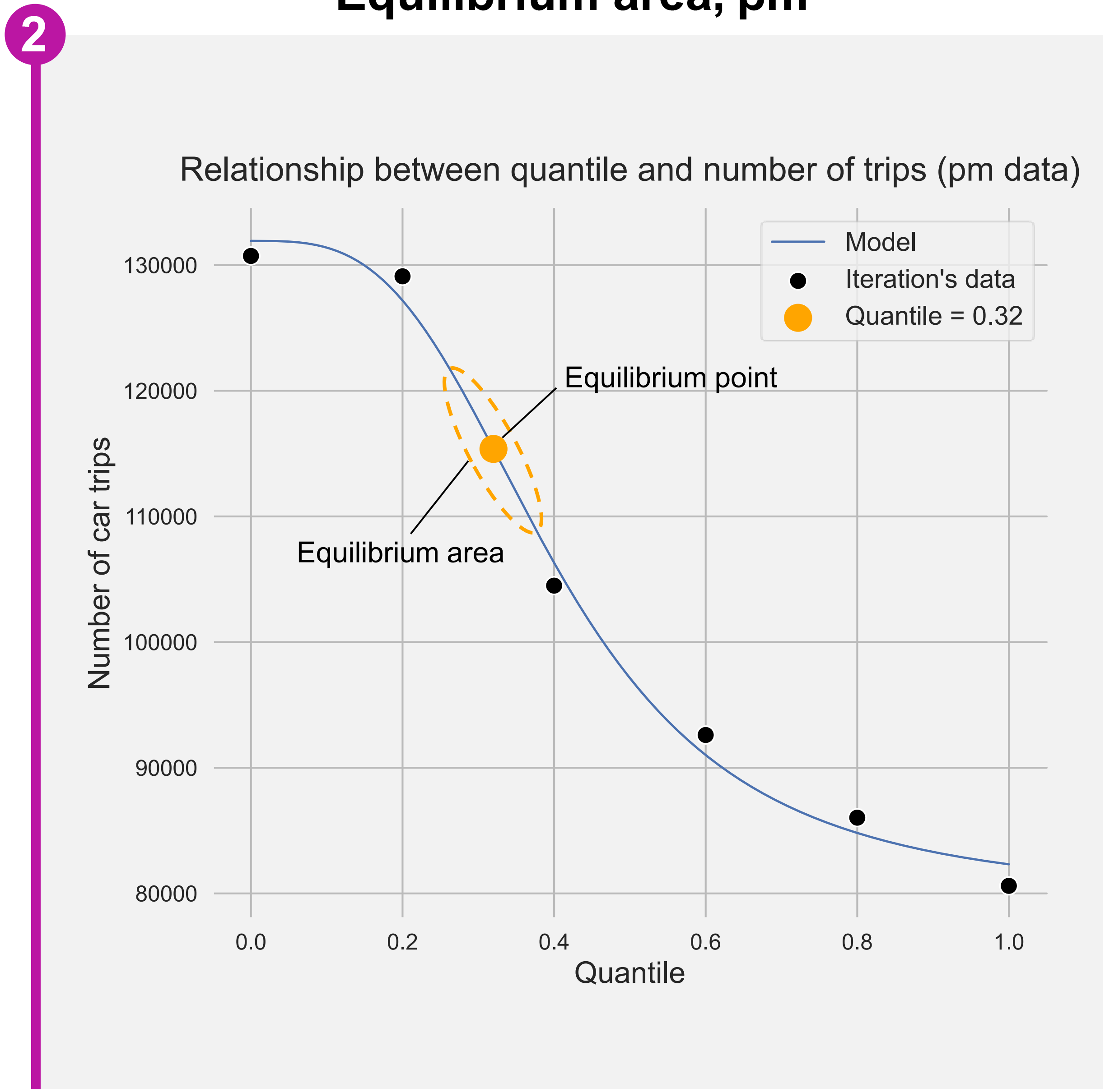


# Identification of optimal quantile

## Equilibrium area, am



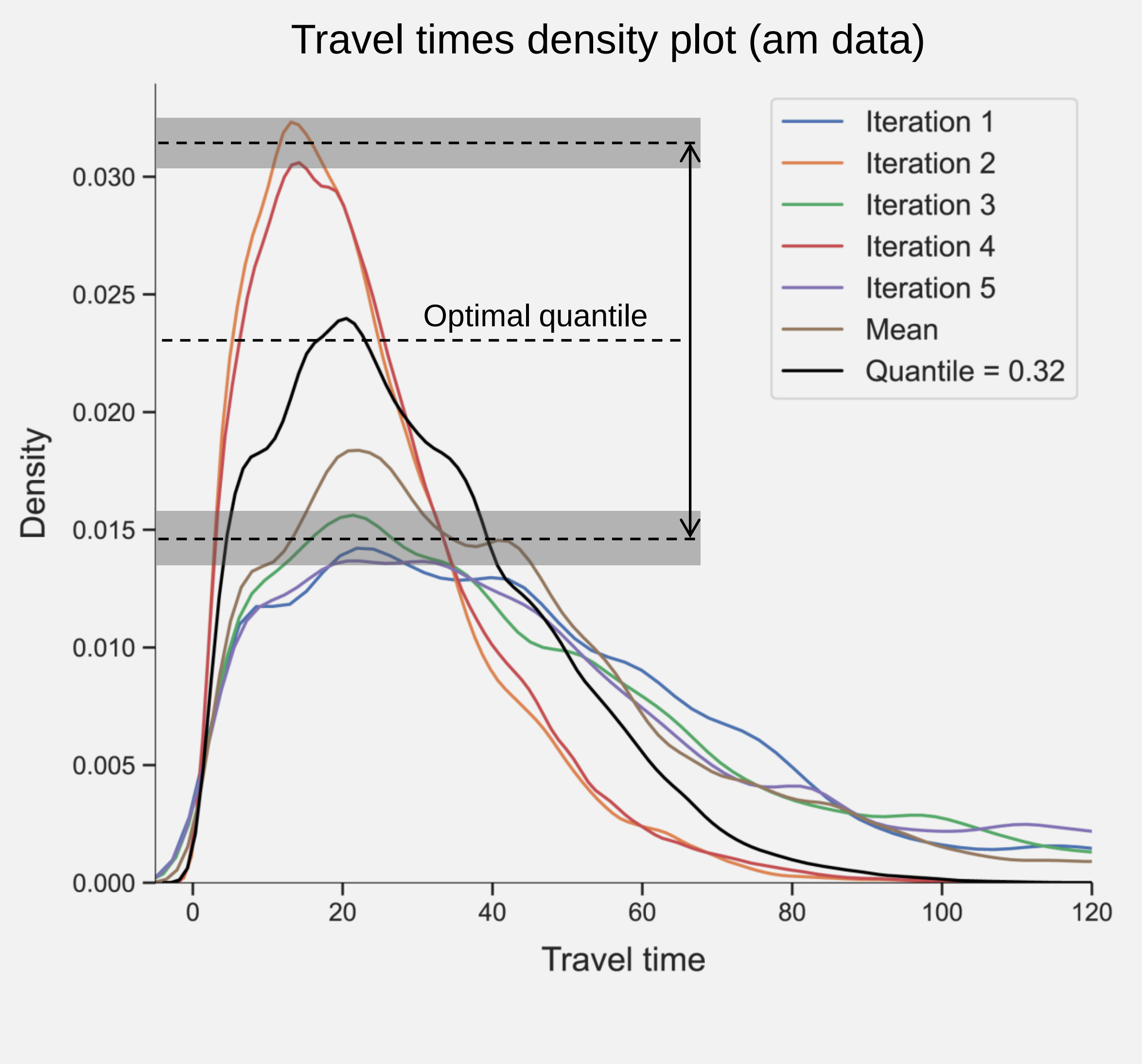

## Equilibrium area, pm



# Identification of optimal quantile

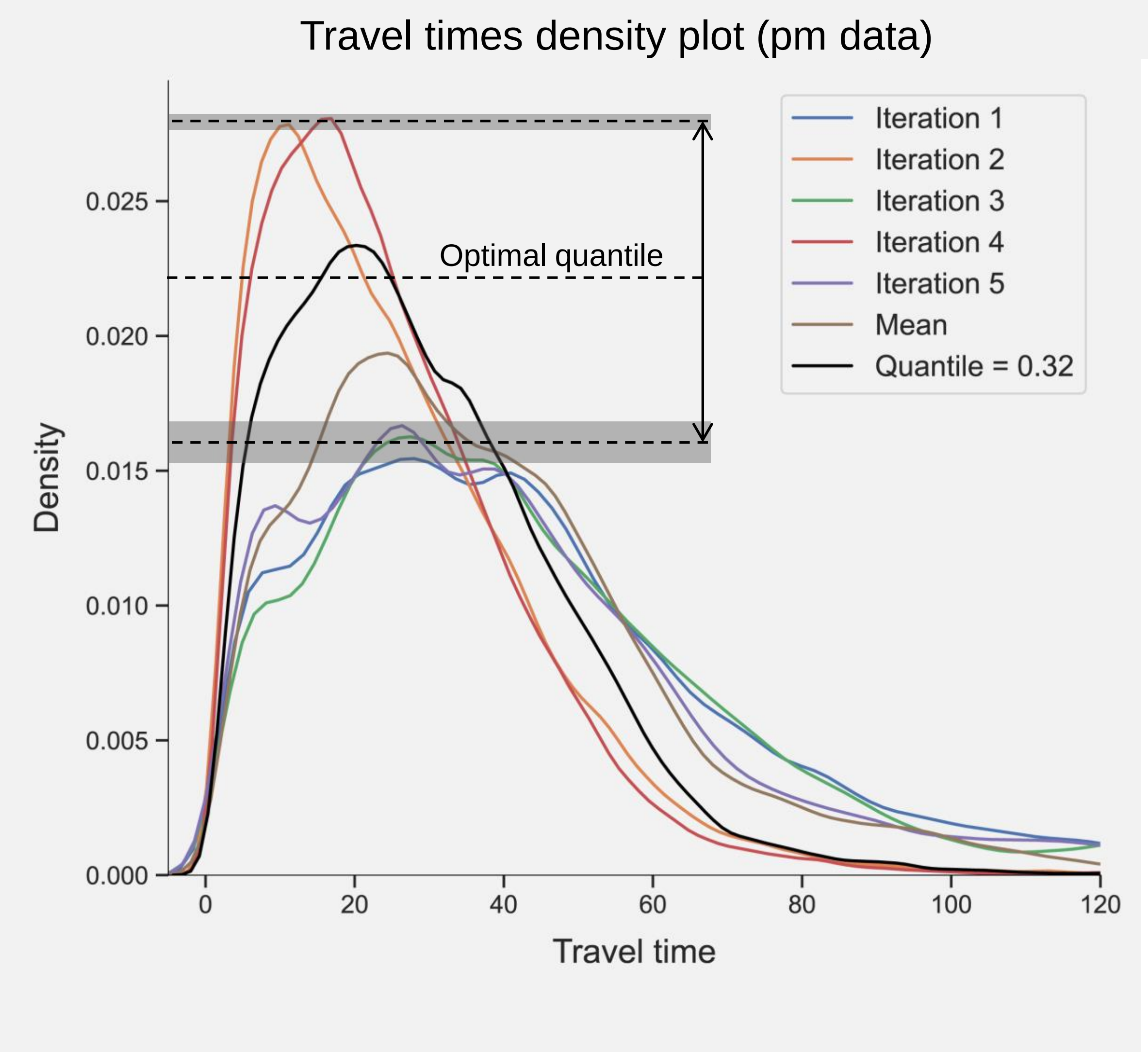
1

## Equilibrium area, am



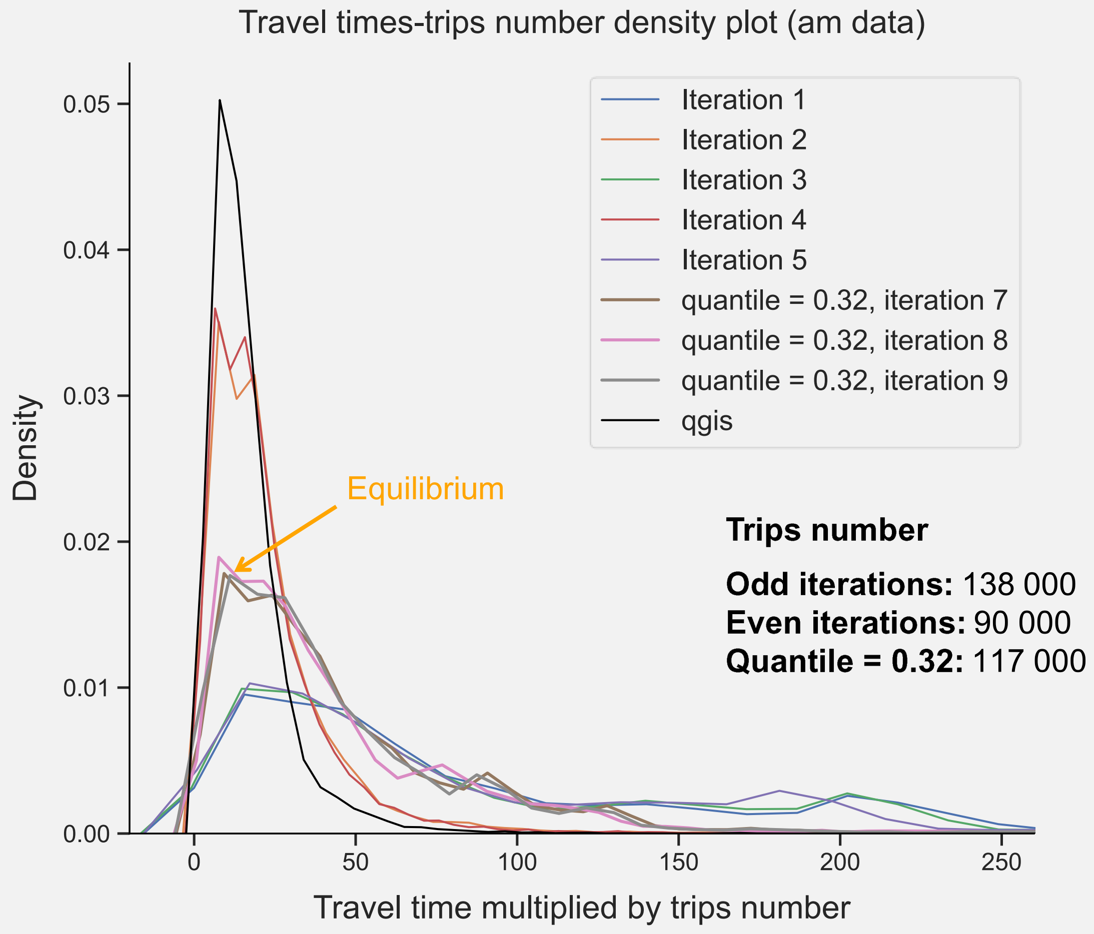

2

## Equilibrium area, pm



# Iterations comparison

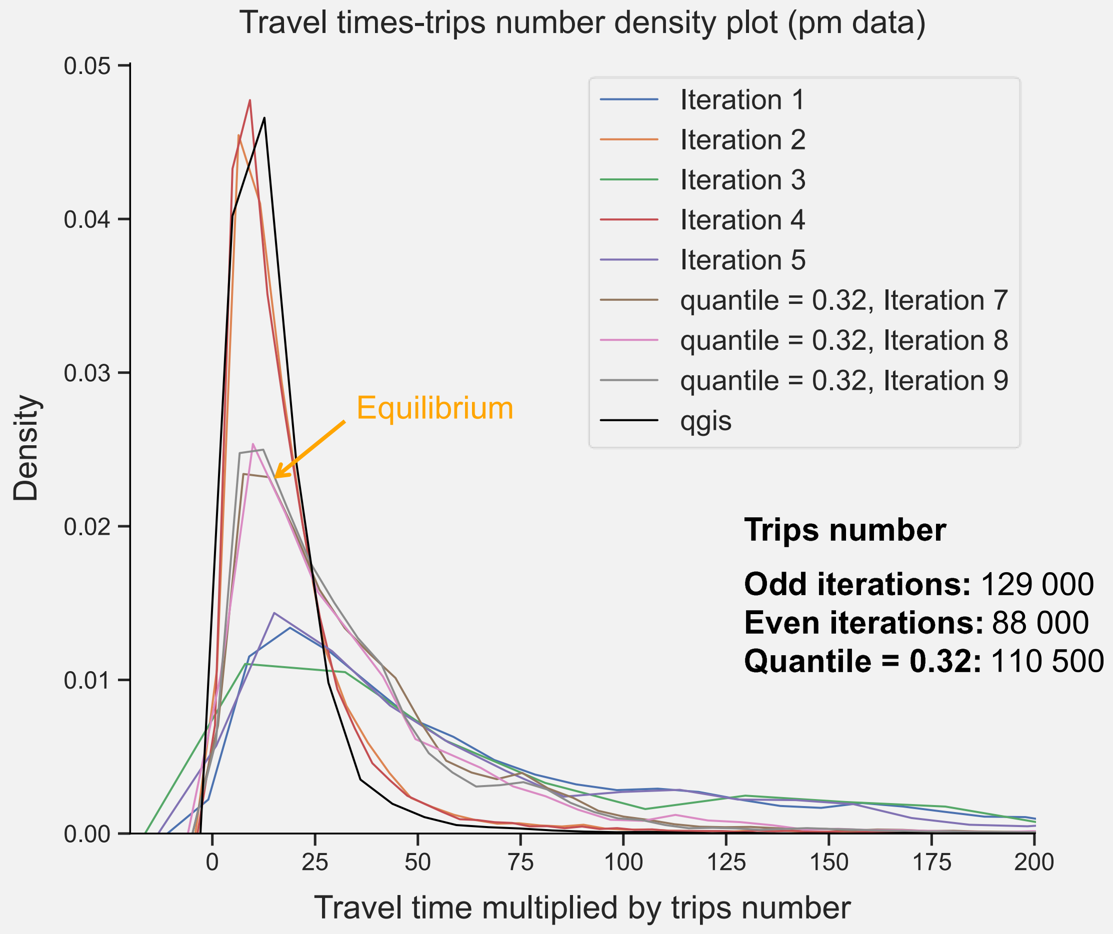
## 1 Iterations density plots and trips number, a.m.




**Trips number**

**Odd iterations:** 138 000
**Even iterations:** 90 000
**Quantile = 0.32:** 117 000

## 2 Iterations density plots and trips number, p.m.



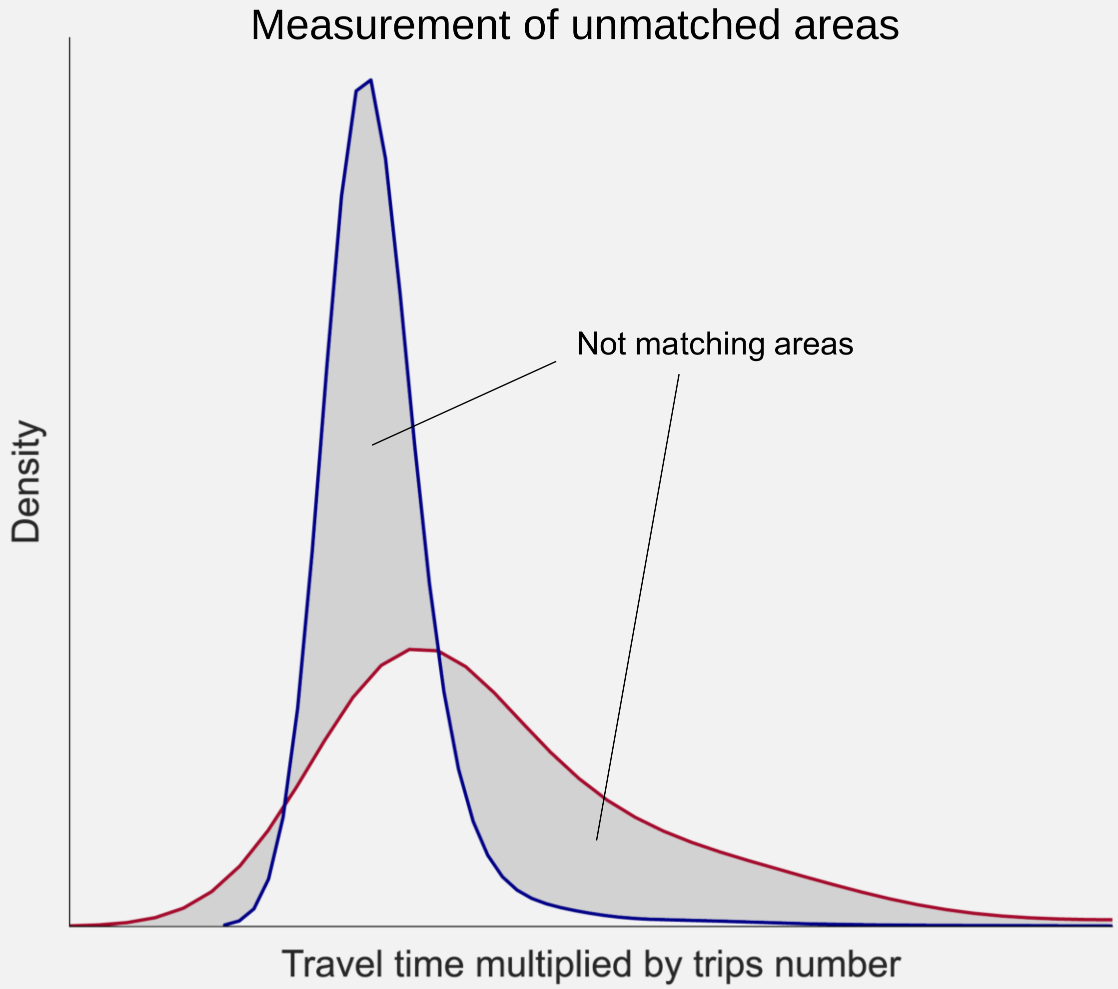

**Trips number**

**Odd iterations:** 129 000
**Even iterations:** 88 000
**Quantile = 0.32:** 110 500

# 1-st method: distribution comparison

## Area difference

1



# 1-st method: distribution comparison

## a.m. data

1


| a.m. | | |
|---|---|---|
| Iteration # | Mean | Standard deviation |
| 1 | 100.12 | 121.69 |
| 2 | 23.15 | 26.94 |
| 3 | 98.95 | 129.62 |
| 4 | 22.11 | 22.33 |
| 5 | 94.02 | 123.06 |
| 7 | 45.51 | 44.59 |
| 8 | 45.03 | 52.70 |
| 9 | 44.45 | 47.71 |

## p.m. data

2


| p.m. | | |
|---|---|---|
| Iteration # | Mean | Standard deviation |
| 1 | 73.28 | 86.51 |
| 2 | 20.94 | 32.32 |
| 3 | 83.08 | 127.47 |
| 4 | 19.77 | 27.52 |
| 5 | 72.01 | 104.11 |
| 7 | 35.87 | 38.15 |
| 8 | 36.51 | 45.88 |
| 9 | 33.85 | 39.90 |

# 2-nd method: Statistic measures

$$\mathrm{RMSN} = \frac{\sqrt{N\sum_{n=1}^{N}\left(Y_n^s - Y_n^o\right)^2}}{\sum_{n=1}^{N} Y_n^o} \qquad (1)$$

**Normalized root-mean-square error** quantifies overall error. Larger value indicates larger errors. RMSN s penalize large errors at a higher rate than small errors

In (1) $N$ = number of observations, $Y_n^o$ = observation, and $Y_n^s$ = simulated value at time n

| a.m. | | |
|---|---|---|
| Iteration # | Travel time | Number of trips |
| 2-1 | 0.94 | 5.00 |
| 3-2 | 1.95 | 5.95 |
| 4-3 | 0.91 | 3.82 |
| 5-4 | 2.01 | 5.84 |
| 7-5 | 0.76 | 3.17 |
| 8-7 | 0.54 | 3.64 |
| 9-8 | 0.56 | 3.86 |

| p.m. | | |
|---|---|---|
| Iteration # | Travel time | Number of trips |
| 2-1 | 0.72 | 7.02 |
| 3-2 | 1.42 | 7.33 |
| 4-3 | 0.79 | 4.87 |
| 5-4 | 1.34 | 7.00 |
| 7-5 | 0.66 | 3.90 |
| 8-7 | 0.43 | 4.55 |
| 9-8 | 0.38 | 4.75 |

## 2-nd method: Statistic measures

$$\text{RMSPE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left[\frac{Y_n^s - Y_n^o}{Y_n^o}\right]^2} \quad (2)$$

**Root-mean-square percentage error** is calculated in percents. Like RMSN it penalizes large errors at a higher rate than small errors

In (2) $N$ = number of observations, $Y_n^o$ = observation, and $Y_n^s$ = simulated value at time n

| a.m. | | |
|---|---|---|
| Iteration # | Travel time | Number of trips |
| 2-1 | 0.54 | RMSPE is not applicable to number of trips because for some O-D couples it's value was 0 |
| 3-2 | 2.95 | |
| 4-3 | 0.55 | |
| 5-4 | 2.85 | |
| 7-5 | 0.65 | |
| 8-7 | 0.73 | |
| 9-8 | 0.77 | |

| p.m. | | |
|---|---|---|
| Iteration # | Travel time | Number of trips |
| 2-1 | 0.75 | RMSPE is not applicable to number of trips because for some O-D couples it's value was 0 |
| 3-2 | 2.62 | |
| 4-3 | 0.50 | |
| 5-4 | 2.42 | |
| 7-5 | 0.65 | |
| 8-7 | 0.56 | |
| 9-8 | 1.67 | |

# 2-nd method: Statistic measures

$$U = \frac{\sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(Y_n^s - Y_n^o\right)^2}}{\sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(Y_n^s\right)^2} + \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(Y_n^o\right)^2}} \quad (3)$$

**Theil's inequality coefficient** provides information on the relative error. It is bounded between 0 and 1 (U = 0 implies perfect fit between observed and simulated measurements)

In (3) $N$ = number of observations, $Y_n^o$ = observation, and $Y_n^s$ = simulated value at time n

In the literature (1), values of $U < 0.5$ are acceptable whereas values of $U$ less than 0.1 are excellent

Theil's $U$ may be decomposed into 3 proportions of inequality: $U^M$ (indicates the systematic error of a model and the preferred values are close to 0), the variance $U^S$ and the covariance $U^C$ (indicates the avoidable random error of a model)

| a.m. | |
|---|---|
| Iteration # | U statistic |
| 2-1 | 0.84 |
| 3-2 | 0.83 |
| 4-3 | 0.85 |
| 5-4 | 0.83 |
| 7-5 | 0.62 |
| 8-7 | 0.53 |
| 9-8 | 0.57 |

| p.m. | |
|---|---|
| Iteration # | U statistic |
| 2-1 | 0.80 |
| 3-2 | 0.83 |
| 4-3 | 0.84 |
| 5-4 | 0.81 |
| 7-5 | 0.65 |
| 8-7 | 0.53 |
| 9-8 | 0.56 |

1) Kolidakis, S., Botzoris, G., Profillidis, V., & Kokkalis, A. (2019). Real-time Intraday Traffic Volume forecasting – a hybrid application using singular spectrum analysis and artificial neural networks. Periodica Polytechnica Transportation Engineering, 48(3), 226–235.

# 2-nd method: Statistic measures

$$U^M = \frac{\left(\bar{Y}^s - \bar{Y}^o\right)^2}{\frac{1}{N}\sum_{n=1}^{N}\left(Y_n^s - Y_n^o\right)^2} \quad (4)$$

In (4) $N$ = number of observations, $Y_n^o$ = observation, and $Y_n^s$ = simulated value at time n

$$U^S = \frac{\left(s^s - s^o\right)^2}{\frac{1}{N}\sum_{n=1}^{N}\left(Y_n^s - Y_n^o\right)^2} \quad (5)$$

In (5) $s^s$ and $s^o$ are standard deviations of the average simulated and observed measurements, $N$ = number of observations, $Y_n^o$ = observation, and $Y_n^s$ = simulated value at time n

| a.m. | | |
|---|---|---|
| Iteration # | U m statistic | U s statistic |
| 2-1 | 0.03 | 0.68 |
| 3-2 | 0.02 | 0.70 |
| 4-3 | 0.02 | 0.73 |
| 5-4 | 0.02 | 0.73 |
| 7-5 | 0.01 | 0.50 |
| 8-7 | 7.02E-05 | 4.04E-03 |
| 9-8 | 1.41E-06 | 2.74E-03 |

| p.m. | | |
|---|---|---|
| Iteration # | U m statistic | U s statistic |
| 2-1 | 0.02 | 0.51 |
| 3-2 | 0.02 | 0.64 |
| 4-3 | 0.02 | 0.68 |
| 5-4 | 0.02 | 0.63 |
| 7-5 | 0.01 | 0.46 |
| 8-7 | 5.88E-06 | 8.66E-03 |
| 9-8 | 9.00E-05 | 9.54E-03 |

## 3-rd method: Correlation coefficient

**Spearman correlation coefficient** was used because relationship between variables was not linear

Spearman's correlation measures the strength and direction of **monotonic relationship** between two variables. Monotonicity is "less restrictive" than linear relationship

$|r| \geq 0.7$ indicates that it's a **strong correlation**, $0.5 \leq |r| < 0.7$ tells about **moderate correlation**

| a.m. | |
|---|---|
| Iteration # | Correlation |
| 2-1 | 0.42 |
| 3-2 | 0.49 |
| 4-3 | 0.48 |
| 5-4 | 0.50 |
| 7-5 | 0.63 |
| 8-7 | 0.58 |
| 9-8 | 0.57 |

| p.m. | |
|---|---|
| Iteration # | Correlation |
| 2-1 | 0.31 |
| 3-2 | 0.39 |
| 4-3 | 0.38 |
| 5-4 | 0.40 |
| 7-5 | 0.50 |
| 8-7 | 0.47 |
| 9-8 | 0.47 |

# 4-th method: Statistical tests

Firstly it was checked whether the data is normally distributed with the help of **Shapiro–Wilk test**. In all iterations data was non-normally distributed

Then **Wilcoxon rank sum test** has been applied to verify the hypothesis whether two independent samples are drawn from the same distribution

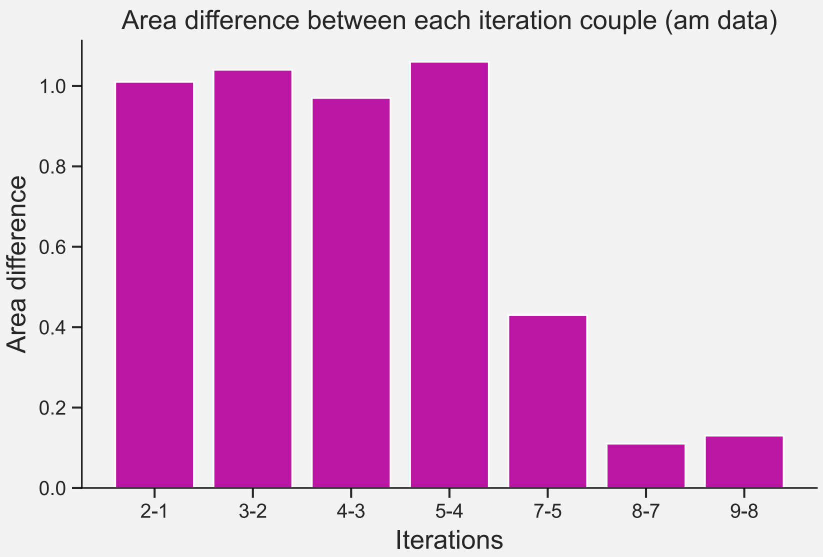
| a.m. | |
|---|---|
| Iteration # | p-value |
| 2-1 | 3.05E-148 |
| 3-2 | 8.82E-145 |
| 4-3 | 1.54E-147 |
| 5-4 | 1.89E-139 |
| 7-5 | 5.62E-20 |
| 8-7 | 0.03 |
| 9-8 | 0.75 |

| p.m. | |
|---|---|
| Iteration # | p-value |
| 2-1 | 9.63E-116 |
| 3-2 | 1.09E-123 |
| 4-3 | 3.78E-123 |
| 5-4 | 1.77E-114 |
| 7-5 | 1.05E-18 |
| 8-7 | 0.10 |
| 9-8 | 0.71 |

# 1-st method: distribution comparison

## a.m. data

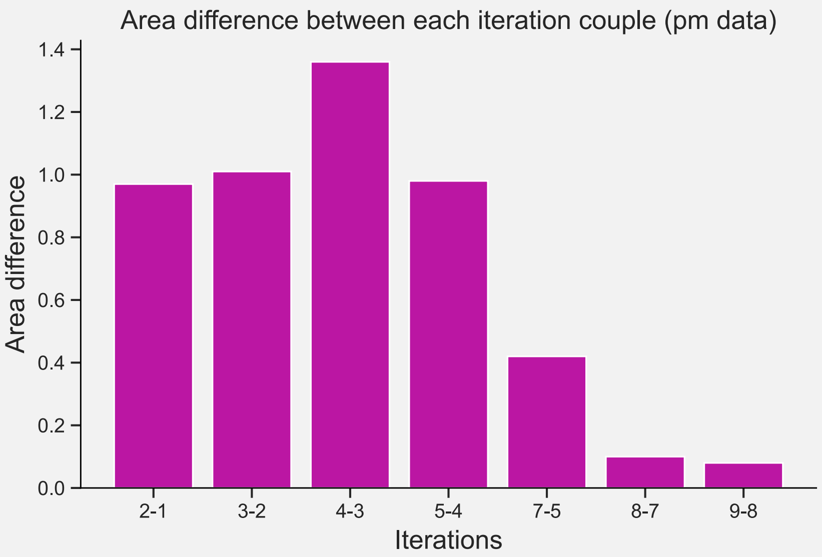
1




| a.m. | | |
|---|---|---|
| Iteration # | Mean | Standard deviation |
| 1 | 100.12 | 121.69 |
| 2 | 23.15 | 26.94 |
| 3 | 98.95 | 129.62 |
| 4 | 22.11 | 22.33 |
| 5 | 94.02 | 123.06 |
| 7 | 45.51 | 44.59 |
| 8 | 45.03 | 52.70 |
| 9 | 44.45 | 47.71 |

## p.m. data

2




| p.m. | | |
|---|---|---|
| Iteration # | Mean | Standard deviation |
| 1 | 73.28 | 86.51 |
| 2 | 20.94 | 32.32 |
| 3 | 83.08 | 127.47 |
| 4 | 19.77 | 27.52 |
| 5 | 72.01 | 104.11 |
| 7 | 35.87 | 38.15 |
| 8 | 36.51 | 45.88 |
| 9 | 33.85 | 39.90 |

# Remote working

## Share of remote workers by occupation

1